ACADEMIC
PRESS

Using prosody to avoid ambiguity: Effects of speaker awareness and referential context

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

[a] Department of Psychology, 1136 William James Hall, Harvard University, 33 Kirkland Street, Cambridge, MA 02138, USA
[b] University of Pennsylvania, USA

Abstract

In three experiments, a referential communication task was used to determine the conditions under which speakers produce and listeners use prosodic cues to distinguish alternative meanings of a syntactically ambiguous phrase. Analyses of the actions and utterances from Experiments 1 and 2 indicated that Speakers chose to produce effective prosodic cues to disambiguation only when the referential scene provided support for both interpretations of the phrase. In Experiment 3, on-line measures of parsing commitments were obtained by recording the Listener's eye movements to objects as the Speaker gave the instructions. Results supported the previous experiments but also showed that the Speaker's prosody affected the Listener's interpretation prior to the onset of the ambiguous phrase, thus demonstrating that prosodic cues not only influence initial parsing but can also be used to predict material which has yet to be spoken. The findings suggest that informative prosodic cues depend upon speakers' knowledge of the situation: speakers provide prosodic cues when needed; listeners use these prosodic cues when present.
© 2002 Elsevier Science (USA). All rights reserved.

Keywords: Prosody; Language production; Language comprehension; Ambiguity; Parsing

Much of the prior research on prosody and syntactic ambiguity has focused either on the speaker or on the listener, but only rarely on the interaction between the two. This division of labor has led to some important advances in our understanding of prosody. For instance, we know that listeners can, under certain circumstances, use the prosodic organization of an utterance to guide their interpretation of a phrase that has a global or local syntactic ambiguity (e.g., Beach, 1991; Beach, Katz, & Skowronski, 1996; Carlson, Clifton, & Frazier, 2001; Kjelgaard & Speer, 1999; Marslen-Wilson, Tyler, Warren, Grenier, & Lee, 1992; Nagel, Shapiro, & Nawy, 1994; Nagel, Shapiro, Tuller, & Nawy, 1996; Price, Ostendorf, Shattuck-Hufnagel, & Fong, 1991; Pynte & Prieur, 1996; Schafer, Carter, Clifton, & Frazier, 1996;

Warren, Grabe, & Nolan, 1995 among others). Likewise, studies of language production have found that the prosodic grouping of an utterance can reflect its syntactic structure (Cooper & Paccia-Cooper, 1980; Klatt, 1975; Lehiste, 1972). In addition, informed speakers can mark different meanings on an ambiguous string through alterations of the prosodic grouping (Allbritton, McKoon, & Ratcliff, 1996; Lehiste, 1973a; Price et al., 1991).

Taken together, these studies indicate that users of a language share some implicit knowledge about the relationship between prosody and syntax, and are capable of using this knowledge to guide linguistic choices. However, studies of this sort, which have examined separately the behavior of the speaker or listener—usually in relatively impoverished referential settings—leave open several important questions about the conditions under which speakers typically produce and listeners typically

---

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

use these cues to structure. Most studies of comprehension have relied upon artificial manipulations of prosodic information to test its effect on syntactic choice. In particular, prosody has typically been manipulated by splicing silent pauses into speech to indicate clause boundaries (e.g, Pynte & Prieur, 1996), manipulating synthesized speech (e.g., Beach, 1991; Lehiste, 1973a), or asking trained speakers to produce particular prosodic variants of an utterance (for discussions see Allbritton et al., 1996; Fox Tree & Meijer, 2000). In many cases, the listener is then asked to select between two alternative meanings, allowing him to focus on this particular linguistic choice, rather than going about the business of assigning meaning relative to the current context (e.g., Lehiste, 1973a; Price et al., 1991). Similarly, studies of the speaker have tended to rely on artificial manipulations of ambiguity to examine its effect on prosody. Most studies have collected data from trained speakers (such as radio announcers) who have been explicitly instructed to contrast the alternate interpretations of an ambiguous sentence (see e.g., Klatt, 1975; Price et al., 1991). It is possible that uninformed speakers, who are engaged in the job of communicating an utterance that is linked to a referential task, may behave differently.

Most of these concerns have been noted by others (Allbritton et al., 1996; Fox Tree & Meijer, 2000; Schafer, Speer, Warren, & White, 2000a), but only a few studies of prosody and syntax have examined how untrained listeners respond to the speech of untrained speakers (Allbritton et al., 1996; Lehiste, 1973a; Schafer et al., 2000a; Wales & Toner, 1979). Fewer still have explored this in paradigms where the listener and speaker are in the same room and have a common task, conditions that are more typical of naturally occurring speech (Keysar & Henly, 1998; Schafer et al., 2000a). Studies of this sort are of interest if we would like to understand how and when listeners use the prosodic cues that are actually available to them. This is especially important when one considers that prosodic variation is influenced by several factors other than syntactic structure, including the length and stress pattern of words, speech rate, the presence of contrastive or emphatic stress, and the prosodic marking of discourse focus (e.g., Ferreira, 1993; Gee & Grosjean, 1983; Kaisse, 1985; Nespor & Vogel, 1986; for reviews see Beckman, 1996; Cutler, Dahan, & van Donselaar, 1997; Fernald & McRoberts, 1996; Shattuck-Hufnagel & Turk, 1996; Warren, 2000).

Consistent with these concerns, experiments employing naïve speakers have found that they produce less reliable prosodic cues for syntactic disambiguation than the informed speakers who are typically used in comprehension experiments (Allbritton et al., 1996; Fox Tree & Meijer, 2000; Lehiste, 1973a; Wales & Toner, 1979). In the most relevant of these studies, Allbritton et al. (1996) compared situations in which speakers were

uninformed or explicitly informed about the ambiguities. In the uninformed condition, untrained speakers (undergraduate students) and professional speakers (radio announcers) were asked to read paragraphs containing globally ambiguous sentences (e.g., "They rose early in May") that had been disambiguated by the prior context. In the informed condition, the professional speakers were provided with the same globally ambiguous sentences, except this time without any context. They were told that the goal of the study was find out how "differences in pronunciation could affect the meaning of ambiguous sentences" (Allbritton et al., 1996, p. 725). Both of the possible meanings were explained to the speaker, and she/he was asked to read each sentence twice, once for each meaning.

The critical utterances were spliced out of context and analyzed by a separate group of subjects who were given unambiguous paraphrases of the two possible meanings and asked to circle the one that they believed the speaker had intended to convey. The findings from this judgment task revealed, in the words of the authors, that "most speakers trained or not, did not produce prosodically disambiguated utterances for most sentences. Trained, professional speakers reliably produced appropriate disambiguating prosody only when they were shown the two meanings of the sentence side by side and were explicitly asked to pronounce the sentence twice, once with each meaning" (Allbritton et al., 1996, p. 731). Such observations suggest that speakers are unable to utilize their prosodic–syntactic knowledge 'on the fly' to convey the intended meaning of an ambiguous utterance.

Recently, Schafer et al. (2000a) have presented data which challenge the Allbritton et al. conclusions. They elicited prosodic variants of temporary and global ambiguities from uninstructed subjects by having them play a game that used a set of scripted commands. These utterances were submitted to acoustic and phonological analyses and a judgment task parallel to that conducted by Allbritton et al. (1996). In all three analyses Schafer and colleagues found evidence that speakers produced consistent prosodic cues to the intended structure. They attribute the divergent findings to differences in the tasks that were used, concluding that while speakers produce reliable prosodic cues when they have clear communicative goals, readers in experimental studies may not always do so.

There is, however, another possible explanation for the Allbritton findings, one which would indicate a more limited role for prosody in syntactic ambiguity resolution. It is possible that speakers only produce cues when the surrounding context does not disambiguate or strongly bias the interpretation of the utterance (Lieberman, 1963; see also Straub, 1997). In the Allbritton et al. study, prosodic cues were inconsistent only when the speaker produced the utterance in an unambiguous,

or highly biasing, context. Perhaps the failure to produce reliable cues is attributable to an unusual type of 'cue trade-off,' in which speakers do not bother with prosodic cues if other cues are present to disambiguate structure. In the Schafer et al. experiment, contextual factors may not have strongly supported the intended alternative for the speaker. In particular, the speakers had limited knowledge of what the listener could see. And, on at least some trials, the critical sentences were ambiguous in context.

Thus, the previous research suggests that the relation between prosody and syntax may be mediated by the context in which an utterance is used. Speakers may provide reliable prosodic cues to syntax only when the context does not strongly support their intended meaning, or they become aware of the potential for ambiguity. The current paper explores these possibilities by examining the effect of referential context and awareness on both the production and comprehension of prosodic cues to structure. In particular, we examine the situations under which untrained speakers produce prosodic cues that will allow listeners to resolve attachment ambiguities. The critical sentences are ones that contain globally ambiguous prepositional phrase attachments, such as "Tap the frog with the flower." Out of context, the phrase "with the flower" can be taken either as Instrument (VP-attachment) indicating what to use to do the tapping, or it can be taken as a Modifier (NP-attachment) indicating which frog to tap. This particular ambiguity was chosen for two reasons. First, PP-attachment is a well studied ambiguity in the comprehension field, allowing us to make comparisons to existing findings. Second, we wished to examine an ambiguity that did not hinge upon the placement of a clause boundary. Prosodic cues to clause boundaries have been studied extensively, and some evidence suggests that they may be marked more strongly and more consistently than phrasal attachments (Cooper & Paccia-Cooper, 1980; Marcus & Hindle, 1990; Price et al., 1991). Less is known about the extent to which speakers can successfully disambiguate ambiguities of phrasal attachment within a clause (for discussion see Pynte & Prieur, 1996; Warren, 1985).

These experiments were conducted using a referential communication task, in which a Speaker and a Listener were separated by a divider allowing for only vocal communication between the two participants (Glucksberg & Krauss, 1967). The Speaker conveyed instructions to the Listener so that she could perform specific actions with objects on her side of the screen. This situation offered two advantages for exploring the relationship between prosody and context. First, the referential context was highly salient, and was defined by the sets of objects in front of the Speaker and Listener. This reduces the role of memory in maintaining information about the context of the utterance, since the

relevant context is defined not by a story which must be stored and consulted, but rather by a physical display that is present during the production task (see Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1996 for a related discussion). Second, separation of the two participants allowed us to manipulate independently the referential context of the Speaker and the Listener, allowing us to disentangle referential effects on the task of production and the task of comprehension.

Assessment of the effectiveness of prosodic cues was tested in the following manner. On each trial in the experiment, the Experimenter demonstrated an action to the Speaker, which was not seen by the Listener. The Speaker was then asked to produce a scripted sentence describing this action, which on target trials was syntactically ambiguous (e.g., "Tap the frog with the flower"). Successful communication was defined as replication of the Experimenter's action by the Listner. Above-chance performance on the part of the Listener would indicate that prosody was playing a role in ambiguity resolution.

In Experiment 1, we examined the use of prosodic cues when the referential context of the Speaker provided support for both meanings of the target sentence. In Experiment 2, we examined prosodic cues when the referential context of the Speaker strongly favored the intended meaning of the utterance. If prosodic choices are affected by the Speaker's knowledge of the referential context, we would expect to see a decreased use of helpful prosodic cues when the referential context provides other cues to disambiguate the utterance. If, on the other hand, knowledge of the referential context is not relevant, we should expect similar performance across the two experiments. Finally, in Experiment 3, we recorded the Listeners' eye movements as they participated in the referential communication task. This more fine-grained measure of the Listener's language interpretation process may disentangle early, perhaps more automatic, contributions of prosody from later, possibly more strategic, uses of this information.

Experiment 1

Method

Participants

Thirty-two pairs of volunteers from the University of Pennsylvania community received extra course credit or were paid for participation. In each pair, one participant played the role of Speaker and the other played the role of Listener. All of the Speakers were female. This was done to eliminate a potential source of variability. While it may limit the generality of our findings we have no reason to believe that the effects that we found are influenced by gender. Half the Listeners were male and

half were female. All participants were native speakers of English.

Procedure

During the experiment, the Speaker and Listener sat on opposite sides of a vertical screen. This barrier ensured that all communication between the pair would be verbal. At the start of each trial, the Speaker and Listener were given identical bags of toys, which they laid out on the trays in front of them. Participants were told that the initial position of the toys would not be relevant to the task and that they could place the objects where they wished. As the participants removed the toys from their bags, the Experimenter introduced each one using indefinite noun phrases (e.g., "this bag contains a dog, a fan..."). This ensured that participants had the same labels for toys and that subsequent reference to the objects using definite noun phrases (e.g., "the dog") sounded natural. When a toy animal had a small object attached to it, it was introduced by listing the animal and the attribute-object separately (e.g., a frog carrying a flower would be introduced as "a frog, a flower"). This was done to avoid biasing subjects by using either a prepositional phrase or a relative clause to describe the attribute-possessor relation.

The Speaker then watched the Experimenter demonstrate an action using the toys. This action could not be seen by the Listener. Next, the Speaker received a card containing a written sentence describing the action. Speakers were instructed to take as much time as needed to memorize the sentence. When the Speaker was ready, the Experimenter took away the card and demonstrated the action again. The Experimenter's second demonstration of the action and this verbal exchange created a delay of approximately 10 s between the time the Speaker last saw the sentence and the time at which she produced it. After the Speaker produced the sentence, the Listener performed the action with his or her own set of toys. Speakers had been told that the primary goal of the experiment was to say each sentence in such a way as to get the Listener to perform the same action on the other side of the screen. Listeners had been told that their job was to perform the action that they believed had been demonstrated to the Speaker.

While the Speaker was memorizing the sentence, the Listener drew a map of where she/he had placed each toy. This was done to familiarize the Listeners with the toys, and to occupy them while the action was demonstrated to the Speaker. Throughout the experiment, interaction between the participants was limited. The Speaker was instructed to ask the Listener if he was ready and wait for a response before producing the sentence. Once the Speaker produced the sentence, the Listener could not ask for clarification, and had to perform the action with his own set of toys. The Speaker's utterances were audiotaped, and the Listener's

actions were videotaped. After the study was completed the participants were separated, and each was interviewed to assess their awareness of the experimental manipulation and the ambiguity in the critical items.

While our testing situation is more like ordinary language use than the typical reading study, it still differs from natural speech because it is scripted. Less constraining test conditions would more closely approximate common speech situations, but they would not allow us to compare pairs of utterances that vary only in their structure. For our purposes, the critical question is not whether the task is natural but whether it distorts the relations between prosody, syntax, and referential context. One could plausibly argue that our task weakens these effects by severing the processing links between syntax and prosodic planning. But one could also argue that it strengthens prosodic disambiguation by giving speakers time to plan appropriate prosody. Ultimately, the sensitivity of this task is one of the empirical questions addressed in these experiments. If our attempts to create a naturalistic communicative context failed completely, then this is essentially a prepared speech task (Sternberg, Monsell, Knoll, & Wright, 1978), a paradigm that has been found to be sensitive to both prosodic and syntactic complexity (Ferreira, 1991).

Stimuli

On critical trials, the target sentence contained an ambiguous prepositional phrase attachment, as in (1a) and (1b) below. Identical sets of objects were given to both participants, and the objects in the bags were the same across all conditions. The objects for Example 1 appear in Fig. 1. On each trial the bag contained: (1) a Target Instrument, a full scale object that could be used to carry out the action (in Fig. 1, the large flower); (2) a Marked Animal, a stuffed animal carrying a small replica of the instrument (the frog holding a little flower); (3) an Unmarked Animal (the empty-handed frog); (4) a Distractor Animal, an unrelated animal wearing or carrying a different miniature object (the giraffe in
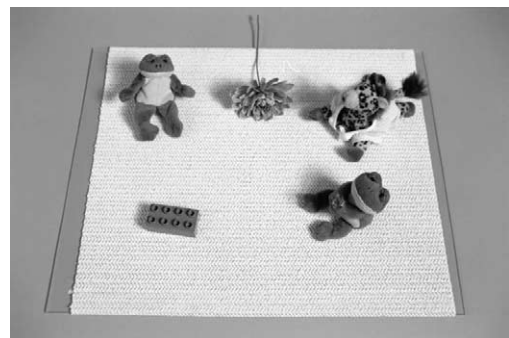

Fig. 1. Referential context for both Listeners and Speakers in Experiment 1 for the instruction "Tap the frog with the flower."

pajamas); and (5) a Distractor Object, an unrelated full-scale object (the lego block). The set of toys supported both interpretations of the ambiguous sentence by providing a potential instrument (large flower) and two possible direct objects (the frogs) for the VP-attachment and a potential direct object for the NP-attachment (frog holding flower).

One of two actions was demonstrated to the Speaker by the Experimenter. The Instrument Demonstration involved the Instrument and the Unmarked Animal (e.g., the Experimenter picked up the large flower and tapped the frog that was not holding anything). The Modifier Demonstration involved the Marked Animal and did not involve the Instrument (e.g., using her hand, the Experimenter tapped the frog that had the small flower). Ambiguous sentences were compared with unambiguous sentences (1c and 1d).

1. (a) *Tap the frog with the flower.*
   (Ambiguous, Instrument Demonstration)
   Action involves the unmarked frog and the instrument.
   (b) *Tap the frog with the flower.*
   (Ambiguous, Modifier Demonstration)
   Action involves the marked frog but not the instrument.
   (c) *Tap the frog by using the flower.*
   (Unambiguous, Instrument Demonstration)
   Action involves the unmarked frog and the instrument.
   (d) *Tap the frog that has the flower.*
   (Unambiguous, Modifier Demonstration)
   Action involves the marked frog but not the instrument.

Note that these referential contexts should favor the modifier interpretation of the utterance. Without the modifier, no single frog is uniquely identified and the definite determiner is infelicitous. We chose these contexts for two reasons. First, we realized that any particular context would favor one interpretation (e.g., a one-frog context would make the modifier interpretation, awkward and over-informative). Second, to maximize our chances of finding effects of prosody on interpretation, we wanted to create stimuli (sentence-scene pairs) which did not have a strong bias toward either interpretation. By using contexts that pushed the listener toward a modifier interpretation, we hoped to offset the effects of lexical biases favoring an instrument interpretation (see Spivey-Knowlton & Sedivy, 1995).

Four presentation lists were constructed by combining 16 target trials with 30 fillers. Filler trials contained a variety of objects and sentence types. Within a list, each target trial appeared in only one of the four conditions illustrated in Example 1, resulting in four target trials in each condition. Targets and fillers appeared in a pseudo-random order, with the only constraint being that at

least one filler appeared between target trials. Each target trial was then rotated through these four conditions, generating four lists each with the same order. Four reverse-order lists were also generated, to control for possible order effects. Each participant-pair was assigned to one of the eight presentation lists. The experiment, including instructions, lasted between 50 and 70 min.

Coding

Actions. Coders, who were blind to the condition of each trial, classified silent videotapes of Listeners' actions into four categories: (1) Instrument Response: Listener used the Target Instrument to perform action on the Marked or Unmarked Animal; (2) Mini-Instrument Response: Listener used the miniature object attached to the Marked Animal to perform action on the Marked or Unmarked Animal; (3) Modifier Response: Listener used his/her hand to perform action on the Marked Animal; (4) Other-Object Modifier Response: Listener used the Distractor Object to perform action on the Marked Animal. Three trials were excluded because the Speaker did not produce the target utterance. In the analyses reported below, we grouped Mini-Instrument Responses with Instrument Responses and Other-Object Modifier Responses with Modifier Responses because of their distribution in the unambiguous trials. Since only 5% of the responses fell into to either of these categories, this coding decision is not critical to the findings described below.

Instrument Responses were coded as such regardless of the animal used. One could argue that instrument actions on a Marked Animal are ambiguous. We rejected this hypothesis because such actions were frequent in response to unambiguous instrument sentences (13.3% of the trials) and rare for unambiguous modifiers (0.7%). Instead, we conclude that these actions signal a VP-attachment of the ambiguous phrase. While a modifier uniquely identifies the direct object, an instrument interpretation leaves the referent ambiguous. Subjects are forced to rely on the pragmatics of the situation to resolve the reference and the basis on which they do so varies. Overall only 7% of responses involved an Instrument Action on the Marked Animal.

Speakers' prosody. Acoustic and phonological analyses of the ambiguous target sentences were performed using speech waveform displays of the Speakers' target utterances.

For the acoustic analyses coders, who were blind to the condition, measured the duration of: the verb, the pause after the verb (if any), the direct object noun, the pause after the noun (if any), and the prepositional phrase. Word onsets or offsets were initially estimated by using visual information from the speech waveform display and then revised by listening to gated regions of the waveform. The acoustic analyses focused solely on

pauses and word durations for two reasons. First, duration appears to have a straightforward monotonic relation with the underlying prosodic structure; stronger prosodic boundaries are marked by greater pre-boundary lengthening or longer pauses (Ferreira, 1993; Liberman & Sproat (1992)). Second, the association between changes in duration and syntax is well studied. Many production studies have found effects of syntax on syllable duration and pause length (Cooper & Paccia-Cooper, 1980; Klatt, 1975; Lehiste, 1973b). Conversely comprehension studies have found effects of syllable or pause length on interpretation in the absence of pitch variation (e.g., Lehiste, Olive, & Streeter, 1976; Scott, 1982). When duration and pitch contour are pitted against one another, duration appears to play a greater role in the interpretation of a structurally ambiguous string (Streeter, 1978; Beach et al., 1996).

Our phonological analyses were conducted by a single highly trained coder who was blind to the experimental condition. This coder used the ToBI labeling system (Silverman et al., 1992) which represents the relative prominence of words in an utterance and their prosodic grouping. According to the prosodic theory underlying the ToBI coding system, there are two levels of prosodic structure between the level of the utterance and the prosodic word, the intermediate phrase and the intonational phrase (Beckman & Pierrehumbert, 1986). Each intermediate phrase contains at least one pitch accent and ends in a high or low phrase tone. Intermediate phrases are grouped together into intonational phrases. An intonational phrase contains at least one intermediate phrase and ends in a high or low boundary tone (which follows the phrase tone of the final intermediate phrase). While prosodic theories vary in the number of hierarchical levels that they recognize, most include levels that roughly correspond to the intermediate and intonational phrase (see e.g., Nespor & Vogel, 1986; Selkirk, 1986).

In our analyses, the coder first noted the pitch accents and boundary tones for the entire utterance and then determined: (1) the index value for the break following the verb; (2) the index value for the break following the direct object noun; and (3) whether there was a pitch accent on the preposition. The break indices could be used to determine the type of prosodic break between two words. An index of 4 indicates an intonational phrase boundary, 3 indicates an intermediate phrase boundary, 2 is used for boundaries that are ambiguous between an intermediate phrase and a word (because tonal and durational information conflict), while 1 indicates a simple word boundary (Beckman & Hirschberg, 1994).

Results and discussion

Actions

The percentage of Instrument actions for each condition is presented in Fig. 2. Perfect performance on this task would consist of performing an Instrument action for all of the trials with Instrument Demonstrations and none of the trials with Modifier Demonstrations. As can be seen in the figure, Listeners' actions in response to ambiguous instructions were affected by the action demonstrated to the Speaker. When an instrument action had been demonstrated to the Speaker, Listeners produced an instrument action 66% of the time. When a modifier action had been demonstrated, Listeners produced an Instrument Response only 24% of the time. Thus, Listeners chose the correct meaning of the ambiguous phrase on about 70% of the trials. If we assume that prosodic cues are categorical (either present or absent, heeded or ignored), then this level of performance would indicate that the utterance was disambiguated on 41% of the trials. If we assume instead that the prosodic cues which differentiate high and low attachments are graded and normally distributed, then this performance
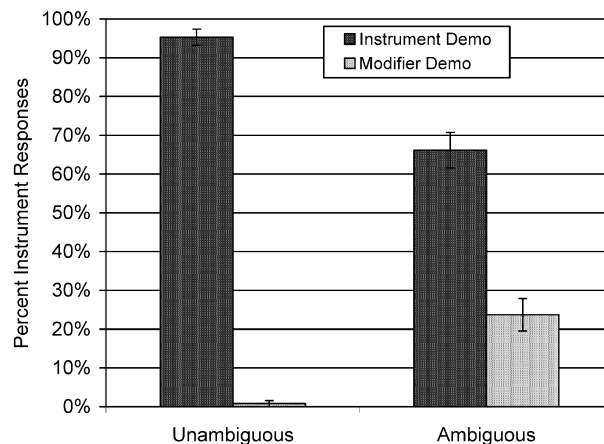

Fig. 2. Listeners' actions in Experiment 1, percentage of Instrument Responses to ambiguous and unambiguous target sentences.

would indicate a moderate level of discriminability ($d' = 1.12$) with little response bias ($\beta = 1.18$). We also found, as expected, that unambiguous instructions (the left-hand portion of Fig. 2) resulted in extremely accurate performance by Listeners.

Effects of Demonstration. Subject and item means were computed for the percentage of Instrument Responses. An analysis of variance (ANOVA) was conducted on the subject means containing three between-subject factors (List, Order, and Gender of Listener) and two within-subject factors (Demonstration and Ambiguity). An equivalent ANOVA was conducted on item means containing one between-item factor (Item Group) and four within-item factors (Order, Gender of Listener, Demonstration, and Ambiguity).

When Listeners were forced to rely on prosody to interpret ambiguous sentences, their performance was above chance but reliably lower than their performance on syntactically unambiguous structures. This resulted in: (1) a reliable interaction between Ambiguity and Demonstration ($F1(1, 16) = 81.91$, $p < .001$; $F2(1, 12) = 113.71$, $p < .001$) and (2) reliable effects of Demonstration for simple effects analyses on both the Unambiguous ($F1(1, 16) = 2928.20$, $p < .001$; $F2(1, 12) = 2311.74$, $p < .001$) and the Ambiguous trials ($F1(1, 16) = 63.42$, $p < .001$; $F2(1, 12) = 77.31$, $p < .001$). As expected from inspecting Fig. 2, ANOVAs also yielded a significant main effect of Demonstration when collapsing across the other factors ($F1(1, 16) = 628.67$, $p < .001$; $F2(1, 12) = 677.82$, $p < .001$). We found no gender differences in the Listeners' ability to interpret the utterances (all F's < 1.5, all p's $> .2$). In the item analysis there was also a significant interaction involving all five variables, including control variables such as item-group; the parallel interaction was not reliable in the subject analysis. In many of the analyses that we will report, there were similar complex interactions or main

effects of the control variables. Because we believe that they have no bearing on the proposals that we will be making, these effects will not be discussed further.

Performance across the experiment. To determine whether the Listener's performance improved as she/he gained experience with the task and the Speaker, we conducted a subject ANOVA of the Ambiguous trials from the first and last half of the experiment. There was a marginal effect of Experiment Half ($F(1, 24) = 3.90$, $p < .06$), suggesting that subjects' response biases changed during the course of the experiment. However, the effect of Demonstration did not interact with Experiment Half ($F(1, 24) = 1.92$, $p > .15$) indicating that the difference between the responses to Instrument and Modifier Demonstrations was stable across the two halves of the experiment. We explored this further by analyzing the Listener's response to the very first ambiguous item they heard. For this item, a χ^2 test revealed that Demonstration had a reliable effect on Response Type ($\chi^2(1) = 4.10$, $p < .05$) with Listeners responding correctly 69% of the time. Apparently subjects did not have to encounter repeated examples of these utterances to disambiguate the attachment.

Speaker's prosody: Acoustic analyses

To determine how our Listeners managed to infer the Speaker's intended meaning, we measured the duration of words and pauses in the ambiguous target sentences. Fig. 3 graphs the mean values of these measurements for the Ambiguous Modifier and Instrument trials. Each section of the bars stands for the duration of a word, pause or phrase in the target sentence. The value along the *x*-axis indicates the cumulative duration from the onset of the utterance. Under each of the bars is a sample sentence, which is aligned to indicate the relative duration of different regions of the target sentence.
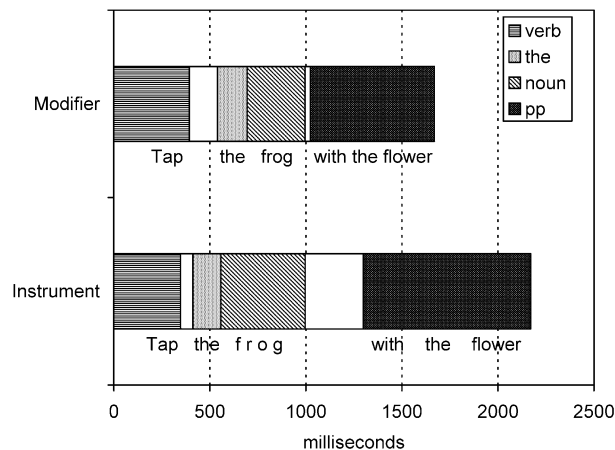

Fig. 3. Time course of the target utterances for Experiment 1 (ambiguous referential contexts).

Overall, when Speakers saw Instrument Demonstrations, they tended to lengthen the direct object noun and they paused between the noun and the with-phrase on 68% of the trials. This prosodic pattern suggests that the major phrase boundary is located between the direct object and the prepositional phrase and is thus consistent with a verb-phrase attachment of the prepositional phrase (instrument interpretation) but not with a noun-phrase attachment (modifier interpretation). In contrast, when Speakers saw Modifier Demonstrations, they tended to lengthen the verb and paused after it 40% of the time. This prosodic pattern suggests that the major phrase boundary is located between the verb and the direct object noun phrase and is more consistent with a noun-phrase attachment. The length of the prepositional phrase also tended to be shorter for these Modifier sentences, perhaps because this prosodic pattern forced the subject to cram most of the utterance into a single prosodic unit.

Effects of demonstration. We conducted a subject and an item ANOVA for the mean duration of the each of the following regions: the verb, the post-verbal pause, the Verb Composite (verb plus the post-verbal pause), the direct object noun, the pause following the direct object noun, the Noun Composite (the direct object noun plus the following pause), and the prepositional phrase. We also analyzed the duration of the article preceding the direct object noun. Since we did not expect to see any effect of Demonstration on this word, it provides a comparison for effects in the other regions. The subject analyses included two between-subject variables, List and Order, and one within-subject variable, Demonstration. The item analyses, contained one between-item variable, Item Group, and

two within-item variables, Order and Demonstration. The results of these analyses are summarized in Table 1 and clearly confirm the patterns suggested by Fig. 3. Demonstration had a large and reliable effect on all the critical measures. Only the duration of the determiner was unaffected by the Speaker's intended meaning.

Performance across the experiment. To examine whether the effects of Demonstration resulted solely from practice and cross-trial comparisons, we conducted subject ANOVAs of the utterances produced on the first target trial. These analyses have considerably less power than the preceding ones, since only half of our subjects received ambiguous sentences on the first trial and only two items appeared in this position. Nonetheless, Demonstration had a reliable effect on the length of the Noun Composite and prepositional phrase ($F(1, 12) = 6.24$, $p < .05$ and $F(1, 12) = 17.93$, $p < .001$) and marginal effect on the length of the direct object noun and the postnominal pause ($F(1, 12) = 4.24$, $p = .062$ and $F(1, 12) = 4.60$, $p < .053$ respectively). Thus we conclude that the intended meaning had an impact on the Speakers' prosody from the outset.

Speaker's prosody: Phonological analyses

The strong effect Demonstration on the Speakers' prosody was confirmed in the ToBI analyses. Subjects producing modifier sentences often placed an intonational phrase break after the verb but rarely placed one after the direct object noun. In contrast, speakers producing instrument sentences typically ended their first intonational phrase after the direct object noun and then placed a pitch accent on preposition. We calculated the average break index in the postverbal and postnominal

Table 1
Effects of Demonstration in the duration analyses, Experiment 1 (ambiguous referential contexts)

Dependent variable	Mean for instrument (ms)	Mean for modifier (ms)	Subject analysis	Item analysis
Verb length	348	393	$F1(1, 24) = 6.55$ $p < .05^*$	$F2(1, 12) = 26.23$ $p < .001^{**}$
Verb pause	63	147	$F1(1, 24) = 13.40$ $p < .001^{**}$	$F2(1, 12) = 30.07$ $p < .001^{**}$
Verb Composite	411	540	$F1(1, 24) = 12.92$ $p < .001^{**}$	$F2(1, 12) = 50.59$ $p < .001^{**}$
Direct object noun	440	302	$F1(1, 24) = 39.92$ $p < .001^{**}$	$F2(1, 12) = 121.84$ $p < .001^{**}$
Noun pause	302	27	$F1(1, 24) = 47.96$ $p < .001^{**}$	$F2(1, 12) = 169.91$ $p < .001^{**}$
Noun Composite	742	329	$F1(1, 24) = 52.71$ $p < .001^{**}$	$F2(1, 12) = 290.42$ $p < .001^{**}$
Prepositional phrase	872	646	$F1(1, 24) = 44.93$ $p < .001^{**}$	$F2(1, 12) = 51.83$ $p < .001^{**}$
Determiner for direct object noun	146	154	$F1(1, 24) < 1$ $p > .5$	$F2(1, 12) = 1.73$ $p > .2$

Table 2
Effects of Demonstration in the phonological analyses, Experiment 1 (ambiguous referential contexts)

Dependent variable	Mean for instrument	Mean for modifier	Subject analysis	Item analysis
Verb break index	1.89	2.76	$F1(1,24) = 18.58$ $p < .001^{**}$	$F2(1,12) = 48.25$ $p < .001^{**}$
IP boundary after verb	13.3%	50.1%	$F1(1,24) = 32.77$ $p < .001^{**}$	$F2(1,12) = 97.91$ $p < .001^{**}$
Noun break index	3.59	1.48	$F1(1,24) = 83.85$ $p < .001^{**}$	$F2(1,12) = 626.45$ $p < .001^{**}$
IP boundary after direct object noun	75.5%	8.3%	$F1(1,24) = 73.18$ $p < .001^{**}$	$F2(1,12) = 201.85$ $p < .001^{**}$
Accented "with"	74.7%	14.8%	$F1(1,24) = 61.27$ $p < .001^{**}$	$F2(1,12) = 125.78$ $p < .001^{**}$
Break indices prediction (% inst)	86.6%	29.3%	$F1(1,24) = 65.96$ $p < .001^{**}$	$F2(1,12) = 509.81$ $p < .001^{**}$

position, the percentage of intonational phrase boundaries in these positions and the percentage of sentences where the preposition had a pitch accent and then conducted a subject and an item ANOVA for each of these cues.[1] As Table 2 illustrates all of these cues showed a reliable effect of Demonstration.

Recently, several researchers have found that the relation between syntax and prosody is best captured by examining the entire prosodic structure of the utterance, rather than individual prosodic boundary points (Carlson et al., 2001; Pynte & Prieur, 1996; Schafer et al., 2000a). In the case of attachment ambiguities, the critical comparison appears to be between the prosodic boundary immediately before the ambiguous phrase and the boundary preceding the phrase that could serve as the lower attachment site (Carlson et al., 2001). Thus for our target trials, instrument utterances should have a larger break before the prepositional phrase (coded here as the noun break index), whereas modifier utterances should have a larger break before the direct object noun phrase (coded here as the verb break index). Following Schafer, Warren, Speer, White, and Sokol (2000b), we compared the break indices at these locations and classified the utterances as having instrument prosody (noun break > verb break, coded as 1), neutral prosody (noun break = verb break, coded as .5 to represent chance in a two-alternative forced choice task) or modifier prosody (noun break < verb break, coded as 0). The results of these analyses are given in Table 2 and illustrated in Fig. 4. Overall 68% of the utterances were spoken with appropriate and disambiguating phrasing,

about 22% of the utterances were spoken with neutral prosodic phrasing, and only 10% of the utterances were spoken with phrasing that was more appropriate for the alternate interpretation. An ideal observer who interpreted the sentences on the basis of this coding scheme, would show a higher level of discriminability than our untrained subjects ($d' = 1.68$ for the coding scheme vs. $d' = 1.12$ for the Listeners) and would have a moderate bias for instrument interpretations ($\beta = .62$). Relying solely on the presence of an intonational phrase break after the noun would further improve discriminability ($d' = 2.11$) and reverse the bias in favor of modifier interpretations ($\beta = 2.09$).

Awareness of ambiguity

To measure our subjects' level of ambiguity awareness, 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. 97% of the Speakers and 91% of Listeners did so, thus it is entirely possible that prosodic disambiguation arose because Speakers were aware of the alternative meanings of the critical items.

Summary of Experiment 1

Listeners' actions in response to ambiguous instructions suggest that prosodic cues were a highly effective but imperfect means of syntactic disambiguation. When Speakers produced an ambiguous utterance like "Tap the frog with the flower," Listeners correctly reproduced the action demonstrated on the other side of the divider roughly 70% of the time. Acoustic and phonological

---

[1] The digital speech files for one subject were corrupted prior to the phonological analysis. In the subject ANOVAs, these values were replaced by the cell mean for those conditions. Nine additional speech files, from four different subjects were also corrupted. These data points were simply excluded from the subject and item averages.
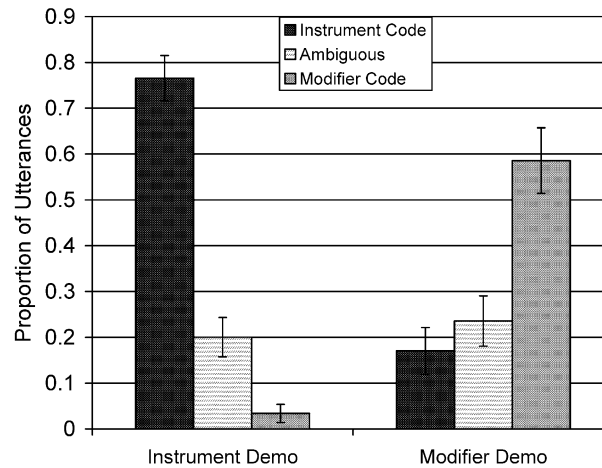
Fig. 4. Prosodic phrasing of the target utterances. Coding system is based on a comparison of the relative size of the boundaries after the verb and the direct object noun (Experiment 1, ambiguous referential contexts).

analyses demonstrated that the placement of prosodic breaks provided the information that the Listeners needed to infer the Speakers' intention. However, the results of the post-experiment interviews raised some concerns about the generality of these findings. Almost all of the Speakers and Listeners reported being aware of the ambiguity. As mentioned earlier, Allbritton et al. (1996) found that ambiguity awareness affected radio announcers' ability to generate useful prosody. Although our participants were not trained speakers, we thought it necessary to explore whether ambiguity awareness, and more generally knowledge of the referential situation, were influencing the kinds of prosodic choices that they made.

Experiment 2

In this experiment, we attempted to decrease Speaker awareness of ambiguity. To do this we, made two changes to the previous experiment. First, we altered the Speakers' referential context, such that only the intended meaning of the ambiguous phrase was supported. This was done in hopes that the alternative interpretation would not be considered by Speakers if the context supported only the intended meaning. Second, we made the type of Demonstration a between-subject variable. This was done because we were concerned that the presence of the two alternative meanings across trials would lead subjects to notice the ambiguity. We arrived at this design by piloting various changes on small groups of subjects. Neither change alone seemed to decrease the Speaker's awareness of ambiguity.

In most other respects, Experiment 2 was the same as Experiment 1. It is particularly important to note that the Listeners' context was the same ambiguous context

used in Experiment 1 (see Fig. 1). Also, participants were told in advance that the Speaker and Listener would receive an identical set of toys on each trial, just as they had been told in Experiment 1. However, in Experiment 2 this was a deception, to be explained at the end of the study.

Methods

Participants

Thirty-two pairs of volunteers from the University of Pennsylvania community received extra course credit or were paid for their participation. All Speakers were female, and 17 of the Listeners were male. All participants were native speakers of English and none had participated in Experiment 1. Two additional pairs participated but were not included in the analyses because of experimenter error (1) or failure to follow instructions on two or more test trials (1).

After the initial 32 pairs of subjects were completed, an additional 10 pairs were tested so that we could separately examine the performance of the subset of pairs in which the Speaker was not aware of the ambiguity ("unaware pairs") and compare the performance of this group with that of the remaining ("aware") pairs. The additional pairs were assigned to those cells in which there were fewer than 4 unaware pairs. Unless otherwise noted the analyses below are based on the initial set of 32 subject pairs.

Procedure

The procedure was the same as Experiment 1 except that the Experimenter did not announce the contents of the bags. This prevented the subjects from discovering that the two bags contained different sets of objects. Instead, a card listing the objects was included in each
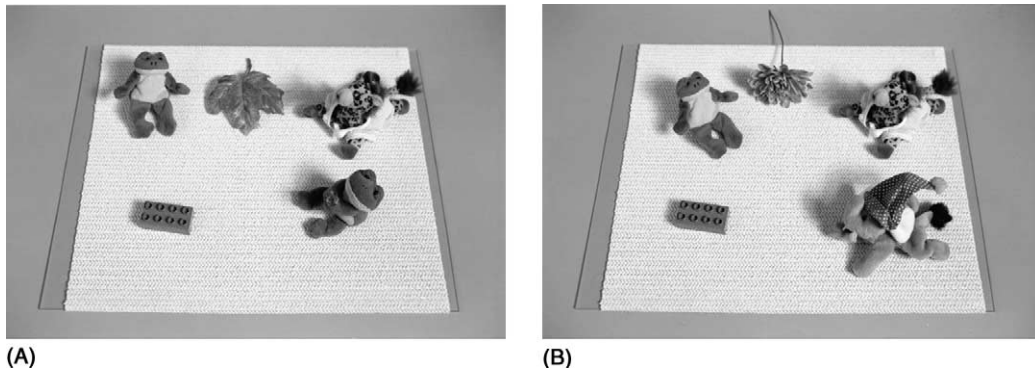
Fig. 5. Referential contexts for the Speakers in Experiment 2. (A) The context for Speakers in the Modifier Condition for the "Tap the frog with the flower." (B) The context for the Speakers in the Instrument Condition for this sentence. All Listeners had the referential context shown in Fig. 1.

bag of toys. Both the Listener and the Speaker were asked to check the contents of the bags against the card to ensure that all of the toys were present. After the experiment, the Speaker and Listener were separated and given the same ambiguity interview as in Experiment 1.

Stimuli

The stimuli and experimental design were the same as Experiment 1, with the following exceptions. First, the type of Demonstration was manipulated between subjects. Thus, for a given Speaker, target items involved either all Instrument Demonstrations or all Modifier Demonstrations. Second, the Speaker's referential context, but not the Listener's, was modified so that it would not support the alternative meaning of the ambiguous sentence. Thus in the Modifier Condition, the Experimenter performed Modifier Demonstrations for all of the target trials and the Speaker always had a Modifier Context (e.g., see Fig. 5A). In these contexts, the Target Instrument from Experiment 1 (the large flower in Fig. 1) was replaced with an unrelated object (e.g., a leaf), making the instrument interpretation of "with the flower" less available to Speakers in this context. In the Instrument Condition the Experimenter performed only Instrument Demonstrations and the Speaker always had an Instrument Context (e.g., see Fig. 5B). The Instrument Contexts were constructed by replacing the Marked Animal from Experiment 1 (the frog holding the flower in Fig. 1) with another unrelated animal (e.g., an elephant wearing a hat), making the modifier interpretation of "the frog with the flower" unavailable to the Speaker.

Finally, we excluded the syntactically unambiguous conditions from this experiment, because Listeners' performance on these conditions in the last study was (quite predictably) nearly perfect. Thus all the target trials in Experiment 2 consisted of sentences containing "with" prepositional phrase attachment ambiguities. To equalize the number of ambiguous sentences that subjects received in each experiment, we divided the 16 critical sentences into two lists. The items on each list appeared in a pseudo-random order embedded in the same 24 filler trials. In addition, reverse-order lists were generated. Presentation List and Demonstration Condition were fully crossed and balanced within the initial set of 32 of subject pairs.

Coding

The videotapes of the Listener's actions were coded in the manner described above. Three test trials in which the Speaker did not produce the target sentence were excluded from further analysis. A fourth trial, in which the Listener received the wrong toys was dropped from the action analyses. All of the audiotapes were digitized and the durations of words and pauses were measured using the same coders and procedures as in Experiment 1. The digitized recordings for the initial set of 32 subjects were also transcribed using the ToBI system by the same trained coder who had analyzed the recordings for Experiment 1.[2]

Results

As alluded to in the method section, a subset of the Speakers in this experiment were still aware that the target sentences were potentially ambiguous. We postpone comparison of data from aware and unaware Speakers until after we present the data patterns independent of speaker awareness.

Actions

The percentage of Instrument Responses for the two conditions appear on the right-hand side of Fig. 6. For comparison, Listener performance on the Ambiguous trials of Experiment 1 is presented on the left-hand side

---

[2] One sound file was corrupted prior to the ToBI analyses.
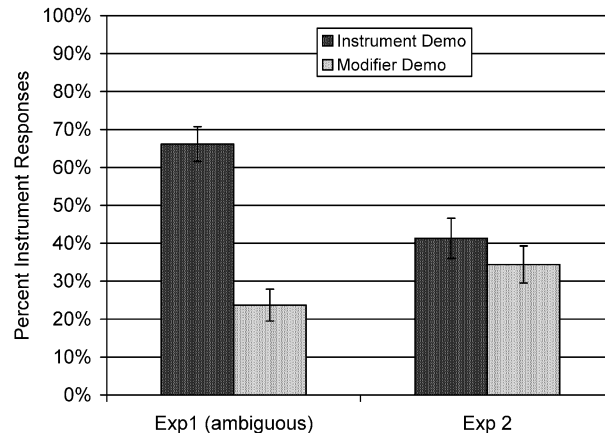
Fig. 6. Listeners' actions, the percentage of Instrument Responses to the ambiguous sentences in Experiments 1 and 2.

of Fig. 6. As seen in the figure, Listeners in Experiment 2 were unaffected by the type of Demonstration performed by the Experimenter. Listeners performed instrument actions 41% of the time when the Speaker had seen an Instrument Demonstration and 34% of the time when the Speaker saw a Modifier Demonstration, indicating that they were unable, as a group, to distinguish between the two interpretations on the basis of prosody ($d' = .18$, $\beta = 1.06$).

Subject and item ANOVAs were conducted on the percentage of instrument Responses with three between-subject factors (List, Order, and Demonstration), one between-item factor (Item Group) and two within-item factors (Order and Demonstration). No significant effects or interactions were observed. Critically, Listeners were unaffected by the type of Demonstration ($F1(1, 24) < 1$, $p > .3$; $F2(1, 12) = 1.88$, $p > .15$). This pattern persisted in the expanded set of 42 subject pairs. The mean percentage of Instrument Responses was 41% for pairs in the Instrument Condition and 39% for those in the Modifier Condition and there was no effect of Demonstration on response ($F1(1, 40) = .10$, $p > .75$; $F2(1, 15) = .27$, $p > .5$).[3]

To compare Experiments 1 and 2, we conducted an item ANOVA on the conditions shown in Fig. 6 with Experiment and Demonstration as within-item factors. There was a main effect of Demonstration ($F(1, 15) = 38.87$, $p < .001$). But, critically, there was

also a strong interaction between Demonstration and Experiment ($F(1, 15) = 19.71$, $p < .001$). A parallel subject ANOVA on the percentage of Instrument responses could not be conducted because Demonstration type was, necessarily, a within-subject variable in Experiment 1 and a between-subject variable in Experiment 2. However, there was a reliable difference in the total number of correct responses by subject in Experiments 1 and 2 ($t(30) = 3.43$; $p < .005$). This difference in the Listener's performance across the two experiments could reflect either a difference in the strength of prosodic cues that were available, or a difference in the Listener's ability to use those cues. Remember that in Experiment 2, each Speaker was asked to convey the same type of meaning for every one of the target trials. In contrast, the Speakers in Experiment 1 saw both Instrument and Modifier Demonstrations. If our Listeners have a strong bias to assume that both meanings will appear equally often, or if they need to hear both types of utterances to make use of the Speakers prosody, then we would expect poorer performance in Experiment 2 regardless of cue strength. However, we would also expect the difference between the experiments to emerge only after the Listener had already heard other ambiguous target sentences.

To explore this possibility, we performed a χ^2 test on the first ambiguous test trial for each of the subjects in both experiments. Responses were coded as correct or incorrect and the test was conducted on the relation between Experiment and Response Type. There was a reliable difference in the distribution of responses between the Experiments ($\chi^2(1) = 4.04$, $p < .05$). While Listeners in Experiment 1 were correct for 69% of initial target trials, Listeners in Experiment 2 were right for only 41% of these trials. Thus, we conclude that our Listener's inability to use prosody in Experiment 2 is not simply an artifact of the between-subject manipulation of Demonstration.

---

[3] In Experiment 2, subjects were more likely to inflict an Instrument action on the Marked Animal than they were in Experiment 1 (13% of the responses in the Instrument Condition, 11% of those in the Modifier Condition). As mentioned above, we do not view these responses as ambiguous actions because they occur in response to unambiguous instrument sentences. But even if we exclude these responses or code them as ambiguous there are no reliable effects of Demonstration (all p's > .05).
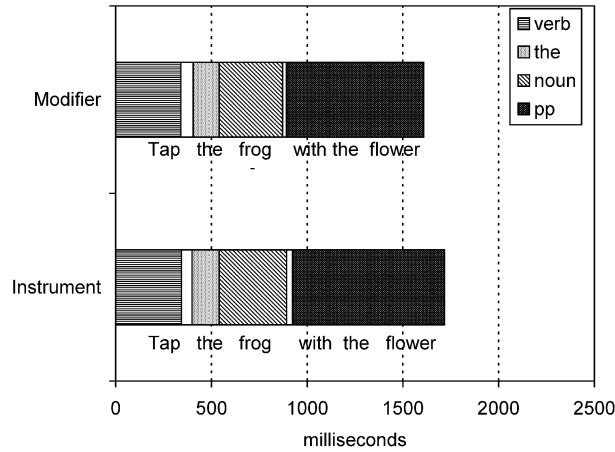
Fig. 7. Time course of the target utterances for Experiment 2 (unambiguous referential contexts).

Speaker's prosody: Acoustic analyses

Fig. 7 shows the cumulative duration of the utterances in both conditions. The results of the ANOVAs on the duration variables are summarized in Table 3. As Fig. 7 suggests, the differences between the conditions are small and for the most part unreliable. There was a significant effect of Demonstration on the length of the prepositional phrase but even this effect was only a third of the size it had been in Experiment 1. Critically, there were no reliable differences at the verb, where we might expect to find a prosodic break for modifier sentences. The analyses of the direct object durations are more equivocal: Demonstration has a small but reliable effect in the item analysis and no effect in the subject analysis. This might be attributable to the greater sensitivity of an analysis which necessarily corrects for differences in the intrinsic lengths of words. Or it could indicate that this

effect is being driven by a small subset of the subjects, perhaps those who were aware of the ambiguity.

Experiments 1 and 2 were compared by conducting item ANOVAs on the acoustic variables, with Experiment and Demonstration as within-item factors. Strong interactions occurred between Demonstration and Experiment for all seven critical measures (all $F's(1,15) > 10$, all $p's < .005$) but not for the length of the determiner ($F(1,15) = 1.24$, $p > .2$).

Speaker's prosody: Phonological analyses

In Experiment 1, the modal response of speakers was to produce utterances with two intonational phrases for both the instrument and the modifier sentences. The placement of this intonational phrase boundary provided a reasonably reliable indicator of the intended meaning of the utterance. In Experiment 2, the modal

Table 3
Effects of Demonstration in the duration analyses, Experiment 2 (unambiguous referential contexts)

Dependent variable	Mean for instrument (ms)	Mean for modifier (ms)	Subject analysis	Item analysis
Verb length	343	341	$F1(1,24) < 1$ $p > .9$	$F2(1,14) < 1$ $p > .8$
Verb pause	56	63	$F1(1,24) < 1$ $p > .3$	$F2(1,14) = 2.61$ $p > .1$
Verb Composite	399	404	$F1(1,24) < 1$ $p > .8$	$F2(1,14) < 1$ $p > .4$
Direct object noun	353	331	$F1(1,24) = 1.72$ $p > .2$	$F2(1,14) = 8.53$ $p < .05^*$
Noun pause	30	19	$F1(1,24) < 1$ $p > .3$	$F2(1,14) = 3.33$ $p = .09$
Noun Composite	383	350	$F1(1,24) = 1.66$ $p > .2$	$F2(1,14) = 7.71$ $p < .05^*$
Prepositional phrase	794	717	$F1(1,24) = 7.58$ $p < .05^*$	$F2(1,14) = 80.64$ $p < .001^{**}$
Determiner for direct object noun	141	137	$F1(1,24) < 1$ $p > .5$	$F2(1,14) = 1.20$ $p > .3$

response in both conditions was an utterance with just one intonational phrase. Consequently the relation between particular prosodic cues and syntactic structure was weak and probabilistic (Fig. 8). These observations were formalized by conducting subject and item ANOVAs on the average break index after the verb and direct object noun, the percentage of intonational phrase boundaries in these positions, the percentage of accented prepositions, and the proportion of sentences predicted to have instrument interpretations based on a comparison of the boundaries following the verb and direct object noun. Table 4 summarizes the results.

The effects of Demonstration condition on prosodic phrasing are generally small and reliable only in the item analyses. Fig. 8 gives the percentage of utterances which were coded as having instrument, modifier or ambiguous prosody based on the relative size of the break indices following the verb and noun (following Schafer et al., 2000b). Although the Instrument Condition shows clear distinctions between the rate of the three coding categories, the information value of such a distribution is low because the summed total proportion of ambiguous and incorrect (modifier) prosodic phrasings is roughly the same as the correct (instrument) phrasing. That is, the rate of uninformative and misleading prosody matches the rate of helpful prosody. In the Modifier Condition the trend is if anything in the wrong direction: the rate of uninformative and misleading prosody is greater than the rate of helpful prosody. An ideal observer who used the boundary comparison as a basis for interpreting the sentences, would have only a marginal ability to discriminate the two sentence types ($d' = .44$, $\beta = .90$). A simpler model which bases its prediction on the presence of an intonational phrase break after the noun would have greater discriminability ($d' = .67$) but a bias toward modifier responses ($\beta = 1.82$). This is approxi-

mately the same level of prosodic disambiguation provided by the untrained, naïve speakers in the Allbritton study, who were given an unambiguous linguistic context ($d' = .56$, $\beta = .77$).

Thus speakers in Experiment 2 were quite poor at producing the prosodic cues that are believed to signal the location of phrasal attachment. This result contrasts sharply with the findings of Experiment 1 (see Figs. 4 and 8). To compare the two experiments, we conducted item ANOVAs on each of the variables listed above with Experiment and Demonstration as within-item factors. We found strong interactions between Demonstration and Experiment for all comparisons (all $F's(1, 15) > 7$, all $p's < .05$).

Ambiguity awareness

The Listeners and the Speakers were given the awareness questionnaire used in Experiment 1. The percentage of aware participants for both experiments appears in Fig. 9. It may be surprising that Speakers in unambiguous contexts would ever report being aware of ambiguity, since no referential ambiguity existed. However, our questions probed for awareness of the semantic ambiguity, independent of context. Thus, Speakers in Experiment 2 could be, and often were, coded as aware. Typically, aware Speakers noted the sentence could mean more than one thing but indicated only one interpretation was plausible given the context.

Listeners in Experiment 2 were just as likely to notice the ambiguity as those in Experiment 1. This is because the same referential contexts were presented to Listeners in both experiments. In contrast, our manipulation drastically decreased the Speakers' awareness of ambiguity. In the initial set of participants only one Speaker out of 16 in the Instrument Condition (6%) reported being aware of the ambiguity, while nine of the Speakers
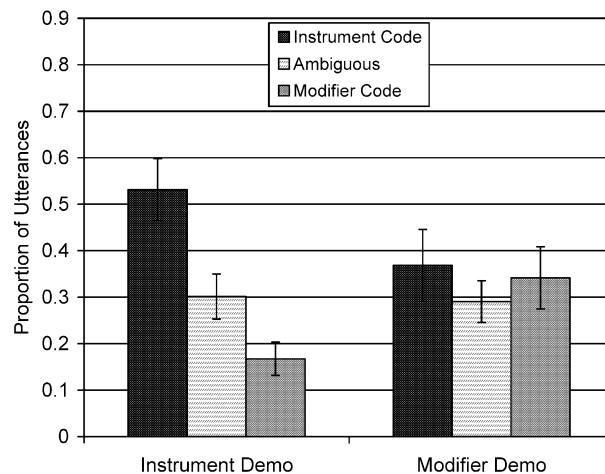

Fig. 8. Prosodic phrasing of the target utterances. Coding system is based on a comparison of the relative size of the boundaries after the verb and the direct object noun (Experiment 2, unambiguous referential contexts).

Table 4
Effects of Demonstration in the phonological analyses, Experiment 2 (unambiguous referential contexts)

| Dependent variable	Mean for instrument	Mean for modifier	Subject analysis	Item analysis
Verb break index	2.07	2.27	$F1(1, 24) = .73$ $p > .3$	$F2(1, 14) = 2.51$ $p > .1$
IP boundary after verb	13.2%	20.0%	$F1(1, 24) = .98$ $p > .3$	$F2(1, 14) = 2.84$ $p > .1$
Noun break index	2.81	2.34	$F1(1, 24) = 3.96$ $p = .058$	$F2(1, 14) = 29.63$ $p < .001**$
IP boundary after direct object noun	29.4%	10.9%	$F1(1, 24) = 3.50$ $p = .073$	$F2(1, 14) = 28.08$ $p < .001**$
Accented "with"	19.4%	4.7%	$F1(1, 24) = 3.81$ $p = .063$	$F2(1, 14) = 29.03$ $p < .001**$
Break indices prediction (% inst)	68.2%	51.3%	$F1(1, 24) = 3.50$ $p = .073$	$F2(1, 14) = 15.20$ $p < .005**$
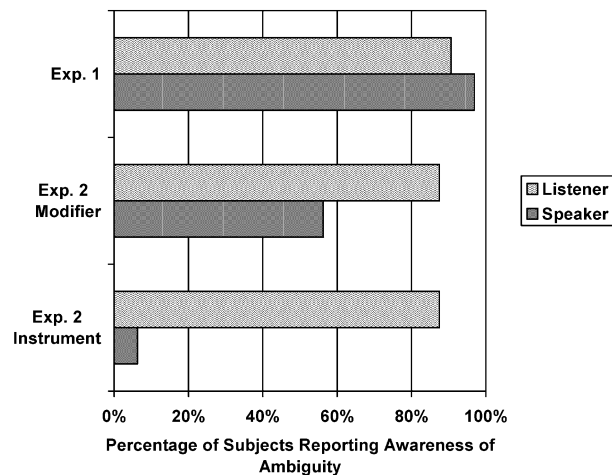
Fig. 9. The percentage of subjects who reported awareness of syntactic ambiguity in Experiments 1 and 2.

(56%) did so in the Modifier Condition. We attribute this asymmetry to verb biases: readers typically prefer instrument attachments when action verbs are used in the materials (Spivey-Knowlton & Sedivy, 1995; Taraban & McClelland, 1988; see also Rayner, Carlson, & Frazier, 1983). Perhaps conflicts between the verb and referential cues in the Modifier Condition led some Speakers to notice the ambiguity. While unexpected, this result is serendipitous, because it allows us to compare within the Modifier Condition the performance of the Listeners who heard utterances from either an aware or unaware Speaker. The decreased performance of Listeners in this experiment could be attributable to one of two factors: the general decrease in Speaker awareness of the ambiguity (Allbritton et al., 1996) or the change in the referential context for the Speakers. By comparing Listener performance from aware and unaware speakers we can test if ambiguity awareness per se is driving the results of Experiment 2.

Awareness and Listener's performance

Because only one Speaker in the Instrument Condition noticed the syntactic ambiguity, our analyses of awareness focused on pairs in the Modifier Condition. As mentioned earlier, we ran an additional 10 subject-pairs to increase the power in performing tests of awareness. In this expanded data set, there were a total of 25 Speakers in the Modifier Condition. Seventeen of them were aware of the ambiguity. The proportion of Instrument Responses on the part of Listeners appears to be unrelated to the Speaker's awareness ($F1(1, 23) < 1$, $p > .8$; $F2(1, 12) < 1$, $p > .99$). If anything, Listeners did slightly worse when their Speaker was aware of the ambiguity ($M = 38\%$, $M = 40\%$; for unaware and aware, respectively). Thus Speaker awareness alone does not determine prosodic disambiguation.

We cannot make the same aware–unaware comparisons for the Instrument Condition, because almost all Speakers in this condition were unaware of the ambi-
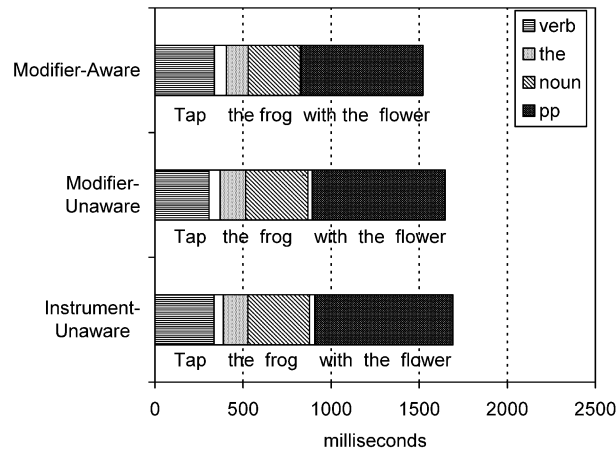
Fig. 10. Comparison of utterances from Aware and Unaware Speakers in the Modifier Condition and Unaware Speakers in the Instrument Condition, Experiment 2.

guity. However, Speakers in Experiment 1 were aware of the ambiguity, and when we make the comparison across the experiments we find that performance of Listeners in the Instrument Condition was indeed better in Experiment 1 (66% Instrument actions) as compared to Experiment 2 (41% Instrument actions). Thus, this across experiment comparison suggests that Speaker awareness may actually have an effect, which stands in contrast to what we just concluded. This apparent conflict in the data can easily be resolved, however, if one takes into account a key difference between Experiment 1 and Experiment 2: the referential context was different for Speakers in these two situations. In Experiment 1, the Speaker's referential context supported both interpretations of the "with" phrase and thus the sentence was, in the absence of prosodic cues, ambiguous in context. In Experiment 2, the Speaker's referential context supported only the relevant interpretation, disambiguating the sentence and making strong prosodic cues unnecessary. In conjunction with the Modifier Condition findings, therefore, the data support the hypothesis that speakers only produce reliable disambiguating prosody when the context doesn't do the work for them.[4]

Awareness and duration

We also examined whether Speaker awareness of ambiguity had any effect on the acoustic measures of

---

[4] While ambiguity awareness is not a sufficient condition for prosodic disambiguation, we cannot rule out the possibility that it is a necessary condition. In our pilot studies, we were unable to create circumstance in which referential ambiguity was present but the speakers remained unaware of the ambiguity. Thus it is possible that both awareness and referential ambiguity are necessary for prosodic disambiguation.

prosody, using the larger set of 42 subject pairs. Given the lack of an effect of Speaker awareness on the Listeners' actions, we might expect prosodic cues to be similar for aware and unaware speakers. Indeed this was the case. Fig. 10 plots mean word and pause durations for the three utterance types from which we have sufficient numbers of subjects: modifier utterances from Aware Speakers ($N = 17$); modifier utterances from Unaware Speakers ($N = 8$); and instrument utterances from Unaware Speakers ($N = 16$). Space limitations preclude a full discussion, but as one can see, durational differences between the groups are quite small. There were small but reliable differences between the Aware and Unaware Modifier utterances at the noun, the noun pause and prepositional phrase. However, Listeners were not sensitive to these differences and Speakers were not making the sorts of prosodic breaks found Experiment 1, where the context induced a need to disambiguate. There were no reliable differences between the Unaware Modifier and Unaware Instrument utterances.

First trial analysis

While we would like to conclude that the differences between the Speakers' utterances in the two experiments are attributable to the manipulation of the referential scene, another possibility remains. In Experiment 1, an individual Speaker saw both Instrument and Modifier Demonstrations, while in Experiment 2 the Speaker saw the same type of Demonstration on every target trial. Consequently, the Speakers in Experiment 1 may have decided that the role of the prepositional phrase was new information across trials and used their prosodic resources to highlight this critical contrast. For the Speakers in Experiment 2 the role of the prepositional phrase was old information and they may have chosen

to use prosody to mark the appearance of a new animal or object.[5]

If Speakers' cross-trial comparisons account for the differences between the two studies, then we should find that the Speakers in the two experiments produce similar utterances on the initial target trial—before there is any difference in the discourse context. Recall that the acoustic analyses of Experiment 1 revealed reliable effects of Demonstration on the first target utterances. Parallel analyses of Experiment 2, however, found no reliable effects of Demonstration, despite the fact that these analyses included twice as many subjects as the analysis of Experiment 1 (where half of the subjects had an unambiguous sentence on the first target trial). Critically there was no difference in the duration of the Verb Composite or the Noun Composite ($F(1, 24) < 1$; $p > .5$ for both). Thus Speakers in Experiment 2 failed to disambiguate the utterances on very first test trial, ruling out the possibility that this inability was related to the contrast set that was created across trials.

Experiment 3

Together Experiments 1 and 2 suggest that the production of informative prosodic cues depends upon the speaker's knowledge of the situation: speakers provide prosodic cues when needed; listeners use these prosodic cues when present. This is somewhat surprising. Cue strength should and usually does depend on cue reliability. If strong prosodic cues are provided only when the referential context is ambiguous, and perhaps only when speakers become aware of this, then we might expect that such cues would be infrequent in ordinary conversations. Afterall, as speakers, we are rarely aware of referential ambiguity and yet temporary syntactic ambiguity is pervasive. Nevertheless the Speakers and Listeners in Experiment 1 were able to produce and use these cues despite the absence of feedback.

Still more puzzling, are the growing number of studies that demonstrate that prosody can rapidly shape online parsing (Kjelgaard & Speer, 1999; Marslen-Wilson et al., 1992; Nagel et al., 1996; Pynte & Prieur, 1996; Steinhauer, Alter, & Friederici, 1999). If speakers only provide reliable prosodic cues when they are aware of a referential ambiguity, then there should be little advantage to incorporating prosodic information into initial processing, undermining either an evolutionary or an experiential explanation for an early processing link.

Interpreting the literature on prosody and online parsing is complicated by the limitations of the online experiments. While the majority of these studies demonstrate that prosody can shape interpretation within a

---

[5] We thank Shari Speer and Amy Schafer for this suggestion.

second or two of encountering an ambiguously attached phrase, it is still unclear at what point prosody is used and how prevalent these effects are. The most common paradigms that are employed are cross-modal lexical decision, cross-modal naming, and speeded judgment tasks, all of which have been criticized for their poor temporal resolution and artificiality (Carlson et al., 2001). Even those experimenters who use measures that do not require metalinguistic judgments or the sudden interruption of an utterance (e.g. ERPs, Steinhauer et al., 1999) employ designs which provide limited information about the time course of prosodic influence. The experiments to date manipulate the consistency of the prosodic contour with subsequent morpho-syntactic information, and then measure effects of prosody at or after the disambiguation point (3–10 syllables after the onset of the ambiguity). Thus it is difficult to tell from these studies whether prosody guides initial syntactic processing (as suggested by Kjelgaard & Speer, 1999; Marslen-Wilson et al., 1992; Nagel et al., 1996) or whether it plays a role at a later processing stage (Marcus & Hindle, 1990; Pynte & Prieur, 1996).

As a first step toward unraveling these mysteries, we combined the real-world eye-gaze paradigm with our referential communication task, to see whether the prosodic cues produced by our Speakers could shape online interpretation. In this experiment, we placed the Speakers in ambiguous referential contexts (as in Experiment 1) and monitored the Listener's eye movements as she/he listened to the target utterances and carried out the task. This study extends the literature on prosody and online parsing in three ways. First, it allows us to examine the use of prosody in the absence of a secondary task. Second, it is the first online study we know of which uses the unaltered utterances of untrained speakers. Finally, because eye movements provide nearly continuous information about the evolving interpretation, this technique could potentially reveal effects of prosody on the initial interpretation of the ambiguous prepositional phrase. Because our sentences are never disambiguated by morpho-syntactic or referential cues, we can also examine how the influence of prosody changes over time in the absence of other sources of information.

Methods

Participants

Twenty-four pairs of volunteers from the University of Pennsylvania community received extra course credit or were paid for their participation. As before, all of the Speakers were female. Fifteen of the Listeners were male. All participants were native speakers of English and none had participated in Experiments 1 or 2. Three additional pairs participated but were not included in the analyses because of experimenter error.

Procedure and equipment

The procedure was the same as Experiment 1 with the following exceptions. First, the Listener's eye movements were recorded using an ISCAN eye-tracking visor (for details see Trueswell, Sekerina, Hill, & Logrip, 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 IS-CAN 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. A microphone, which was placed next to the Speaker, was connected to the audio input of the video recorder so that the utterances produced by the Speaker would be time-locked to the Listener's eye movements. Periodically throughout the experiment, the calibration of the tracker was checked by asking the Listener to fixate on individual objects, and if necessary a recalibration was performed.

To ensure that the Listener began each trial by looking at a point that was roughly equidistant from all of the objects, we placed a fixation point (a red star) at the center of the table and marked out five locations where the Listener was to place the toys. The Speaker was instructed to begin each trial by first asking the Listener if she/he was ready (as in the earlier studies) and then telling her/him to "look at the star." This led to slight increase in the delay between the time when the Speaker read the sentence and the time when she produced it.

Stimuli

The stimuli in this experiment were in most respects like those in Experiment 1. Speakers and Listeners were both given the ambiguous referential contexts that were used in the first experiment and each Speaker saw both Modifier and Instrument Demonstrations. However, the unambiguous conditions from Experiment 1 were not included in this experiment. This was done for two reasons. First, performance in these conditions had been uniformly excellent in Experiment 1. Second, we wanted to rule out the possibility the absence of the unambiguous targets in Experiment 2 contributed to the changes in Speaker prosody and the drop in Listener performance. Perhaps the unambiguous targets highlighted the salience of the two interpretations in Experiment 1 encouraging subjects to distinguish them.

To keep the number of ambiguous sentences constant, we again divided the 16 target items into the two stimulus lists that were used in Experiment 2. Within

each of these stimulus lists, every target trial was rotated through the two Demonstration conditions to generate four presentation lists. Within a presentation list, each target trial appeared in one of the two Demonstration types, resulting in four target trials in each condition. In each list the target trials were combined with the 24 filler trials from Experiment 2 and were presented in same psuedo-random order as before. Four additional lists were generated by reversing the order of trials in each list. Each participant-pair was assigned to one of the eight lists.

Coding

The videotapes of the Listener's actions were coded in the same manner as before. Four test trials in which the Speaker did not produce the target utterance were excluded from further analysis. The durations of target words and pauses were coded from the digital videotape that contained both the utterance and the Listener's eye movements. Coders, blind to condition, determined the frame on which a word began and the frame on which it ended and the duration was calculated by multiplying the difference by the sampling rate of the video record (1/30th of a second).

The Listener's eye movements were coded by hand from the digital videotape, using the frame-by-frame viewing on a digital VCR. A trial began at the onset of the target utterance and ended when the Listener touched one of the objects. Coders recorded the onset of each fixation and the object that was fixated. If the eye-tracker failed to determine the position of the eye or if the eye position information across frames was contradictory, then that frame was coded as track loss (5% of all frames). Every trial was coded by two research assistants and all disagreements were resolved by a third coder (8% of all frames).

Results

Actions, prosody, and awareness

Unsurprisingly, the Listeners and the Speakers in Experiment 3 performed similarly to those in Experiment 1. First, the Listeners' responses to the ambiguous sentences reflected the intentions of the Speaker ($F1(1, 16) = 18.23$, $p < .001$; $F2(1, 12) = 15.53$, $p < .005$). When the Speaker had seen an Instrument Demonstration, Listeners produced an Instrument Response 68% of the time. For Modifier Demonstrations, Listeners produced an Instrument Response only 39% of the time. Second, the Speakers' prosody clearly varied with intended structure. We conducted a subject and an item ANOVA for the mean duration of each of the critical regions examined in the previous studies (the verb, the post-verbal pause, the Verb Composite, the direct object noun, the post-nominal pause, the Noun Composite, and the prepositional phrase). All analyses produced a

reliable effect of Demonstration (all $F1(1, 16) > 7$, $p < .05$; $F2(1, 12) > 14$, $p < .005$). In every case this effect was at least as large as it had been in Experiment 1.

Thus we replicate the findings of Experiment 1. In ambiguous referential contexts, untrained and uninformed speakers produce strong prosodic cues which untrained listeners use to recover the structure of an otherwise ambiguous utterance. But as we noted earlier, under these conditions Speakers also tended to be aware of the two potential interpretations of the ambiguous sentences. This finding was also replicated in Experiment 3. In the post-experimental interview, 92% of the Speakers and 96% of the Listeners indicated that they had noticed the ambiguous utterances.

Online interpretation

Listener's eye movements during the target sentences were used to determine when prosody affects online interpretation. Prosody's influence could manifest itself in two ways. First, it could influence the perceived referent of the direct object. If the prepositional phrase is inter-

preted as an instrument, then there is no information in the sentence to determine which frog is being referred to, and Listeners should be equally likely to look at the Marked and Unmarked Animals. In contrast, if the prepositional phrase is interpreted as a modifier, then the entire NP must refer to the Marked Animal and the Listener should be less likely to look at the irrelevant Unmarked Animal. Second, to the extent that prosody influences the interpretation of the prepositional phrase, Listeners should look more to the full-size Target Instrument in the Instrument Condition and more to the Marked Animal (or the small instrument that it is holding) in the Modifier Condition.

The timing of these effects depend both on when the prosodic information appears in the utterance and the rapidity with which prosody influences parsing. We expect no effects before the onset of the direct object because prior to that time the utterance could refer to any of the objects. Yet we could see differences in the interpretation of the direct object noun before the ambiguous prepositional phrase. In our stimulus set, post-
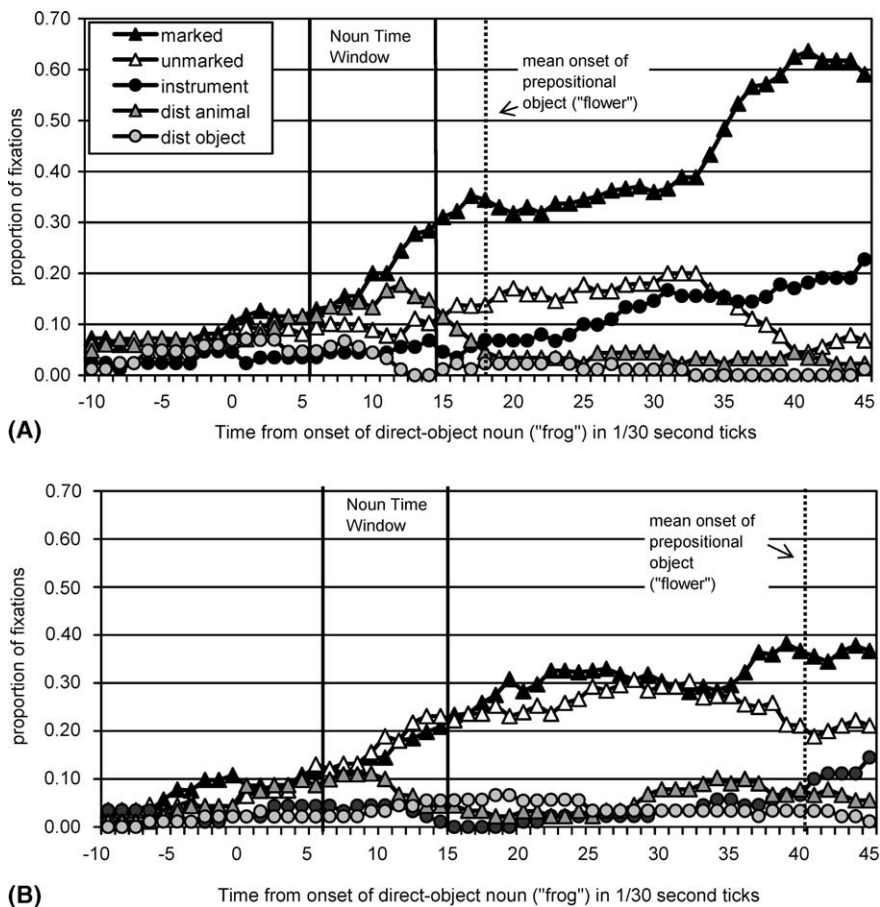

Fig. 11. The proportion of fixations to objects of each type over time (relative to onset of the direct object noun) for (A) trials on which the Speaker saw a Modifier Demonstration and (B) trials on which the Speaker saw an Instrument Demonstration.

nominal modifiers referred only to animals that held small objects (the Marked Animal and Distractor Animal). In fact, it would be awkward to describe the Unmarked Animal in this way: the Speaker cannot see where the Listener placed it and thus it has no salient or unique features. If the Listeners in the Modifier Condition can use prosody to recognize that the first noun is part of a complex noun phrase, then this feature of our stimuli could allow them to infer that the Marked Animal is the referent as soon as they hear the direct object noun. If Listeners fail to make this inference, then prosody will not constrain interpretation until the Listener encounters the ambiguous prepositional phrase. In this case, the effects of prosody on establishing the referent of the direct object noun and the referent of the prepositional object may be superimposed on one another.

Our analyses of the eye movements were complicated by systematic differences in the duration of words across conditions: utterances in Instrument Condition had a considerably shorter Verb Composite and a longer Noun Composite than the utterances in the Modifier Condition. We dealt with this problem by re-synchronizing the utterances at each word and conducting our analyses on small time windows. This ensures that we are comparing regions in which the subjects in the two conditions have heard similar portions of the utterance. In all figures and analyses, we group the trials by the meaning that the Speaker was trying to convey, not by the response that the Listener gave or the prosodic contour that the Speaker produced. This maximizes the power of the experiment and maintains our counterbalancing across the conditions. If we analyzed only those instances in which the subject performed the correct action, we might create a spurious effect of Demonstration on early eye movements. This could happen if visual attention influences the ultimate interpretation independent of prosody. For instance, Listeners who happen to look at the Marked Animal early might be more likely to perform a modifier action (for a similar effect in young children, see Trueswell et al., 1999).

The direct object noun. The first time window that we examined was synchronized at the onset of the direct object noun (Noun Onset). Figs. 11A and B plot the fixation probabilities of each object type over time for the two Demonstration Conditions. At the beginning of the trial, most of the subjects were still looking at the central fixation point, so the total of the fixation probabilities is well below one. The most notable difference between Figs. 11A and B is in the proportion of looks to the Unmarked Animal. Shortly after the onset of the direct object noun, the Listeners in the Instrument Condition began looking at both the Marked and Unmarked Animals, indicating that both are considered potential referents for the noun. At about the same

time, the Listeners in the Modifier Condition were mostly looking at the Marked Animal. This suggests that they have used prosody to eliminate the Unmarked Animal as the potential referent of the unfolding noun phrase.[6]

Our analysis focused on a time window that began 200 ms after the onset of the direct object noun and ended 500 ms after the noun onset. We began the analysis at this point because of previous research demonstrating that lexical information does not begin to influence eye movements until at least 200 ms after word onset (Allopenna, Magnuson, & Tanenhaus, 1998). This delay in response appears to be due in part to the time needed to program and execute an eye movement which has been estimated to be as great as 150 ms (see e.g., Matin, Shao, & Boff, 1993). The time window was limited to 300 ms to ensure that subjects in both in conditions would have heard little or nothing beyond the direct object noun. The median length of the direct object noun in the Modifier Condition was 8 frames or 267 ms. The Direct Object time window ended approximately 200 ms after this to ensure that subjects in both conditions had equivalent lexical information.

The proportion of looking time to the Unmarked Animal during this time window was calculated for each trial. A trial was discarded from the analysis if there was track loss on more than half of the frames in the relevant time slice. In all the analyses that follow, fewer than 5% of the trials were removed due to track loss. The proportion of looking time was averaged by subjects and items and entered into separate ANOVAs.[7] The subject ANOVA had two between-subject factors (List and Order) and one within-subject factor (Demonstration). The item ANOVA had two within-item factors (Order and Demonstration) and one between-item factor (Item Group). In the Direct Object time window, the effect of Demonstration was reliable in the subject analysis

---

[6] The reader may be surprised by the lack of an effect of Demonstration on the absolute percentage of looks to the Marked Animal. If Listeners in the Instrument Condition are splitting their time between the two animals, you might expect that looking time to the Marked Animal would be lower here than it is in the Modifier Condition. But instead, subjects in this condition are simply more likely to begin looking around after the direct object noun. The referential ambiguity in the Instrument condition may lead to more eye movements—subjects in the Instrument Condition make more fixations ($F(1, 16) = 9.37$, $p < .01$) and most of these additional fixations are on the Unmarked and Marked Animals.

[7] In all cases, we performed the analyses both on the raw proportions and on their arc-sin transforms (for raw proportions arcsin $[2x − 1]$, for the difference scores arcsin x). Unless otherwise noted, the results of the two analyses were the same. In all cases the F and p values will be taken from the analysis of the transforms and the means from the untransformed data.

$(F1(1, 16) = 7.12,\ p < .05;\ F2(1, 12) = 2.94,\ p = .11)$. On Instrument trials, Listeners spent 16.7% of their time looking at the Unmarked Animal, while on Modifier trials this number dropped to 9.4%.

To understand the relative time course of the use of prosodic information, we analyzed three smaller time windows within this same range (200–300, 300–400 and 400–500 ms after the onset of the noun). The results of this analysis are given in Table 5. As can be seen in the table, the effect of Demonstration becomes reliable in the 400–500 ms time slice. Given that it takes about 150 ms to program an eye movement, this suggests that prosody has a reliable influence on interpretation within 250 ms of the onset of the direct object noun.

Clearly, the prosody of the utterance is having a rapid influence on the Listeners interpretation of the first noun. To understand how the timing of this process relates to other phenomena in language comprehension, we compared it with the influence of phonological information on Listeners' interpretation of the direct object noun. Studies of spoken word recognition employing similar paradigms have demonstrated that listeners incrementally restrict the reference of a word as its phonological form unfolds (Allopenna et al., 1998; Swingley & Aslin, 2000; Tanenhaus, Magnuson, Dahan, & Chambers, 2000). We examined the time course of

phonological reference restriction in the present experiment by computing the difference in the looking time to the Marked Animal and the Distractor Animal (Animal Identification) in the same three time slices. When this number is reliably greater than zero it tells us that the subjects are spending more time looking at the kind of animal that is being named, indicating that they have used information from the lexicon to restrict the reference of the noun that is being spoken. The results of these analyses appear in Table 6. We chose to compare the Distractor Animal to the Marked Animal because these two objects were roughly equivalent in perceptual complexity. This measure of animal identification also has the advantage of providing similar results across the two Demonstration conditions. If we use the Unmarked Animal or the average of the two, there is a slight delay in the reliability of Animal Identification in the Modifier Condition.

A comparison of Tables 5 and 6 indicates that prosodic effects emerge at about the same time as phonologically driven effects. In the first time slice neither is reliable, in the second time slice both are marginal, but in the third time slice there is clear evidence for both Animal Identification and an effect of Demonstration. Between 400 and 500 ms after the noun onset, Listeners are using the unfolding phonological information to

Table 5
Effects of Demonstration on looking time to the Unmarked Animal, Experiment 3

Time relative to noun onset	Mean % unmarked for Instrument Demo	Mean % unmarked for Modifier Demo	Subject analysis	Item analysis
Overall (200–500 ms)	17.6	9.5	$F1(1, 16) = 7.12$ $p < .05^*$	$F2(1, 12) = 2.94$ $p = .11$
Time slice 1 (200–300 ms)	12.5	9.0	$F1(1, 16) = 1.32$ $p > .2$	$F2(1, 12) < 1$ $p > .8$
Time slice 2 (300–400 ms)	16.7	8.6	$F1(1, 16) = 9.34$ $p < .01^*$	$F2(1, 12) = 3.74$ $p = .078$
Time slice 3 (400–500 ms)	22.2	10.0	$F1(1, 16) = 8.72$ $p < .01^*$	$F2(1, 12) = 7.41$ $p < .05^*$

Table 6
The reliability of Animal Identification (looking time to Marked Animal – looking time to Distractor Animal), Experiment 3

Time relative to noun onset	Mean% to Marked Animal	Mean% to Distractor Animal	Subject analysis	Item analysis
Overall (200–500 ms)	17.4	11.4	$F1(1, 16) = 6.91$ $p < .01^*$	$F2(1, 12) = 9.61$ $p < .01^*$
Time slice 1 (200–300 ms)	12.6	11.4	$F1(1, 16) < 1$ $p > .7$	$F2(1, 12) < 1$ $p > .6$
Time slice 2 (300–400 ms)	17.0	11.6	$F1(1, 16) = 3.83$ $p = .068$	$F2(1, 12) = 4.21$ $p = .063$
Time slice 3 (400–500 ms)	23.3	10.4	$F1(1, 16) = 15.87$ $p < .001^{**}$	$F2(1, 12) = 16.16$ $p < .005^{**}$

identify the word and focus in on the correct type of animal. During this same time period, they are already using prosodic information to eliminate the Unmarked Animal as a possible potential referent when the Speaker has seen a Modifier Demonstration.

The discovery that prosody is influencing interpretation prior to the ambiguous region is unexpected but by no means mysterious. The modifier utterances were typically produced with a sizeable pause after the verb, allowing them to be easily distinguished from the instrument utterances. This pause, and the short duration of the direct object noun, could have allowed our subjects to predict a complex noun-phrase before they even identified the noun. Given the pragmatic constraints of the situation, this complex NP could be used to refer to the Marked Animal and the Distractor Animal but not the Unmarked Animal. Consistent with this, the subjects in the Modifier Condition show an early rise in looks to both the Distractor Animal and the Marked Animal and then sudden drop in looks to the Distractor Animal as the identity of the noun becomes clear (see Fig. 11A).

The prepositional object. The second time window was synchronized at the onset of the prepositional object (PP Object Onset). Figs. 12A and B plot the fixation probabilities of each object type over time relative to the PP Object Onset for each of the Demonstration Conditions. Two differences between the two panels suggest that the Listeners are using prosody to rapidly determine the role of the prepositional phrase. First, soon after the PP Object Onset, Listeners in the Instrument Condition began looking at the Target Instrument (the big flower). Second, at approximately the same time, there seems to be an increase in the looks to the Marked Animal in the Modifier Condition.

We explored this pattern of effects by looking at the mean proportion of looking time to the Marked Animal and the Target Instrument in a time window that began 200 ms after the PP Object onset and ended 800 ms later. A larger time window was possible in this analysis because there were no words following the prepositional object. Separate item and subject ANOVAs were conducted for both the dependent variables, with Order, Demonstration and Item Group/List as factors. There
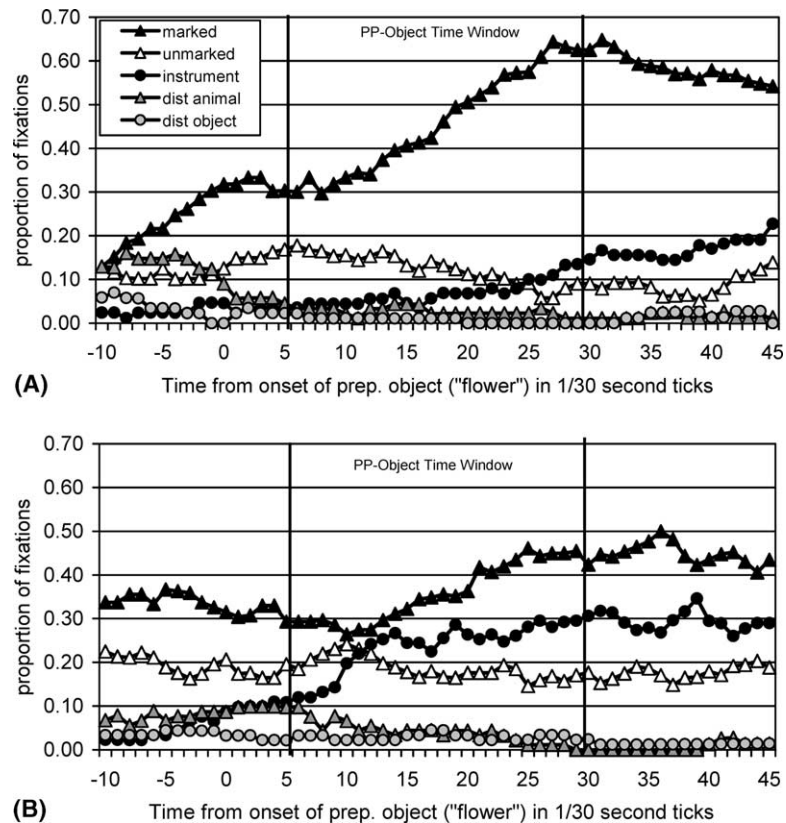

Fig. 12. The proportion of fixations to objects of each type over time (relative to onset of the prepositional object) for (A) trials on which the Speaker saw a Modifier Demonstration and (B) trials on which the Speaker saw an Instrument Demonstration.

was a reliable effect of Demonstration on the proportion of time spent looking at the target Instrument ($F1(1, 16) = 10.30$, $p < .005$; $F2(1, 12) = 5.44$, $p < .05$). On Instrument trials, Listeners looked at the Instrument for 25% of the time while Modifier Trials they spent only 16% of their time here. However, the effect of Demonstration on looking time to the Marked Animal was marginal ($F1(1, 16) = 3.53$, $p = .08$; $F2(1, 12) = 5.92$, $p < .05$). Perhaps because the time spent on the Marked Animal was so high in both conditions ($M = 35\%$ and $M = 45\%$ in the Instrument and Modifier Conditions, respectively).

General discussion

In this paper, we have suggested that a speaker's knowledge of the referential situation affects whether she uses prosody to disambiguate an otherwise ambiguous utterance. When a speaker observes a need to disambiguate, because the context itself does not strongly support the intended meaning of the utterance, the speaker will attempt to do so via changes in prosodic grouping. In particular, when the speaker's referential scene supported both the modifier and the instrument interpretation of an ambiguous PP-attachment (i.e., Experiment 1), speakers significantly altered their production of the utterance in ways that were consistent with the intended structure (e.g., providing a prosodic break after the direct object noun when conveying VP-attachment). Listeners in this situation were sensitive to these cues, even on the first critical trial. In Experiment 2, the referential scene of the speaker but not the listener was changed, such that the speaker's scene supported only the intended meaning of the syntactically ambiguous utterance. Here speakers showed little prosodic variation between the two interpretations, even on the first trial. As a result, listeners were at a loss when resolving the ambiguity. In Experiment 3, speakers were again placed in ambiguous situations and again attempted to disambiguate prosodically, replicating the acoustic and action results of Experiment 1. Eye movement analyses further revealed that a speaker's prosodic variation affected the listener's interpretation in real-time, as soon as these cues could be gleaned from the utterance.

The role of prosody in disambiguation

The findings suggest that speakers and listeners have a fairly good understanding of the correspondence between prosody and phrasal structure, but speakers use such information sparingly, providing reliable cues only when the situation does not provide other relevant information about the intended structural organization of the utterance. This conclusion is consistent with our interpretation of Allbritton et al. (1996), in which reliable prosodic disambiguation occurred only when trained speakers were explicitly contrasting two meanings of the same sentence. Given that we observe naïve speakers disambiguating utterances in a context-contingent manner, it seems likely that the untrained speakers in the Allbritton study were failing to disambiguate their utterances because they appeared in supportive/disambiguating discourse contexts.

But recent findings from Schafer, Speer, Warren and colleagues appear to be in conflict with these conclusions. These researchers independently developed a cooperative game task which shares many properties with our own, and used it to examine the prosodic disambiguation of major clause boundaries (Schafer et al., 2000a). This work did not address the issue of ambiguity awareness or the disambiguation of phrasal ambiguities like PP-attachment. However, data collected from these same experiments, and reported in a series of working papers (Schafer et al., 2000b; Warren, Achafer, Speer, & White, 2000), suggest that prosodic disambiguation of PP-attachment ambiguities occurs even under contextually unambiguous conditions (target sentences are given in 2). VP-attachment utterances were more likely to contain a prosodic break after the direct object noun than NP-attachment utterances, regardless of whether the context singled out a single plausible interpretation.

2 (a) I want to change the position of the square with the triangle.
　(b) I am able to confirm the move of the square with the triangle.

We attribute these discrepant results to at least three specific properties of the Schafer et al. game that may have inadvertently inflated the prosodic differences between the two interpretations and minimized any effects of the referential context. First, because the modifier utterances all came from repeated use of the same complex noun phrase, it is likely that this phrase became a fixed, lexicalized, unit. Specifically, at the start of game subjects memorized the names of the four pieces that were to be used in the game, i.e., "the square," "the triangle," "the cylinder," and "the square with the triangle," with the last name referring to a house-like structure. During the game, subjects inserted these names into previously memorized sentence frames, such as "I'd like to change the position of _____." Crucially, all examples of NP-attachment utterances came from speakers inserting the single established name "the square with the triangle" into two frames.[8] It is likely that the reduced pronunciation of NP-attachment arose from the fact that "the square with a triangle" had be-

---

[8] We thank Sarah Brown-Schmidt and Michael Tanenhaus for pointing out this critical property of the Schafer et al. (2000b) materials.

come lexicalized—an established name repeatedly used in the study. Indeed, lexicalized phrases are notable for their reduced pronunciation and shift in stress (e.g., blackbird vs. black bird, or "The Cat in the Hat;" see Liberman & Sproat (1992), for a discussion of lexicalization and NP + PP constructions). After the phrase has been lexicalized, we would expect that pronunciation of the new name would be shortened further upon repeated use (e.g., Bard et al., 2000; Fowler & Housum, 1987). Thus repetition of the same complex noun phrase could on its own account for the differences between the modifier and instrument utterances.

Second, the contextual cues to disambiguation that were used by Schafer et al. were far subtler than our own referential manipulations. Schafer et al. defined an utterance's context as unambiguous if the alternative action was not possible given the rules of the game. For instance, the utterance "…move the square with the triangle" was considered unambiguous if the house-like structure was blocked because other objects would have to be moved first, or if the triangle, the potential instrument, could not be moved immediately. In contrast, in our corresponding unambiguous condition the other possible referents were not even present in the speaker's scene (Fig. 5). Thus, the added complexity of tracking possible moves in the Schafer et al. task, in conjunction with the possibility that a speaker may not have been convinced that the listener knew the rules of the game well enough to compute only legal moves, may have increased the level of perceived ambiguity, and hence increased the frequency of prosodic disambiguation.

Finally, the sentences used by Schafer and colleagues were considerably longer and more complex than those used in the current experiments. Our targets were six to nine syllables in length, short enough to produce as a single intonational phrase, if the speaker wished. The targets in the Schafer study were 16 and 17 syllables making them difficult to say without pausing. This increase in length is partially due to the decision to nest the critical noun ("square") inside a prepositional phrase, which in turn is nested inside of the direct object noun phrase ("the move of the square…"). Because of the complexity of the noun phrases used in the Schafer sentences, any place where the subject pauses from the onset of the verb through the completion of the sentence would either disambiguate the prepositional phrase attachment or require another pause further downstream to maintain the ambiguity. If this analysis is valid, then each of these sets of findings can only be generalized to sentences of the length and type that were tested. The ecological validity of these two sets of studies would then depend on relative frequency with which these sentences are encountered. Given that the mean length of an utterance in casual conversations is about 6.8 words or roughly 8.6 syllables (Jurafsky, 2001 personal communication), we suspect that the lengths of the utterances used in our study are

reasonably representative of the lengths of utterances that occur in conversational settings.[9]

Prosody and parsing

A number of researchers have demonstrated that prosody influences early syntactic interpretation (Pynte & Prieur, 1996; Kjelgaard & Speer, 1999; Steinhauer et al., 1999, among others). However, all of the prior findings are compatible with both truly interactive models, in which prosody affects the initial incorporation of a word into the syntactic structure, and more modular models, in which the initial analysis is based solely on lexical or syntactic information. The results of Experiment 3 strongly favor the interactive account. We find that the prosodic form of the utterance begins to influence interpretation shortly after the direct object noun and prior to the onset of the ambiguously attached preposition. Listeners in the Modifier condition appear to be using prosodic information to predict that the speaker is going to produce a complex noun phrase, much as they might use semantic and syntactic information to predict an upcoming noun or verb (Altmann & Kamide, 1999; Kamide, Scheepers, Altmann, & Crocker, 2002). Furthermore, to map this unspoken modifier onto the Marked Animal, our listeners have to rely on constraints that are specific to this experimental situation to rule out the Unmarked Animal as a potential referent. Outside of this experiment unencumbered objects can often be identified with a postnominal modifier ("the frog next to the lego") but the presence of the screen eliminates this option.[10]

---

[9] These estimates were provided by Dan Jurafsky. The estimate of words per utterance was based on 158,482 utterance sample from the Switchboard corpus. An utterance was defined by the criteria given in Bell et al. (submitted) except that all fragments were excluded from the analysis. Single word and multi-clause utterances were included. The average number of syllables per word was calculated from a 38,000 word sample of the Switchboard corpus which has been coded for syllables. Our estimate of syllables per utterance makes the simplifying assumption that these two numbers are independent.

[10] An anonymous reviewer noted that similar experiments focusing on the use of referential constraints in online processing have not found predictive effects at the direct object noun (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; Trueswell et al., 1999). While this difference could reflect a processing advantage for prosodic information relative to referential information, we believe that it is attributable to differences in the way that the props are presented in these experiments. In the referential studies, the location of each toy is knowledge that is shared by the speaker and listener, allowing either of the potential referents to be felicitously described by its location (e.g., "the apple to the left of the box"). Subjects in these studies may have used referential information to predict that a post-nominal modifier was coming but they could not use that information to predict which of the two objects would be referred to.

It is tempting to dismiss this effect as an uninteresting artifact of a particular experimental design. The perfect correlation between complex noun phrases and animals with attributes in our materials allows our subjects to make use of prosody in a way that would not ordinarily be available to them. But the very existence of these effects speaks to the plasticity of human sentence processing. Within 300 ms of the onset of the direct object noun, our subjects are able to employ an experimentally specific strategy of using prosodic information to restrict the reference of this noun, a strategy which emerges after only a few trials. These findings suggest that language processing is dynamic and flexible, a proposal that is unsurprising in light of recent evidence that even seemingly automatic processes, such as the rapid deployment of visual attention, can be shaped by task specific constraints and implicit learning (Folk, Remington, & Johnston, 1992; Lambert, Norris, Naikar, & Aitken, 2000).

Learning prosodic cues to syntax

Our results raise an important question about how listeners learn the relevant prosodic cues to structure. Experiments 1 and 2 suggest that these cues are unreliable, appearing primarily in situations in which other information does not disambiguate the utterance. Nevertheless, our subjects had clearly learned something about them prior to participating in the study. How else could they have used prosody to communicate in the absence of feedback? Furthermore, this knowledge was robust enough to affect the listener's initial interpretation of utterance (Experiment 3). Why are listeners placing so much weight on information that is evidently often unreliable?

To make sense of this apparent anomaly it is critical to distinguish between cue frequency and cue validity. Our statistical analyses simply indicated that the modifier and instrument sentences were different from one another in ambiguous contexts but quite similar in unambiguous contexts. This could happen in one of two ways. The relevant cues could be used consistently in ambiguous contexts but randomly in the unambiguous ones. In this case, they would be frequent but invalid. Or these cues could be used consistently in ambiguous contexts but simply omitted in the unambiguous ones. In this case, the cues would be less frequent but quite valid. A closer look at the productions in Experiments 1 and 2 suggests that this later description is more accurate.

Across studies, one of the most reliable cues to prepositional phrase attachment is the presence or absence of a pause after the direct object noun (Cooper & Paccia-Cooper, 1980; Price et al., 1991; Warren, 1985). In Experiment 1, when the referential context was ambiguous, subjects produced post-nominal pauses of 200 ms or more on 30% of the trials. This cue accurately signaled the intended structure on 96% of these trials. When the context was unambiguous, speakers only produced post-nominal pauses 4% of the time. Nevertheless, these pauses continued to be a reasonably valid cue to structure—76% of them occurred during instrument utterances. Post-nominal IP breaks and prepositions with pitch accents are also cues which were valid in both conditions but simply more frequent in the ambiguous contexts. In contrast, the cue validity of an IP break after the verb is substantially reduced in unambiguous contexts (from 79% in Experiment 1 to just 60% in Experiment 2).

It is very likely however that more is going on than probabilistic mapping between phrase structure and acoustic grouping. Our acoustic analyses and those of other studies on phrasal ambiguity suggest that speakers may be marking these boundaries by 'borrowing' prosodic cues that are typically used at clause boundaries or in longer utterances. That is, speakers vary the pitch and provide a silent pause at the phrase boundary when they would otherwise not pronounce such a grouping in that way. Listeners already know from the clausal contexts that these cues mark a break and therefore need only learn that they can be applied in a different situation. Hence, the 'infrequency' of possible learning events may not be at issue here at all, since the relevant prosodic markings arise at the end of nearly every clause, which coincides with the end of least one phrase.

Ambiguity avoidance

Our findings on the limitations of prosodic disambiguation are consistent with the broader literature on ambiguity avoidance. The most well known work on avoiding syntactic ambiguity has focused on speakers' insertion of function words to disambiguate temporary syntactic ambiguities (Elsness, 1984; Ferreira & Dell, 2000). These studies have found that writers (Elsness, 1984) and speakers (Ferreira & Dell, 2000) often do not bother to disambiguate temporary ambiguities by inserting function words (e.g., they often choose to say "The chef knew the man would..." rather than "The chef knew that the man would..."). Function word insertion is best predicted by accessibility of upcoming material for speaker and not by the likelihood that the listener will temporarily misunderstand the utterance (Ferreira & Dell, 2000). Unpublished work from our own lab replicates these effects and finds that speakers seem to have a weak tendency to use prosody to disambiguate temporary syntactic ambiguity, although these effects were not always significant (Mims & Trueswell, 1999). Ferreira & Dell (2000) suggest that speakers fail to disambiguate because they are unaware of temporary ambiguity or because the ambiguity is quickly resolved by subsequent material in the utterance (for a similar set of conclusions involving temporary ambiguities with alternating datives, see Arnold, Wasow,

Asudeh, & Alrenga, submitted). Such an account is consistent with the results of our experiments. Speakers appear to choose disambiguation only when the context fails to resolve a globally ambiguous utterance, perhaps because these contexts lead them to become aware of the ambiguity (although awareness alone is not enough, see discussion of Experiment 2).

Given these observations, it seems likely that ambiguity awareness, and attempts to disambiguate ambiguity, may be tied to the level of representation at which the ambiguity arises. One should expect that non-linguistically trained speakers will be far more aware of referential ambiguity than they are of syntactic ambiguity. At fairly young ages, children will use post-nominal modification to distinguish two otherwise identical referents (e.g., McKee, McDaniel, & Snedeker, 1998; Hurewitz, Brown-Schmidt, Thorpe, Gleitman, & Trueswell, 2000), indicating a sensitivity to the potential for referential ambiguity. Yet, speakers frequently produce 'technically ambiguous' referential expressions, such as pronouns, and syntactically ambiguous utterances presumably because the context clearly disambiguates the utterance from the speaker's perspective.

It would seem that syntactic ambiguities are avoided or prosodically disambiguated only when the syntactic alternatives have feasible and incompatible referential implications. For instance, the speakers in our study may not have been explicitly distinguishing between VP-attachment and NP-attachment, but rather between reference to a potential instrument (the large feather) or reference to an object that could be used to distinguish between two animals (the small feather that the frog held) which implicitly correspond to VP- and NP-attachment.

Summary

These studies demonstrate that untrained, uninstructed speakers can provide prosodic cues to the structure of an utterance that would otherwise be ambiguous. However, when the situational context disambiguates the sentence, these cues all but disappear. When prosodic cues are present, untrained and uninstructed listeners can use them, though their performance clearly suffers when compared with utterances that are lexically disambiguated. Nevertheless, listeners are able to rapidly integrate prosodic and lexical information to determine the referent of the direct object noun phrase and the role of the ambiguous prepositional phrase. We found no evidence that lexical information has a privileged role in determining reference; the effects of prosody are visible as early as the effects of phonological form. We conclude that while strong prosodic cues to syntax may be infrequent, they appear to be valid. English speakers learn them because they are predictive, but produce them primarily when they appear to be necessary for clear communication.

References

Allbritton, D., McKoon, G., & Ratcliff, R. (1996). Reliability of prosodic cues for resolving syntactic ambiguity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 714–735.

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

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73*, 247–264.

Arnold, J., Wasow, T., Asudeh, A., & Alrenga, P. (submitted). Avoiding attachment ambiguities: The role of constituent ordering.

Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., & Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language, 42*, 1–22.

Beach, C. M. (1991). The interpretation of prosodic patterns at points of syntactic structure ambiguity: Evidence for cue trading relations. *Journal of Memory and Language, 30*, 644–663.

Beach, C. M., Katz, W. F., & Skowronski, A. (1996). Children's processing of prosodic cues for phrasal interpretation. *Journal of the Acoustical Society of America, 99*(2), 1148–1160.

Beckman, M. E. (1996). The parsing of prosody. *Language and Cognitive Processes, 11*, 17–67.

Beckman, M. E., & Hirschberg, J. (1994). *The ToBI annotation conventions*. Columbus, OH: Ohio State University.

Beckman, M. E., & Pierrehumbert, J. B. (1986). Intonational Structure in Japanese and English. *Phonology Yearbook, 3*, 255–309.

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M., & Gildea, D. (submitted). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation.

Carlson, K., Clifton, C., & Frazier, L. (2001). Prosodic boundaries in adjunct attachment. *Journal of Memory and Language, 45*, 58–81.

Cooper, W. E., & Paccia-Cooper, J. (1980). *Syntax and speech*. Cambridge, MA: Harvard University Press.

Cutler, A., Dahan, D., & van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language & Speech, 40*(2), 141–201.

Elsness, J. (1984). That or zero? A look at the choice of object clause connective in a corpus of American English. *English Studies, 65*, 519–533.

Fernald, A., & McRoberts, G. (1996). Prosodic bootstrapping: A critical analysis of the argument and the evidence. In J. Morgan & K. Demuth (Eds.), *Signal to syntax*. Mahwah, NJ: Erlbaum.

Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. *Journal of Memory and Language, 30*, 210–233.

Ferreira, F. (1993). The creation of prosody during sentence production. *Psychological Review, 100*, 233–253.

Ferreira, V., & Dell, G. (2000). Effects of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology, 40*, 296–340.

Folk, C. L., Remington, R., & Johnston, J. C. (1992). Involuntary orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance, 18*, 1030–1044.

Fox Tree, J. E., & Meijer, P. J. A. (2000). Untrained speakers' use of prosody in syntactic disambiguation and listeners' interpretations. *Psychological Research, 63*, 1–13.

Fowler, C. A., & Housum, J. (1987). Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language, 26*, 489–504.

Gee, J. P., & Grosjean, F. (1983). Performance structures: A psycholinguistic and linguistic appraisal. *Cognitive Psychology, 15*, 411–458.

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.

Hurewitz, F., Brown-Schmidt, S., Thorpe, K., Gleitman, L. R., & Trueswell, J. C. (2000). One frog, two frog, red frog, blue frog: Factors affecting children's syntactic choices in production and comprehension. *Journal of Psycholinguistic Research, 29*(2), 597–626.

Kaisse, E. (1985). *Connected speech: The interaction of syntax and phonology*. Orlando: Academic Press.

Kamide, Y., Scheepers, C., Altmann, G., & Crocker, M., (2002). *Integration of syntactic and semantic information in predictive processing: Anticipatory eye-movements in German*. Paper presented at the 14th Annual CUNY Conference on Human Sentence Processing, New York, March 2002.

Keysar, B., & Henly, A. (1998). *Speakers overestimate the transparency of their intention*. Paper presented at the 39th Annual Meeting of the Psychonomic Society, Dallas, November 1998.

Kjelgaard, M. M., & Speer, S. R. (1999). Prosodic facilitaion and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language, 40*, 153–194.

Klatt, D. H. (1975). Vowel lengthening is syntactically determined in a connected discourse. *Journal of Phonetics, 3*, 129–140.

Lambert, A., Norris, A., Naikar, N., & Aitken, V. (2000). Effects of informative peripheral cues on eye movements. *Visual Cognition, 7*, 545–569.

Lehiste, I. (1972). Timing of utterances and linguistic boundaries. *Journal of the Accoustical Society of America, 51*, 2018–2024.

Lehiste, I. (1973a). Phonetic disambiguation of syntactic ambiguity. *Glossa, 7*, 102–122.

Lehiste, I. (1973b). Rhythmic units and syntactic units in production and perception. *The Journal of the Acoustical Society of America, 54*(5), 1228–1234.

Lehiste, I., Olive, J. P., & Streeter, L. A. (1976). Role of duration in disambiguating syntactically ambiguous sentences. *Journal of the Acoustical Society of America, 60*, 1199–1202.

Liberman, M., & Sproat, R. (1992). The stress and structure of modified noun phrases in English. In I. A. Sag & A. Szabolcsi (Eds.), *Lexical matters*. Palo Alto, CA: Stanford University.

Lieberman, P. (1963). Some effects of semantic and grammatical context on the production and perception of speech. *Language & Speech, 6*, 172–187.

Marcus, M., & Hindle, D. (1990). Description theory and intonation boundaries. In G. Altmann (Ed.), *Cognitive models of speech processing: Computational and psycholinguistic perspectives*. Cambridge, MA: MIT Press.

Marslen-Wilson, W. D., Tyler, L. K., Warren, P., Grenier, P., & Lee, C. S. (1992). Prosodic effects in minimal attachment. *Quarterly Journal of Experimental Psychology, 45A*, 73–87.

Matin, E., Shao, K. C., & Boff, K. R. (1993). Information-processing time with and without saccades. *Perception & Psychophysics, 53*(2), 372–380.

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

Mims, K., & Trueswell, J. (1999). *Do speakers help listeners? Lexical cues, prosodic cues, and ambiguity avoidance*. Poster presented at the 12th Annual CUNY Sentence Processing Conference, New York, February, 1999.

Nagel, H. N., Shapiro, L. P., & Nawy, R. (1994). Prosody and the processing of filler-gap sentences. *Journal of Psycholinguistic Research, 23*, 473–485.

Nagel, H. N., Shapiro, L. P., Tuller, B., & Nawy, R. (1996). Prosodic influences on the resolution of temporary ambiguity during on-line sentence processing. *Journal of Psycholinguistic Research, 25*, 319–344.

Nespor, M. A., & Vogel, I. (1986). *Prosodic Phonology*. Boston: Kluwer.

Price, P., Ostendorf, M., Shattuck-Hufnagel, S., & Fong, C. (1991). The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America, 90*, 2956–2970.

Pynte, J., & Prieur, B. (1996). Prosodic breaks and attachment decisions in sentence parsing. *Language and Cognitive Processes, 11*, 165–192.

Rayner, K., Carlson, K., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal Behavior, 22*(2), 358–374.

Schafer, A., Carter, J., Clifton, C., & Frazier, L. (1996). Focus in relative clause construal. *Language and Cognitive Processes, 11*(1–2), 135–163.

Schafer, A., Speer, S., Warren, P., & White, S. (2000a). Intonational disambiguation in sentence production and comprehension. *Journal of Psycholinguistic Research, 29*, 169–182.

Schafer, A. J., Warren, P., Speer, S. R., White, S. D., & Sokol, S. B. (2000b). Prosodic disambiguation in ambiguous and unambiguous situations. In *Annual meeting of the Linguistic Society of America, Chicago, IL.*

Scott, D. (1982). Duration as a cue to the perception of a phrase boundary. *Journal of the Acoustical Society of America, 71*, 996–1007.

Selkirk, E. O. (1986). On derived domains in sentence phonology. *Phonology Yearbook, 3*, 371–405.

Shattuck-Hufnagel, S., & Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research, 25*, 193–247.

Silverman, K., Beckman, M. E., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., & Hirschberg, J. (1992). *ToBI: A standard for labeling English prosody*. Paper presented at the Second International Conference on Spoken Language Processing, Banff, Canada.

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

Steinhauer, K., Alter, K., & Friederici, A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience, 2*(2), 191–196.

Sternberg, S., Monsell, S., Knoll, R. L., & Wright, C. E. (1978). The latency and duration of rapid movement sequences: Comparisons of speech and typewriting. In G. E. Stelmach (Ed.), *Information processing in motor control and learning* (pp. 117–152). New York: Academic Press.

Straub, K. A. (1997). *The Production of Prosodic Cues and their Role in the Comprehension of Syntactically Ambiguous Sentences*. Unpublished doctoral dissertation, Rochester, NY: University of Rochester.

Streeter, L. A. (1978). Acoustic determinants of phrase boundary perception. *Journal of the Acoustical Society of America, 64*, 1582–1592.

Swingley, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition, 76*, 147–166.

Tanenhaus, M. K., Magnuson, J. S., Dahan, D., & Chambers, C. (2000). Eye movements and lexical access in spoken-language comprehension: Evaluating a linking hypothesis between fixations and linguistic processing. *Journal of Psycholinguistic Research, 29*, 557–580.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1996). Using eye movements to study spoken language comprehension: Evidence for visually mediated incremental interpretation. In T. Inui & McClelland (Eds.), *Attention and performance 16: Information integration in perception and communication* (pp. 457–478). Cambridge, MA: MIT Press.

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

Taraban, R., & McClelland, J. (1988). Constituent attachment and thematic role assignment in sentence processing: Influences of content-based expectations. *Journal of Memory and Language, 27*, 597–632.

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

Wales, R., & Toner, H. (1979). Intonation and ambiguity. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett*. Hillsdale, NJ: Erlbaum.

Warren, P. (1985). *The Temporal Organization and Perception of Speech*. Unpublished doctoral dissertation, UK: University of Cambridge.

Warren, P. (2000). Prosody and sentence processing. In S. Garrod & M. Pickering (Eds.), *Language processing* (pp. 155–188). Hove, UK: Psychology Press.

Warren, P., Grabe, E., & Nolan, F. (1995). Prosody, phonology, and parsing in closure ambiguities. *Language and Cognitive Processes, 10*, 457–486.

Warren, P., Schafer, A., Speer, S., & White, S. (1999). Prosodic resolution of prepositional phrase ambiguity in ambiguous and unambiguous situations. *UCLA Working Papers in Phonetics, 99*, 5–33.