Available online at www.sciencedirect.com
SCIENCE DIRECT®
Cognitive Psychology 49 (2004) 238–299
Cognitive Psychology
www.elsevier.com/locate/cogpsych

The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing☆

Jesse Snedeker[a,*] and John C. Trueswell[b]

[a] *Department of Psychology, Harvard University, 1136 William James Hall, 33 Kirkland Street, Cambridge, MA 02138, USA*
[b] *Department of Psychology, University of Pennsylvania, 3401 Walnut Street, Room 410C, Philadelphia, PA 19104, USA*

Accepted 24 March 2004
Available online 25 May 2004

Abstract

Two striking contrasts currently exist in the sentence processing literature. First, whereas adult readers rely heavily on lexical information in the generation of syntactic alternatives, adult listeners in world-situated eye-gaze studies appear to allow referential evidence to override strong countervailing lexical biases (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). Second, in contrast to adults, children in similar listening studies fail to use this referential information and appear to rely exclusively on verb biases or perhaps syntactically based parsing principles (Trueswell, Sekerina, Hill, & Logrip, 1999). We explore these contrasts by fully crossing verb bias and referential manipulations in a study using the eye-gaze listening technique with adults (Experiment 1) and five-year-olds (Experiment 2). Results indicate that adults combine lexical and referential information to determine syntactic choice. Children rely

☆ A portion of this work was presented in proceedings to the 23rd Annual Conference of the Cognitive Science Society. The ideas in this paper owe much to our conversations with Lila Gleitman and to the comments of the many audiences who heard preliminary reports of this research. We thank Kirsten Thorpe for her assistance with testing, coding, and participant recruitment and Sylvia Yuan for her assistance in data analysis. We also gratefully acknowledge Tracy Dardick who carried out the norming studies and Jared Novick and David January who assisted in comparisons between head-mounted eye-tracking and our procedure. This work was supported by NIH Grant 1-R01-HD37507 to the second author and a National Science Foundation Science and Technology grant to the Institute for Research in Cognitive Science at the University of Pennsylvania (NSF-STC Cooperative Agreement No. SBR-89-20230).

* Corresponding author.
E-mail address: snedeker@wjh.harvard.edu (J. Snedeker).

doi:10.1016/j.cogpsych.2004.03.001

exclusively on verb bias in their ultimate interpretation. However, their eye movements reveal an emerging sensitivity to referential constraints. The observed changes in information use over ontogenetic time best support a constraint-based lexicalist account of parsing development, which posits that highly reliable cues to structure, like lexical biases, will emerge earlier during development and more robustly than less reliable cues.

Keywords: Syntactic ambiguity resolution; Spoken language comprehension; Language development; Verbs; Eye movements

1. Introduction

One of the central goals in the study of human language comprehension has been to understand how we recover the syntactic structure of a sentence from a string of words. As in many areas of cognitive psychology, research has focused on the relative contribution of lower level local cues (words) and higher level global cues (context). Questions about the time course of processing have dominated the field: Are initial structural hypotheses influenced by information about the current situation or discourse? Or, do constraints on the architecture of the comprehension system or the organization of the information source force us to exclude these nonlinguistic facts during the early stages of processing? In general, mental chronometry is not the ultimate goal of this work; it is used as a tool to carve the mind at its joints.

An alternate approach to studying the organization of the language system is to trace the development of language comprehension in ontogenetic time. In the absence of a blueprint, we may be able to learn something about the underlying architecture of sentence processing by watching the edifice go up. Until recently however, little work in developmental psycholinguistics has addressed the central issues of sentence processing. This is partially attributable to technological limitations on collecting real-time measures of processing from children and partially attributable to theoretical frameworks which have emphasized knowledge rather than performance (see Trueswell, Sekerina, Hill & Logrip, 1999). In this paper, we explore the use of contextual and lexical cues on both the ontogenetic and the chronometric time scales. We hope it will become clear that the combined examination of both time scales not only links the fields of adult and child comprehension more seamlessly but provides us with a better understanding of how linguistic discovery and linguistic use interact and mutually constrain one another.

1.1. Contextual and lexical contributions to ambiguity resolution

In the adult sentence processing literature questions about the role of contextual and lexical information in language comprehension have played themselves out in a series of studies examining the way readers initially interpret and misinterpret syntactically ambiguous phrases (for reviews, see Clifton, Frazier, & Rayner, 1994; Tanenhaus &

Trueswell, 1995). To summarize these findings, consider the sentence fragment (1) below which contains a syntactic ambiguity.

(1) The chef poked the pastry with the...

The prepositional phrase (PP) beginning with *with* is temporarily ambiguous because it could be linked to the verb *poked* (verb phrase (VP)-attachment), indicating an Instrument (e.g., *with the fork*); or it could be linked to the definite noun phrase *the pastry* (noun phrase (NP)-attachment) indicating a Modifier (e.g., *with the flaky crust*).

Crain and Steedman (1985) hypothesized that readers initially resolve ambiguities like this by taking into account the referential context of the sentence. Specifically, it was proposed that the syntactic analysis with the fewest unsatisfied presuppositions ought to be preferred. If one assumes that a definite NP like *the pastry* requires a unique referent, interpreting the *with*-phrase as a restrictive modifier would require the reader to presuppose the presence of two or more pastries, one of which has the distinguishing property (e.g., a flaky crust). An Instrument interpretation presupposes a single unique pastry and asserts the use of an instrument. Thus it was predicted that in a context containing two possible referents (two-referent contexts) readers would pursue a modifier (NP-attachment) analysis, but in a one-referent context, or even a null context, readers would pursue an Instrument (VP-attachment) analysis since the presupposition of multiple referents is not satisfied. Indeed, a number of reading studies have found precisely this pattern (e.g., Altmann & Steedman, 1988; van Berkum, Brown, & Hagoort, 1999, among many). However, other similar studies failed to find any such effects of context on ambiguity resolution (e.g., Ferreira & Clifton, 1986; Rayner, Garrod, & Perfetti, 1992), finding instead a general preference for VP-attachment. These authors suggested the VP-attachment preference reflected a general parsing heuristic to prefer the syntactically simpler alternative (minimal attachment), but this account had difficulty explaining the positive effects of context found in the literature.

One plausible explanation of these conflicting findings comes from lexicalist or constraint-based theories of parsing (e.g., MacDonald, Pearlmutter, & Seidenberg, 1994; Trueswell & Tanenhaus, 1994). Such theories predict that referential effects should be weakened or eliminated when semantic or structural properties of the words in the sentence strongly support a single syntactic analysis. Thus the conflicting findings on the efficacy of referential context manipulations could be attributable to differences in the verbs and prepositions that were used in the various experiments. Indeed, several reading studies have found that such lexical facts play an important role in initial parsing preferences (e.g., Taraban & McClelland, 1988; Trueswell, Tanenhaus, & Kello, 1993) and that referential effects are weakened or even eliminated when a verb strongly prefers a single analysis (e.g., Britt, 1994; Spivey-Knowlton & Sedivy, 1995). For instance, using materials like "*Susan put/dropped the book on the civil war onto the table*," Britt (1994) found that 2-book vs. 1-book contexts failed to guide parsing preferences when the verb strongly preferred a PP argument. For verbs like *put*, readers initially pursued VP-attachment regardless of context, but for verbs like *dropped*, context guided parsing. Similar interactions between lexical biases and referential or semantic manipulations have been found with other syntactic ambiguities (e.g., Boland, 1997; Boland, Tanenhaus, Garnsey, & Carlson, 1995; Garnsey, Pearlmutter, Myers, & Lotocky, 1997; MacDonald, 1994; MacDonald et al., 1994; Spivey-Knowlton, Trueswell, & Tanenhaus, 1993; Spivey & Tanenhaus, 1998; Trueswell, 1996).

Taken as a whole, these reading studies suggest that context has its greatest influence in the absence of strong lexical biases. This has led some researchers to contend that lexical information plays the privileged role of proposing syntactic analyses, which are compared against context only at a later stage (e.g., Boland & Cutler, 1996). Those advocating more interactive constraint-satisfaction theories of language processing have argued against such "propose-and-select" models of language processing by pointing out that the dominance of bottom-up information follows as a natural consequence of interactive theories if the lower level evidence is more reliable than higher level evidence (see MacDonald et al., 1994; as well as Kawamoto, 1993). If lexical regularities are better predictors of structure than contextual regularities (an assumption we will discuss in more detail later in this paper), such effects would emerge even in a system that does not a priori give special status bottom-up sources. Regardless of the explanation however, reading research supports a central role for lexical biases in parsing decisions, even to the extent of masking or overriding referential contributions.

1.2. Lexical and referential information revisited: World-situated listening studies

Recent work on syntactic ambiguity resolution in spoken language comprehension has however raised questions about the relative contributions of contextual and lexical information. Tanenhaus, Spivey and colleagues (Spivey, Tanenhaus, Eberhard, & Sedivy, 2002; Tanenhaus, Spivey-Knowlton, Eberhard & Sedivy, 1995) have found that situation-specific contextual information can guide parsing decisions even in the presence of verb biases that support a competing syntactic alternative. In their studies, participants were given spoken instructions to move objects about on a table while their eye movements were recorded. Target instructions, like (2) below, contained a temporary PP-attachment ambiguity, in which the verb's argument preferences strongly supported an initial VP-attachment analysis of *on the napkin.*

(2) Put the apple on the napkin in the box.

The verb *put* requires an argument to fill the role of goal or destination, and this argument usually appears as a prepositional phrase that follows the direct-object noun. The results of the reading studies (e.g., Britt, 1994) might lead one to predict that this strong lexical bias would cause listeners to initially misanalyze the first prepositional phrase (*on the napkin*) as destination regardless of the referential context. But in fact, the two-referent context was sufficient to allow listeners to avoid this garden path. In particular, scenes containing two apples, one of which was on a napkin, eliminated early and late looks to an incorrect destination object (e.g., an empty napkin), resulting in eye movements similar to unambiguous controls (e.g., *Put the apple that's on the napkin…*). Scenes with one apple resulted in large numbers of early and late looks to the incorrect destination for ambiguous items like (2) above. These findings do not support propose-and-select models, in which the initial proposed alternatives are frequency weighted, since it appears that the two-referent scene was able to guide initial parsing toward a modifier interpretation with ease despite the strong argument preferences of the verb, put. As the authors note however, such powerful effects of context are also unexpected under most interactive, constraint-satisfaction theories of parsing, given the overwhelming structural bias of this verb. Spivey and colleagues plausibly suggest that the effectiveness of their contexts was due in part to the salience and specificity of the referential scene as a constraint on interpreting the instructions. That is, when referential cues are salient, co-present with the linguistic utterance, and hence easy to maintain in memory, they may prevail over strong countervailing lexical biases. The weaker effects of the referential context in reading tasks using similar materials (e.g., Britt, 1994) are thus partially attributed to the contexts being less salient, and more difficult to maintain in memory.

1.3. The development of real time parsing abilities

Recently, Trueswell and colleagues have begun using eye gaze techniques to examine the development of sentence processing in children (Hurewitz, Brown-Schmidt, Thorpe, Gleitman, & Trueswell, 2000; Trueswell et al., 1999). The first of these studies (Trueswell et al., 1999; henceforth TSHL) was modeled on the Spivey and Tanenhaus study described above (Tanenhaus et al., 1995). Children heard instructions like *Put the frog on the napkin in the box*, in both two-referent and one-referent contexts. Five-year-olds, in striking contrast with older children and adults, blindly pursued the VP-attachment analysis, ignoring referential information. In both one-referent and two-referent contexts, children frequently looked at the incorrect destination (the empty napkin). Moreover, their actions indicated that they never revised this initial misanalysis. On 60% of the trials the children performed an action that involved the incorrect destination (e.g., moving a frog to the empty napkin before putting it in the box). In contrast to these ambiguous sentences, children's performance on unambiguous controls (*Put the frog that's on the napkin in the box*) was nearly perfect indicating that the difficulty was attributable to ambiguity rather than complexity alone. By age 8, most children acted like adults in this task, using referential context to guide their parsing decisions about ambiguous phrases.

TSHL offered two plausible explanations for the overwhelming VP-attachment preference on the part of young children. First, children's parsing preferences could be driven by their statistical knowledge of the verb *put*, which strongly supports the presence of a PP-argument.[1] This explanation would be consistent with lexicalist theories and constraint-satisfaction theories more generally. By this account, children's

[1] As TSHL note, the verb *put* is quite common in child-directed speech, giving ample opportunity for children to learn its structural preferences, namely, that when a parent utters *put* followed soon after by *on* or *in*, the PP is almost always linked as a verb argument indicating the destination of the moved object.

gradual recruitment of the referential principle over development could reflect a slower learning curve for this discourse-syntactic regularity, which is arguably less consistent and more difficult to spot during comprehension than lexical contingencies. Detecting the need for modification from a scene or conversation requires tracking the relevant domain of reference as well as understanding how focus changes over the course of the conversation. In contrast, detecting lexical contingencies merely requires hearing the word and noting the semantic or syntactic relation that is conveyed.

Second, it is possible instead that the children in TSHL were exhibiting a general structural preference for VP-attachment. Certain acquisition theories make this prediction, proposing that the child-parser might avoid complex syntactic structures (i.e., a Minimal Attachment strategy, Frazier & Fodor, 1978; Goodluck & Tavakolian, 1982) or might ban complex syntactic operations entirely (e.g., the No-Adjoin principle, Frank, 1998). Parsing revisions that are based on lexical and/or referential sources may simply get *faster* over the course of development (Goodluck & Tavakolian, 1982), to the point where the predicted erroneous attachment, although still made, is undetectable to the experimenter who is measuring adult parsing behavior (Frazier & Clifton, 1996).

Importantly, the children's failure to use referential context in TSHL is accounted for differently under these two theories. Under syntactically driven accounts, these results could be taken as evidence that children cannot use visual referential context to guide the resolution of PP-attachment ambiguities, perhaps because their processing system cannot be influenced by non-syntactic representations. However, the presence of strong lexical biases in the critical sentences raises another possibility. As we noted above, adults often fail to use referential information to rapidly guide online interpretation when the test materials contain strong lexical biases (Britt, 1994). Perhaps five-year-old children are capable of using referential information, but failed to do so in the TSHL study because of the strong destination preference of the target verb.

1.4. The role of lexical information in grammatical development

The question of whether children use lexical biases in online interpretation is intimately tied to questions about the relationship between lexical generalizations and syntactic representations in language development. In the field of language acquisition, as in sentence processing, these issues have inspired innumerable studies and fueled a lively debate (see e.g., Fisher, 2002; Tomasello, 2000a). For our purposes the critical distinction is between theories in which major syntactic categories are primitive and innate and theories in which lexical distribution is the stuff of which categories are made.

Theories of this second kind are obviously compatible with the precocious use of lexical distribution. For example, both constructivists (such as Tomasello, 1992) and advocates of distributional learning (e.g., Maratsos, 1982; Maratsos & Chalkley, 1980) have argued that children's early combinatorial speech reflects small scale generalizations about the distribution of individual predicates. By noting overlap in the contexts in which words appear, children form increasingly broad categories: first building subcategories of words that appear in similar contexts (contact actions and mental states) and eventually creating the major syntactic categories like verb or preposition. Since the youngest children have, by hypothesis, not yet formed broad syntactic categories, they would have to rely on lexically specific generalizations to guide online comprehension. In fact, given this limited system of representation, even cues that are not specific to a lexical item (prosodic phrasing, referential context) would have to be stored with and conditioned upon the particular verb with which they occurred. Tomasello and colleagues claim that children construct the syntactic category verb between 3 and 4 years of age but generalizations at the level of individual words and lexical subclasses continue to play a role in the grammar on into adulthood (Tomasello, 2000b). Thus on this account there is no principled reason to expect that slightly older children would stop using lexical distribution in online interpretation.

In contrast, theories in which syntactic categories are innate primitives will vary in the predictions that they make about children's early knowledge of lexical distribution. For example, semantic bootstrapping proposals claim that children identify syntactic categories in the input by using innate semantic–syntactic correspondences. Whether these models predict the early creation of verb subcategories depends on the set of correspondences that they postulate. Pinker's (1984) correspondences for grammatical functions provide the necessary primitives for tracking the complementation privileges of verbs, while the more limited set of semantic–syntactic correspondences proposed in Macnamara (1982) do not.

None of these accounts however rules out the possibility that children collect lexical statistics for some purpose other than syntactic category formation. For example, Gleitman and colleagues have argued that learning the meaning of a verb often requires information about the set of syntactic environments in which that verb appears (Gleitman, 1990; Snedeker & Gleitman, 2004). Research on this syntactic bootstrapping hypothesis has demonstrated: that the structural contexts of verb use in parental speech are systematically related to verb meaning (Fisher, Gleitman, & Gleitman, 1991); that toddlers are able to use a single syntactic context as a cue to verb meaning (Fisher, 1996; Fisher & Snedeker, 2002; Naigles, 1990); and that under some circumstances children may use the range of frames in which a verb is used to further constrain its meaning (Naigles, 1996). While these studies do not demonstrate that children use this distributional information in parsing, they do indicate that children are able to detect and represent this information. We will return to these issues when we evaluate the findings described below.

1.5. Goals of the present study

In this introduction, we have outlined two contrasts that currently exist in the sentence processing literature. First, adult reading studies suggest an important or even privileged role for lexical information in proposing syntactic alternatives, whereas world-situated eye-gaze studies suggest that referential cues play a decisive role. Second, in contrast with adults, similar eye-gaze studies of children show that their parsing decisions are driven almost exclusively by lexical biases or perhaps structurally based parsing principles. Although explanations have been provided individually for each of these contrasts, a unified account of the entire data pattern remains elusive. The current set of experiments was designed to shed some light on these issues.

With this goal in mind, we followed the lead of the prior adult reading studies that have, in a single experiment, fully crossed lexical preferences with manipulations of referential context, except we performed these manipulations in the world-situated eye-gaze task of Tanenhaus and colleagues. Such manipulations should reveal the relative contributions of these factors under all possible combinations. In Experiment 1, we explore the role of lexical biases in adult spoken language comprehension. Does lexical information play a role in the presence of a rich visual context? Do verb biases play a role even when visual scene information strongly supports a particular analysis? In Experiment 2, we collect similar observations in five-year-olds, to observe the full pattern of information combination in this age group. This experiment seeks to characterize the developmental change observed by TSHL, distinguishing in particular between the lexicalist and structural accounts of child parsing preferences given above.

2. Experiment 1

In this experiment, adults heard instructions containing a PP-attachment ambiguity (e.g., "Feel the frog with the feather") in both two-referent and one-referent contexts. We compared verbs with different structural preferences, ranging from verbs that frequently appear with an instrument phrase to those that rarely do so. Target instructions were globally ambiguous sentences rather than the temporarily ambiguous sentences typically used in comprehension studies. This was done for two reasons. First, we wanted to use simpler sentences than those used in TSHL (i.e., ones without a second preposition) to avoid confusing children with uncommon sentence types (Experiment 2). Although the children in TSHL performed well on the complex unambiguous sentences, it is possible that complexity interacts with ambiguity. For example, children may have trouble resolving a PP-attachment ambiguity when it appears in a long or complex sentence. Second, we were concerned that the previous listening studies (Tanenhaus et al., 1995; TSHL) may have failed to find evidence that a VP-analysis was being considered in two-referent contexts simply because the disambiguating prepositional phrase occurred so soon after the introduction of the ambiguous phrase. The widespread use of structurally ambiguous *with*-phrases was verified with a corpus analysis of child-directed speech (see Appendix A).

Because the sentences used in this study are never definitively disambiguated, we should expect continuity between the listeners' online attachment preferences and their ultimate interpretations. If listeners rely entirely on the visual context, then in two-referent contexts they should interpret the ambiguous phrase as a modifier, regardless of verb bias. This preference should be reflected in both their eye movements and their actions. In contrast, if listeners simultaneously consider both lexical and contextual information then we would expect to find: (1) an effect of verb bias in both the one- and two-referent contexts and (2) an effect of referential context in some or all of the verb classes. Eye movement patterns should reveal any temporal dissociations between the uses of these two cues.

2.1. Methods

2.1.1. Participants

Thirty-six students at the University of Pennsylvania volunteered for the experiment (twelve in each of the verb bias conditions). They received course credit or were paid for their participation. Twelve were male and all were native speakers of English.

2.1.2. Procedure

Participants were told that they were going to follow spoken instructions and that their responses would serve as a point of comparison for a study of how children follow directions. Each participant sat in front of an inclined podium. At the center of the podium was a hole for a camera which was focused on the participant's face. In each quadrant of the podium was a shelf where one of the props could be placed. At the beginning of a trial, one experimenter laid out the props and introduced each one using indefinite noun phrases. For instance, the objects shown in Fig. 1B would have been introduced by paying "This bag contains a candle, a feather, a frog, another candle, another frog and another feather." This procedure ensured that participant knew the labels for toys and that subsequent reference to the objects using definite noun phrases (e.g., "the frog") sounded natural. Any object held by a toy animal was always introduced separately and not mentioned as part of a complex NP (i.e., phrases like "a frog with a candle" were not used to introduce objects). This was done to avoid structurally biasing or priming subjects for the use of a complex NP in the target utterance.

The second experimenter played prerecorded sound files from a laptop computer connected to external speakers. The trial began with an instruction to look at a fixation point at the center of the display. Then the participant was given two single sentence commands involving the props. The participant heard the first command, performed that action, and then heard the second command. Participants signaled that an action was completed by saying "done." A second camera, placed behind the participant, recorded his/her actions and the locations of the props.[2] The first experimenter stepped behind the participant before the first sentence began and remained out of sight until the final action was completed. The second experimenter was separated from the participant by the computer screen and the platform and was instructed not to look at the participant during the trial. On a few occasions the participant asked for clarification or requested that the instruction be repeated. The experimenter responded by playing the sound file again but the eye movements were only coded for the initial presentation of sentence. After the study was completed the

[2] Much of the previous research on this topic has used head-mounted cameras and computer eye-tracking algorithms to measure the fixations that accompany a spoken sentence (see e.g., Tanenhaus et al., 1995; Trueswell et al., 1999). Our justification for this change in methods and a comparison between data collected with an eye-tracking system and data collected with the hidden camera is given in Appendix D.

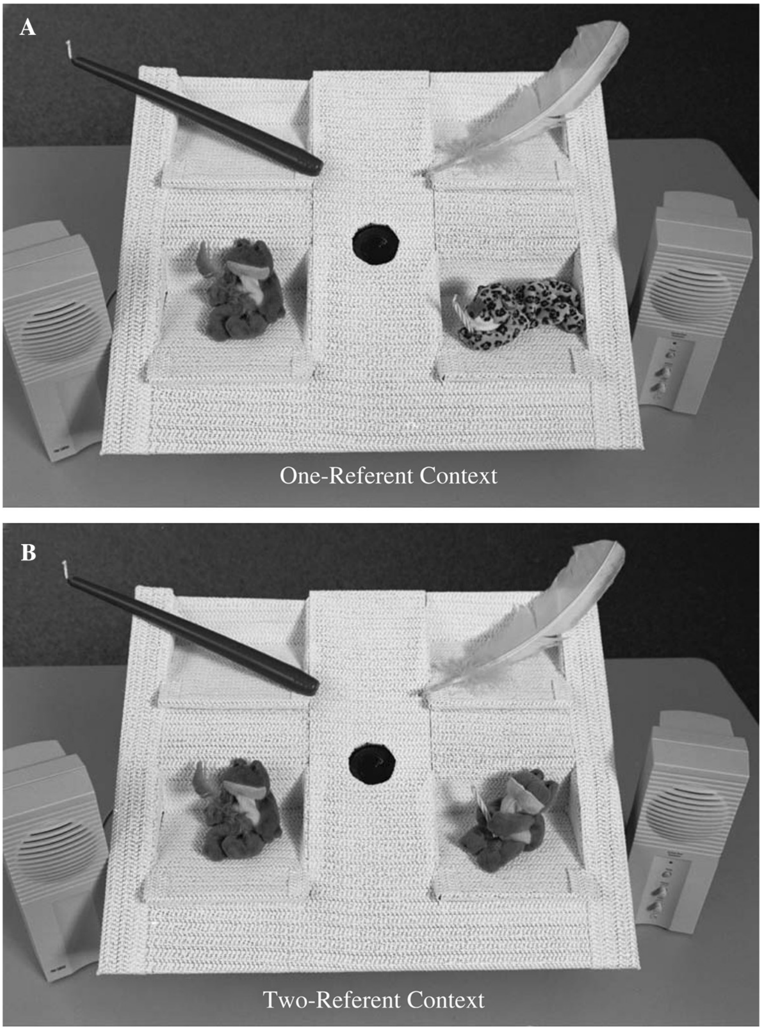


Fig. 1. Example of: (A) a One-Referent and (B) a Two-Referent context in Experiments 1 and 2 for the target sentence "Feel the frog with the feather."

participants were interviewed to assess their awareness of the experimental manipulations and the ambiguity in the critical items.

2.1.3. Stimuli

On the critical trials, the first command contained an ambiguous prepositional phrase. Examples of the three different types of verbs were used in this study are given in (3a–c). The verbs were identified in an earlier sentence completion study (see Appendix B), in which adult participants were asked to complete sentence fragments that ended with the ambiguously attached preposition (e.g., "Touch the teddy bear with…"). The verbs in the Modifier Bias condition were ones for which modifier completions (e.g., "the big brown eyes") were at least three times as frequent as instrument completions (e.g., "your pointer"). For the Instrument Bias verbs the opposite rule applied. Equi Bias verbs were those that fell somewhere in between ($M = 52\%$ instrument completions).

(3) (a) Choose the cow with the stick. (Modifier Bias),
(b) Feel the frog with the feather (Equi Bias),
(c) Tickle the pig with the fan (Instrument Bias).

The set of toys that accompanied these sentences contained the following objects: (1) a Target Instrument, a full scale object that could be used to carry out the action (e.g., for 3b a large feather); (2) a Target Animal, a stuffed animal carrying a small replica of the Target Instrument (e.g., a frog holding a little feather); (3) a Distractor Instrument; a second full scale object (e.g., a candle); and (4) A Distractor Animal, a stuffed animal carrying a replica of the Distractor Instrument. For Two-Referent trials the Distractor Animal and Target Animal were of the same kind (e.g., both frogs) while for the One-Referent trials the Distractor Animal was of a different kind (e.g., a leopard carrying a candle). Fig. 1 illustrates the One-Referent and Two-Referent Contexts for the sentence "Feel the frog with the feather." The Target Instruments for each sentence were chosen on the basis of a prior norming study in which participants were asked to rate objects as potential instruments for performing particular actions (see Appendix C). We selected objects that were rated as being moderately plausible and balanced the ratings across the three Verb Bias conditions.

Verb Bias was manipulated between participants. This was done to minimize the number of trials per participant to ensure that children could complete the same study. Referential Context was manipulated within participants but was blocked. Two presentation lists were constructed for each Verb Bias condition, so that each of the eight target trials appeared in only one of the Referential Context conditions on a given list but appeared in both conditions across lists (resulting in four target trials in each condition per participant). The first half of one list contained all One-Referent Contexts while the first half of the other list contained just Two-Referent Contexts. The critical trials were interspersed with 24 filler trials. The prop sets for the filler trials were similar to those used in the target trials: the attributes of the animals were matched to the large objects, and animals of the same kind were used in half of the filler prop sets. Both filler and target trials included two commands; the second command was always an unambiguous filler sentence. Thus, each subject heard 56 unambiguous sentences (the first instruction of the 24 filler trials and the second instruction of all 32 trials) and eight ambiguous ones. Each list was presented in two orders (forward and reverse).

2.1.4. Coding

Trained coders watched the videotape of the participant's actions and coded them into three separate categories: (1) Instrument Response: participant used the Target Instrument to perform action on Target or Distractor Animal; (2) Mini-Instrument Response: participant used miniature object attached to the Target Animal to perform action on Target Animal; (3) Modifier Response: participant performed the action on Target Animal without using the Target or Mini-Instrument. Based on previous studies with unambiguous instrument sentences we would have expected to get Mini-Instrument responses on about 10% of the trials where participants interpreted the ambiguous phrase as an instrument (Snedeker & Trueswell, 2003). Because Mini-Instrument responses might lead to less interpretable eye-movement patterns, we explicitly discouraged participants from touching the miniature objects.[3] As a result, Mini-Instrument responses were vanishingly rare in this experiment (1 in 288 trials). Two test trials were excluded from further analysis due to experimental errors.

Eye movements were coded from the videotape of the participant's face, using frame by frame viewing on a digital VCR. Both the camera and VCR used SONY DVCAM digital video tapes with audio-lock recording. The coding was conducted by undergraduate assistants, post-baccalaureate research assistants and the first author, who also trained the other coders. The first coder noted the onset of the verb and the direct-object noun and the onset of each change in gaze and the direction of the subsequent fixation. The direction of a fixation was coded as being in one of the quadrants, at center, or away from the display. If the subject's eyes were closed or not visible, the frame was coded as missing and the data were excluded from the analysis (only 1.7% of the coded frames were missing). A trial began at the onset of the target utterance and ended when the participant began the target action. When the beginning of the action could not be verified (e.g., for the verb "listen") the trial ended two seconds after the sentence ended. Sentence onset and the direction of each fixation were verified by a second coder who was given information about the onset of the eye movements. The two coders agreed on the direction of fixation for 94.4% of the coded frames. Disagreements were resolved by a third coder.

To further validate our measure, we performed a direct comparison of data collected with the hidden camera and data collected with a head-mounted eye-tracking system. For two individuals we simultaneously recorded eye data from our hidden camera as well as an ISCAN head-mounted eye-tracker. Then independently, one coder recorded the direction of eye fixations from the hidden camera, while another coder recorded eye fixation locations from eye-tracker scene image output in the manner normally done for this device (see Appendix D for details). Agreement on eye position across all trials was high (93.2%). In fact the degree of agreement between the methods is just below what we have observed for two independent coders within either of the methods.

2.2. Results and discussion

The results are divided into three sections below. First, we present participants' actions in response to the target instructions, analyzing whether an instrument was

[3] Because Mini-Instrument and Modifier interpretations should both lead to exclusive fixation on the Target Animal, these responses would weaken any effect of attachment on eye movements. At the beginning of the study participants were told not to touch the little objects because they might become unglued and they were reminded of this if they strayed during the filler trials. No feedback was given in response to any of the target sentences.

used to carry out an action. This offline measure provides an assessment of participants' final interpretation of the ambiguous phrase as an instrument. Second, we briefly present data on the proportion of eye movements to the instrument object within a large 'coarse-grain' temporal window. This gives us a second offline measure of participants' interpretation of the ambiguous phrase. Online measures are then presented in the third section, where we analyze looks to the potential instrument and other objects in the scene over time. For all measures, the reliability of effects were determined by conducting an analysis of variance (ANOVA) on the participant means with three between-participant factors (Verb Type, List, and Order) and one within-participant factor (Referential Context). Equivalent ANOVAs were conducted on item means with two between-item factors (Verb Type and Item Group) and two within-item factors (Referential Context and Order).

2.2.1. Offline-measure: Actions

Fig. 2A plots by condition the proportion of trials in which the participants performed their action using an instrument. Participants' instrument actions were strongly influenced by the type of verb used in the instruction ($F1(2,24) = 36.54$, $p < .001$; $F2(2,18) = 69.99$, $p < .001$). Collapsing across the referential context conditions, participants performed instrument actions 77% of the time for instructions containing Instrument-Biased verbs, 21% of the time for Equi-Biased verbs, and only 7% of the time for Modifier-Biased verbs.

Referential Context also had a strong effect on performance ($F1(1,24) = 10.81$, $p < .005$; $F2(1,18) = 15.99$, $p < .001$). In One-Referent Contexts 42% of the responses involved the Target Instrument, in Two-Referent Contexts only 27% did so.

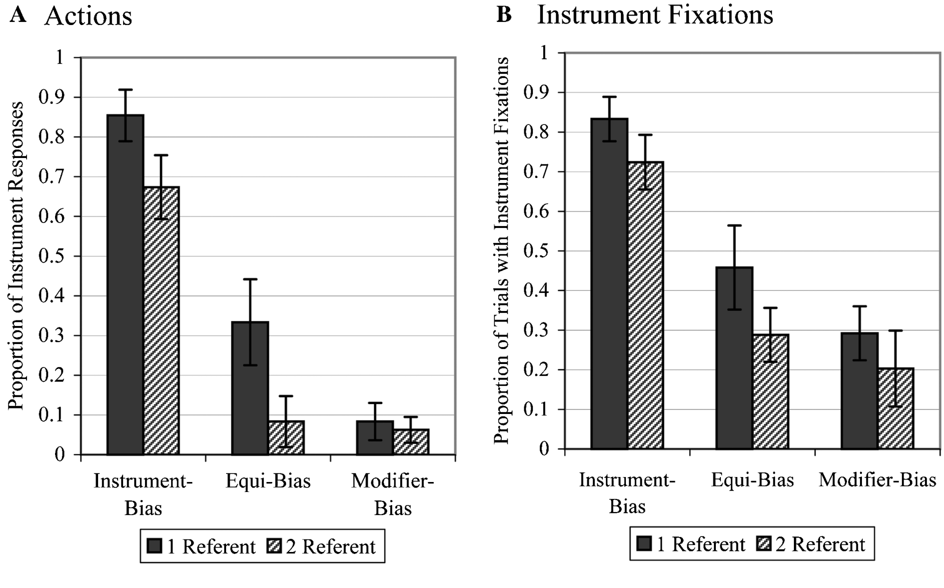


Fig. 2. Proportion of: (A) Instrument responses and (B) trials with fixations on the Target Instrument from sentence onset to the onset of the action, Experiment 1 (adults).

Although the interaction between Verb Type and Referential Context was not reliable ($F1(2,24)=2.20$, $p<.2$; $F2(2,18)=3.28$, $p=.06$), the effect of context appeared to be isolated to the Equi Biased Verbs ($F1(1,8)=5.33$, $p<.05$; $F2(1,6)=11.39$, $p<.05$) and Instrument Biased Verbs ($F1(1,8)=5.59$, $p<.05$; $F2(1,6)=4.74$, $p=.07$). There was no effect of Referential Context for the Modifier Biased Verbs ($F1(1,8)<1$, $p>.5$; $F2(1,6)=1.00$, $p>.3$).

Of particular interest here is the fact that two-referent scenes did not completely eliminate instrument actions for adults, as we would expect if such scenes required the ambiguous PP (*with the stick*) to be interpreted as a restrictive modifier. Instead, 65% of trials in the Two-Referent Instrument Condition resulted in instrument actions. Moreover, when participants in this condition performed an instrument action, they were just as likely to act on the Distractor Animal (e.g., the pig with the leaf) as they were to act on the Target Animal (e.g., the pig with the fan), indicating that they did not interpret the PP as a modifier.

If these were children performing the task, one might conclude that they did not understand that a definite NP, such as "the frog," requires a unique referent in the presence of multiple referents. However, given that these are adults, the most logical conclusion is that the restrictive modifier interpretation of "with the stick" was outweighed by the need to take this phrase as an instrument for the verb. That is, using constraint-satisfaction terminology, the accessibility of the modifier interpretation was greatly diminished by the verb's support for a competing interpretation of the PP.

This particular finding stands in contrast to the earlier *put*-studies on adults (Tanenhaus et al., 1995; TSHL) where it was observed that two-referent scenes eliminated the goal (VP-attached) interpretation of the ambiguous phrase. We suspect that multiple differences between these studies conspire to produce such disparate findings. We postpone discussion of these potential differences until the end of the paper. We only note here that two-referent scenes are clearly not sufficient for selection of the modifier interpretation, but rather serve as one of several cues to interpretation. Another potent cue is also revealed here, namely the syntactic and semantic preferences of the individual verb.

2.2.2. *Offline-measure: Coarse grained analysis of eye movements*

For each trial we determined whether the participant looked at the Target Instrument from the onset of the sentence until the beginning of the action. This measure was chosen for the initial analysis of the eye movements because it directly parallels earlier measures of VP-attachment in the *put*-studies (looks to the Incorrect Destination, Spivey et al., 2002; Tanenhaus et al., 1995; TSHL). Fig. 2B shows the proportion of trials with instrument fixations in each of the six conditions. Participants tended to look at the Target Instrument when they were going to use it to perform the action but seldom fixated on it otherwise. Thus the results for the fixation analysis closely echo those of the action analysis.

Fixations during the ambiguous instructions were strongly affected by the type of verb in the sentence ($F1(2,24)=36.54$, $p<.001$; $F2(2,18)=69.55$, $p<.001$). Subjects who heard Instrument Biased verbs looked at the Target Instrument on 78% of the trials, indicating that they were considering the VP-attachment. Those who were given Modifier Biased verbs looked at the Target Instrument on only 25% of the trials. Referential Context also had a reliable effect on performance ($F1(1,24) = 10.81$, $p < .005$; $F2(1,18) = 15.98$, $p = .001$). When the ambiguous sentence occurred in a Two-Referent Context, 41% of the trials included an instrument fixation; while in One-Referent Contexts 53% of the trials did so. The interaction between Verb Type and Referential Context was not reliable ($F1(2,24) = 2.12$, $p > .05$; $F2(2,18) = 3.27$, $p > .05$).

2.2.3. Online measures

While the instrument fixation measure (Fig. 2B) provides a quick summary of the participants' eye movements, it gives us no information about when those eye movements occurred. To explore the relation between the unfolding utterances and the participant's evolving interpretation, we analyzed how the distribution of eye movements changes over time (see Figs. 3 and 4). In each figure, time is displayed along the x axis in increments of 1/30th of a second (equivalent to a single video frame). Time is measured relative to the onset of a word in the sentence. The lines represent the four types of objects that the subject could look at: the Target Animal, the Distractor Animal, the Target Instrument, and the Distractor Instrument. The y axis gives the proportion of trials on which subjects were looking at each object.

We hypothesized that the participants' interpretation of the target sentences would have two effects on their pattern of eye movements. First, the syntactic attachment of the ambiguous phrase determines the perceived referent of the prepositional object ("the flower"). If the listener interprets the prepositional phrase as an instrument we should see looks to the Target Instrument. In contrast, if the listener interprets the phrase as modifier then there is no particular reason for looking at the Target Instrument (instead the listener might look at the small instrument that the Target Animal is holding). Second, the listener's interpretation can determine the perceived referent of the direct-object noun phrase. In the One-Referent Condition, the referent of this phrase is disambiguated by the noun itself, but in the Two-Referent Condition the referent depends upon the interpretation of the prepositional phrase. If a listener interprets the prepositional phrase as a modifier then the entire noun phrase must refer to the Target Animal and s/he should be less likely to look at the irrelevant Distractor Animal. In contrast, if the listener interprets the prepositional phrase as an instrument, then she has no information about which animal is being referred to and we would expect her to look at both animals. Thus, an effect of verb bias on parsing should influence both looks to the Target Instrument and measures of direct-object interpretation. Because the effects on direct-object interpretation are only detectable in the Two-Referent Condition, the effect of referential context can only be explored by examining looks to the Target Instrument.

To test these hypotheses we examined the proportion of fixations on these objects as the utterances unfolded over time. Because the stimulus sentences varied slightly in length, we re-synchronized the utterances at each word to ensure that we are comparing regions in which the participants in the two conditions had heard more or less equivalent portions of the utterance. We begin our analyses 200 ms (six video

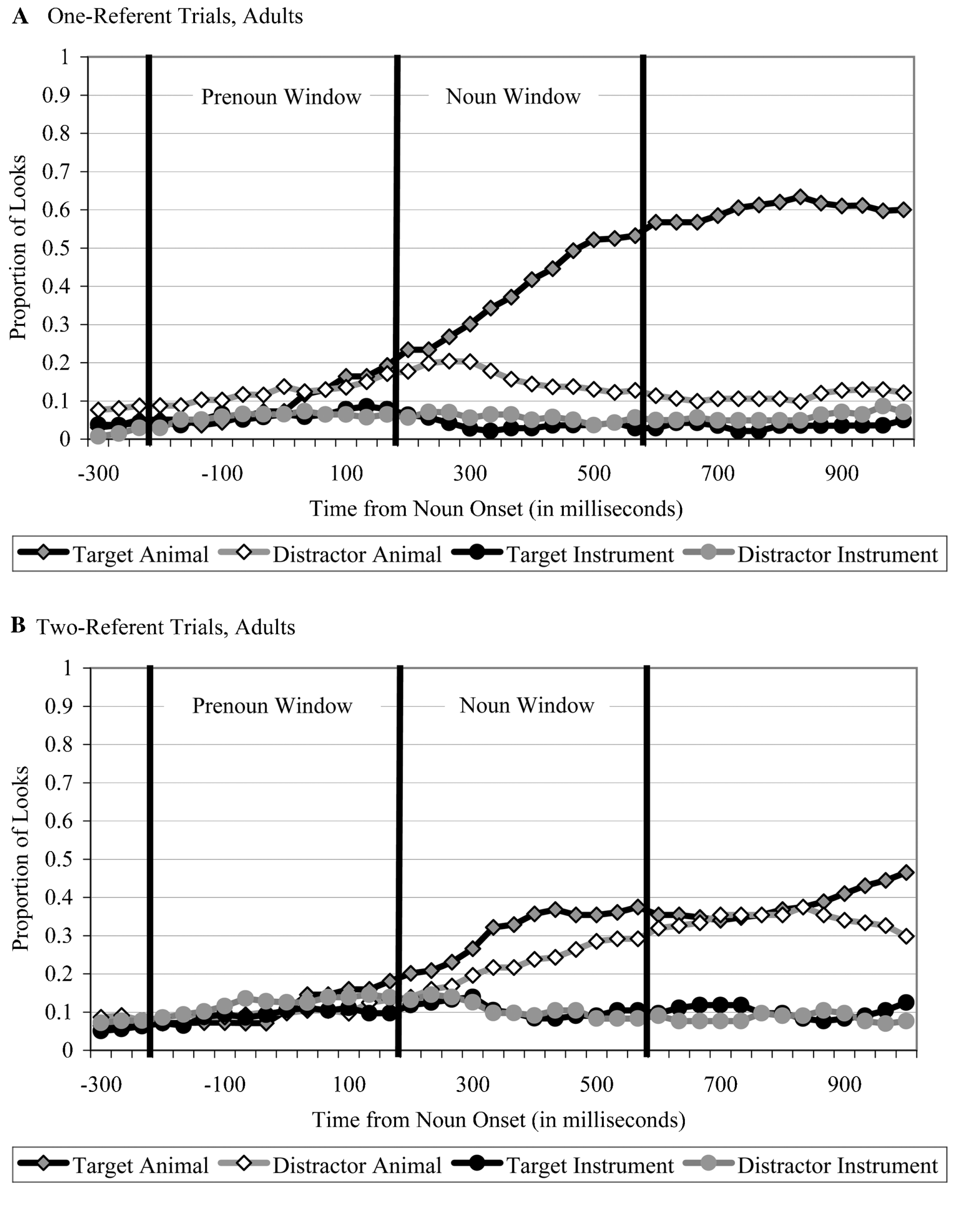


Fig. 3. Probability of fixating objects of each type over time (relative to onset of the direct object noun) Experiment 1 (adults).

frames) after the onset of the critical words because of previous research demonstrating that lexical information does not begin to influence eye movements until this time (Allopenna, Magnuson, & Tanenhaus, 1998). This lag is attributable in large part to the time that it takes to program an eye movement, which has been estimated to be as great as 150 ms (see e.g., Matin, Shao, & Boff, 1993).

2.2.3.1. Direct object noun. To demonstrate the sensitivity of our new method for measuring eye movements, we began by replicating a well-studied phenomenon: how word recognition processes influence eye movements (see e.g., Allopenna et al., 1998; Dahan, Magnuson, & Tanenhaus, 2001). The goals were to determine when the lexical content of the direct-object noun (e.g., 'frog') begins to influence looks to the Target Animal and to demonstrate that this is mediated by the number of potential referents available in the visual scene. Figs. 3A and B plot the fixation probabilities of each object type over time (relative to the Noun Onset) for the One-Referent and Two-Referent Conditions, collapsing across the three types of verbs. At the beginning of the trial, most of the participants were still looking at the central fixation point, so the total of the fixation probabilities is well below one. Shortly after the onset of the direct-object noun (e.g., "frog") the Listeners in the Two-Referent Conditions began looking at both animals (which are both frogs) while the Listeners in the One-Referent Conditions primarily look to the Target Animal (a frog) and ignore the Distractor Animal (e.g., a giraffe). Although not shown in the graph, fixation proportions are largely uninfluenced by Verb Bias in this time window, and there are few looks to the Target Instrument in any condition.

These observations were formalized by conducting statistical analyses in two time windows. The Pre-Noun window (200 ms before the Noun Onset to 167 ms after Noun Onset) contains fixations that were presumably programmed prior to perceiving the phonemic content of the noun and serves as a baseline for effects associated with the noun. The Noun window begins 200 ms after Noun Onset and ends 367 ms later (167 ms after the earliest onset time for the prepositional object). For each time window we calculated the difference in looking time to the Target Animal and the Distractor Animal (the Animal Difference Score) and conducted subject and item ANOVAs using the independent variables given above.[4]

In the Pre-Noun window there were no reliable effects of Verb Bias or Referential Context on the Animal Difference Score (all $Fs < 1$, all $ps > .3$). Participants in all conditions appeared to look at the two animals equally often. In the Noun window, however, there is a clear effect of Referential Context ($F1(1,24) = 19.11$, $p < .001$; $F2(2,18) = 18.89$, $p < .001$) verifying that the participants are using the identity of the noun to close in on the correct animal in the One-Referent Context but not the Two-Referent Context. There was no effect of Verb Bias in this time window ($F1(2,24) = 1.15$, $p > .3$; $F2(2,18) < 1$, $p > .5$).

2.2.3.2. Prepositional object. Figs. 4A–F plot the fixation probabilities of each object type over time (relative to the PP-Object Onset) for all conditions. In the Instrument Conditions, looks to the Target Instrument begin rising roughly 300 ms after the

[4] Whenever an ANOVA was conducted on a difference score or proportion of a looking time, a parallel ANOVA was conducted on an arcsine transformation of the data [for difference scores arcsine x, for proportions arcsine $(2x - 1)$]. This was done to adjust for the fact that a proportion is bounded at 0 and 1. Throughout the paper, we report the F values and p values from the transformed data (for sake of accuracy) and the means from the untransformed data (for sake of clarity). Unless otherwise noted, any effect that was reliable in the transformed data was also reliable in the untransformed data.

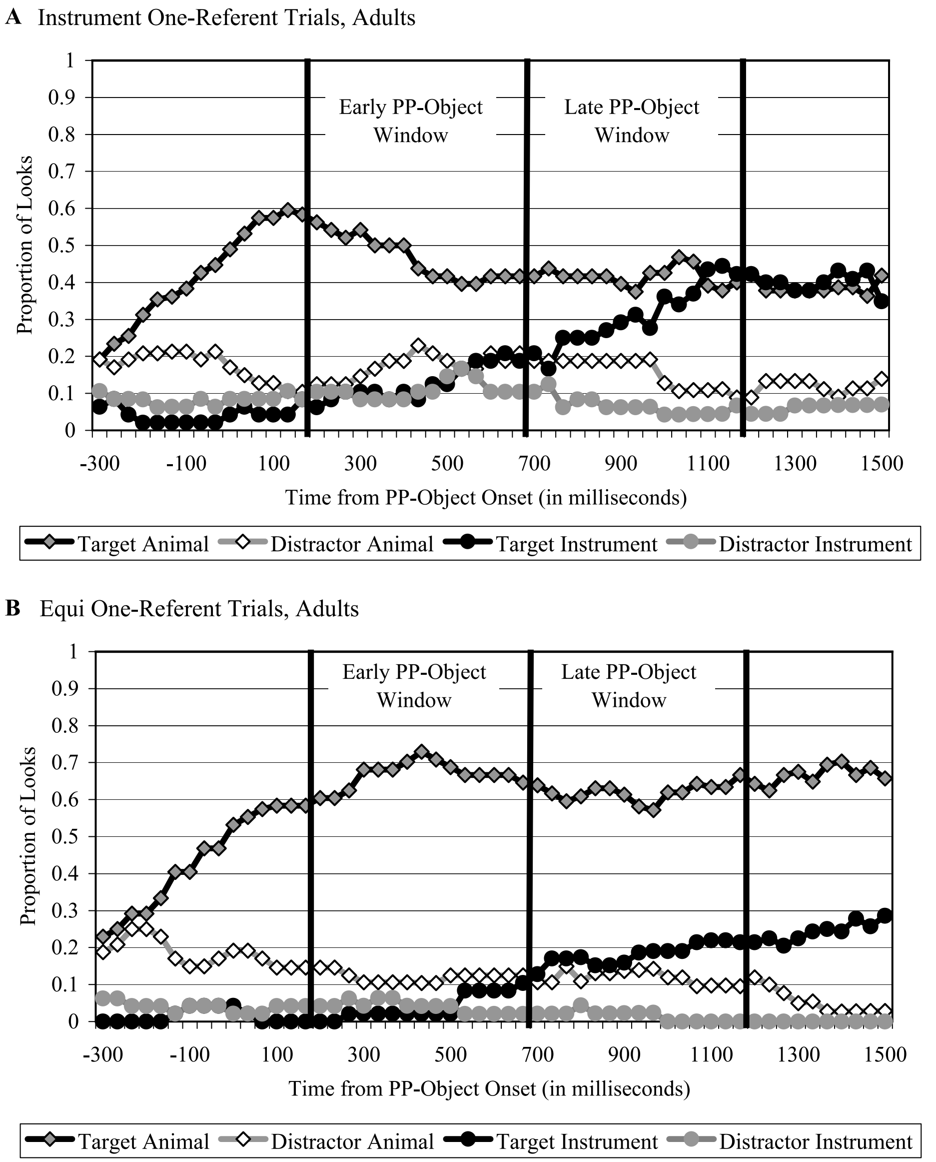
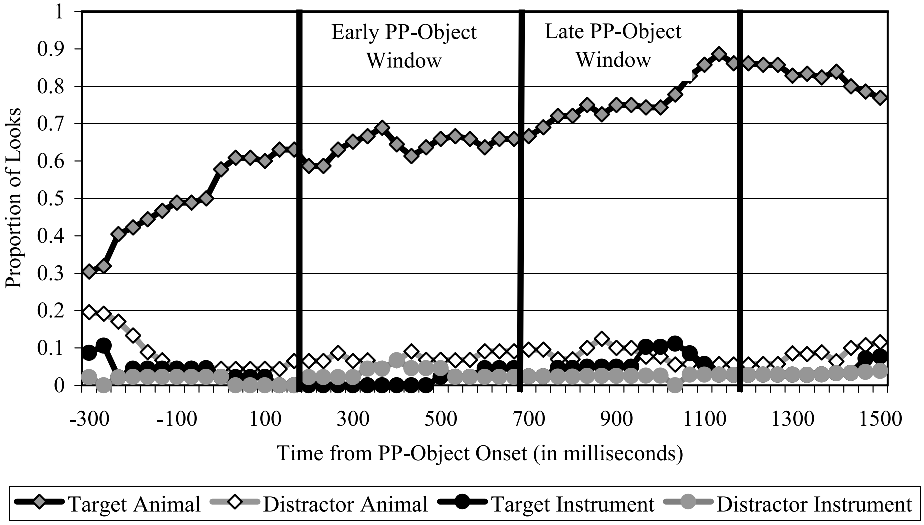


Fig. 4. Probability of fixating objects of each type over time (relative to onset of the prepositional object) for each condition, Experiment 1 (adults).

PP-Object Onset and peak about 900 ms later. In contrast, participants in the Modifier Conditions rarely look at the Target Instrument. In the One-Referent Context, Target Instrument looks in the Equi-Bias condition appear to be intermediate between the Instrument and Modifier conditions while the Two-Referent Equi Condition patterns with the Modifier Conditions. There is also a clear difference in the

C Modifier One-Referent Trials, Adults



D Instrument Two-Referent Trials, Adults

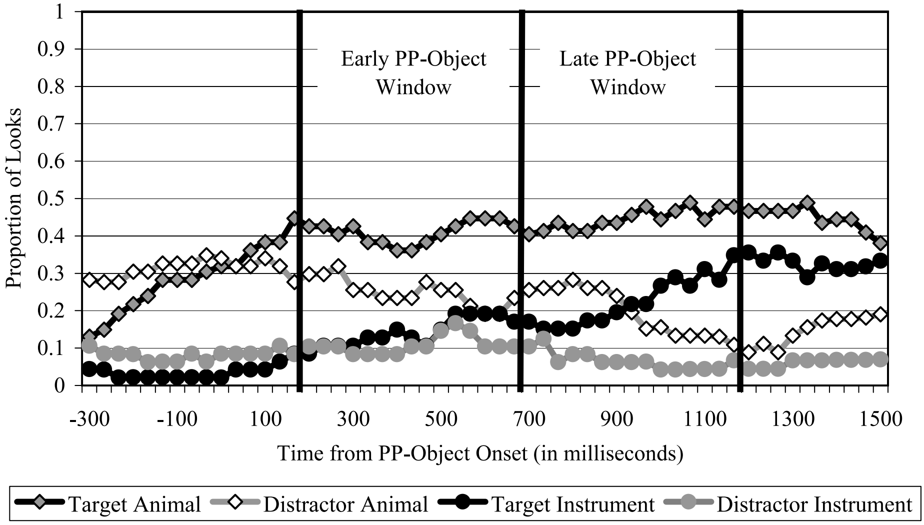
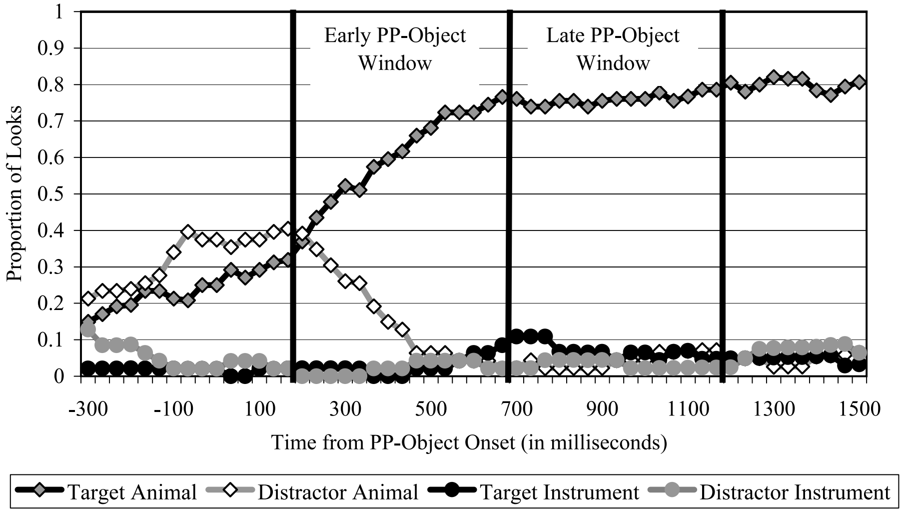


Fig. 4. (*continued*)

looks to the Distractor and Target Animals. Until about 200 ms after the PP-Object, participants in all of the Two-Referent Conditions are looking equally at the two animals, while participants in the One-Referent Conditions have already used lexical information to rule out the Distractor Animal. Shortly after the onset of the PP-Object, participants in the Equi and Modifier Two-Referent Conditions abruptly abandon the Distractor Animal in favor of the Target Animal, demonstrating that

E Equi Two-Referent Trials, Adults



F Modifier Two-Referent Trials, Adults

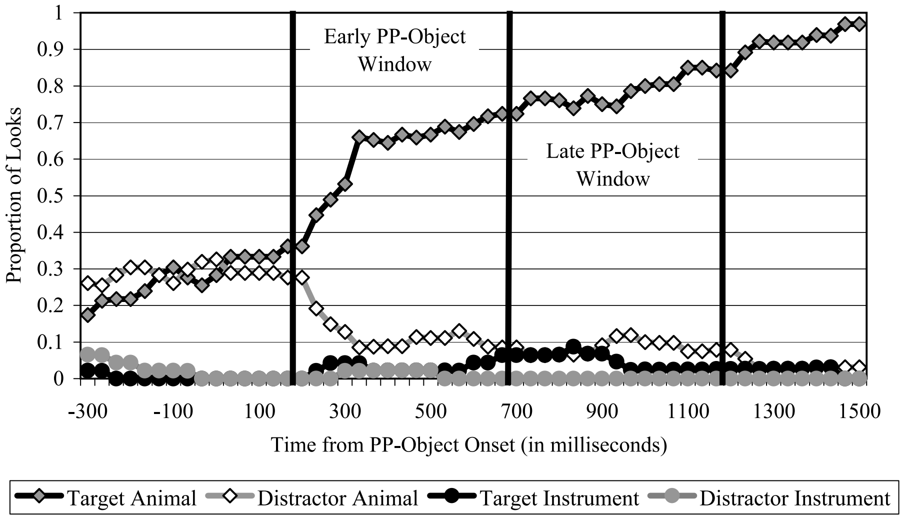


Fig. 4. (*continued*)

they have used the prepositional phrase to restrict the reference of the noun phrase. In contrast, while participants in the Instrument Two-Referent Condition show a gradual and increasing preference for the Target Animal, they continue to ogle the Distractor Animal until the end of the trial.

We validated these observations by examining two 500 ms time windows, an Early PP-Object window beginning 200 ms after PP-Object Onset, and a Late PP-Object window beginning 700 ms after the onset. Our primary analysis focused on the proportion of looking time to the Target Instrument. In the ANOVAs of the Early PP-Object window there was a reliable effect of Verb Bias ($F1(2,24) = 15.82$, $p < .001$; $F2(2,18) = 6.40, p < .01$) but no effect of Referential Context ($F1(1,24) < 1$, $p > .7$; $F2(1,18) < 1, p > .8$) and no interaction between them ($F1(2,24) < 1, p > .8$; $F2(2,18) < 1, p > .5$). Participants in the Instrument Condition devoted 13% of their looking time to the Target Instrument while participants in the Equi and Modifier Condition essentially ignored it (3 and 2%, respectively).

The effect of Verb Bias remained strong in the Late PP-Object window ($F1(2,24) = 13.19$, $p < .001$; $F2(2,18) = 18.23$, $p < .001$), largely driven by the high proportion of looks to the Target Instrument in the Instrument Condition (27%) relative to the Equi and Modifier Conditions (13 and 5%, respectively). In this later time window the effect of Referential Context was reliable in the item analysis and marginal in the participant analysis ($F1(1,24) = 4.11$, $p = .054$; $F2(1,18) = 5.34$, $p < .05$). Participants looked at the Target Instrument more when there was a unique referent for the Direct Object Noun (18% in the One-Referent Condition, 12% in the Two-Referent Condition). Again there was no interaction between Verb Bias and Referential Context ($F1(2,24) = 1.31, p > .2$; $F2(2,18) = 1.15, p > .3$).

As noted earlier, in the Two-Referent Condition the interpretation of the prepositional phrase should also have an effect on the perceived referent of the direct-object noun—a modifier interpretation specifies the Target Animal while an instrument interpretation does not specify the referent. We explored this by calculating looking time to the Target Animal as a proportion of looking time to both animals (Target Animal Preference) in the two time windows.[5] In the Early PP-Object Window, participants in the Modifier and Equi Conditions show a preference for the Target Animal (86 and 81%, respectively) while those in the Instrument Condition do not (54%), resulting in a reliable effect of Verb Bias ($F1(2,24) = 6.21$, $p < .01$; $F2(2,18) = 3.77$, $p > .05$). All groups show a preference for the Target Animal in the Late PP-Object Window but the size of the preference is considerably smaller in the Instrument Condition (65%) than in the Modifier and Equi Conditions (94 and 87%) ($F1(2,24) = 4.62, p < .05$; $F2(2,18) = 5.68, p < .01$).

In short, Verb Bias clearly shapes the interpretation of the prepositional phrase within 500 ms of the PP-Object Onset, while there is no sign of an effect of Referential Context until 500 ms later. Given this pattern of findings it is tempting to conclude that referential information has a delayed effect on interpretation relative to lexical information. But this conclusion is not warranted. In all measures and in all time

[5] We did not use Target Animal Preference as a dependant measure in the Pre-Noun and Noun time slices because many participants did not look at either animal during this time. This would have resulted in a large number of observations with 0 in the denominator. We chose not to use the Animal Difference Score in the PP-Object slices because it would not be independent of the looking time to the Target Instrument. The absolute size of the Difference Score is proportional to both the relative preference for the Target Animal and the total proportion of the time spent looking at both animals. Participants in the Instrument Condition are frequently looking at the Target Instrument and therefore have less time to look at animals. As is it turns out, the same pattern of results emerges regardless of which measure is used.

windows the effect size for Referential Context is considerably smaller than the effect size for Verb Bias. If we assume that the two effects remain roughly proportional over time, then the predicted effect in the Early PP-Object Window would be too small to be detected in this experiment.[6]

2.2.4. Analyses of the first block

In the previous analyses the effects of Verb Bias are larger, earlier and more robust than those of Referential Context. While this asymmetry could be attributable to real differences in the influence of these two information sources, it is also possible that it merely reflects an asymmetry in the experimental design. While Verb Bias was manipulated between subjects, Referential Context was manipulated within subjects. Because the Referential Context manipulation was blocked and counterbalanced for order, we can put the two variables on more equal footing by analyzing only the first block of the experiment. The results for the first block should be identical to an experiment in which both variables were manipulated between subjects. These analyses produce a pattern of findings quite similar to those reported above. The actions still show a substantial effect of Verb Bias ($F1(2,24)=25.11$, $p<.001$; $F2(2,18)=3.77, p<.05$) and a smaller effect of Referential Context ($F1(1,24)=4.12$, $p=.052$; $F2(1,18)=8.14$, $p<.05$) with no reliable interaction between the two ($F1(2,24)=1.48$, $p>.05$; $F2(2,18)=2.97$, $p>.05$). In the Early PP-Object window, looks to the Target Instrument are still affected by Verb Bias ($F1(2,24)=6.86$, $p<.005$; $F2(2,18)=4.64, p<.05$) but not Referential Context ($F1(1,24)<1, p>.5$; $F2(1,18)<1, p>.5$). This influence of Verb Bias persists through the Late PP-Object window ($F1(2,24)=10.22, p<.001$; $F2(2,18)=9.68, p<.001$). In fact, in this analysis the effect of Referential Context on Target Instrument looks is no longer reliable in the Late PP-Object window ($F1(1,24)<1, p>.4$; $F2(1,18)=1.13, p>.3$). This raises the possibility that referential effects were if anything enhanced when the adult participants had the opportunity to compare One-Referent and Two-Referent Contexts.

2.2.5. Post-experimental interview

To measure our subjects' awareness of ambiguity and the experimental manipulations, we administered a post-experiment questionnaire that contained increasingly leading questions about the purpose of the study (beginning with "What did you think the experiment was about?" and ending with "Did you notice that some of the sentences could mean more than one thing?"). We coded subjects as aware of the ambiguity if, in answer to any of the questions, they mentioned that the *with*-phrase could have two meanings or remembered a particular ambiguous target item. Ambiguity awareness varied with the Verb Bias condition ($\chi(2)=6.98, p<.05$). Subjects in

[6] To be more precise, the predicted effect size for Referential Context in the Early PP-Object time window was estimated based on: (1) the observed difference between the Modifier and Instrument Conditions and (2) the ratio of the Modifier-Instrument difference to the One Referent – Two Referent Difference in the Late PP-Object time window. This effect was somewhat smaller than the critical difference needed to find a reliable effect of reference. However, it was also seven times larger than the observed effect size. Thus these data can neither support nor refute the possibility of a delay in the use of Referential Context.

the Modifier Condition typically produced only modifier responses and rarely realized that the sentence could have more than one meaning (25% aware). Subjects in the Equi-Bias and Instrument Conditions often performed both types of responses and were typically aware of the two interpretations (75 and 67%, respectively). Within the three Verb Bias conditions there was no consistent relationship between ambiguity awareness and performance: aware subjects in the Modifier Condition were slightly more likely to have given instrument responses in One Referent Contexts ($p < .05$), while aware subjects in the Instrument Condition were marginally more likely to have given modifier responses in the Two Referent Contexts ($p < .10$). Few subjects noticed the potential ambiguity of the direct object noun (17%) and there were no systematic differences across the conditions ($\chi(2) = 1.20$, $p > .5$). None of the subjects thought that the study was about the types of verbs that were used or the number of potential referents.

2.3. Summary of Experiment 1

In Experiment 1, we found that adults interpret an ambiguous "*with*-phrase" as an instrument or a modifier depending on the evidence provided by both the referential scene and the type of verb. In fact, the influence of these two factors was largely additive. One referent scenes as compared with two referent scenes increased measures of the instrument interpretation and decreased measures of the modifier interpretation. Likewise, as the tendency of the verb to appear with an instrument phrase increased, measures of an instrument interpretation increased and measures of a modifier interpretation decreased. With relatively few exceptions, these factors manifested themselves in online measures at the very point one might expect the linguistic input to trigger their influence, suggesting rapid use of both evidential sources in the ongoing interpretation process.

Perhaps somewhat more surprising was the observation that two referent scenes alone do not appear to be sufficient to induce a modifier interpretation of an ambiguous phrase. Rather, they provide partial support which must be weighed against other factors. This finding is at odds with prior eye-gaze studies of temporary ambiguity (Tanenhaus et al., 1995; TSHL), but in line with the predictions of constraint-satisfaction theories of parsing and with earlier findings from adult reading (e.g., Britt, 1994). We return to this issue at the end of the paper, but emphasize here that the referential scene does exert an effect on adult parsing preferences.

3. Experiment 2

A very similar experiment was conducted with five-year-old children. By examining the use of referential scene information and verb-bias information in this age range, we will be able to distinguish between certain developmental accounts of parsing. Recall that TSHL found an overwhelming VP-attachment bias in five-year-olds: in response to '*Put the frog on the napkin into the box*' children overwhelming interpreted *on the napkin* as the goal, moving a frog to the empty napkin despite the temporary nature of the ambiguity. This finding could be the result of an ontogenetically early reliance on lexical information or it could be evidence that children begin by using a general structural parsing principle such as minimal attachment. In the present study, a lexically based theory would predict that children's attachment preferences would be guided by verb information. An explanation based on structural simplicity would predict a VP-attachment preference independent of verb type. In addition, manipulating verb type allows us to see whether children's failure to use referential context is limited to strongly biased verbs. Children might prove to be sensitive to context when given sentences with Equi-Biased verbs.

3.1. The linguistic expertise and cognitive limitations of five year olds

We have chosen to focus on five-year-old children for two reasons. First, this experiment seeks to build on and illuminate the findings of TSHL with this age group. Second, because of their linguistic strengths and cognitive weaknesses, five-year-olds provide an intriguing starting point for exploring the development of language comprehension. It is arguably the first age at which one can study the comprehension of sentences of this complexity without having concerns about whether the children have the grammatical knowledge required to represent them. By four, children have a strong grasp of the core syntax of their language including post-nominal modifiers and multi-clausal sentences (Brown, 1973; Crain & Thornton, 1998; Limber, 1973; McKee, McDaniel, & Snedeker, 1998). To the naked eye, their language comprehension and production appears almost adult-like. Up to age four children's utterances steadily grow longer and more complex. But by five this growth decelerates and utterance length depends more on context than age or grammatical sophistication (Chabon, Kent-Udolf, & Egolf, 1982; Klee, Schaffer, May, Membrino, & Mougey, 1989; Meline & Meline, 1981; Scarborough, Wyckoff, & Davidson, 1986).

Yet on a number of other cognitive dimensions, four and five-year-olds are radically different from adults. They perform more poorly than adults and older children on tasks that involve the inhibition of dominant responses or competing representations of an event (Hughes & Graham, 2002; Passler, Isaac, & Hynd, 1985; Perner & Wimmer, 1985; Welsh, Pennington, & Groisser, 1991). Processing speed on many cognitive tasks increases radically from early childhood to adolescence (Kail, 1991; Kail & Salthouse, 1994). Critically, young children have a much smaller memory span than adults or older children (for reviews see Dempster, 1981, 1985; Schneider & Bjorklund, 1998). Similar developmental patterns are found in tests of verbal and spatial working memory (Gaulin & Campbell, 1994; Jenkins, Myerson, Hale, & Fry, 1999). In adults, individual differences in working memory performance are correlated with performance in online reading tasks (Just & Carpenter, 1992; King & Just, 1991; MacDonald, Just, & Carpenter, 1992), suggesting that parallel limitations in children might shape their spoken language comprehension.

While five-year-olds are linguistically sophisticated in most respects, their performance on communication tasks is less impressive. They are poor at constructing unambiguous messages, detecting ambiguity in the speech of another, or judging the adequacy of a message (Glucksberg & Krauss, 1967; Lloyd, Camaioni, & Ercolani, 1995; Robinson & Robinson, 1982). Finally, most children of this age are illiterate. Thus their experiences with language are considerably different from those of the well-educated adults that are typically studied.

3.2. Methods

3.2.1. Participants

Thirty-six children between 4;6 and 5;10 participated in the study ($M = 5;1$). Parents were contacted from Philadelphia area preschools and a commercial mailing list. Four additional children participated but were not included in the analyses because they: refused to cooperate (1) were bilingual (2), or had been identified as developmentally delayed (3). Half of the children were male. Sex and age were balanced across the Verb Bias conditions and Lists.

3.2.2. Procedure and stimuli

Children were tested in the lab or in an unoccupied room at their preschool. They were told that they were going to play a game about following instructions. The procedure was identical to Experiment 1 with the following exceptions. First, the children were told the names of each object twice. Second, the children were not asked to tell us when they had finished performing each action. Instead the experimenter who introduced the toys waited until the child finished moving the toys or looked at her and then praised the child for her response regardless of her action.[7] Third, the number of filler trials was reduced from 24 to 10 so that the experiment could be completed before the children lost interest. Thus each child heard 28 globally unambiguous sentences (the first instructions from the 10 filler trials and the unambiguous second sentences from all 18 trials) plus 8 ambiguous target instructions. One experimenter interacted with the child and monitored the camera that was recording the eye movements, while the second experimenter sat behind the child and played the sound files. On every trial, before the first sentence began, the first experimenter either moved behind the child or moved behind the platform to "check" the camera. In either case the experimenter's face was out of sight during the time that the instruction was played. If the child refused to respond or asked what to do, the sound file was played again, but the eye movements were always coded from the initial presentation of the sentence. Children were not given the post-experimental interview.

3.2.3. Coding

The children's data were coded in the manner described above. Children were somewhat more likely to produce Mini-Instrument responses but the proportion was still very low (17 out of 288). There were also three actions in which the Target Instrument was used as a destination for one of the animals, suggesting the child had

[7] There was one exception to this rule. If the children moved the miniature objects on a filler trial, they were reminded that the small objects might break and told to only move the big objects and the animals. This same feedback was given to the adults as well.

interpreted the *with*-phrase as a location (like "Put the end table with the sofa"). Since this analysis implies a VP-attachment of the ambiguous phrase, these responses were treated as instrument responses for the purposes of the action analyses. The two eye-movement coders agreed on the direction of gaze for 93.6% of the coded frames and disagreements were resolved by a third coder. Only .5% of the frames were coded as missing—the coding category used when the child's eyes were closed or were not visible.

3.3. Results

3.3.1. Offline measure: Actions

The proportion of instrument actions in each condition is presented in Fig. 5A. While the adult's actions were influenced by both Verb Bias and Referential Context, the children appear to rely solely on lexical information. When the participants heard an Instrument Biased verb, they performed an instrument action 96% of the time. When they heard Modifier Biased verbs, they did so only 9% of the time, with Equi Biased verbs falling in the middle ($F1(2,24)=58.21$, $p < .001$; $F2(2,18)=83.67$, $p < .001$). In contrast, Referential Context had no effect on the children's actions ($F1(1,24)=1.15$, $p > .2$; $F2(1,18)=1.46$, $p > .2$). In One-Referent Contexts instruments were used on 47% of the trials, in Two-Referent Contexts they were used on 51% of the trials. There was no interaction between Referential Context and Verb Bias ($F1(2,24) < 1$, $p > .5$; $F2(2,18) < 1$, $p > .4$). In the One-Referent Context, children virtually always performed the action on the Target Animal (97%), indicating that they were able to identify the animals and understand the digitized speech. In

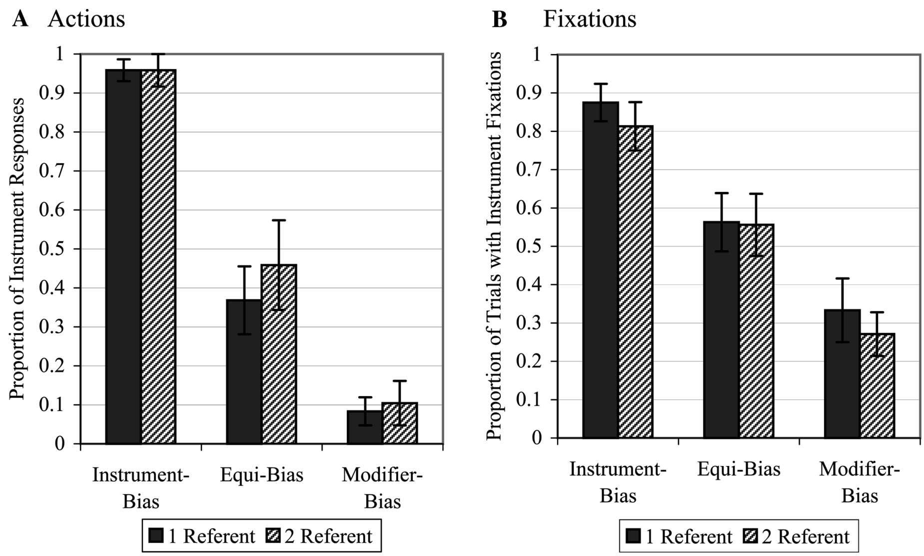


Fig. 5. Percentage of: (A) instrument responses and (B) trials with fixations on the Target Instrument, from sentence onset to the onset of the action Experiment 2 (five-year olds).

the Two-Referent Context, animal use depended on instrument use. When the children used the Target or Mini Instrument to perform the action, they showed only a slight bias towards selecting the Target Animal (62%).[8] In contrast when the instrument was not used, participants always selected the Target Animal. Thus the children, like the adults appear to be assigning just one interpretation to the ambiguous phrase.

A direct comparison of Experiments 1 and 2 reveals that five-year-old children are more likely to give instrument responses than adults ($F1(1,48) = 8.46$, $p < .005$; $F2(1,18) = 15.77$, $p < .001$). However this difference is limited to the Two-Referent Condition, resulting in an Age Group by Referential Context interaction ($F1(1,48) = 10.69, p < .005$; $F2(1,18) = 15.05, p < .001$). In the One-Referent Context, there was no effect of Age Group nor an Age Group by Verb Bias interaction (all Fs < 1.1, all p's $> .3$), indicating that the children and adults were equally sensitive to the bias of the verb. In the Two-Referent Contexts, there was both an effect of Age Group ($F1(1,48) = 20.08, p < .001$; $F2(1,18) = 28.12, p < .001$) and an Age Group by Verb Bias interaction ($F1(2,48) = 3.64$, $p < .05$; $F2(2,18) = 5.12$, $p < .05$). Children gave more instrument responses, especially in the Instrument-Biased and Equi-Biased conditions. Like adults, the children's responses in the One-Referent Conditions are guided by the combinatorial properties of the verb. Unlike adults, children fail to use the referential constraint provided in the Two-Referent Context.

3.3.2. Offline-measure: Coarse grained analysis of eye movements

Fig. 5B shows the proportion of trials with Target Instrument fixations in each of the six conditions. Like the adults, the children's fixations mirrored their actions and were strongly affected by the type of verb in the sentence ($F1(2,24) = 43.39, p < .001$; $F2(2,18) = 16.25$, $p < .001$). Participants who heard Instrument-Biased verbs looked to the Target Instrument on 84% of the trials, while those who heard the Modifier-Biased verbs did so only 30% of the time. In contrast, Referential Context had no significant effect on the children's instrument fixations; there were Target Instrument looks on 59% of the One-Referent trials and 55% of the Two-Referent ones ($F1(1,24) < 1$, $p > .4$; $F2(1,18) = 1.21$, $p > .2$). There was no significant interaction between Verb Type and Referential Context ($F1(2,24) > 1$, $p > .8$; $F2(2,18) > 1$, $p > .5$). In all three Verb-Bias Conditions, Referential Context appeared to play no role in determining the attachment of the ambiguous phrase.

3.3.3. Online interpretation

3.3.3.1. Direct object noun. To demonstrate that our technique is sensitive to children's online language comprehension, we analyzed fixations relative to the

[8] We do not attribute this slight bias to select the Target Animal to confusion over the attachment of the prepositional phrase. Unlike the adults, many of the children seemed to view the experiment as a matching game. Several of them attempted to rearrange the toys into matched pairs before the start of a trial. These children typically chose to use a matching animal and instrument for filler instructions where they were given free choice.

direct-object noun. A number of researchers have demonstrated that children can rapidly use phonological information to focus in on the referent of a noun (e.g., Fernald, Pinto, Swingley, Weinberg, & McRoberts, 1998; Swingley, Pinto, & Fernald, 1999; TSHL). These analyses simply replicate this finding. Figs. 6A and B plot the fixation probabilities of each object type over time (relative to the Noun Onset) for

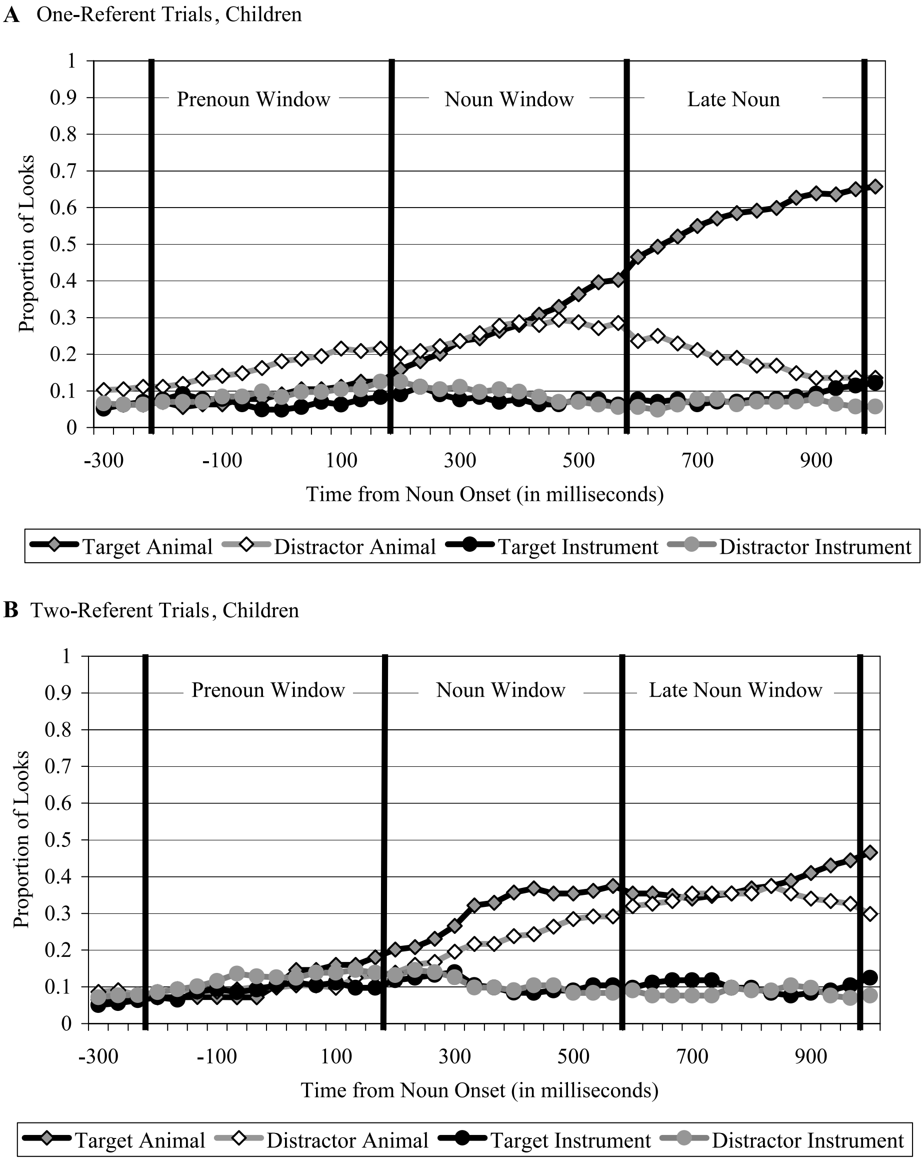


Fig. 6. Probability of fixating objects of each type over time (relative to onset of the direct object noun) Experiment 2 (five-year olds).

the One-Referent and Two-Referent Conditions, collapsing across the three types of verbs. Before hearing the direct-object noun, children in all conditions examined both the Target and the Distractor Animals. About 500 ms after the onset of the direct-object noun (e.g., "frog"), the kids in the One-Referent Conditions began switching from the Distractor Animal (e.g., a giraffe) to the Target Animal (a frog). Meanwhile those in the Two-Referent Conditions continued to look at both animals (which are both frogs) until about 900 ms after the direct-object noun, when information from the prepositional phrase began to influence fixations.

We examined the pattern of fixations in three time windows synchronized to the onset of the direct-object noun (Noun Onset). The Pre-Noun window (200 ms before the Noun Onset to 167 ms after Noun Onset) serves as a baseline for effects associated with the noun. Eye movements made during the Early Noun window (200–567 ms after Noun Onset) would have been programmed while the direct-object noun was unfolding. The Late Noun window includes looks that were programmed shortly after the offset of the direct-object noun (600–967 ms). For each time window we calculated the difference in looking time to the Target Animal and the Distractor Animal (the Animal Difference Score) and conducted ANOVAs (using the same independent variables given above).

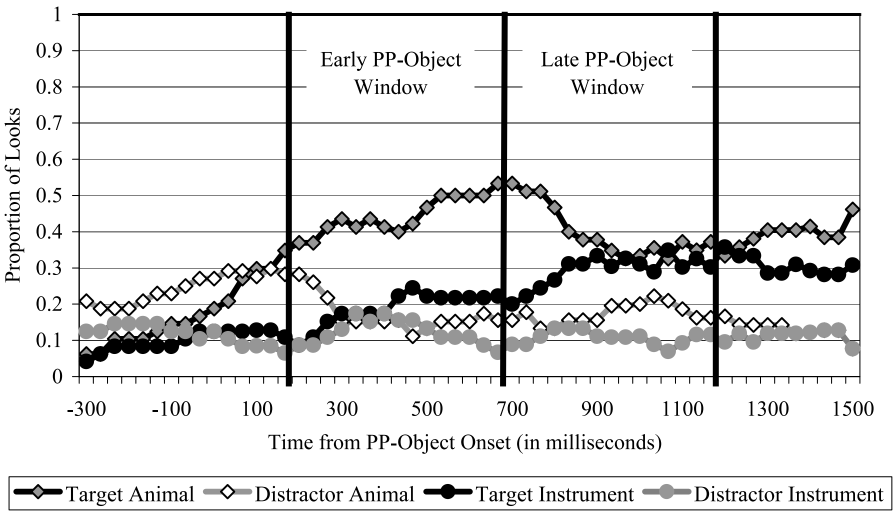
In the Pre-Noun time window there are apparent differences in the salience of the two animals in each condition. This results in a marginal effect of Referential Context on the Animal Difference Score during the Pre-Noun window ($F1(1,24)=2.57$, $p=.12$; $F2(1,18)=7.42, p<.05$). Subjects in the One-Referent Condition show a reliable preference for the Distractor Animal ($F1(1,24)=4.26$, $p<.05$; $F2(1,18)=7.32$, $p<.05$) while in the Two-Referent Condition neither animal is favored ($F1(1,24)<1, p>.5$; $F2(1,18)<1, p>.5$).[9] This initial bias and the lexical information appear to cancel each other out in the Noun time window, resulting in no effect of Referential Context on the Animal Difference Score ($F1(1,24)<1$, $p>.4$; $F2(1,18)<1, p>.4$).[10] Only in the Late Noun window do the children unequivocally demonstrate that they are using the identity of the noun to close in on the correct animal in the One-Referent Context but not the Two-Referent Context ($F1(1,24)=33.93, p<.001$; $F2(1,18)=12.67, p<.005$). While the children's response to the direct-object noun is a bit sluggish when compared with adult performance (Experiment 1), it is comparable to that of the five-year olds in the ambiguous conditions of TSHL.

[9] Note that this pattern is different from what we would expect if the experimenters were cueing the children in some way. Accurate cueing should result in an across the board preference for the Target Animal. The observed difference appears to be related to the types of animals that were used. The Distractor Animals in the One-Referent Condition were drawn from a larger set than the Target Animals because there were no constraints on the number of syllables in the word or its frequency in child-directed speech. They tended to be more exotic animals with high contrast patterns (e.g., zebras, tigers and giraffes). The Distractor Animals in the Two-Referent Condition were of the same type as the Target Animals (e.g., dogs, bears, frogs).

[10] The effect of Verb Bias on the Animal Difference Score was reliable in the participants analysis ($F1(2,24)=7.63, p>.2$; $F2(2,18)=1.39, p>.2$). Participants in the Modifier Condition appeared to prefer the Target Animal while those in other Conditions do not.

3.3.3.2. Prepositional object. The children's interpretation of the prepositional phrase was explored by examining fixations relative to the onset of the PP-Object (Figs. 7A–F). In the Instrument Conditions, the children tended to look at the Target Instrument throughout the trial. These looks have no clear onset or inflection point but reach a plateau about 800 ms after the PP-Object Onset. In contrast, participants in the Modifier Conditions rarely look at the Target Instrument, while the response to

A Instrument One-Referent Trials, Children



B Equi One-Referent Trials, Children

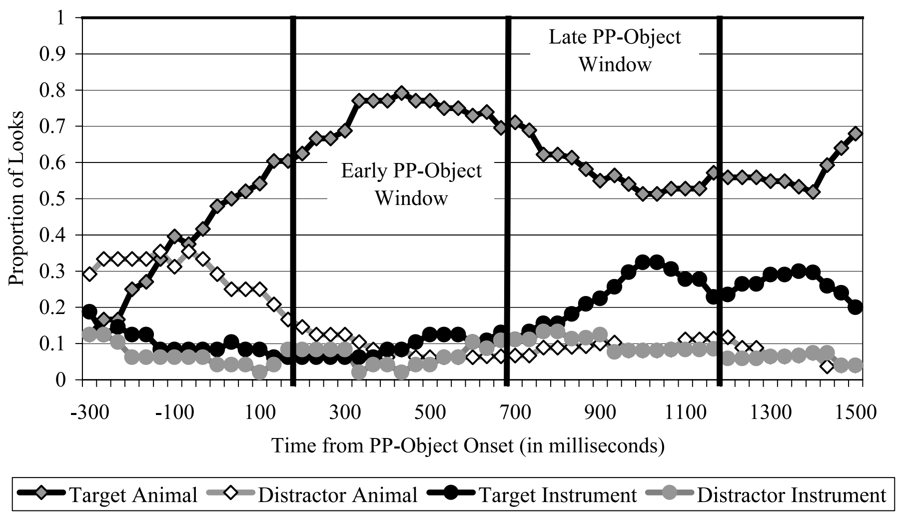
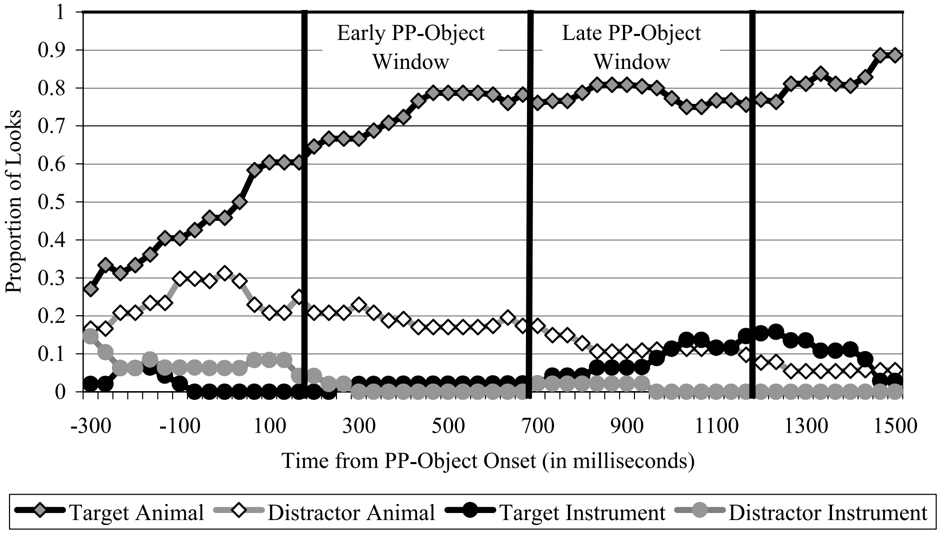


Fig. 7. Probability of fixating objects of each type over time (relative to onset of the prepositional object) for each condition, Experiment 2 (five-year olds).

the Equi Bias sentences is somewhere in between. There is also a clear difference in the looks to the Distractor and Target Animals across Verb-Bias conditions. By the time the PP-Object begins, the children in the One-Referent Conditions have begun using information from the direct-object noun to focus in on the Target Animal. In contrast, the participants in the Two-Referent Conditions are looking

C Modifier One-Referent Trials, Children



D Instrument Two-Referent Trials, Children

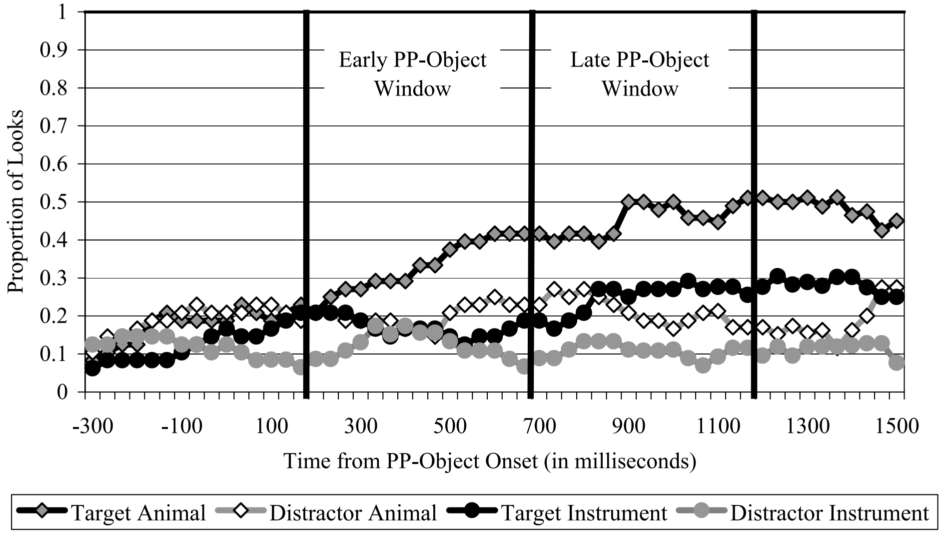
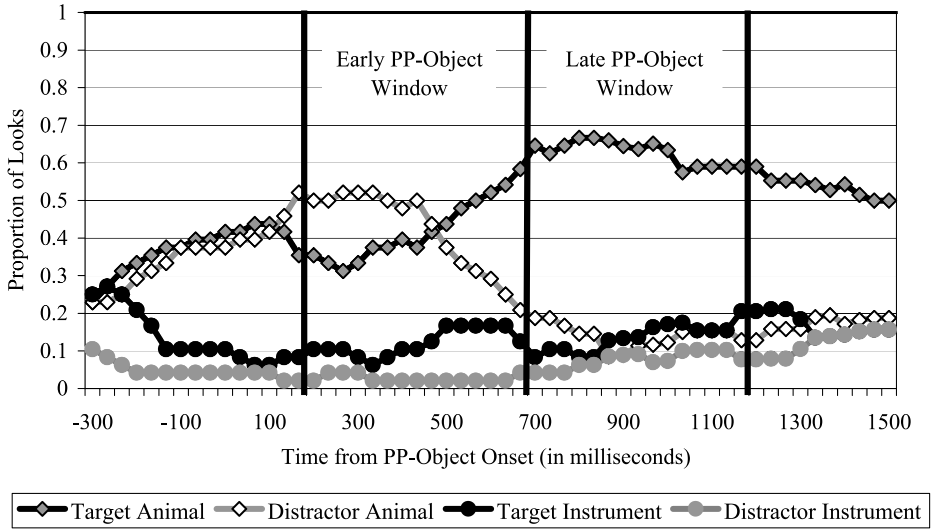


Fig. 7. (*continued*)

equally often at the two animals. About 400 ms after the PP-Object onset, participants in the Equi and Modifier Two-Referent Conditions abruptly abandon the Distractor Animal in favor of the Target Animal, demonstrating that they have used the prepositional phrase to restrict the reference of the noun. In contrast, the participants in the Instrument Two-Referent Condition continue to look at the Distractor Animal, and only gradually develop a bias for the Target Animal. In fact, some children in

E Equi Two-Referent Trials, Children



F Modifier Two-Referent Trials, Children

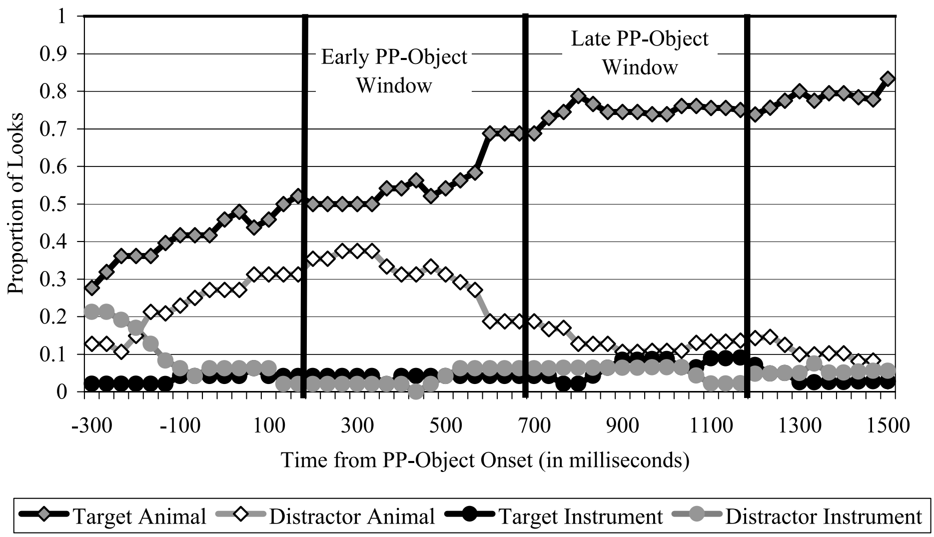


Fig. 7. (*continued*)

this condition asked which animal they should act on, indicating that they noticed the referential ambiguity but had failed to resolve it by analyzing the prepositional phrase as a modifier.

As in Experiment 1, we divided this period into two time windows for our analyses: an Early PP-Object window (200–667 ms after PP-Object Onset) and a Late PP-Object window (700–1167 ms). In the Early PP-Object window there was a reliable effect of Verb Bias on the proportion of looking time to the Target Instrument ($F1(2,24)=6.02$, $p < .01$; $F2(2,18)=6.77$, $p < .01$) but no effect of Referential Context ($F1(1,24) < 1$, $p > .8$; $F2(1,18) < 1$, $p > .6$) and no interaction between them ($F1(2,24) < 1$, $p > .8$; $F2(2,18) < 1$, $p > .6$). Participants in the Instrument and Equi Conditions often looked at the Target Instrument (18 and 11% of their looking time, respectively) while those in the Modifier Condition essentially ignored it (3%).

The effect of Verb Bias persisted in the Late PP-Object window ($F1(2,24)=10.33$, $p < .001$; $F2(2,18)=10.58$, $p < .001$; $M=27\%$, $M=20\%$, and $M=7\%$ in the Instrument, Equi, and Modifier Conditions, respectively). The effect of Referential Context was reliable in the subject analysis but not the item analysis ($F1(1,24)=4.95$, $p < .05$; $F2(1,18) < 1$, $p > .3$), suggesting that context may be influencing interpretation on a small number of items. This is curious because Referential Context seems to have no effect on the children's actions or early eye movements. While there was no interaction between Verb Bias and Referential Context ($F1(2,24)=1.52$, $p > .2$; $F2(2,18) < 1$, $p > .8$), the largest differences between One and Two-Referent Contexts appeared to be in the Equi Bias condition. As noted earlier, adults in reading studies typically only show sensitivity to referential manipulations in the absence of strong lexical biases. If children had some ability to use referential constraints but these constraints were lightly weighted or the ability itself was fragile, we might expect that it would only be detectable when other constraints on interpretation were weak. But the strength of the verb bias fails to predict the effect of Referential Context on Target Instrument looking time ($F(1,22) < 1$; $p > .8$).[11]

In the Two-Referent Conditions, participants can only identify a unique referent for the noun phrase if they interpret the prepositional phrase as a noun phrase modifier. Thus a preference for looking at the Target Animal in Two-Referent Condition provides evidence for a modifier analysis. In the Early PP-Object Window, participants in the Modifier Condition show a reliable preference for the Target Animal (64%) while those in the Equi and Instrument Conditions do not (48 and 60%, respectively). The effect of Verb Bias, however, is not reliable ($F1(2,24)=1.80$, $p > .1$; $F2(2,18) < 1$, $p > .4$). All groups show a preference for the Target Animal in the Late PP-Object Window but the size of the preference is smaller in the Instrument Condition (69%) than in the Modifier and Equi Conditions (85 and 81%), though the effect of verb bias is only marginal ($F1(2,24)=2.43$, $p=.11$; $F2(2,18)=3.53$, $p=.051$). Thus while the children clearly use verb bias to evaluate the instrument interpretation, the

[11] The dependant variable in this analysis was the difference in the proportion of looking time to Target Instrument between the reference conditions (Two Referent – One Referent). The independent variable was $|p - .5|$ where p is the proportion of instrument responses as measured in the verb norming study.

evidence that they use verb bias to evaluate the likelihood of a modifier analysis is less compelling.

3.3.4. Analyses of the first block

Young children are notorious for the tendency to perseverate, both in their actions and in their cognitive representations (Levin, Culhane, Hartmann, & Evankovich, 1991; Passler et al., 1985; Sato, Tanaka, Hosokawa, & Murai, 1998; Zelazo, Muller, Frye, & Marcovitch, 2003; Zelazo, Reznick, & Spinazzola, 1998). Given our experimental design, any perseveration across trials would have counteracted the effects of the Referential Context because this factor was manipulated within subjects. Likewise, effects of Verb Bias would have been magnified because verb type was manipulated between subjects. To put the two variables on more equal footing we conducted separate analyses of the actions and eye movements on the first block of trials. Most of these analyses simply confirm our earlier findings. The actions still showed a substantial effect of Verb Bias ($F1(2,24)=75.44$, $p<.001$; $F2(2,18)=3.77$, $p<.05$), no effect of Referential Context ($F1(1,24)<1, p>.4$; $F2(1,18)<1, p>.4$), and no interaction between Context and Bias ($F1(2,24)<1$, $p>.5$; $F2(2,18)<1$, $p>.5$). The pattern of findings for the Early PP-Object window is also unchanged: looks to the Target Instrument are affected by Verb Bias ($F1(2,24)=7.10$, $p<.005$; $F2(2,18)=5.95$, $p<.01$) but not Referential Context ($F1(1,24)<1$, $p>.5$; $F2(1,18)<1, p>.5$). The influence of Verb Bias persists through the Late PP-Object window ($F1(2,24)=12.91$, $p<.001$; $F2(2,18)=7.82$, $p<.005$). However, when we limit our analysis to the first block, the effect of Referential Context on gazes to the Target Instrument is substantial and reliable in the Late PP-Object window ($F1(1,24)=13.18$, $p<.001$; $F2(1,18)=7.58$, $p<.05$). Children spent 27% of their time looking at the Target Instrument in the 1-Referent Conditions but did so only 10% of the time in the 2-Referent Conditions. While there was no reliable interaction between Verb Bias and Referential Context ($F1(2,24)<1$, $p>.5$; $F2(2,18)<1$, $p>.5$), the difference between the context conditions appeared to be greatest for the Equi-Bias verbs ($M=34\%$ and $M=5\%$ for One-Referent and Two-Referent Conditions, respectively).

This effect of context on late eye movements has several possible interpretations. First, it could indicate that children have an emerging ability to use context to resolve attachment ambiguity. This knowledge would have to be strong enough to influence shifts in attention but not strong enough to have a reliable effect on the children's actions. This explanation runs into a minor hitch: the effect size for Referential Context in the Late PP-Object Window is, if anything, larger than that of Verb Bias. If these effects are straight-forward indices of the influence of the constraints on parsing at that point in time, then there is no obvious explanation for why Verb Bias has a substantial influence on the participants' actions while Referential Context does not. Furthermore, in the first block the effect of context on eye gaze is larger and more robust in the children than it is in the adults. Yet the adults' actions are influenced by context while the children's are not. One interpretation of this data pattern is that children commit to an analysis and plan their actions during the Early PP-Object Window, when lexical information is more potent than referential cues, and are then unable to revise these plans in light of late arriving information about referential context. Children's difficulty with reanalysis is supported by TSHL's finding that children failed to revise their initial interpretation of the local ambiguity in light of the second, disambiguating, prepositional phrase. This interpretation is also consistent with recent work by Crain and Meroni (2002) suggesting that children's inability to use referential context in the put-studies, may be linked to difficulties in inhibiting or revising action plans.

Alternately, one could argue that the effect of Referential Context in the Late PP-Object window reflects difficulty resolving the referential ambiguity that exists when the prepositional phrase is not interpreted as a modifier. The presence of referential ambiguity in the Two-Referent contexts could potentially result in repeated eye movements to the Target Animal and Distractor Animal. Since participants can gaze at only one object at a time, these looks could squeeze out fixations to the Target Instrument. An analysis of number of looks to the Target and Distractor Animals provides some support for this interpretation. In the first block of trials, there is an interaction between Referential Context and Verb Bias ($F1(2,24) = 7.64$, $p < .005$; $F2(2,18) = 7.38$, $p < .005$). In Instrument and Equi Bias conditions participants shift their gaze to the Target Animal and Distractor Animal more often in Two-Referent Contexts, suggesting that the referential ambiguity has not been resolved and opening up the possibility that looks to the animals could limit looking time to the Target Instrument. In contrast, in the Modifier Condition the number of looks to the two animals is actually lower in Two Referent Contexts than it is in One Referent Contexts.

3.4. Summary of Experiment 2

One implication of the second experiment is clear: children have formed parsing strategies that derive largely from their syntactic/semantic knowledge of individual verbs. These effects emerge shortly after the beginning of the ambiguous phrase, demonstrating that children like adults engage in incremental interpretation and make rapid use of lexical cues. The role of referential specificity in guiding children's parsing is less clear. Referential context played no role in the children's final interpretation, even for verbs that have no strong attachment preferences which might override the effects of context. However referential context clearly affected the allocation of attention toward the end of the trial. This demonstrates that young children are sensitive to the referential specificity of the direct-object noun. But it is less clear whether it demonstrates that this sensitivity plays a role in online interpretation of the ambiguous prepositional phrase.

4. General discussion

Several important observations emerge from this work, all of which, we will argue, point toward greater developmental continuity in parsing abilities than one might have inferred from earlier results (TSHL; Tanenhaus et al., 1995). First, lexical biases play an important role in both child and adult parsing preferences in a world-situated task. Children in our targeted age range (5 year olds) show an impressive sensitivity to the individual semantic/syntactic preferences of known verbs, and they can use this information to guide online parsing decisions. Likewise, adults in our task are sensitive to verb biases, a finding that is consistent with the literature on ambiguity resolution in reading (e.g., Britt, 1994; Garnsey et al., 1997; Trueswell et al., 1993). Right away, these observations rule out an account of the earlier TSHL child parsing results that appeals to the existence of a general parsing heuristic (e.g., minimal attachment) that diminishes with age or experience. Rather than being 'mini-minimal attachers,' child parsers appear to be 'little lexicalists,' relying on a database of lexico-syntactic and semantic knowledge. These effects are robust; since we began this work, two additional offline studies have demonstrated that young children use verb-specific information to resolve prepositional phrase attachment ambiguity (Hurewitz, Brown-Schmidt, Trueswell, & Gleitman, in progress; Kidd, 2003).[12]

Second, these experiments demonstrate that children's and adult's use of referential context may not be as categorically distinct as earlier work suggested. Previous studies suggested that there was a radical developmental discontinuity in the role of referential constraints on parsing. While five-year-old children showed no sensitivity to their referential scene manipulations this information was nearly deterministic for adults and even eight-year-olds (Hurewitz et al., 2000; Tanenhaus et al., 1995; Trueswell et al., 1999). For the adults in the present study, the mere presence of two potential referents for a definite NP (e.g., "the frog") was not sufficient to induce a restrictive modifier interpretation of an immediately following PP (e.g., "with the feather"), rather it only increased the odds of this interpretation. In children, we found that this manipulation had no effect on ultimate attachment preferences. However, the presence of two potential referents did reduce consideration of the instrument during the late time window, raising the possibility that the emergence of referential constraints may be more gradual than we originally believed.

The remainder of this discussion is organized as follows. We begin by exploring the implications these results have for our understanding of parsing development and the architecture of the language comprehension system. In doing so, we defend a particular account of our data, which assumes that multiple evidential sources are at work even during the earliest stages of the development of parsing development. Next, with this account in mind, we examine apparent contradictions in the current literature and how they might be reconciled. Finally, we turn to a specific but urgent matter for our theorizing, namely, our understanding of what sort of information underlies effects of verb-bias in online interpretation, and how thinking on this matter connects to our understanding of verb learning.

[12] Kidd (2003) also explored children's online processing of ambiguous prepositional phrase attachments with an auditory moving window technique. He found that 7- and 9-year-old children make rapid use of verb information but he found no effects of verb type in 5-year-olds. We attribute this difference in findings to differences in the sensitivity of the two tasks. The moving window technique requires young children to make novel overt responses, looking paradigms do not. Consistent with this, the decision times for the 5-year-olds in Kidd's study were more than twice as long as those for older children.

4.1. Implications for theories of parsing and their development

Thus far, our experimental work (TSHL and the results above) has demonstrated a near-exclusive role of lexical evidence in informing children's parsing decisions. As we see it, at least two viable and distinguishable developmental accounts exist for these findings.

The modular/single-cue hypothesis. First, it is possible that the observed differences between children and adults reflect changes or expansions in processing ability. For instance, a limited, single-cue, or encapsulated parsing system might become more interactive as processing ability grows with age. Indeed, several current modular theories of parsing grant an architectural privilege to lexical cues. For example, Boland and colleagues have argued that the lexicon alone proposes syntactic and semantic structures while other cues are used at a later stage to select between the proposed analyses (Boland & Blodgett, 2001; Boland & Cutler, 1996). The 5-year-old preference to use lexical cues might very well reflect an encapsulated, lexicalist parsing system 'in the raw' which becomes increasingly interactive as processing power increases. Similarly, it is possible that children control multiple cues, but can only use a single cue at a time due to an early inability to coordinate multiple information sources. This later account provides a particularly satisfying explanation for the fickle effects of referential context: children's early eye movements demonstrate their ability to use lexical biases to guide interpretation. The later eye movements also demonstrate their sensitivity to referential constraints. But their failure to use referential context to guide their actions could indicate that they are unable to coordinate these information sources in a timely and stable fashion.

Multiple cue system from the start. In contrast, our own account assumes a probabilistic multiple-cue comprehension system from the start, with the ordering of cue use over development reflecting each cue's relative reliability. As various evidential databases are built and found to be informative, they come into use in the comprehension process. Under this account, the child parsing system shows an earlier reliance on lexical sources (above and beyond other relevant information sources such as referential sources) precisely because of the high-degree of reliability of lexical information for syntactic structuring. By age 5, the child has learned a great deal about the possible and probable syntactic/semantic environments in which particular verbs appear—especially for the common verbs we have tested thus far. Other sources, such as referential scene constraints on syntax, might simply take longer to acquire and use because they are less reliable over the input database as a whole, and arguably more difficult to track in particular instances of use than lexico-syntactic contingencies. Thus, the child parsing system can in principle use multiple evidential sources to guide a syntactic choice. But in practice the usefulness of particular sources of evidence is a matter of discovery, and hence changes with experience. On this account the fragile effects of referential constraints on late eye movements would be interpreted as signs of an emerging sensitivity to the predictive power of the referential scene.

This developmental account, like all constraint-based lexicalist theories, is in many ways reminiscent of the Bates and MacWhinney (1987) "Competition Model." For instance, both theories assume constraint-satisfaction mechanisms for language discovery and use, and therefore emphasize information reliability when accounting for developmental patterns. However, a crucial difference between these theories is that the constraint-based lexicalist models assume a central role for detailed linguistic representations in language use along multiple, partially independent dimensions (phonology, syntax, and semantics). Representational modularity in the presence of interactive processing, a key assumption of these models, is crucial for accounting for a broader range of phenomena (Trueswell & Tanenhaus, 1994; see also Jackendoff, 2002).

If this multiple cue account is correct, there seem to be several demands of proof that we have not yet satisfied. First, we must be able to provide reasonable arguments (and preferably experimental evidence) that lexical predictors of PP-attachment decisions are more reliable and easier to track than referential scene predictors. Such observations would be consistent with an evidential discovery account of the developmental asymmetry reported here. Second, we should expect to find that children in the targeted age range (five years) can indeed simultaneously use multiple sources of evidence to resolve syntactic ambiguity, if the evidential sources in question are highly predictive of the intended structure. We believe we can make tentative claims on all these grounds, sketched below, leading us to conclude that a multiple constraint parsing system is at work over the entire course of language development.

4.1.1. Lexical and referential predictors of restrictive modifiers

Do lexical predictors of structure precede referential predictors because of their relative reliability? First, there should be little doubt in the minds of most psycholinguists that lexical cues are highly predictive of local structure if tracked in a statistical fashion. Work in computational linguistics on statistical natural language processing systems provides perhaps the best existing proof that such cues when tracked constrain structure quite well, so well in fact that the most impressive advances in parsing systems have been achieved via the tracking of lexico-syntactic contingencies and little else, achieving high parsing precision within context-independent systems (e.g., Collins, 1996, 1997; Marcus, 1994). Indeed, Collins and Brooks (1995) specifically looked at PP-attachment ambiguity and the contribution lexical statistics can make to attachment choice. This work showed that relatively high accuracy (about 85% correct) can be obtained by tracking the attachment biases of verbs, prepositions and nouns.

There is also good evidence that very young children track the distribution of a variety of linguistic elements, even in the absence of meaning. After just 2–3 min exposure to syllable sequences generated by an artificial language, infants can detect both structural patterns and statistical tendencies (e.g., Gomez & Gerken, 1999; Hudson & Newport, 2003; Marcus, 2000; Saffran, Newport, & Aslin, 1996). It seems plausible therefore, especially in the light of adult processing findings above, that statistical tracking is accomplished at multiple levels of utterance representation, up to and including syntactic elements, implicating a deep continuity between learning and comprehension processes over the course of the language system's development. Moreover, experimental observations attest to the reliability of the particular knowledge of interest here, verb-specific preferences. Indeed, the range of complement privileges for common verbs is just about fully attested within single hour-long mother–infant conversations (Geyer, 1997; Lederer, Gleitman, & Gleitman, 1995; Li, 1994). Maternal usage "errors" in this regard (e.g., saying "Don't come your toys into the living room" or "Put your apple.") are rare to nonexistent. Thus the information is both bountiful and reliable.

What about the referential scene as a predictor of structure? Can a child listener deduce from scene information alone the need for referential specificity? That is, can the child look out into the visual world and easily anticipate a speaker's need to utter *the little star* (and not *the star*), *the toy closest to you* (and not *the toy*), or *the frog on the napkin* (and not *the frog*)? The assumption thus far in the literature has been that a speaker's use of a bare definite NP (*the frog*) in the presence of multiple entities of that sort (multiple frogs) ought to be a near-perfect predictor of the need for further linguistic specification, i.e., a post-NP restrictive modifier (e.g., *the frog you caught yesterday*). But is this the case? A recent adult-to-adult referential communication study suggests that there would be only sporadic evidence for this scene-contingent inference (Brown-Schmidt, Campana, & Tanenhaus, 2002, in press). It was observed that adults do not utter restrictive modifiers every time there is more than one potential referent. In particular, nearly half of all definite NPs uttered (48%) did not have a unique referent in the scene (e.g., "Okay, pick up the square" might be uttered in the presence of multiple squares). However, conversants' eye movements, actions and vocal responses all showed that they routinely achieved referential success under these conditions. Obviously, this success is not evidence for psychic abilities on the part of the conversants. Rather, success occurred because the shape of the discourse and the goals of the task had narrowed the field of possible referents down to one (e.g., only one of the squares was currently a plausible referent). Definite NPs containing restrictive modifiers were uttered only when more than one potential referent was currently under discussion.

These findings are consistent with an extensive literature on the pragmatics of definite reference (e.g., Hawkins, 1978; Prince, 1981, 1992; Stone & Webber, 1998). Although adults understand that a definite NP almost always requires a unique (and agreed-upon) referent, disambiguation of the referent need not be accomplished linguistically; the local discourse and/or the current goals of the interlocutors often do the job instead. For instance, "Pass the salt" in the presence of multiple salt shakers does not usually generate a request for further specification (e.g., Lyons, 1999).[13]

These facts have important implications for understanding how children might discover that modification is required to determine the referent of a definite NP. In the presence of multiple objects of the same kind (three books), children will often hear bare singular definite NPs (e.g., *the book*) which for most conversants will uniquely determine the referent. What this means is that the discourse and its goals

[13] It is of course even more complicated this. Bridging inferences are often required to establish reference of a bare definite NP (consider, "I was on a bus the other day and we got into an accident because the/a driver was drunk," Prince, 1981). It has even been noted that agreement between interlocutors on a referent can be optional, leaving the exact referent unspecified: "He reads the newspaper every day"; "He always misses the bus" (for a full discussion with many examples, Birner & Ward, 1994).

relative to the scene, rather than the scene itself, ought to be a far better predictor of restrictive modifier use for the younger child. In prior referential-scene parsing studies (Spivey et al., 2002; TSHL), adults and older children might have used scene evidence as *a proxy for the discourse*, because such a discourse was absent. Or perhaps humans develop an understanding of how scene cues partially predict structure but this understanding emerges gradually over developmental time, given the sporadic nature of the cue. Indeed, these conclusions are consistent with what is known about children's understanding of definite reference (e.g., Karmiloff-Smith, 1979; Maratsos, 1976). This literature shows that, like adults, young children (3–7 years) typically use definite NPs when there is a unique or especially salient referent in the situation and use indefinite NPs when they have no unique referent in mind. But unlike adults, they will sometimes use definite NPs to refer to entities that cannot be easily identified by their interlocutor. We interpret this pattern as evidence that children know the meaning of the definite determiner but can have difficulty interpreting and using definite NPs precisely because they have trouble computing the referential domain intended by their interlocutor.[14]

We therefore have good reason to believe that lexical evidence for resolving this ambiguity can be discovered more easily by the child than the relevant referential scene information, especially when one considers how and when bare definite NPs get used in naturally occurring settings (Brown-Schmidt et al., 2002; Lyons, 1999; Prince, 1981, 1992).

4.1.2. The use of multiple constraints in preschool parsing

The multiple cue account which we are defending here also predicts that children in this age range ought to use multiple sources of information to guide parsing commitments, if those sources are good predictors of structure. Indeed, some recent evidence suggests that this is the case. First, given the discussion above, one might expect that five-year-olds will be sensitive to referential contextual manipulations if they are supported by the structure of a preceding discourse, specifically a discourse that points out the need to contrast two potential referents in the scene. Recent work in the lab of the second author has begun to explore this issue (Hurewitz, Brown-Schmidt, Trueswell & Gleitman, in progress). Here, a preceding discourse, conducted by two conversing puppets, establishes the goal to contrast multiple referents in the scene prior to hearing an ambiguous PP. This was done by having the puppets produce questions that either supported a modifier interpretation of the ambiguous answer (Question: *Which cat did the turtle tickle?* Answer: *I know, the turtle tickled the cat on the fence*) or did not (Question: *Can you tell me something about the story?*

[14] Wexler (2003) pointed out the connection between definite reference acquisition and the child parsing phenomenon described here. In contrast to the present account, Wexler (2003) proposes that five-year-olds do not yet know the meaning of the word "the" (specifically, that it applies maximally to the speaker's intended referential domain). While there are findings consistent with this account, a more detailed consideration of the definite reference acquisition literature suggests that children do not lack this knowledge but simply have difficulty coordinating their referential domain with that of their conversational partner (see Trueswell, Papafragou, & Choi, in press, & references therein).

Answer: *I know, the turtle tickled the cat on the fence.*) Five-year-olds evaluated the truth of these answers, and corrected them when they were deemed erroneous; their corrections revealed their interpretation of the ambiguous phrase. The results were clear: contrastive questions greatly increased the probability of children interpreting the ambiguous phrase "on the fence" as an NP modifier as compared to non-contrastive questions. Moreover, this contextual effect was found to combine with a second evidential source we manipulated: verb bias. Hence, the pattern taken together suggests that children do not have a ban on contextual/referential factors when making parsing decisions, rather they need further support from the conversation to help shape the need to specify such a referent. However, because eye-movement data were not collected in this paradigm, these results cannot tell us when (in chronometric time) the discourse influenced ambiguity resolution.

Evidence for the use of multiple cues in children's online parsing comes from the first author's ongoing work on the role of prosody and verb bias in ambiguity resolution (Snedeker & Yuan, 2003). These studies used the same ambiguous *with*-attachments used in the current study. We recorded two versions of each sentence, one with instrument prosody (an intonational phrase break after noun) and one with modifier prosody (an intonational phrase break after the verb). When verb bias was neutralized, both adults and children (4–6) were able to rapidly use prosody to interpret the ambiguous utterance. When biased verbs were used, prosody had a smaller but still reliable effect on online interpretation and actions. Critically, the effects of prosody emerged in the same early time window where the effects of verb bias appear in the present experiment. There was no evidence of an initial stage in which only lexical information was used.

Curiously, in prior studies we found that adult speakers frequently fail to provide prosodic cues to prepositional phrase attachment, producing both NP and VP attachment structures with neutral phrasing (Snedeker & Trueswell, 2003). If this is the sort of input children get, why then do children have early access to this parsing constraint? We see several reasons why prosodic cues might become available earlier in ontogenetic time than the referential scene information explored in the current experiments. First, while prosodic cues are often absent, when they are present they are highly reliable, making them a useful predictor of structure (Snedeker & Trueswell, 2003). Second, while referential context exerts a top-down influence on syntactic parsing, prosodic cues are arguably a bottom-up constraint. By this we merely mean that the meaning of the sentence, and thus the referents of the noun phrases within it, depends upon its syntactic structure. In contrast, from the perspective of the listener the prosodic structure may constrain syntactic structure but it is not dependent upon it. This asymmetry could influence either the salience of this representation as a potential information source for online interpretation or the parser's ability to gain timely access to this information during comprehension. Finally, the two representational systems may develop at different rates, which in turn might influence the age at which they become integrated with online parsing. Many have argued that five-year-olds are still struggling to understand the referential demands and goals of various communication situations (Glucksberg & Krauss, 1967; Robinson & Robinson, 1982). In contrast, even young infants show a well-developed sensitivity to the prosodic structure of their language. Newborns discriminate between languages on the basis of their rhythmic properties (Jusczyk, Friederici, Wessels, Svenkerud, & Jusczyk, 1993; Mehler et al., 1988; Nazzi, Bertoncini, & Mehler, 1998; Nazzi, Jusczyk, & Johnson, 2000). Half a year later, infants rely heavily on prosodic cues to segment the speech stream into words (Johnson & Jusczyk, 2001; Jusczyk, Culter, & Redanz, 1993; Morgan & Saffran, 1995; Morgan, 1996). Critically, by nine months of age, infants are sensitive to the coalition of cues that mark the prosodic boundaries between groups of words (Hirsh-Pasek et al., 1987; Jusczyk et al., 1992).

4.2. Comparing the put-studies and the with-studies: When does context overwhelm lexical biases

The data from our adult study also indicate a greater continuity between the reading and listening than previous results might suggest. Like Britt (1994), we observe contributions of both lexical and referential sources to online parsing commitments. However, we observe these patterns in auditory language processing, in the presence of potent visual cues to the referential context. Why then did the previous *put*-studies show no consideration of the VP-attachment in the two referent contexts, despite strong lexical cues supporting this analysis? We see three plausible explanations.

First, additional sources of information present in the *put*-sentences may have further reduced consideration of VP-attachment. For example, the second (disambiguating) prepositional phrase (*Put the frog on the napkin* *into the box*) is uttered at the very moment that eye movements should show consideration of VP-attachment. This rapid disambiguation may have served as a post-ambiguity cue that further reduced any consideration of this parse. In addition, the resolution of the ambiguity in earlier trials may have influenced subjects' interpretation of the temporarily ambiguous phrase in later trials. Finally, prosodic cues may have provided evidence during the first PP that a second potential argument was forthcoming. Prosody was held constant across conditions in the *put*-studies, but the neutral prosody that the experimenters aimed for may have revealed that the utterance would continue, and hence supported the NP-attachment interpretation. Our own studies of prosody, which used a similar task and measure, suggest that differences of this kind can influence parsing as rapidly as lexical information (Snedeker & Trueswell, 2003). In the *put*-studies, all of these additional cues support a modifier analysis. It is possible that these cues in combination with the context manipulation were able to completely eliminate the VP-attachment analysis in the two-referent condition. However, they may not have been strong enough to overwhelm the combination of lexical and contextual cues favoring the VP-attachment in the one-referent condition.

A more intriguing explanation of the differences between the current study and the *put*-studies comes from considering the kinds of thematic/semantic contrasts that exist for these two different PP-attachment ambiguities.[15] In the current study, the

[15] We thank Michael Tanenhaus for suggesting this explanation as well as the one in the following paragraph.

competing interpretations of the syntactic ambiguity do not overlap nearly as much as they do for the *put*-sentences. That is, the interpretation of *with the feather* as an Instrument carries a very different meaning than the interpretation of this same phrase as a Noun Phrase Attribute. Under one analysis, the ambiguous phrase refers to a means by which one can cause the Theme to undergo the action (e.g., use the stick to cause the frog to have been tickled), whereas the other analysis refers to a current state of the Theme (e.g., the entity that is the Theme is holding said entity). In contrast, the interpretation of *on the napkin* as a Goal/Destination shares numerous properties with the competing Noun Phrase (Locative) Attribute interpretation, because under both interpretations a spatial relationship must hold between the Theme (the frog) and the object of the prepositional phrase. The phrase either specifies the *new location* of the Theme (in the VP-attachment interpretation) or the *current location* of the Theme (in the NP-attachment interpretation). Thus, in the case of *put*-sentences fewer aspects of the semantic interpretation may be in question, and far less 'work' may be required to change one's interpretation (see also Ferreira, Christianson, & Hollingworth, 2001, for discussion of rescinding interpretations in garden-path sentences). Thus, two-referent scenes may be more likely to 'tempt' listeners away from VP-attachment toward NP-attachment for *put-on*-sentences than for *tickle-with*-sentences precisely because much more is shared between the two interpretations in the former case than in the latter.[16]

Finally, differences between the actions that are required in the two studies may have affected the impact of the referential ambiguity on interpretation. In the *put*-studies subjects almost always pick up the animal specified by the direct object and then move it to the destination. Disambiguiting the referent of the direct object noun phrase is required to plan the initial movement. In contrast, in the *with*-studies subjects who perform an instrument action typically pick up the instrument first and move it to the animal. They must determine whether the prepositional phrase denotes an instrument before beginning the action, but they are not forced to resolve the referential ambiguity until later. These different task demands could potentially influence both the relative strength of the two analyses and the time course of cue activation. Clearly, further experimentation would be needed to elucidate these subtle differences. But these unresolved questions do not undermine the primary finding of our experiment with adults: lexical biases play a substantial role in uncovering the syntactic structure of spoken utterances.

4.3. What is a lexical bias?

In these experiments we manipulated the lexical bias of the verbs in the ambiguous instructions. We followed a common practice from the adult sentence processing literature and measured bias by asking subjects to complete ambiguous sentence fragments (see Appendix B). To be precise, verb bias was operationally defined as

[16] Note that nothing in this explanation is particular to auditorily presented sentences or the visually presented contexts. This raises the question of why reading studies using highly biased locative verbs have failed to find effects of the number of referents (see e.g., Britt, 1994).

the relative frequency of instrument and modifier completions for each verb. This task was validated by comparing the verb-bias measure with actual frequencies of instrument and modifier *with*-phrases in a corpus of child directed speech (see Appendix A). By defining and validating the measure in this way, we have perhaps implied that the information guiding the children's responses is also distributional in nature. This conclusion is not warranted. There are at least three types of information—plausibility, lexical semantics, and lexical distribution—which could account for the verb bias effect (and these are by no means mutually exclusive). These factors are undoubtedly confounded in this study, just as they are confounded in the real world. Particular verbs are frequently heard with instrument phrases because they belong to particular semantic categories that describe events in which instrument use is plausible.

We attempted to avoid effects of plausibility that were unrelated to verb bias by norming and controlling the fit between the verb and the target instrument (Appendix C) and by holding the agent and the affected entities constant. But we did not control for the global plausibility of the event. Since the modifier actions are rarely carried out with instruments, we assume that global plausibility and verb bias were confounded. However, we suspect that event plausibility was not the primary factor behind our findings. In the filler trials children were asked to enact many implausible events (e.g., make elephants climb bottles or frogs swim in buckets), ensuring that real world plausibility was a poor predictor of sentence meaning. Because we selected instruments that were only moderately plausible for the verb in question, it is unlikely that our subjects had experience with the exact events denoted by the instrument interpretations of any of the target sentences. Thus, any effect of plausibility would have to be based on some coarser level of generalization. If these generalizations are at roughly the same grain as verb concepts, then this explanation may be indistinguishable from one based on lexical semantics. In adult reading studies the roles of plausibility and lexical bias have been explored with factorial manipulations of the distributional profile of the verb and the fit between the verb and the potential argument (Garnsey et al., 1997). Both factors influence early reading times. A similar approach could be equally informative in the world-situated comprehension studies.

Lexical distribution and lexical semantics appear to be more plausible candidates. These factors are intrinsically related. If the meaning of a verb logically requires particular numbers and types of semantic roles, then that verb will appear in structures that meet these obligations (see e.g. Baker, 2001; Jackendoff, 2002). Yet it is an empirical possibility that parsing relies on only one of these two sources of information. Unsurprisingly, meaning and distribution are tidily confounded in both our materials (Appendix B, Table 5) and in the speech that children hear and produce (Appendix A, Tables 3 and 4). Thus our subjects could be basing their initial interpretation on the semantic subclass that the target verb belongs to (contact verbs frequently appear with instruments, verbs of perception rarely do) or the syntactic environments in which it appears (*hit* often occurs with VP-attached *with*-phrases, *choose* does not). In either case, participants could be relying on categorical information (membership in a class, possession of a feature) or fine grained information about frequency of occurrence. Attempting to disentangle these factors with known words appears futile: examples in which meaning and structure are unconfounded are few and far between (the dative alternation being the parade case) and sets of words in which these factors have been teased apart inevitably vary along some third dimension (e.g., for datives frequency, syllable number, and phonological form). Here is another case where the cross-fertilization of language acquisition and language processing could be productive. While we cannot orthogonally manipulate what subjects know about real words, we can systematically vary what we teach them about novel words. By providing distinct distributional profiles for verbs with the same meaning, or identical distributional profiles for verbs with different meanings, we may be able to find out more about the knowledge that underlies adults' and children's online parsing.

How then do children discover lexical biases? At first glance the answer might appear to be transparently linked to the nature of the representation underlying the bias: if it is semantic, the bias might be learned as a part of word learning, if it's distributional then the bias would have to be acquired as a post hoc supplement to guide online interpretation. But this simple equivalence depends on the assumption that word learning is prior to and independent of the acquisition of syntax and the development of online comprehension. This may be a viable account of the early acquisition of nouns; the situation in which a noun is used is often sufficient for identifying the meaning of the word (Gillette, Gleitman, Gleitman, & Lederer, 1999; Snedeker, Gleitman, & Brent, 1999). However, nonlinguistic context provides misleading and incomplete information about the meanings of many verbs (Gillette et al., 1999; Snedeker, 2000). Gleitman and her colleagues have argued that efficient verb learning requires a distributional analysis of the syntactic contexts in which a verb occurs (Fisher et al., 1991; Gleitman, 1990; Snedeker & Gleitman, 2004). Toddlers, preschoolers and adults can all use structural context to interpret the meaning of a novel verb (Fisher, Hall, Rakowitz, & Gleitman, 1994; Naigles, 1990). Thus the representation of lexical distribution and the creation of language specific semantic–syntactic subclasses may well be the by-product of the word learning machinery.

We therefore conclude by suggesting a deep connection between word learning strategies and ambiguity resolutions strategies, at all stages of development. In both cases, various sources of evidence can be brought to bear on the problem, including the physical context, the ongoing discourse and the distributional profile of individual words. In both cases, the learner's reliance on these cues will change over development, as he/she develops the relevant knowledge base and the ability to employ it efficiently. In both cases, while multiple sources of evidence are potentially available only some of these sources are likely to be informative for a particular item or class of items. For instance, while it is relatively easy to inspect the world and notice whether a cat is present when the word *cat* is uttered, it is considerably more difficult to observe when *thinking* is taking place. Thus the observed scene is more informative for concrete words, while syntactic distribution is particularly useful for identifying more abstract words (Gillette et al., 1999; Snedeker & Gleitman, 2004).

Seen in this light, the problem faced by a child or adult who is attempting to resolve the meaning of a prepositional phrase looks similar to the problem faced by the language user who is attempting to discover the intended meaning of an abstract word. Clearly the intended meaning of the ambiguous phrase is often related to the current situation: a postnominal modifier is used because there is need to contrast the properties of one object with another (just as *thinking* is typically uttered when thinking is relevant in some way to the activity at hand). However, it is not easy to look out into the world and determine when such meanings arise, because relationships like these are typically in the eye of the beholder, who is in this case the speaker. In contrast, lexical and distributional facts in the sentence itself are there for the taking and constrain meaning in the relevant ways. As a result, lexical information ought to, and indeed does, exert an early influence on the child parsing machinery.

Appendix A. A corpus analysis of with-phrases in children's speech and child-directed speech

To learn more about the *with*-phrases in children's input and production we conducted an analysis of the Adam, Eve, and Sarah transcripts (Brown, 1973) from the CHILDES corpus (MacWhinney, 2000). Our goals were: (1) to determine the relative frequency of NP and VP attached *with*-phrases in speech to young children; (2) determine the relative frequency of different semantic types of *with*-phrases in this speech; (3) determine whether children spontaneously produce both instrument and modifier *with*-phrases and (4) identify potential verbs for use in the subsequent experiments. Brown's transcripts were selected because they capture naturalistic language use in a wide variety of settings and are extensive enough to provide detailed information about the range of contexts in which a given word appears. The children ranged in age from 1;6 to 5;1 and were taped approximately every week (Sarah) or every other week (Adam and Eve). This corpus included about 365,000 words of adult speech (primarily child-directed) and 295,000 words of children's speech.

A.1. Methods

All utterances containing "with" were extracted from these transcripts ($n = 2536$) and coded along six dimensions. Each utterance was coded once by a research assistant and then checked by the first author. The coding scheme is given in Table 1. The preposition "with" has many related and overlapping senses. We based our coding categories for the semantic role of the *with*-phrase on an early version of the classification scheme of McKercher (2001), which in turn drew on the work of Nilsen (1973) and Schlesinger (1995). Preliminary use of this system revealed the need for two additional categories: (1) Nominal Role: NP-attached *with*-phrases which are not modifiers but instead get semantic roles from the head noun (most of tokens in this category were instances of idiomatic phrases like "What's the matter with you?") and (2) Task: VP-attached *with*-phrases that describe tasks or events and are not included in any other category ("She's done with her exercise class").

For many of the analyses we wanted to examine both the syntactic and semantic coding categories by first dividing the items by the type of attachment and then breaking them down into larger semantic subclasses. The description of this system is given in Table 2.

A.2. Results

The children produced both NP-attached and VP-attached *with*-phrases in proportions roughly similar to those of the adults around them. All three children used *with*-phrases productively before age 3 (1;10 for Eve, 2;7 for Adam; 2;9 for Sarah). For all three children the first several uses of *with* were verb-phrase

Table 1
Coding scheme for the corpus analysis

Coding dimension	Categories	Explanation
(A) Speaker	Parent	Child's mother or father only
	Child	Target child
	Other adult	Experimenter or friend of family
(B) Verb preceding *with*	e.g., "see," "eat," "go"	Main verb, citation form
	Blank	Used if no verb in utterance
(C) Syntactic attachment of *with*-phrase	VP-attached	
	NP-attached	
	Other attachment	Higher attachments or unattached phrase
	Ambiguous/uncodeable	
(D) Semantic role of *with*-phrase	Modifier	"The one with the pointy ears"
	Nominal role	"The problem with elevators is…"
	Instrument	"Tickle him with a feather"
	Ancillary instrument	"Paint the ceiling with a ladder"
	Proper part	"He hit the tree with his fender"
	Undergoer	"Don't play with these blocks"
	Objective	"The garden is swarming with bees"
	Material	"We made the pastry with lard"
	Accompaniment	"He danced with the bear"
	Cause	"I was delighted with the present"
	Location	"I left the hat with the gloves"
	Manner	"Say it with more emotion"
	Task	"When you're done with your homework"
	VP-unknown	"What will you do with it?"
	Condition	"With Burt gone, the meeting went quickly"
	Ambiguous/uncodeable	
(E) Is there a noun phrase in between the verb and the *with*-phrase?	Yes, with a single noun	"Cover the doll with the blanket"
	Yes, a noun inside a PP	"He went to the store with me"
	Yes, other	"I read the story about the dog with Mommy"
	No	"Don't talk with your mouth full"
(F) Can the nearest noun be modified?	No	Pronouns, proper names, "home" and "school"
	Yes	All other nouns

attachments (primarily instruments, undergoers and co-agents).[17] But within a few months all of the children had produced *with*-modifiers as well. Clearly, by five years of age children have had extensive practice producing both NP and VP-attached *with*-phrases.

Overall, the syntactic attachment and semantic content of the children's *with*-phrases closely paralleled that of their parents (see Figs. 8A and B). For both groups VP attachments were far more common than NP

[17] Because *with*-modifiers are relatively infrequent, we would expect that they would be less likely to appear in any given speech sample even if the child was already using this construction. For each child we tested this possibility by counting the number of *with*-phrases produced before the first modifier and then calculating the probability of getting this distribution given the proportion of attribute phrases in all of the child's transcripts. This hypothesis could only be rejected for Adam who produced over 50 *with*-phrases before producing an attribute ($p < .05$). For the other two children there is no strong evidence for a delay.

Table 2
Syntactic categories and broad semantic categories used in the corpus analysis

Syntactic category	Broad semantic category	Semantic coding category
NP-attached	Modifier	Modifier
	Other	Nominal role
VP-attached	Instrument	Instrument
	Instrument-like	Ancillary instrument
		Proper part
		Undergoer
		Objective
	Other VP	Material
		Accompaniment
		Cause
		Location
		Manner
		Task
		VP-unknown
Other	Other	Condition
		Ambiguous/uncodeable
		Other roles with high attachment,
		No attachment

attachments (86 vs. 8% for parents and 88 vs. 8%, for the children). Instrument phrases were also more frequent than modifier phrases though the disparity was considerably smaller (11 vs. 5% for parents and 12 vs. 7%, for the children). Most of these *with*-phrases were used in sentence contexts where NP attachment is extremely unlikely or impossible (e.g., after intransitive verbs or at the beginning of an utterance). To examine the distribution of *with*-phrases in potentially ambiguous contexts, we separated out a set of utterances that was structurally similar to the ones that we would be using in our online experiments. This set included all utterances that contained a "...V...N...with..." sequence where the intervening noun was potentially modifiable (all pronouns, proper names and unique place names were excluded). Within this subset of utterances (see Figs. 9A and B) the bias for VP attachments is greatly reduced (64 vs. 33% for parents and 66 vs. 28% for the children) and modifier interpretations were as common as instruments for parents (18 vs. 18%) and even more common than instrument interpretations for children (23 vs. 13%). Thus, while *with* has a strong global bias toward a VP-attachment, instrument and modifier interpretations are equally likely when the *with*-phrase follows a postverbal noun.

The final goal of the corpus analysis was to identify potential target verbs for the sentence norming study. Unfortunately, few verbs appeared often enough to allow us to get a reasonable estimate of attachment bias and only a handful of those verbs met the other criteria for use in the experiment. (See Appendix B for information on the selection of the experimental verbs.) However, the corpus was large enough to provide an insight into the types of verbs that are likely to appear with instrument and modifier phrases. To explore the interaction between verb meaning and the meaning and attachment of *with*-phrases, we divided the verbs into three broad semantic classes: (1) perceptual and psychological verbs (psych verbs); (2) light verbs and (3) action verbs.[18] Tables 3 and 4 list the number of modifier and instrument phrases for verbs in each of these three categories.

[18] The verbs were coded into categories by the first author. The light verb category included the verbs: *do*, *get*, *make*, *take* and *be*. The action verb category included all verbs that did not fit into either of the other categories. Most of these verbs described actions (e.g., *chew*, *dance*, *go*) but several did not (*end up*, *happen*, *rhyme*).

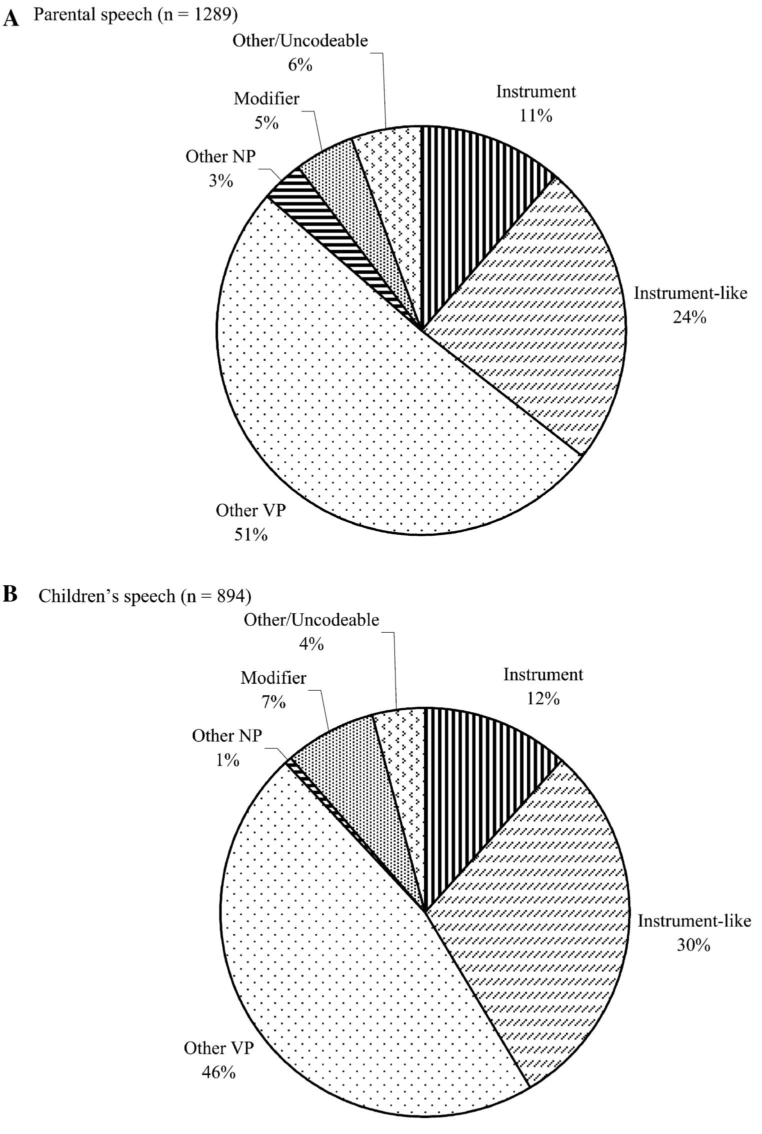


Fig. 8. All of the *with*-phrases from the corpus divided into broad semantic categories.

Our analysis of the verb classes focused in on two distinctions. First, we contrasted action and light verbs with psych verbs. In an analysis of text directed at adults, Spivey-Knowlton and Sedivy (1995) found that *with*-phrases following psych verbs were more likely to be NP-attached than those following action verbs. A parallel pattern emerged in our analysis of children's speech and child-directed speech. For both groups, *with*-phrases used with psych verbs were more likely to be modifiers and to be NP-attached (all χ^2's(1) > 30, all p's $< .001$).

Second, we suspected that there might be a tradeoff between the semantic weight of a verb and the weight of the noun. More specifically, if speakers have a rough tendency to equalize the semantic contribution

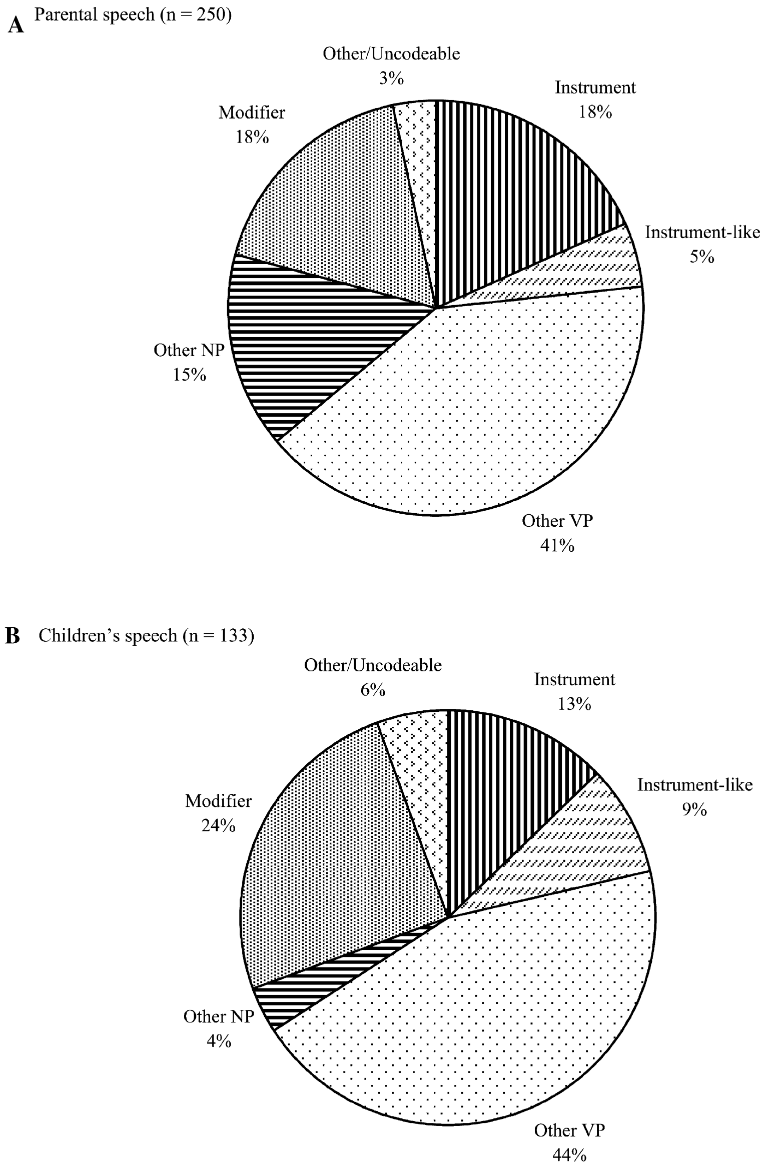


Fig. 9. The potentially ambiguous *with*-phrases from the corpus divided into broad semantic categories.

of their utterances then we might expect that they would be more likely to use a complex noun phrase when they have chosen to use a light verb or alternately to use a light verb when they need to use a complex noun phrase. We explored this by comparing the light verbs with the action verbs. Both the children and the adults were more likely to produce modifier and NP-attached *with*-phrases when they were using a light verb (all χ^2's $(1) > 30$, all p's $< .001$). In fact the children and their parents appeared to use the verbs in quite similar ways. For the 40 verbs that appeared at least twice for each group, there were strong

Table 3
Parental *with*-phrases by type of verb

Verb type	Semantic code			Syntactic code		
	# of Modifiers	# of Instrument	% Instruments[a]	# NP	# VP	% VP[b]
Psych	14	2	12.5	14	10	41.7
Light	25	19	43.2	64	310	83.9
Action	7	443	98.4	7	784	99.1

[a] Instrument sentences as a percentage of Modifier + Instrument.
[b] VP-attached sentences as a percentage of NP-attached + VP-attached.

Table 4
Child *with*-phrases by type of verb

Verb type	Semantic code			Syntactic code		
	# of Modifiers	# of Instrument	% Instruments[a]	# NP	# VP	% VP[b]
Psych	16	1	5.9	16	10	38.5
Light	18	3	14.3	20	89	81.7
Action	6	95	94.1	6	666	99.1

[a] Instrument sentences as a percentage of Modifier + Instrument.
[b] VP-attached sentences as a percentage of NP-attached + VP-attached.

correlations between the children's and the adult's percentage of instruments ($r = .79$) and modifiers ($r = .76$). This strongly suggests that children have access to the information, be it semantic or distributional, that underlies adult lexical biases.

Appendix B. A sentence completion study to determine verb bias

Because the corpus analysis did not provide sufficient data on a large enough set of verbs, we conducted a sentence completion study to select the target verbs for Experiments 1 and 2.

B.1. Method

B.1.1. Participants

Eighty-four students at the University of Pennsylvania volunteered for the experiment. Some received extra course credit or payment for their participation. All the participants were native speakers of English.

B.1.2. Stimuli

Forty verbs were selected using the following criteria: (1) all verbs appeared in the Brown transcripts (see Appendix A) and were judged to be familiar to young children; (2) all verbs frequently appear with either a direct object (e.g., *touch the dog*) or a prepositional compliment (*look at the cat*) and no other arguments (this was verified in the corpus search); (3) the verbs could be acted out on small stuffed animals without risk to the participant or destruction of the experimental props; (4) we chose verbs from a variety of semantic classes in order to increase the chances of getting a range of bias differences in the completion of the *with*-phrase. A sentence fragment was constructed for each verb consisting of a proper name, the verb (plus the preposition if any), the word "the," a direct-object noun, and the word "with" (see 1 below).

The subject and the object were always names of toys that depict animals or people (Barbie, the baby doll, etc.).[19]

(1) Grover pinches the teddy bear with...

The target sentences were divided into 4 lists of 10 verbs each. The ten target fragments in each list were mixed with the same 40 filler fragments. Half of these fragments were ambiguous (syntactic category ambiguities and relative-clause/main clause ambiguities). The other half were very similar to the target fragments but ended in unambiguous prepositions (half VP-attached, half NP-attached).

B.1.3. Procedure

Participants were given the list of sentence fragments and were asked to generate endings for each one as rapidly as they could. They were told that the endings need not be elegant or sensible so long as they were possible English sentences. Participants were told to complete the items in the order they appeared on the sheet and not to return to an item that they had completed.

B.1.4. Coding

Each completion was coded as a modifier, instrument, other VP-attachment, a higher attachment or an ambiguous attachment. These codes were based on the same criteria as in the corpus analysis (see Table 2), except that ancillary instrument, undergoer, and objective responses were coded as instruments. This was done both because it was impossible to reliably code distinctions between these categories without a context and because they involved very similar PP-objects and relations. Obviously, most completions could be construed as "ambiguous" by an adequately creative coder. Our coder was instructed to label the sentence as ambiguous only if she was actually uncertain about what the writer had intended.

B.2. Results

For each of the verbs we calculated two composite measures, Relevance and Bias (see Table 5). Relevance was a measure of the proportion of responses that fell into our two target categories ((Instrument + Modifier)/Total). Bias measured the proportion of these responses that encoded instruments (Instrument/(Instrument + Modifier)). We chose to use a bias measure based solely on instrument and modifier responses because we believed that the other interpretations would not be under consideration in the experimental paradigm that we would be using. For example, most of the VP-other responses involved manner ("sing to the teddy bear with great emotion") or accompaniment ("sing to the teddy bear with me") both of which are implausible if a concrete and inanimate object is in the *with*-phrase. The Bias scores were used to select eight verbs for each of our three bias classes. Modifier verbs had a Bias less than 25%, Instrument verbs had a Bias greater than 75%, and Equi verbs fell somewhere in between. When more than eight verbs met the criterion for the bias class, we selected those with the greatest Relevance and those which would maximize the Bias difference across classes.

In the final set of verbs, marked in Table 5, there was a reliable and strong difference in Bias across classes ($F(2,18) = 132.67$, $p < .0001$). The Modifier verbs had a mean bias score of 12% while the Equi and Instrument bias verbs had scores of 52 and 97%, respectively. Relevance was reasonably well balanced across classes ($M = 37\%$, $M = 24\%$, and $M = 36\%$ for Modifier, Equi, and Instrument respectively, $F(2,18) = 2.74$, $p > .05$). The results of the sentence completion study also support the findings from the

[19] Later we decided to use commands in the eye-tracking experiments. To ensure that this change in sentence form did not radically alter the attachment preferences of individual verbs, we created a new version of the sentence completion experiment using parallel commands (*Pinch the teddy bear with...*). 6 of the 21 participants tested on each list were given the command fragments. Since no systematic pattern of differences was found, we combined the data from the two tests so that we would have a more fine-grained measure of bias for each verb.

Table 5
Verbs from the sentence completion study

Verb	Modifier (%)	Instrument (%)	VP-other (%)	Amb/higher (%)	Relevance[a] (%)	Bias[b] (%)	Verb type
choose	81	0	10	10	81	0	M
look at	19	0	71	10	19	0	M
grab	14	0	86	0	14	0	
whisper to	10	0	86	5	10	0	
move	5	0	90	5	5	0	
stare at	5	0	90	5	5	0	
listen to	38	5	57	0	43	11	M
yell at	30	5	65	0	35	14	M
sing to	29	5	62	5	33	14	M
find	24	5	62	10	29	17	M
talk to	24	5	71	0	29	17	M
hug	19	5	76	0	24	20	M
throw	10	5	80	5	15	33	E
drag	14	10	67	10	24	40	E
point at	14	10	76	0	24	40	E
feel	14	14	67	5	29	50	E
turn over	14	14	67	5	29	50	E
pinch	10	14	76	0	24	60	E
blow on	10	19	67	5	29	67	E
scratch	5	14	81	0	19	75	E
twist	5	24	67	5	29	83	
surprise	5	29	62	5	33	86	
clean	5	38	57	0	43	89	I
brush	5	43	52	0	48	90	I
cover	0	43	48	10	43	100	I
feed	0	43	57	0	43	100	I
bop	0	38	62	0	38	100	I
poke	0	29	71	0	29	100	I
tickle	0	24	76	0	24	100	I
hit	0	24	76	0	24	100	I
knock down	0	14	86	0	14	100	
touch	0	14	86	0	14	100	
scare	0	14	86	0	14	100	
hold	0	10	90	0	10	100	
push	0	10	86	5	10	100	
shake	0	10	90	0	10	100	
reach for	0	5	95	0	5	100	
claw	0	5	86	10	5	100	
spin	0	5	76	19	5	100	
wiggle	0	0	100	0	0	X	

[a] Relevance is the percentage of responses that were coded as modifier or instruments.
[b] Bias is instrument responses as a percentage of the relevant responses (instrument or modifier).

corpus analysis; for the verbs which appeared in both studies, there was a moderately large correlation between the two Bias measures ($r = .61$).[20]

[20] This analysis is based on the thirteen verbs which were used in the sentence completion study and appeared at least twice in the analysis of parental speech.

Subjects were asked to complete fragments that included both the verb and the direct-object noun. This raises the possibility that these completions reflect the tendency of the noun to appear with a modifier instead of or in addition to the tendency of the verb to appear with an instrument PP or a heavy noun phrase.[21] There are two reasons why we consider this a manipulation of verb bias rather than noun bias or a combination of the two. First, the nouns that were used in the sentence completion study were all drawn from a narrow semantic class (toys depicting animate entities). The similarity of the nouns would presumably limit their influence on variation in attachment preferences across fragments. Second, the nouns that were used in the completion were different from those used in Experiments 1 and 2. Thus we see stable differences in attachment preferences between bias classes despite variation in the nouns. While these two features of the design minimize the possibility that the nouns alone are accounting for the differences across verbs, they do not address the possibility that we are measuring the degree to which particular verbs appear with instrument phrases when they have a direct object from this semantic class. In other words the bias ratings could reflect an interaction between the verb and the noun. We addressed this concern with a second sentence completion study in which all of the direct object nouns were replaced with novel words ("Feel the blicket with ....").[22] Since subjects have no experience with these particular nouns, interpretation with the *with*-phrase would have to be based entirely on the verb, or perhaps some central tendency of nouns as a class. The results of this completion study closely parallel the first one. An items analysis revealed that subjects were equally likely to produce modifier completions in the two experiments ($t(23) < 1$, $p > .3$), but more likely to produce instrument completions when given the fragments with nonsense words ($t(23) = 4.80$, $p < .001$). Critically, there was a large and reliable correlation between Bias in the two experiments ($r = .85$, $p < .001$). If we assigned 8 items to each Verb Bias class on the basis of the second completion study, 16 of the 24 items would have remained in the same class and no item would have switched from the Modifier Condition to Instrument Condition or vice versa.

Appendix C. An object norming study to balance the fit between the instrument and verb

These experiments focus on the role of verb biases in guiding online parsing. Like all experimenters we sought to hold all other variables constant. Critically we were concerned with the role that the prepositional objects might play in interpreting these ambiguities. Some objects make better instruments and others make better attributes (compare *hammer* and *moustache*). One way to attempt to control this would be to select a single set of potential instruments/modifiers and use it with all three sets of verbs. We rejected this solution because the suitability of an instrument varies with the verb that is being used (compare *point at the bear with the stick* and *blow on the bear with the stick*). Worse yet, a randomly selected object is more likely to be a suitable instrument for an Instrument verb than for a Modifier verb, introducing a potential confound between verb class and the fit of the instrument. For these reasons we chose to control for the fit of the instrument by measuring it beforehand and balancing it across verb classes.

C.1. Method

C.1.1. Participants

Nineteen students at the University of Pennsylvania volunteered for the experiment. Some received extra course credit for their participation. All the participants were native speakers of English.

[21] We thank an anonymous reviewer for this suggestion.

[22] Forty-five subjects participated. The 24 target words were divided into two lists of 12 words each. These target sentences were embedded in the same set of 40 filler fragments used in the primary norming study. To make the target items less noticeable, one word in each of the filler fragments was replaced with a nonsense word. Responses were coded in the same manner as before.

C.1.2. Stimuli

For each of the 40 verbs used in the sentence completion study, we selected seven objects to be rated as potential instruments for carrying out that action. We attempted to choose objects that spanned the range of plausibility from tools that were especially designed for carrying out the action to objects that could not possibly be used effectively. Special care was taken to select several objects that were unusual but viable instruments for the verb. For example for the verb tickle we selected: a feather, a toothbrush, a paper fan, a stick, a sock, a salt shaker, and a rock.

C.1.3. Procedure

Participants were tested in two large groups (7 and 12 persons each). They were told that their job would be to rate whether an object could be used as an instrument to perform a particular action. An instrument was defined as something you use to carry out an action (in contrast to the thing that you act upon). Participants were given two examples to clarify this distinction ("it seems sensible to say *use a pen to write* but silly to say *use a lego to write*") and had the opportunity to ask questions.

We wanted to control for the fit between the verb and the specific instrument but not for the overall likelihood of using an instrument for the action. So we instructed participants to "imagine that [they] had to do that activity by using an instrument of some kind" and then rate whether the particular object "would be a relatively good one or a relatively poor one for performing that action." To encourage them to focus on the fit of the object for a particular verb, we blocked the presentation by verb. For each block the experimenter announced the verb, then held up the first object and labeled it. Participants were given about seven seconds to rate this object before the experimenter moved on to the next one. Objects were

Table 6
Ratings for target instruments

Verb	Verb type	Instrument	Rating[a]
choose	M	fork	3.22
sing to	M	funnel	3.21
look at	M	glass	2.79
listen to	M	tube	3.58
yell at	M	funnel	4.32
find	M	stick	2.95
talk to	M	tube	3.95
hug	M	blanket	4.79
throw	E	cup	3.47
drag	E	pipe cleaner	3.47
point at	E	flower	3.42
feel	E	feather	2.74
turn over	E	stick	3.79
pinch	E	barrette	5.05
blow on	E	fan	4.68
scratch	E	paper	2.58
brush	I	sponge	3.63
clean	I	t-shirt	3.68
hit	I	flower	2.67
tickle	I	fan	3.16
poke	I	feather	3.05
bop	I	ball	4.26
cover	I	book	3.16
feed	I	glass	5.32

[a] Rating is the mean instrument rating on a scale of 1–7 where 1 is the worst instrument and 7 is the best.

rated on a scale of 1–7 (where 1 was the worse possible instrument imaginable and 7 was the best). The order of the seven objects and the order of the verbs were randomized.

C.2. Results

We calculated the average instrument rating for each object and the variance in the ratings. For each of the 24 target verbs we selected an instrument, subject to the following constraints: (1) the instrument must have a mean rating of between 2.5 and 5.5; (2) the same instrument could not be used for more than two verbs; (3) the mean rating was balanced as closely as possible across bias classes and (4) instruments with a low variance in their ratings were preferred. The target instrument for each verb and the mean instrument rating is given in Table 6. By applying these criteria we succeeding in balancing instrument plausibility across the verb classes ($M = 3.6$, $M = 3.7$, $M = 3.6$ for Modifier, Equi, and Instrument, respectively, $F(2, 18) < 1$, $p > .99$). The Target Instruments in Experiments 1 and 2 were the same objects that participants had rated in the norming study or exact replicas of those objects.

Appendix D. Validating the hidden camera as a method of recording eye movements

Prior work on these questions has used head-mounted cameras and computer eye-tracking algorithms to measure fixation patterns in response to spoken sentences (Hurewitz et al., 2000; Tanenhaus et al., 1995; Trueswell et al., 1999). We chose instead to videotape the subject's face and code the eye movements from the videotapes. There were two reasons for this decision. First, we believed that more daycares, parents, and children would be willing to participate in the study if it used less invasive technology. Second, we were interested in exploring the hidden camera technique because it is cheaper, more portable, and can be used with younger children than head-mounted eye-tracking.

Although new to sentence processing, this technique is simply a variant of the preferential looking paradigms which are widely applied in developmental psychology (Fagan, 1970; Fantz, 1961; Spelke, 1979). Intermodal preferential looking studies typically show very high inter-coder reliability especially when frame-by-frame coding is employed, as it is in the present experiments (Hirsh-Pasek & Golinkoff, 1996). Many of these intermodal studies have looked at children's comprehension of spoken language (Golinkoff, Hirsh-Pasek, Cauley, & Gordon, 1987; Hirsh-Pasek & Golinkoff, 1996; Reznick, 1990). When frame-by-frame coding is synchronized with a speech stimulus, the paradigm is quite similar to the eye-tracking paradigms used in sentence processing. These techniques have proven to be sensitive enough to explore the resolution of pronouns in preschoolers (Song & Fisher, 2002) and improvements in the speed word identification between 15 and 24 months (Fernald et al., 1998; Swingley et al., 1999).

There are several aspects of our testing procedure which were designed to ensure the accuracy of the eye-movement coding. First, the room was well lit and the camera was tightly focused on the participant's face, allowing the coders to see the iris and thus determine eye position. Next, subjects were placed close to the display and their chair was positioned so that their gaze was centered at the location of the camera. This ensured that gazes to each of the four object locations were typically distinguished by the direction in which the eyes rotated and not merely by the extent to which they did so. The image of the participant's face on the hidden camera was monitored throughout the experiment to ensure that the subject remained properly positioned. To ensure that coders received frequent feedback about the relation between eye position and gaze direction, we elicited a predictable sequence of gazes from the participants at the beginning of each trial by laying out the props in a consistent order (clockwise from the upper left) and drawing their attention to each one. Finally, by beginning each trial with a central fixation, we gave the coder a clear reference point for subsequent eye movements. All eye movements were coded by two independent coders who agreed on the gaze direction for 94.4% of all coded frames for adults (Cohen's $\kappa = .92$) and 93.6% of all coded frames for children (Cohen's $\kappa = .91$).

Three features of the present experiments validate our version of this technique. First, our analyses of fixations after the Direct-Object Noun demonstrate that our methods produce findings which closely parallel prior studies using head-mounted eye-tracking. In the One Referent Conditions, our adult participants begin to fixate more on the Target Animal than the Distractor Animal approximately 300 ms after the onset of the noun, suggesting that identification of the noun is influencing the allocation of attention. In the parallel condition of TSH this effect emerges at almost exactly the same point in time. Similarly, both the five-year-olds in TSH and the children in our One Referent Conditions begin demonstrating a Target Animal preference 500 ms after noun onset. Second, the observed effects of Verb Bias and Referential Context, demonstrate that our coding techniques are sensitive to effects of the variables of interest. Finally, our measures show very high inter-coder reliability.

Furthermore, we have directly compared data collected with the hidden camera and data collected with a head-mounted eye-tracking system. Two adult participants participated in a study similar to Experiment 1. The props were placed on the four quadrants of the podium and a camera in the center of the podium filmed the participant's face as he or she listened to the instructions. At the same time the participant's eye movements were recorded using an ISCAN eye-tracking visor (for details see Trueswell et al., 1999). The visor had two miniature cameras attached to a monocle. One of the cameras (the scene camera) recorded the visual scene from the perspective of the subject's left eye, while the other recorded a close-up image of this eye. The ISCAN tracker analyzed this eye image and determined the position of the center of the pupil and the corneal reflection. The computer then used this information (along with data from a point-of-light calibration procedure) to calculate the eye position in real time. The eye position was displayed as a cross-hair superimposed on the video recorded by the scene camera. This image was recorded to tape using a frame-accurate digital video recorder with audio lock.

The video of the participant's face was coded by a single coder who followed the procedure detailed in the coding section of Experiment 1. A second researcher coded the scene video from the eye-tracker. This coder viewed the video frame-by-frame and noted the onset of each eye-movement and the location of the subsequent fixation. Location was coded as being in one of the four quadrants, at the center of the display or at some other location. Track loss was also noted. Both coders were trained by the second author and had prior experience coding data of this kind.

The two coding methods produced strikingly similar results. In both cases the amount of lost data was small (2.4% of frames were coded as track loss for eye-tracker, 3.3% were coded as missing for the face camera). The data sets were re-aligned at the beginning of each trial, creating a pair of locations for each video frame from the sentence onset until the end of the trial. Because there was little overlap in the missing frames (only .7% missing in both sets), we excluded pairs with missing data before comparing the two data sets. The two methods resulted in the same fixation location for 93.2% of all coded frames (Cohen's $\kappa = .91$). This is almost as high as the inter-coder reliability of the hidden camera method (94.4% for adult participants). In this situation at least, the two methods produce essentially equivalent data.

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38*(4), 419–439.

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

Baker, M. C. (2001). *The atoms of language: The mind's hidden rules of grammar*. Basic Books.

Bates, E. A., & MacWhinney, B. (1987). Competition, variation and language learning. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 157–194). Hillsdale, NJ: Erlbaum.

Birner, B. J., & Ward, G. (1994). Uniqueness, familiarity, and the definite article in English, In *Proceedings of the Twentieth Annual Meeting of the Berkeley Linguistics Society* (pp. 93–102).

Boland, J. E. (1997). Resolving syntactic category ambiguities in discourse context: Probabilistic and discourse constraints. *Journal of Memory and Language, 36*(4), 588–615.

Boland, J. E., & Blodgett, A. (2001). Understanding the constraints on syntactic generation: Lexical bias and discourse congruency effects on eye movements. *Journal of Memory and Language, 45*(3), 391–411.

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

Boland, J. E., Tanenhaus, M. K., Garnsey, S. M., & Carlson, G. N. (1995). Verb argument structure in parsing and interpretation: Evidence from wh-questions. *Journal of Memory and Language, 34*(6), 774–806.

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.

Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.

Brown-Schmidt, S., Campana, E., & Tanenhaus, M. K. (2002). Reference resolution in the wild: On-line circumscription of referential domains in a natural, interactive problem-solving task. In *Proceedings of the 24th Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.

Brown-Schmidt, S., Campana, E., & Tanenhaus, M. K. (in press). Reference resolution in the wild: Circumscription of referential domains by naive participants during an interactive problem solving task. In: J. C. Trueswell & M. K. Tanenhaus (Eds.) *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions*. Cambridge, MA: MIT Press, to appear.

Chabon, S., Kent-Udolf, L., & Egolf, D. (1982). The temporal reliability of Brown's mean length of utterance (MLU-M) measure with post-Stage V children. *Journal of Speech and Hearing Research, 25*, 124–128.

Clifton, Jr.,, C., Frazier, L., & Rayner, K. (Eds.). (1994). *Perspectives on sentence processing*. Hillsdale, NJ: Erlbaum.

Collins, M., & Brooks, J. (1995). Prepositional phrase attachment through a backed-off model. In *Proceedings of the Third Workshop on Very Large Corpora*.

Collins, M. (1996). A new statistical parser based on bigram lexical dependencies. In *Proceedings of the 34th Annual Meeting of the ACL*, Santa Cruz.

Collins, M. (1997). Three generative, lexicalized models for statistical parsing. In *Proceedings of the 35th Annual Meeting of the ACL (jointly with the 8th Conference of the EACL)*, Madrid.

Crain, S., & Meroni, L. (2002). *Children's use of referential context*. Paper presented at the 27th Annual Boston University Conference on Language Development, November 2002, Boston, MA.

Crain, S., & Thornton, R. (1998). *Investigations in universal grammar: A guide to experiments on the acquisition of syntax and semantics*. Cambridge, MA: MIT Press.

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

Dahan, D., Magnuson, J., & Tanenhaus, M. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology, 42*, 317–367.

Dempster, F. N. (1981). Memory span: Sources of individual and developmental differences. *Psychological Bulletin, 89*, 63–100.

Dempster, F. N. (1985). Short-term memory development in childhood and adolescence. In C. J. Brainerd & M. Pressley (Eds.), *Basic processes in memory development* (pp. 208–248). New York: Springer.

Fagan, J. F. (1970). Memory in the infant. *Journal of Experimental Child Psychology, 9*(2), 217–226.

Fantz, R. L. (1961). The origin of form perception. *Scientific American, 204*, 66–72.

Fernald, A., Pinto, J., Swingley, D., Weinberg, A., & McRoberts, G. (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science, 9*(3), 228–231.

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

Ferreira, F., Christianson, K., & Hollingworth, A. (2001). Misinterpretations of garden-path sentences: Implications for models of sentence processing and reanalysis. *Journal of Psycholinguistic Research, 30*(1), 3–20.

Fisher, C. (1996). Structural limits on verb mapping: The role of analogy in children's interpretations of sentences. *Cognitive Psychology, 31*, 41–81.

Fisher, C. (2002). The role of abstract syntactic knowledge in language acquisition: A reply to Tomasello (2000). *Cognition, 82*, 259–276.

Fisher, C., Gleitman, H., & Gleitman, L. R. (1991). On the semantic content of subcategorization frames. *Cognitive Psychology, 23*, 331–392.

Fisher, C., Hall, G., Rakowitz, S., & Gleitman, L. (1994). When it is better to receive than to give: Syntactic and conceptual constraints on vocabulary growth. *Lingua, 92*, 333–375.

Fisher, C., & Snedeker, J. (2002). *Counting the nouns: Simple sentence-structure cues guide verb learning in 21-month-olds*. Paper presented at the Twenty-seventh Boston University Conference on Language Development, Boston, MA, November 2002.

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

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

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

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.

Gaulin, C., & Campbell, T. (1994). Procedure for assessing verbal working memory in normal school-age children: Some preliminary data. *Perceptual and Motor Skills, 79*, 55–64.

Geyer, H. (1997). *Subcategorization as a predictor of verb meaning: Evidence from Modern Hebrew.* Unpublished manuscript University of Pennsylvania.

Gillette, J., Gleitman, L., Gleitman, H., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition, 73*, 153–190.

Gleitman, L. R. (1990). The structural sources of verb meanings. *Language Acquisition, 1*, 3–55.

Glucksberg, S., & Krauss, R. (1967). What do people say after they have learned to talk? Studies in the development of referential communication. *Merrill-Palmer Quarterly, 13*, 309–316.

Golinkoff, R. M., Hirsh-Pasek, K., Cauley, K. M., & Gordon, L. (1987). The eyes have it: Lexical and syntactic comprehension in a new paradigm. *Journal of Child Language, 14*, 23–45.

Gomez, R., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition, 70*, 109–135.

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

Hawkins, J. A. (1978). *Definiteness and indefiniteness: A study in reference and grammaticality prediction*. Atlantic Highlands, NJ: Humanities Press.

Hirsh-Pasek, K., & Golinkoff, R. M. (1996). The intermodal preferential looking paradigm: A window onto emerging language comprehension. In D. McDaniel, & C. McKee (Eds.), *Methods for assessing children's syntax. Language, speech, and communication* (pp. 105–124).

Hirsh-Pasek, K., Kemler Nelson, D. G., Jusczyk, P. W., Wright-Cassidy, K., Druss, B., & Kennedy, L. (1987). Clauses are perceptual units for young infants. *Cognition, 26*, 269–286.

Hudson, C. L., & Newport, E. L. (2003). *Regularizing unpredictable variation in the transition from pidgin to creole: An experimental investigation of the role of adult learners in pidgin/creole development.* Manuscript submitted for publication.

Hughes, C., & Graham, A. (2002). Measuring executive functions in childhood: Problems and solutions?. *Child and Adolescent Mental Health, 7*(3), 131–142.

Hurewitz, F., Brown-Schmidt, S., Thorpe, K., Gleitman, L., & Trueswell, J. (2000). One frog, two frog, red frog, blue frog: Factors affecting children's syntactic choices in production and comprehension. *Journal of Psycholinguistic Research, 29*, 597–626.

Hurewitz, F., Brown-Schmidt, S., Trueswell, J., & Gleitman, L. (in progress). The contribution of conversation goals and verb-preferences to children's syntactic decisions.

Jackendoff, R. (2002). *Foundations of language*. Oxford: Oxford University Press.

Jenkins, L., Myerson, J., Hale, S., & Fry, A. (1999). Individual and developmental differences in working memory across the life span. *Psychonomic Bulletin & Review, 6*, 28–40.

Johnson, E. K., & Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language, 44*, 548–567.

Jusczyk, P. W., Friederici, A. D., Wessels, J. M. I., Svenkerud, V. Y., & Jusczyk, A. M. (1993). Infants' sensitivity to the sound patterns of native language words. *Journal of Memory and Language, 32*, 402–420.

Jusczyk, P. W., Culter, A., & Redanz, N. J. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development, 64*, 675–687.

Jusczyk, P. W., Hirsh-Pasek, K., Kemler Nelson, D. G., Kennedy, L., Woodward, A., & Piwoz, J. (1992). Perception of acoustic correlates of major phrasal units by young infants. *Cognitive Psychology, 24*, 252–293.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 98*, 122–149.

Karmiloff-Smith, A. (1979). *A functional approach to child language: A study of determiners and reference*. New York: Cambridge University Press.

Kail, R. V. (1991). Development of processing speed in childhood and adolescence. In H. W. Reese (Ed.), *Advances in child development and behavior* (Vol. 25). New York: Academic Press.

Kail, R. V., & Salthouse, T. A. (1994). Processing speed as a mental capacity. *Acta Psychologica, 86*, 199–225.

Kidd, E. J. (2003). An investigation of children's sentence processing: A developmental perspective. Unpublished Doctoral Dissertation, LaTrobe University.

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

Kawamoto, A. H. (1993). Nonlinear dynamics in the resolution of lexical ambiguity: A parallel distributed processing account. *Journal of Memory and Language, 32*(4), 474–516.

Klee, T., Schaffer, M., May, S., Membrino, S., & Mougey, K. (1989). A comparison of the age-MLU relation in normal and specifically language-impaired preschool children. *Journal of Speech and Hearing Research, 54*, 226–233.

Lederer, A., Gleitman, L., & Gleitman, H. (1995). Verbs of a feather flock together: Structural properties of maternal speech. In M. Tomasello & E. Merriam (Eds.), *Beyond names for things: Young children acquisition of verbs*. New York: Academic Press.

Levin, H. S., Culhane, K. A., Hartmann, J., Evankovich, K., et al. (1991). Developmental changes in performance on tests of purported frontal lobe functioning. *Developmental Neuropsychology, 7*(3), 377–395.

Li, P. (1994). Maternal verb usage in Mandarin Chinese. Unpublished manuscript University of Pennsylvania.

Limber, J. (1973). The genesis of complex sentences. In T. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 169–186). New York: Academic Press.

Lloyd, P., Camaioni, L., & Ercolani, P. (1995). Assessing referential communication skills in the primary school years: A comparative study. *British Journal of Developmental Psychology, 13*(1), 13–29.

Lyons, C. (1999). *Definiteness*. Cambridge, UK: Cambridge University Press.

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

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.

Macnamara, J. (1982). *Names for things*. Cambridge, MA: MIT Press.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). Mahwah, NJ: Erlbaum.

Maratsos, M. P. (1976). *The use of definite and indefinite reference in young children: An experimental study of semantic acquisition*. New York: Cambridge University Press.

Maratsos, M. (1982). The child's construction of grammatical categories. In E. Wanner & L. Gleitman (Eds.), *Language acquisition: The state of the art* (pp. 240–266). New York: Cambridge University Press.

Maratsos, M., & Chalkley, M. (1980). The internal language of children's syntax. In K. Nelson (Ed.), *Children's language* (Vol. 2). New York: Gardner Press.

Marcus, G. F. (2000). Pabiku and Ga Ti Ga: Two mechanisms infants use to learn about the world. *Current Directions in Psychological Science, 9*(5), 145–147.

Marcus, M. (1994). New trends in natural language processing: Statistical natural language processing. *Proceedings of the National Academy of Science, 92*, 10052–10059.

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

McKee, C., McDaniel, D., & Snedeker, J. (1998). Relatives children say. *Journal of Psycholinguistic Research, 5*, 573–595.

McKercher, D. M. (2001). The polysemy of with in first language acquisition. Unpublished Doctoral Thesis. Stanford University.

Mehler, J., Jusczyk, P. W., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition, 29*, 143–178.

Meline, T., & Meline, N. (1981). Normal variation and prediction of mean length of utterance from chronological age. *Perceptual and Motor Skills, 53*, 376–378.

Morgan, J. L., & Saffran, J. R. (1995). Emerging integration of sequential and suprasegmental information in preverbal speech segmentation. *Child Development, 66*, 911–936.

Morgan, J. L. (1996). A rhythmic bias in preverbal speech segmentation. *Journal of Memory and Language, 35*, 666–688.

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

Naigles, L. (1996). The use of multiple frames in verb learning via syntactic bootstrapping. *Cognition, 58*, 221–251.

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: Towards an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 756–766.

Nazzi, T., Jusczyk, P. W., & Johnson, E. K. (2000). Language discrimination by English-learning 5-month-olds: Effects of rhythm and familiarity. *Journal of Memory and Language, 43*, 1–19.

Nilsen, D. F. (1973). *The instrumental case in English: Syntactic and semantic considerations*. Number 156 in Janua Linguarum. The Hague: Mouton.

Passler, M. A., Isaac, W., & Hynd, G. D. (1985). Neuropsychological development of behavior attributed to frontal lobe functioning in children. *Developmental Neuropsychology, 1*, 349–370.

Perner, J., & Wimmer, H. (1985). John thinks that Mary thinks that…: Attribution of second-order beliefs by 5- to 10-year-old children. *Journal of Experimental Child Psychology, 39*, 437–471.

Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.

Prince, E. F. (1981). Toward a taxonomy of given-new information. In P. Cole (Ed.), *Radical pragmatics* (pp. 223–256). New York: Academic Press.

Prince, E. F. (1992). The ZPG letter: subjects, definiteness, and information-status. In S. Thompson & W. Mann (Eds.), *Discourse description: Diverse analyses of a fund raising text* (pp. 295–325). Philadelphia/Amsterdam: John Benjamins BV.

Rayner, K., Garrod, S., & Perfetti, C. (1992). Discourse influences during parsing are delayed. *Cognition, 45*, 109–139.

Reznick, S. J. (1990). Visual preference as a test of infant word comprehension. *Applied Psycholinguistics, 11*, 145–166.

Robinson, E. J., & Robinson, W. P. (1982). Knowing when you don't know enough: Children's judgments about ambiguous information. *Cognition, 12*(3), 267–280.

Saffran, J., Newport, E., & Aslin, R. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language, 35*, 606–621.

Sato, R., Tanaka, A., Hosokawa, T., & Murai, N. (1998). Characteristics of perseverative errors by school-age children. *Psychological Reports, 83*(3), 1075–1080.

Scarborough, H., Wyckoff, J., & Davidson, R. (1986). A reconsideration of the relation between age and mean utterance length. *Journal of Speech and Hearing Research, 29*, 394–399.

Schlesinger, I. M. (1995). *Cognitive space and linguistic case: Semantic and syntactic categories in English*. Cambridge, UK: Cambridge University Press.

Schneider, W., & Bjorklund, D. F. (1998). Memory. In D. Kuhn & R. S. Siegler (Eds.), *Cognitive, language, and perceptual development* (Vol. 2, pp. 467–521). In W. Damon (General Editor), *Handbook of child psychology* (5th Ed.), New York: Wiley.

Snedeker, J. (2000). Cross-situational observation and the semantic bootstrapping hypothesis. In E. Clark (Ed.), *Proceedings of the Thirtieth Annual Child Language Research Forum*. Stanford, CA: Center for the Study of Language and Information.

Snedeker, J., & Gleitman, L. (2004). Why it is hard to label our concepts. In Hall & Waxman (Eds.), *Weaving a lexicon* (pp. 255–293). Cambridge, MA: MIT Press.

Snedeker, J., Gleitman, L., & Brent, M. (1999). The successes and failures of word-to-world mapping. In A. Greenhill, M. Hughs, H. Littlefield, & H. Walsh (Eds.), *Proceedings of the Twenty-third Annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press.

Song, H., & Fisher, C. (2002). Young children's sensitivity to discourse cues in on-line pronoun interpretation. *Proceedings of the 26th Annual Boston University Conference on Language Development*. Boston, MA: Cascadilla Press.

Spelke, E. S. (1979). Perceiving bimodally specified events in infancy. *Developmental Psychology, 15*(6), 626–636.

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

Snedeker, J., & Trueswell, J. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language, 48*, 103–130.

Snedeker, J., & Yuan, S. (2003). *Is it the words or the melody? The use of lexical and prosodic information in children's parsing*. Paper presented at CUNY Sentence Processing Conference, Cambridge, MA, March 2003.

Spivey-Knowlton, M. J., Trueswell, J. C., & Tanenhaus, M. K. (1993). Context effects in syntactic ambiguity resolution: Discourse and semantic influences in parsing reduced relative clauses. *Canadian Journal of Experimental Psychology, 47*(2), 276–309.

Spivey, M. J., & Tanenhaus, M. K. (1998). Syntactic ambiguity resolution in discourse: Modeling the effects of referential context and lexical frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(6), 1521–1543.

Spivey, M. J., Tanenhaus, M. K., Eberhard, K. M., & Sedivy, J. C. (2002). Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology, 45*, 447–481.

Stone, M., & Webber, B. (1998). Textual economy through close coupling of syntax and semantics. In *Proceedings of INLG* (pp. 178–187).

Swingley, D., Pinto, J. P., & Fernald, A. (1999). Continuous processing in word recognition at 24 months. *Cognition, 71*(2), 73–108.

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.

Tomasello, M. (1992). *First verbs: A case study in early grammatical development*. Cambridge: Cambridge University Press.

Tomasello, M. (2000a). Do young children have adult syntactic competence?. *Cognition, 74*, 209–253.

Tomasello, M. (2000b). The item-based nature of children's early syntactic development. *Trends in Cognitive Sciences, 4*(4), 156–163.

Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (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 Eimas & Miller (Eds.), *Handbook in perception and cognition, Vol. 11: Speech language and communication*. New York: Academic Press.

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

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

Trueswell, J. C., Papafragou, A., & Choi, Y. (in press). Referential and syntactic processes: What develops? In E. Gibson & N. Pearlmutter (Eds.), *Proceedings to the 16th annual CUNY human sentence processing conference*, to appear.

Trueswell, J. C., & Tanenhaus, M. K. (1994). Toward a lexicalist framework of constraint-based syntactic ambiguity resolution. In Clifton & 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*(3), 528–553.

van Berkum, J., Brown, C., & Hagoort, P. (1999). Early referential context effects in sentence processing: Evidence from event-related brain potentials. *Journal of Memory and Language, 41*, 147–182.

Welsh, M., Pennington, B., & Groisser, D. (1991). A normative-developmental study of executive function: A window on prefrontal function in children. *Developmental Neuropsychology, 7*, 131–149.

Wexler, K. (2003). *Maximal trouble*. Paper presented at CUNY Sentence Processing Conference, Cambridge, MA, March 2003.

Zelazo, P. D., Muller, U., Frye, D., & Marcovitch, S. (2003). Study 3: What do children perseverate on when they perseverate?. *Monographs of the Society for Research in Child Development, 68*(3), 65–72.

Zelazo, P. D., Reznick, J. S., & Spinazzola, J. (1998). Representational flexibility and response control in a multistep multilocation search task. *Developmental Psychology, 34*(2), 203–214.