Prosodic Choice: Effects of Speaker Awareness and Referential Context

Jesse Snedeker (jessned@psych.upenn.edu),
Lila Gleitman (gleitman@ psych.upenn.edu),
Michael Felberbaum (felberbm@sas.upenn.edu),
Nicora Placa (nicorap@sas.upenn.edu),
John Trueswell (trueswel@ psych.upenn.edu)
Institute for Research in Cognitive Science; 3401 Walnut Street, Suite 400A
Philadelphia, PA 19104 USA

Abstract

These experiments were designed to discover whether untrained speakers produce prosodic cues that are sufficient to allow listeners to interpret ambiguous PP-attachments. A referential communication task was used to elicit productions of ambiguous sentences and determine whether listeners could use prosodic cues to correctly interpret these ambiguities in context. In Experiment 1, the referential context supported both potential interpretations of the ambiguity. Acoustic analyses indicated that Speakers produced potentially informative prosodic cues. Listeners' responses to the ambiguous sentences strongly reflected the demonstration the Speaker had seen, indicating that they were able to use this information. However, post-experiment interviews revealed that Speakers were aware of the ambiguous situations. Experiment 2 manipulated Speaker awareness by altering the Speaker's referential context to support only the intended meaning, and by making the resolution of the ambiguity a between subjects variable. Although Listeners' contexts were unchanged from Experiment 1, Listeners now showed no sensitivity to the Speakers' intended meaning. Acoustic analysis indicated that the strong prosodic cues provided in Experiment 1 were absent in Experiment 2. The experiments 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.

Introduction

One of the current challenges for research on prosody and syntactic ambiguity is to bring together what we know about the listener with what we know about the speaker. In doing so, we can begin to understand whether the prosodic cues that are available in speech can influence a listener's interpretation. The research reported here attempts to address this challenge by examining how a speaker uses prosody in the face of ambiguity and whether an accompanying listener is able to interpret the speaker's intended meaning. We will propose from this research that prosodic cues in adult-to-adult speech often depend upon the speakers' knowledge of the referential context. In particular, the choice to provide helpful prosodic cues depends upon whether or not the referential situation furnishes other cues that could help resolve the ambiguity.

Prior research on prosody and syntactic ambiguity has focused almost exclusively on either the speaker or the listener, and only rarely on the interaction between the two.

This division of labor has led to important advances in our understanding of prosody; we know a fair amount about what listeners can do with prosodic cues, and what prosodic cues speakers can produce.

Numerous language comprehension studies have demonstrated that prosodic manipulations of the linguistic input can influence comprehenders' on-line and off-line decisions about syntactic ambiguity (for reviews see, Warren, 1996; Kjelgaard & Speer, 1999). These studies have used a wide variety of experimental techniques, including cross-modal naming, lexical decision, word monitoring, and sentence judgments, and have found effects of prosody on the interpretation of a variety of temporary and global ambiguities. Likewise, studies of language production have found that the prosody of an utterance often reflects its syntactic structure (Cooper & Paccia-Cooper, 1980). Moreover, informed speakers can mark different meanings of an ambiguous string through prosodic grouping (Lehiste, 1976; Allbritton et al., 1996). These studies suggest that speakers of a language share knowledge about prosodic cues to syntax, and can use this information in decisions about production.

Curiously, most comprehension and production studies have relied upon distorted and/or artificial manipulations of prosodic information. In comprehension studies prosody is typically manipulated by splicing silent pauses into speech to indicate clause boundaries, manipulating synthesized speech, or asking trained speakers to produce particular prosodic variants of an utterance. Production studies have relied upon data from trained speakers, such as radio announcers, who have been explicitly instructed to contrast the alternate interpretations of an ambiguous sentence. Notably, few studies of prosody and syntax have examined how untrained listeners respond to the speech of untrained speakers in contexts in which the participants are attempting to communicate about a shared situation.

In naturally occurring speech, syntactic structure is only a weak predictor of prosodic variation (for review see Fernald & McRoberts, 1996) This is because prosodic patterns are affected by many other factors, including the length and stress pattern of words, speech rate and discourse factors such as contrastive stress (Selkirk, 1984). Unsurprisingly, a number of researchers have found that naïve speakers produce less consistent prosodic cues for syntactic disambiguation than the informed speakers typically used in comprehension experiments (Lehiste, 1973; Wales & Toner, 1979; Allbritton, McKoon, & Ratcliffe, 1996). In the most rele-

vant of these studies, Allbritton et al. (1996) compared situations in which speakers were uninformed or explicitly informed about potential ambiguities. In the uninformed case, 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") which had been disambiguated by the prior context. In the informed case, radio announcers were provided with the same globally ambiguous sentences without a disambiguating context, both of the meanings were explained to the speaker, and he/she was asked to read the sentence twice, to convey each of these meanings.

Recordings of these utterances were played for a separate group of subjects who were given both meanings and asked to identify the one that the speaker was attempting 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."

There are three plausible explanations of the Allbritton et al. findings, each of which has different implications for the role of prosody in syntactic ambiguity resolution. First, it is possible, as the authors claim, that speakers only produce reliable cues when instructed to do so. However, this would suggest that prosodic cues to structure are rare in natural speech, raising questions about how listeners become sensitive to these cues. Second, the results could be interpreted as evidence that speakers only produce reliable cues when the surrounding context does not disambiguate the utterance. In the Allbritton et al. study, experimental naïveté and contextual constraint were confounded. Perhaps, as Lieberman (1967) suggested, speakers don't bother to divide up an utterance into informative prosodic chunks if other cues are present to disambiguate structure. Finally, it is possible that speakers do not produce reliable prosodic cues when reading connected text, regardless of whether that text provides a disambiguating context.

Recently, Schafer, Speer, Warren & White (1999) have presented data which challenges the Albritton findings. 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 judgement 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, suggesting that the subjects in the earlier study were reading and had no clear communicative intentions.

The current paper attempts to explore the role of prosodic cues in language production and comprehension. In particular, we examine the situations under which untrained speakers can produce reliable 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 as Instrument (VP-Attachment) indicating what to use for the tapping, or the phrase can be taken as a Modifier (NP-attachment) indicating which frog to tap.

These experiments were conducted using a referential communication task, in which a Speaker and a Listener were separated by a divider, allowing for only verbal communication between the two participants. Under discussion in these studies was the movement of objects, with Speaker attempting to have the Listener perform actions upon an identical set of objects on the other side of the screen. This situation provided two advantages of over other common tasks. First, the referential context was highly salient, and was defined by the set of objects in front of the speaker and listener. Memory considerations for referential factors (e.g., what a speaker remembers about a story) are not relevant in such a task since the reference world is co-present with the production task. Second, the separation of the Listener and Speaker allowed us to manipulate independently the referential context of the Speaker and the Listener, allowing us to disentangle referential affects on the task of production and the task of comprehension.

In Experiment 1, we examined the use of prosodic cues when the referential context of the Speaker supported either meaning of the target sentence. In Experiment 2, we examined prosodic cues when the referential context of the Speaker strongly supported the intended meaning of the utterance. If prosodic choice is affected by Speaker's knowledge of the referential context, we would expect to see decreased use of helpful 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 would expect similar performance across the two experiments.

Experiment 1

Methods

Participants Thirty-two pairs of participants from the University of Pennsylvania community volunteered for the experiment. They received extra course credit or were paid for their participation. In each pair, one participant played the role of Speaker and the other played the role of Listener. All Speakers were female whereas 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. On each trial they were given identical bags containing toys, which they laid out on the trays in front of them. As the Speaker and Listener removed toys from their bags, the Experimenter introduced each toy using indefinite noun phrases (e.g., This bag contains a dog, a fan…).

Next, the Experimenter showed the Speaker a demonstration of the target action. This action could not be seen by the Listener. The Speaker then received a card containing a written sentence describing this action. Speakers memorized the sentence and returned the card to the Experi-

menter. After seeing a second demonstration, the Speaker produced the sentence. The Listener responded by attempting to perform the correct action with his or her own set of toys. Speakers were 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. Each Listener was told that her job was to perform the action that she believed had been demonstrated to the Speaker.

During the course of the experiment, interaction between the Speaker and the Listener was limited. Once the Speaker produced the sentence, the Listener could not ask for any clarification. Listeners' actions were videotaped and the Speakers' utterances were audiotaped. After the study was completed the Listener and Speaker were separated and each was interviewed to assess their awareness of the experimental manipulation and the ambiguity in the critical items.

Stimuli On critical trials, the target sentence contained an ambiguous Prepositional Phrase attachment, as in (1a) and (1b) below. Identical bags of objects were given to both participants. On each trial the bag contained: 1) a Target Instrument, a full scale object that could be used to carry out the action (e.g., a large flower); 2) a Marked Animal, a stuffed animal carrying a small replica of the instrument (e.g., a frog holding a little flower); 3) an Unmarked Animal (e.g., an empty-handed frog); and 4) two unrelated objects (e.g., a giraffe in pajamas and a lego block). The set of toys supported both interpretations of the ambiguous sentence by providing a potential direct object (plain frog) and instrument (large flower) for the VP-attachment and a potential direct object for the NP-attachment (frog holding flower).

The Experimenter demonstrated one of two possible actions: an Instrument action (e.g., the Experimenter picked up the large flower and tapped the plain frog) or a Modifier action (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).

1a. Tap the frog with the flower. (Amb, Inst)
 Action involves the unmarked frog and the instrument.

1b. Tap the frog with the flower. (Amb, Mod)
 Action involves the marked frog and not the instrument.

1c. Tap the frog by using the flower. (Unamb, Inst)
 Action involves the unmarked frog and the instrument.

1d. Tap the frog that has the flower. (Unamb, Mod)
 Action involves the marked frog and not the instrument.

Four presentation lists were constructed so that each of the 16 target trials appeared in only one of the four possible conditions on a given list but appeared in each of the conditions across lists (resulting in four target trials in each condition per subject pair). The target trials were interspersed with thirty distractor trials. Four additional lists were generated by reversing the order of trials in each list.

Coding The videotapes of Listeners' actions were edited to include only the actions on the sixteen target trials, and all audio was removed. Coders, who were blind to the condition of each trial, judged whether the Listener made an Instrument response (performed the target action using the Target Instrument or the miniature instrument).

Results

Listener's Actions The percent of Instrument responses in each of the four conditions is presented in Figure 1. Listeners' actions in response to the ambiguous instructions were affected by the action demonstrated to the Speaker ($F1(1,16) = 63.42$, $p < .001$; $F2(1,12) = 77.31$, $p < .001$). 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 action only 24% of the time.
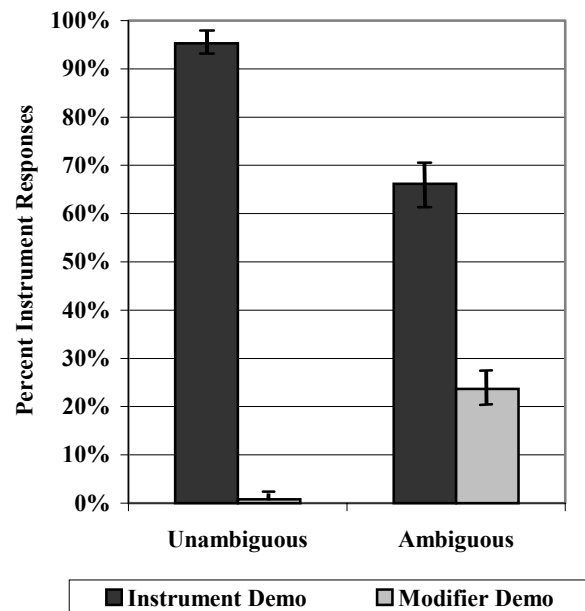

Figure 1: Experiment 1 Listener's Actions

Also, as expected, unambiguous instructions (the left-hand portion of Figure 1) resulted in extremely accurate performance by Listeners. The interaction between Ambiguity and Demonstration Type was reliable ($F1(1,16) = 81.91$, $p < .001$; $F2(1,12) = 113.71$, $p < .001$). As can be seen in the figure, this interaction arose because Listeners were more accurate at reproducing the demonstrated action when the utterance was syntactically Unambiguous than when it was Ambiguous. This pattern suggests that the prosodic cues produced by Speakers were highly informative to Listeners, but not as informative as unambiguous sentences.

Speaker's Prosody To verify that our Listeners were gleaning this information from prosodic cues provided by the Speaker, we conducted acoustic analyses of the ambiguous target sentences. The audio recordings were digitized and a speech waveform display was generated for each target utterance. Coders, who were blind to the condition, measured

the duration of the Verb Composite (verb plus the postverbal pause, if any) and the Noun Composite (the direct object noun plus the following pause, if any). The onset or offset of a word was initially estimated by using visual information from the speech waveform display. This estimate was revised by listening to gated regions of the waveform.
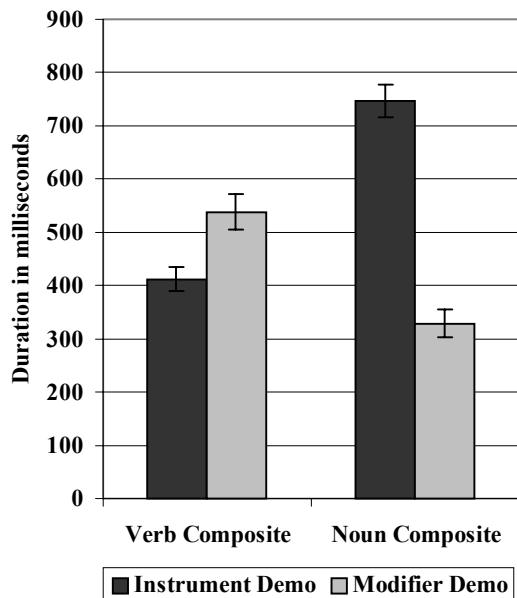

Figure 2: Experiment 1, Mean Durations from Speakers' Utterances

As Figure 2 suggests there is a reliable and substantial effect of demonstration on the mean duration of both the Verb Composite ($F1(1,24) = 12.92$, $p < .001$; $F2(1,12) = 50.59$, $p < .001$) and Noun Composite ($F1(1,24) = 52.71$, $p < .001$; $F2(1,12) = 290.42$, $p < .001$). When Speakers saw Instrument Demonstrations, they tended to lengthen the direct object noun and they paused between the noun that 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 the verb 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.

Ambiguity Awareness Listeners' actions in response to ambiguous instructions suggest that prosodic cues were a highly effective but imperfect means of syntactic disambiguation. However the results of the postexperimental interviews raised some concerns about the generality of these findings. 97% of the Speakers in our experiment and 91% of the Listeners reported being aware of the ambiguity. As mentioned earlier, Allbritton and colleagues (1996) found

that ambiguity awareness affected radio announcers' ability to generate useful prosody. Although our participants were not trained radio announcers, we thought it necessary to explore if ambiguity awareness, and more generally knowledge of the referential situation, were influencing the kinds of prosodic choices made by our Speakers.

Experiment 2

In this experiment, we attempted to decrease Speaker awareness of ambiguity. This was accomplished by making two changes to the previous experiment. First, we altered the Speakers' referential context, so that only the intended meaning of the ambiguous phrase was supported. This was done in hopes that the alternate interpretation would not be considered by the Speakers if it was not suggested by the context itself. Second, we made the type of Demonstration a between subjects variable.

All other aspects of Experiment 2 were the same as Experiment 1. It is especially important to note that the Listeners' context was the same as that used in Experiment 1. And, as in Experiment 1, participants were told in advance that on each trial the Speaker and Listener would receive an identical set of toys. However, in Experiment 2 this was a deception, to be explained at the end of the study.

Methods

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

Procedure The procedure was the same as Experiment 1 except that the contents of the bags were not listed aloud, to prevent the subjects from discovering that their bags contained different sets of objects. Instead a card listing the objects was included in each bag and the participants were told to check the contents of the bags against the card to insure that all of the toys were present.

Stimuli The stimuli and experimental design were the same as in Experiment 1, with the following exceptions. When the Experimenter performed an Instrument Demonstration, the Speaker's bag of toys did not include a Marked animal (e.g., the frog holding the flower) but instead included a second unrelated animal (e.g., an elephant wearing a hat). Hence, a modifier interpretation of the with-phrase should be less available to the Speaker. When the Experimenter performed a Modifier Demonstration, the Speaker's bag of toys did not include the Target Instrument (e.g., the large flower) but instead included a second unrelated object (e.g., a leaf). Hence, the Instrument interpretation of the with-phrase should be less available to Speakers in this context.

In addition, we excluded the unambiguous conditions in this experiment, because these sentences had been uniformly interpreted and coded correctly in Experiment 1. To equalize the number of ambiguous sentences that subjects
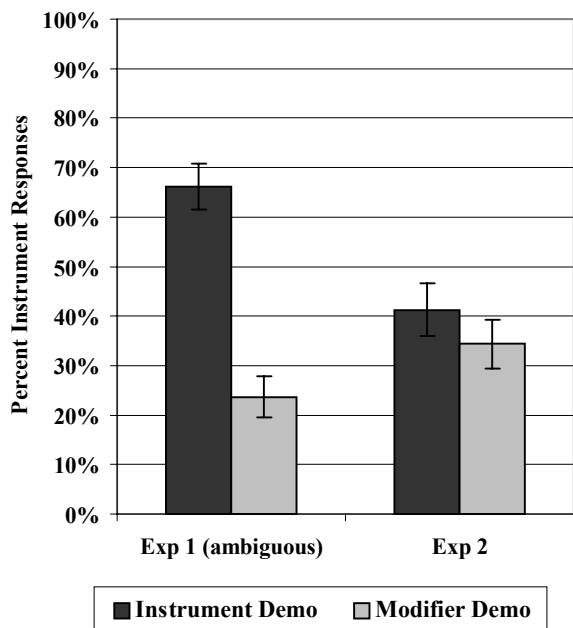
received in each experiment, we divided the 16 critical sentences into two lists. The items on each list appeared in a pseudo-random order embedded in the same twenty-four distractor trials. In addition, reverse-order lists were generated.

Finally, the type of Demonstration was manipulated between subjects. In the Instrument Condition, all target items were ambiguous and involved an Instrument demonstration (and an Instrument context for the Speakers). In the Modifier Condition, all target items were ambiguous and involved a Modifier demonstration (and a Modifier context for the Speakers).

Results

Ambiguity Awareness Listeners in Experiment 2, like those in Experiment 1, usually reported that they were aware of the ambiguity. This is to be expected, given that the same referential contexts were presented to Listeners in both experiments, and it suggests that the between-subjects design does not, by itself, affect ambiguity awareness.

Speaker awareness of ambiguity did change across experiments. In particular, only one speaker in the Instrument condition (6%) reported being aware of the ambiguity. Interestingly, and in contrast, nine of the Speakers in the Modifier condition, or 56%, reported being aware of the ambiguity. This pattern was unexpected; we were hoping that few if any of the Speakers would be aware of the ambiguity. This difference may be related to the fact that Modifier attachments are dispreferred by readers, especially with action verbs (Spivey-Knowlton & Sedivy, 1994). Conflicts between lexical and referential cues in the Modifier condition may brought the ambiguity into awareness. As we

see, this unexpected result is serendipitous, because it allows us to compare the performance of the Listeners who heard utterances from aware and unaware Speakers.

Listener's Performance The percent of Instrument responses for each conditions of Experiment 2 appear on the right hand side of Figure 3. Listeners in Experiment 2 were clearly unaffected by the type of Demonstration performed by the ($F1(1,24) < 1, p > .3$; $F2(1,12) = 1.88, p > .3$), suggesting that Speakers were not effective in helping Listeners resolve the ambiguity. To compare performance in the ambiguous conditions of Experiments 1 and 2, ANOVAs were conducted on the percent correct for items and subjects with Experiment as a between subjects and within items factor. Unsurprisingly, Listeners in Experiment 1 performed significantly better than those in Experiment 2 ($F1(1,62) = 11.76, p < .001$; $F2(1,14) = 19.91, p < .001$).

Speaker's Prosody The audiotapes were digitized and coded in the manner described above. As Figure 4 suggests, there was no reliable effect of condition on the duration of the Verb Composite ($F1(1,24) < 1, p > .6$; $F2(1,12) = 1.69, p > .2$). The effect of condition on the Noun Composite was not significant in the subjects analysis ($F1(1,24) = 1.66, p > .2$) and was small but reliable in the items analysis ($F2(1,12) = 7.71, p < .05$).
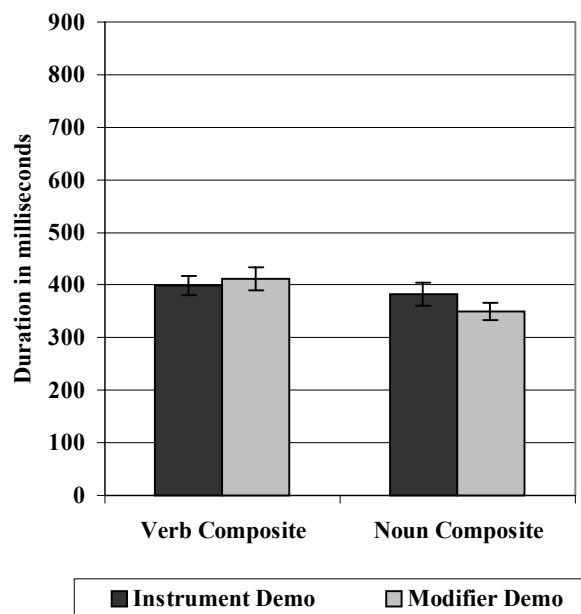

Figure 4: Experiment 2, Mean Durations from Speakers' Utterances

General Discussion

When the Speakers' context strongly supported the intended meaning of an ambiguous utterance, Listeners showed complete insensitivity to the intended meaning of the utterance. This stands in contrast to Experiment 1, where Listeners had the same referential context, but were highly sensitive to the intended meaning. There are two possible explanations to

shall

Figure 3: Listeners' Responses to the Ambiguous Sentences in Experiments 1 & 2

these findings. First, the findings may be attributable to Speakers' awareness of the ambiguity (Allbritton et al, 1996). Speakers in Experiment 1 were almost always aware of the ambiguous sentences while those in Experiment 2 were usually unaware. Second, the findings may be due to the change in the referential context of the Speakers. In the first experiment, Speakers were given a context that supported either meaning of the ambiguous sentence, whereas in Experiment 2 Speakers were provided with a context that supported only the relevant interpretation. Perhaps speakers only produce informative prosody when the context doesn't disambiguate the sentence for them (Lieberman, 1967).

Because a subset of the subjects reported being aware of the ambiguity in the Modifier condition, we can test if ambiguity awareness per se is driving the results of Experiment 2. Ten additional subject pairs were tested to gather sufficient data for this comparison. We found no reliable differences between responses to aware speakers and those to unaware speakers ($F1(1,23) < 1$, $p > .8$; $F2(1,15) < 1$, $p > .9$). When the referential context disambiguated the sentence, awareness of the **potential** for ambiguity did not lead speakers to provide adequate prosodic cues.

The substantial difference in performance across the studies and the absence of an effect of awareness within Experiment 2, suggest that referential context itself is critical in determining whether speakers will produce strong prosodic cues. 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. Under these conditions, Speakers produced prosodic cues that were not only consistent with the intended structure but also inconsistent with the competing interpretation. Listener's were able to use these cues to determine the intended meaning, albeit imperfectly. In Experiment 2, the Speaker's referential context supported only the relevant interpretation, disambiguating the sentence and making strong prosodic cues unnecessary. Listeners, who did not have access to this disambiguating referential context, were able to find nothing in the Speakers' prosody to guide them. These data, therefore, support Lieberman's hypothesis that speakers only produce informative prosody when the context doesn't do the work for them.

This conclusion and these results appear to conflict with those of Schafer and colleagues who find that untrained, uninformed speakers produce consistent prosodic cues regardless of whether the context of the utterance provides disambiguating information (1999). To add to the confusion the tasks appear to be quite similar: both experiments use variants of the referential communication task and ask speakers to produced scripted, memorized commands to achieve concrete results. The two experiments, however varied in several critical respects. First, in the Schafer study, there is a higher degree of uncertainty about the listener's referential context. The speaker knows both that the listener has information about the context that the speaker lacks and that the listener's context will change as the experiment progresses. Second, the participants are given the set of commands, which contains both interpretations of the ambiguity, at the beginning of the study and are exposed to situations in which each meanings is applicable. Thus it seems

likely that these speakers were aware of the globally ambiguous sentences and believed that there was the potential for referential ambiguity.

In this paper, we have suggested that a speaker's knowledge of the referential situation affects her ability to disambiguate otherwise ambiguous utterances. In particular, we propose that when a speaker recognizes that an utterance is ambiguous in context, she will disambiguate it by making prosodic choices that are consistent with the relevant interpretation and inconsistent with the alternatives.

References

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

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

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.

Lehiste, I. (1973). Phonetic disambiguation of syntactic smbiguity. *Glossa*, **7**, 102-122.

Lieberman, P. (1967). *Intonation, perception and language*. Cambridge MA: MIT Press.

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

Schafer, A., Speer, S., Warren, P., & White, D. (1999). Intonational disambiguation in sentence production and comprehension. Paper presented at the Twelfth Annual CUNY Conference on Sentence Processing, New York, NY, March 1999.

Selkirk, E.O. (1984). Phonology and syntax: The relation between sound and structure. Cambridge, MA: MIT Press.

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

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

Warren, P (1996). Prosody and parsing: An introduction. *Language and Cognitive Processes*, **11**, 1-16.