

Running Head: UNTRAINED SPEAKERS' PROSODY

To cite, please refer to: Fox Tree, J. E., & Meijer, P. J. A. (2000). Untrained speakers' use of prosody in syntactic disambiguation and listeners' interpretations. *Psychological Research/Psychologische Forschung*, 63, 1-13.

Untrained Speakers' Use of Prosody in Syntactic Disambiguation  
and Listeners' Interpretations

Jean E. Fox Tree and Paul J. A. Meijer  
University of California, Santa Cruz

Corresponding Author:

Jean E. Fox Tree, [foxtree@cats.ucsc.edu](mailto:foxtree@cats.ucsc.edu)

## Abstract

We investigated how naïvely produced prosody affects listeners' end interpretations of ambiguous utterances. Non-professional speakers who were unaware of any ambiguity produced ambiguous sentences couched in short, unambiguous passages. In a forced-choice task, listeners could not tell which context the isolated ambiguous sentences came from (Experiment 1). But listeners were able to correctly paraphrase the least-ambiguous subset of these utterances, showing that prosody can be used to resolve ambiguity (Experiment 2). Nonetheless, in everyday language use, both prosody and context are available to interpret speech. When the least-ambiguous sentences were cross-spliced into contexts biasing towards their original interpretations or into contexts biasing towards their alternative interpretations, answers to content questions about the ambiguous sentence, confidence ratings, and ratings of naturalness all indicated that prosody is ignored when context is available (Experiment 3). Although listeners can use prosody to interpret ambiguous sentences, they generally don't, and this makes sense in light of the frequent lack of reliable prosodic cues in everyday speech.

## Untrained Speakers' Use of Prosody in Syntactic Disambiguation and Listeners' Interpretations

In conveying information with speech, people do more than select the right words and syntax to express their ideas. They also give their words rhythm, put pauses between words or phrases, and emphasize some words over others. The resulting melody of speech, or prosody, can convey information to listeners above and beyond the words' propositional content. For example, emphasis can make contrasts clearer, as in "take the NEXT right turn, not THIS right turn," or highlight information, as in "there was a party at HERB's house" (Altenberg 1990; Gernsbacher & Jescheniak, 1995; Levelt, 1989). Prosody can also convey emotional content such as approval, disapproval, or humor (Fernald, 1989; Ladd & Cutler, 1983); it can indicate that the literal meaning of an utterance be interpreted figuratively, conveying irony or sarcasm (Scherer & Wallbott, 1985; Winner, Windmueller, Rosenblatt, Bosco, Best, & Gardner, 1987); it can be used for rhetorical effect when giving speeches (Atkinson, 1984; Clemmer, O'Connell, & Loui, 1979); and it can be used to organize information in speech, such as by marking off discourse units (Stenström, 1986; Swerts & Geluykens, 1994). Many of these uses of prosody have been shown to affect listeners' interpretations or their actions. Disapproving prosody can make babies cry, carefully timed prosody in speech can make audiences applaud, a mocking tone of voice can convey sarcasm, and precisely placed emphases can indicate to listeners what information to store in memory for future reference. In this paper, we will explore the use of prosody in resolving syntactic ambiguity; in particular, we will look at how reliable a cue prosody is and whether prosody can lead listeners to the intended interpretation of a syntactically ambiguous utterance.

Though syntactic ambiguity has long been a research topic for many psycholinguists who study sentence comprehension, noticing ambiguous sentences in everyday speech is a rare occurrence. Syntactic ambiguity may go by unnoticed because

the ambiguous sentences are pronounced so that they are unambiguous. For example, she saw a man eating fish may be pronounced as “she saw a man...eating fish,” or it may be pronounced as “she saw...a man eating...fish.” Pause patterns are only one characteristic contributing to the prosody of an utterance; other factors include differences in relative stress and syllable duration, changes in amplitude, and changes in sentences’ pitch contour (Ladd & Cutler, 1983; Levelt, 1989; Selkirk, 1984). Our intent here is not to describe the various uses people can put their voice to, nor to detail precisely what combination of stress, duration, intensity, and pitch leads to a particular interpretation. Instead, we will look at the role of all these factors treated together as a functional unit, in whatever combination the speakers happen to produce in creating the utterances investigated in this series of experiments.

Prosodic information has been shown to provide listeners with valuable information about sentence structure. Several researchers have determined that these cues can be used to select the appropriate interpretation of syntactically ambiguous sentences (Beach, 1991; Marslen-Wilson, Tyler, Warren, Grenier, & Lee, 1992; Price, Ostendorf, Shattuck-Hufnagel & Fong, 1991; Speer, Crowder, & Thomas, 1993; Wales & Toner, 1979). For example, Price et al. (1991) presented listeners with ambiguous sentences such as the old men and women watched television, which can be seen as either (old men) and (women) or old (men and women). The sentences were originally produced as the last sentence in a passage biasing towards one interpretation of the ambiguous sentence. The listeners heard these sentences in isolation and then decided which of two possible contexts would go best with the sentence. Listeners were highly accurate at choosing the context in which the ambiguous sentence had originally been produced. Similarly, Beach (1991) found that listeners can accurately determine which interpretation of an ambiguous sentence was meant after hearing only the initial portions of synthesized sentences. For example, the initial phrase Mary suspected her boyfriend contains enough prosodic information for listeners to correctly choose whether was lying

to her or immediately originally followed it. These results have been replicated for short initial phrases (Stirling and Wales, 1996) and have been observed using a cross-modal naming task (Marslen-Wilson et al., 1992). Other researchers using other tasks have also found that prosodic information can be used by listeners in parsing speech (see Warren, 1996, for a review).

Given this research, one plausible explanation for not noticing ambiguity in everyday talk is that prosody can be counted on to disambiguate the speech. But there are at least two ways that the role of prosody in everyday talk might have been over-emphasized in this research.

One reason the results of these studies might not generalize to everyday speech is that disambiguating context was eliminated. Syntactic ambiguity may go by unnoticed because of the swift resolution provided by context. A garden path sentence like The horse raced past the barn fell when read from visual text is no longer perceived as ambiguous in a context where several horses are mentioned, one of which is racing past a barn (Altmann, 1987; Altmann & Steedman, 1988; Crain & Steedman, 1985). Price et al. (1991) in their speech experiments presented the ambiguous sentences in isolation, and then presented the two possible interpretations for selection. Though they demonstrated that prosody is used in this situation, it is possible that it isn't when context is available before hearing the ambiguous sentence.

Another reason why the results of these studies might not hold in real-life situations is that the speech tested was always carefully prepared, read aloud, or in some cases even artificially created (Beach, 1991). Price et al. (1991) used speech recorded by carefully instructed professional radio newscasters, because these speakers are argued to produce prosodic cues more clearly and consistently. As Price et al. noted (p. 2959), the results obtained with these stimuli do not automatically generalize to spontaneous speech. Wales and Toner (1979) tested read speech, which varies from non-read speech in that it is better articulated, has fewer and shorter pauses, has more pauses related to syntactic

structure, and is more rhythmically predictable (Chawla & Krauss, 1994; Clemmer et al., 1979; Howell & Kadi-Hanifi, 1991; Mehta & Cutler, 1988). The use of carefully prepared speech is common in research on prosody and parsing, and tests are often done with speech from a single trained speaker (e.g., Pynte & Prieur, 1996; Schafer, Carter, Clifton, & Frazier, 1996). With other studies, it's not clear exactly how the speech was produced, and if using ambiguous material, how aware of the ambiguities the speakers were (e.g., Stirling & Wales' 1996 replication of some, but not all, of Beach's 1991 results).

In the current research, we address both the issue of context and the issue of naturalistic speech. We are interested in which potential cues, context or prosody, plays the more important role in determining listeners' final interpretations. There are two components to our investigation: one is whether naïve speakers automatically insert disambiguating prosody into their ambiguous utterances as they say them, and the other is whether listeners use this information. So, though it's been demonstrated that prosody can be used to disambiguate speech in conditions of neutral context or with specialized speech, we'd like to know whether it is generally used. Our interest also lies with what listeners come away with while listening. The question of how prosody and context and natural versus nonnatural speech interact in the on-line speech comprehension process is an interesting one, but not one that we will be addressing in this paper.

The first component of our investigation tests whether ambiguous sentences couched in disambiguating context contain consistently disambiguating prosodic cues when spoken by untrained speakers. This hypothesis was independently and simultaneously tested by Allbritton, McKoon, & Ratcliff (1996), who found that when speakers were unaware of the ambiguities, ambiguous sentences did not contain

consistently disambiguating cues. But when they were aware of the ambiguities, disambiguating cues, such as phrase-final lengthening, could be reliably added to speech. The second component of our investigation looks at how a range of prosodic structures naïvely produced by six speakers affect listeners' interpretations.

With respect to the first component of our investigation, the claim that prosody may not be a reliable cue to syntactic disambiguation has some support in the research literature. Nicol and Pickering (1993) tested how relative and complement clauses are parsed in the auditory speech stream. They played sentences like The receptionist informed the doctor that the journalist had phoned about the events which were pronounced either as relative clauses (that the journalist had phoned describes the doctor) or as complement clauses (that the journalist had phoned about the events describes what the receptionist told the doctor). Listeners then performed an on-line cross-modal lexical decision task to a visually presented relevant word, in this case doctor, at a point later in the speech stream, in this case when phoned was heard. The researchers found that even with intentionally disambiguating prosody, both possible parses were available at the time phoned was heard. So, at least for relative clauses versus complement clauses, listeners did not pick up on prosody to help resolve the ambiguities. They argue that “intonation is only probabilistic; any one pattern is only likely to convey a particular interpretation; there is no one intonation pattern which corresponds to either the relative clause or the complement clause construction” (Nicol & Pickering, 1993, p. 226). Ferreira, Henderson, Anes, Weeks, and McFarlane (1996) and Ferreira, Anes, and Horine (1996) also found that deliberately produced disambiguating prosody is not always successful in disambiguating for listeners. The first study compared active versus reduced

relative ambiguities such as the editor played the tape and agreed the story was big versus the editor played the tape agreed the story was big, and the second compared prepositional phrase attachment ambiguities such as Andrea moved the bottle under the bridge, which can be read as either Andrea moved (the bottle under the bridge) or Andrea moved the bottle (under the bridge).

Our replication began by collecting ambiguous speech spoken by untrained college students who memorized and then said a series of three sentence passages. Although these materials were not spontaneous, they were produced without the visual support of written text to guide the talk. The materials also approach the everyday speaking situation through the use of naïve, human speakers. We needed to cue speakers about what to say in order to collect productions of particular ambiguous sentences and contexts. Recalled materials are closely related to spontaneously produced utterances; as McDonald, Bock, and Kelly (1993, p. 196) observed, “In many experiments, the recall of sentences has been found to be sensitive to factors that characterize natural formulation processes (see Bock, 1982, for a review), probably because the reconstruction that underlies the recall of linguistic materials rests heavily upon normal production mechanisms.”

In half of the passages speakers recalled, the middle sentence could be ambiguous, though the context always favored one interpretation. A new group of students then listened to the ambiguous middle sentences and judged in a forced-choice task which context the sentences came from (Experiment 1). This is similar to a paraphrase-choosing task used by others (Ferreira, Anes, & Horine, 1996). We then selected the sentences that had the most correct responses and examined whether these sentences, when presented in isolation, contained enough prosodic cues to deduce the correct interpretation (Experiment 2). Finally, we examined whether prosody would still be used as a cue when context was also available to resolve ambiguities. Each ambiguous

sentence was cross-spliced into a context biasing towards its original interpretation and a context biasing towards its alternative interpretation. Students were asked a content question about the ambiguous sentence and were asked to rate their confidence in their answer and the naturalness of the prosody of the total passage (Experiment 3).

## Experiment 1

### Method

Subjects. Eighteen students from the University of California at Santa Cruz participated in this experiment as part of their course requirements. Two subjects were eliminated from the analyses for failing to follow the instructions.

Materials and Design. Twelve ambiguous sentences were produced by six speakers, two female and four male, yielding 72 ambiguous stimuli. The 12 ambiguous sentences can be found in the Appendix. Five of these sentences were taken from Price et al. (1991). In these sentences, a mid-sentence word could either attach to the phrase on the left or the phrase on the right. For example, the word gradually in the sentence When you learn gradually you worry more could either describe a way of learning (When you learn gradually, ...) or a way of worrying (..., gradually you worry more). Price et al. (1991) studied these five sentences and found them to be clearly disambiguated by prosody; in more than 94% of the cases listeners were able to correctly guess the intended interpretation when they were presented in isolation.

Each sentence occurred in two passages, each of which biased towards one interpretation of the sentences. The passages consisted of an ambiguous middle sentence which was preceded and followed by disambiguating context, as in the following pair:

(1) Toni went deep sea diving in the Pacific Ocean. She saw a man-eating fish. It scared her.

(2) Jenny went to the Seafood restaurant. She saw a man eating fish. He seemed to like it.

The 24 passages are listed in the Appendix. All 24 passages also had companion passages

where the middle sentences were replaced by unambiguous sentences, as in the following:

(3) Toni went deep sea diving in the Pacific Ocean. She saw a dangerous fish. It scared her.

(4) Jenny went to the Seafood restaurant. She saw a man having dinner. He seemed to like it.

The companion passages were necessary for Experiment 3, and were not used in Experiments 1 and 2.

The 48 passages were divided into two groups of 24. Each group contained one of the two passages containing an ambiguous sentence, and a companion passage with the context of the alternative version. So, for example, one group would receive passages (1) and (4) above, and the other group would receive passages (2) and (3). The speakers produced one of the two groups of 24 short passages, three speakers for each list. They were instructed to read and study each passage carefully and then recite the passage twice without looking at the written text. Their speech was recorded using a Sony ECM 77B microphone, a Sony PCM-2300 digital audio recorder, and a Mackie 1202 mixer. As determined in the debriefing session, none of the speakers were aware that some of the passages contained ambiguous sentences. When specifically asked about the middle sentences, some speakers noticed that they were shorter, but all gave answers unrelated to ambiguity, such as “dealt mostly with emotions.”

The more fluent of the two productions of a given ambiguous sentence for each speaker was digitally excised from the speech stream at 22 kHz using a speech editing program (SoundEdit 16) onto a Macintosh Power PC. By fluent we mean the rendition of the ambiguous sentence with the fewest hesitations, mispronunciations, or stutters. If each rendition were equally acceptable, one was chosen at random. This yielded a total of 72 sentences, 12 from each of the 6 speakers. In addition, 24 visual stimuli were created in which two matched passages appeared on the right and the left of the screen (12

passage pairs, one for each ambiguous sentence, in two versions where position of passages, left-right or right-left, was reversed). The middle sentence in these passages was removed and replaced by dots.

Visual and auditory stimuli were combined into an experiment using the Superlab experiment generating software. The experiments were generated based on two lists of 74 stimuli containing 2 practice trials and 72 experimental trials. The lists were pseudo-randomized such that the six instances of a given sentence were equally divided over the list, with utterances of the same speaker never appearing on consecutive trials. The second list was created by reversing the order of the 72 trials of the first list.

Procedure. Subjects were tested one at a time. They were seated in front of a computer and were given the instructions. After they had read those, two practice trials were administered followed by the 72 experimental trials. A trial had the following structure: First, subjects heard a beep signifying the beginning of a new trial. After a pause of 500 ms an ambiguous sentence was played over Sennheiser headphones, followed by another pause of 500 ms. Then a visual stimulus appeared containing the two passages without the middle sentence. Subjects then judged which passage the auditory stimulus went with. So, for example, subjects would hear the sentence she saw a man eating fish and then press a response key on the keyboard corresponding to the left or right side of the computer screen when they saw the following:

- |                                      |                                       |
|--------------------------------------|---------------------------------------|
| (5) Toni went deep sea diving in the | Jenny went to the Seafood restaurant. |
| Pacific Ocean. (...) It scared her.  | (...) He seemed to like it.           |

Subjects were instructed to pay special attention to the rhythm and melody of the sentence when making this decision. These instructions were intended to bias subjects towards using prosodic cues in making their decisions. The subjects had 30 seconds to read the passage pairs and push one of two buttons corresponding either to the left or the right passage. After making this decision, subjects had 10 seconds to judge the naturalness of the melody and rhythm of each stimulus on a 7-point scale. If the sentence sounded as if it could readily be produced in everyday

speech, it was considered natural (the score 7). If it sounded contrived or stilted, it was unnatural (the score 1). No subject reported any difficulty making these judgements. These off-line tasks tap into people's metalinguistic awareness of prosody. The experiment lasted about 20 minutes.

## Results

Nineteen trials in which subjects failed to respond in time or hit the wrong key on the keyboard were excluded from the analysis (2%). All subjects performed about equally well, missing between zero and three trials. Likewise, all 72 items were about equally difficult, with any particular item having no more than two missing data-points. Missing data were also equally divided over lists.

Overall, listeners could not accurately match an ambiguous sentence to its context. The subjects guessed the correct context in 51% of the cases, which was not significantly different from chance ( $t_1(15) = .49$ ,  $p = n.s.$ ;  $t_2(11) = .33$ ,  $p = n.s.$ ).

In post-hoc analyses, we investigated whether other variables might be influencing the results. None of the analyses reached significance at the 5% level, even before a Bonferroni correction of the alpha level to 1.25%. There was no effect of presentation order (48% versus 53%;  $t_1(14) = 2.01$ ,  $p = n.s.$ ;  $t_2(11) = 1.84$ ,  $p = n.s.$ ). Likewise, the items containing a left versus right attachment ambiguity were guessed at chance (51% correct;  $t_1(15) = .59$ ,  $p = n.s.$ ;  $t_2(4) = .31$ ,  $p = n.s.$ ). Performance was also not better for the renditions of a particular speaker (both  $F_s \leq 1$ ). The average percentages correct by speaker ranged from 43% to 54%.

The lack of an effect for the five attachment ambiguity items contrasts with Price et al.'s finding of a 95% accuracy rate for these same items. But in concert with Price et al., we also found no difference between left attachment versus right attachment (48% and 53% in our data; 94% and 94% in their data). Both sets of data show that there is no inherent bias towards left or right attachment.

Although overall listeners could not match an ambiguous sentence to its context, it is possible that some items were more successful than others. On the basis of the average percentages correct for all 72 stimuli, we selected the best of the six renditions for all 12

ambiguous sentences. The percentage of correct responses for these 12 items ranged between 56% and 88% with an average of 74%, which is significantly different from chance ( $t_1(15) = 6.63, p < .001$ ;  $t_2(11) = 7.61, p < .001$ ). We also selected a matched best set consisting of the 12 ambiguous renditions with the highest percentage correct in the alternative context. The scores for these 12 items ranged between 27% and 87% with an average of 49%, which is not significantly different from chance ( $t_1(15) = -.26, p = n.s.$ ;  $t_2(11) = -.19, p = n.s.$ ). Both the best set and the matched best set contained utterances from every speaker. Although this is fortuitous, it's important to emphasize what this means: the best set was not produced by a single speaker, or perhaps pair of speakers, who happened to provide a lot of disambiguating prosody. It's not the case that some speakers were better at prosodically disambiguating the materials than others. These best set and the matched best set of stimuli were used in Experiments 2 and 3. A summary of the Experiment 1 results for the complete set of stimuli, the best set, and the matched set can be found in Table 1.

-----  
Insert Table 1 About Here  
-----

Like accuracy, naturalness of prosody ratings also differed for the best set as compared to the items as a whole. Across items, subjects rated the stimuli as sounding fairly natural, with an average score of 4.9 on a 7-point scale with 7 being most natural. There was no correlation between the average percentage correct for a given stimulus and its average naturalness score ( $r = .12, p = n.s.$ ), but the 12 renditions of the best set were considered more natural than the other renditions (5.2 compared to 4.8,  $t_1(15) = 3.73, p < .01$ ;  $t_2(11) = 3.45, p < .01$ ). This means that the accuracy of the best set is not the result of unnatural prosodic cues guiding listeners' choices, as might be expected if speakers had produced utterances with unusually long pauses, as in for example "she saw a man ... eating fish." Keep in mind that in this experiment subjects heard only the middle sentences; they read the contexts. The prosodic naturalness ratings could only have applied to the critical middle sentences.

## Discussion

The results of this experiment show that most ambiguous sentences, when produced by naïve, nonprofessional speakers in a disambiguating context, do not contain enough prosodic cues to steer listeners towards the intended interpretations. Whatever prosodic information is contained in them, it is not enough for listeners to accurately choose between potential interpretations of the sentences. It is important to remember that we are not saying that prosody has no effect on comprehension at any time, just that naïve speakers do not in general produce useful prosody for syntactic disambiguation. This finding replicates that of Allbritton et al. (1996).

It is important to remember that our experiment was methodologically identical to Price et al.'s (1991) perceptual experiments, and so should have yielded the same effect were prosody behaving in the same way. For no item in our experiment was listeners' accuracy as high as that of the listeners in Price et al.'s (1991) experiments. There are at least three potential explanations for this difference between our results and Price et al.'s, and we'll discuss each in turn.

One reason might be that the contexts in our experiments were more biasing than in Price et al.'s. If our contexts more strongly biased towards a particular interpretation, then the ambiguous sentences might have been more likely to be linked to a passage than in Price et al.'s experiment, resulting in fewer responses made on the basis of prosody. But we think the contexts were more similar than different, especially for those five items of Price et al. for which we supplied the same lead-in context sentences before the ambiguous sentences. In any case, the differences between contexts are not of the magnitude of neutral context versus biasing context, but rather different kinds of biasing context. The contexts seem to be too close to cause such a huge difference in percentages correct.

A second reason that the percentages correct were so low might be that listeners forgot the specific prosodic information in the process of reading and evaluating the two contexts. At the time they heard the sentence they may have had one idea, but by the time they read the contexts, they may have become distracted by the two equally plausible alternatives, and

guessed. This scenario is unlikely because listeners have been able to retain prosodic information in memory long enough to reveal reliable benefits of prosody in other experiments carried out with exactly the same method (Price et al., 1991).

A third reason might be that using professional newscasters as speakers makes a difference. We believe this to be the correct explanation for the differences in our effects. Professional newscasters are trained to speak clearly and unambiguously, and a comparison of our data with Price et. al.'s shows that they do precisely that. Nonetheless, it's important to remember that most people are not professional newscasters, and so comprehension of everyday speech may not follow the same pattern as found in these earlier studies.

To investigate just how naïve speech differs from trained speech, we looked more closely at those utterances that were most unambiguous in Experiment 1. One possibility is that our naïvely produced materials contain no disambiguating prosodic information; if this is the case, then it is no surprise that listeners were at chance in identifying original contexts. Another possibility is that they contain prosodic information, but that this information plays a minor role in comparison to context when deciding the speaker's intent. Of our original 72 stimuli, we selected the stimuli of the best set and the matched set for the next two experiments. Recall that these materials consisted of the renditions of every ambiguous sentence that had higher percentages correct than the other renditions of the ambiguous sentence spoken in the same original context. We chose these stimuli because if any of our stimuli were to contain prosodic information, it would be these. Alternatively, these items' inclusion in the best set and matched best set may have been purely based on chance.

In Experiment 2, listeners listened to the ambiguous sentences in isolation and paraphrased them in their own words. They were not given a forced choice between possible interpretations, but were allowed to interpret the sentences as they heard them. This task forces subjects to rely solely on the characteristics of the ambiguous sentences, increasing their baseline sensitivity to prosodic features. If people can make use of prosodic information to interpret the ambiguous materials, then they should give semantically different paraphrases for the two

instances of each sentence. However, if they do not use prosody or if the stimuli do not contain enough prosodic cues, paraphrasing should be similar for both versions of the ambiguous sentences.

Another purpose for doing Experiment 2 was to collect information about what causes the best set to be less ambiguous than the matched set. There are at least three possibilities here. One is that people were better at the best set by chance. If this is the case, then testing interpretations again in Experiment 2 should show regression to the mean, with listeners' accuracy dropping to chance. Another possibility is that listeners were biased towards particular interpretations. For example, it may be more common to see a person eating fish than to see a shark, biasing the interpretation of she saw a man eating fish even before listeners saw the two contexts. The best set stimuli may have come from contexts that were more plausible, rather than having clearer disambiguating prosody. In Experiment 2, each ambiguous sentence will be presented in both the best set and the matched set renditions to each subject. If there is an interpretation bias, we should find that both sentences are paraphrased according to the best set interpretation. A third possibility for the best set advantage is that the best set is spoken with more disambiguating prosody; if this is the case, then the best set versions should have more accurate paraphrases than the matched set.

## Experiment 2

### Method

Subjects. Nineteen students from UC Santa Cruz participated in this experiment for course credit, 10 on one list and 9 on the other. Two subjects were eliminated for not following instructions, one from each list. The odd number of remaining subjects per list poses no problem for the interpretation of the results because due to data loss each subject did not have a response for each item. Data were lost when sentences were inaccurately recalled or ambiguously paraphrased. Interpretation was based on remaining useable items, however many per subject that turned out to be.

Materials. The best set of 12 items and the matched best set of 12 items from Experiment 1 were used in this experiment. So there were two instances of each ambiguous sentence, one originally produced in Context A and the other in Context B, with each instance spoken by different speakers.

Design. Two different lists were constructed. In one, the items of the best set were presented first, followed by the items of the matched set. In the other, the items of the matched set were presented first, followed by the items of the best set.

Procedure. Subjects were tested one at a time. They were seated in front of a computer and given instructions. A trial consisted of two hearings of an ambiguous sentence. After the first hearing, subjects wrote down the exact wording of the stimulus. This was to ensure that they had heard the sentence accurately. After the second hearing, subjects paraphrased the sentence in writing. They were instructed to minimize the use of words that occurred in the original sentence and also not to repeat large segments of the original utterance. As an example, they were told that they could paraphrase I like to drink hot coffee as someone says that he enjoys the taste of a hot drink made of beans, which repeats the word hot but avoids the phrase hot coffee. The experiment lasted about 25 minutes.

## Results

Responses were first scored according to the accuracy of the verbatim transcriptions. Of the 408 responses, 9% were inaccurate to the point of eliminating the ambiguity. These responses were excluded because we couldn't be sure that the test sentences were heard correctly. We accepted transcriptions with minor errors of number or agreement (he saw a man eating fish instead of she saw a man eating fish). Responses were then scored according to the useability of the paraphrases. An additional 20% of the paraphrases still contained the ambiguity (a female saw a man eating fish), and a handful of the paraphrases did not correspond to one of the two interpretations we were investigating. These responses were excluded because they could not be scored as either correct or incorrect relative to the stimuli heard. Of the remaining 290 cases (151 in list 1 and 139 in list 2), the correct response was given 56% of the time, a reliable effect over

both subjects and items ( $t_1(16) = 3.23, p < .005$ ;  $t_2(11) = 2.53, p < .025$ ). Although the effect is numerically small, it shows that listeners can pick up the prosodic cues available in the ambiguous sentences. Furthermore, the fact that the effect is valid across items shows that the effect is not driven by a few effective stimuli.

In addition to this effect of prosody, we also found an effect of interpretation bias. To do these analyses, we considered only those cases where listeners provided useable paraphrases to both versions of the ambiguous sentences (117 pairs). Listeners correctly paraphrased both versions 12% of the time. They incorrectly paraphrased both versions 1% of the time. In the remaining 87% of the pairs, listeners gave the same answer for both versions ( $t_1(16) = 10.98, p < .001$ ;  $t_2(11) = 8.54, p < .001$ ). In 61% of these biased pairs, people responded twice with the paraphrase that was correct for the ambiguous sentence of the best set (a best set bias). They responded with the correct paraphrase for the matched set in the other 39% (a matched set bias). There is a tendency for subjects to prefer one type of paraphrase over the other ( $t_1(16) = 2.97, p < .01$ ;  $t_2(11) = 0.64, p = n.s.$ ).

The biases are not influenced by which of the two ambiguous sentences subjects heard first. In list 1, where the items of the best set were followed by the items of the matched set, the best set bias was 65% versus a matched set bias of 35%. In list 2, where the items of the matched set were presented first, the best set bias was 55% versus 45% for the matched set. The order of presentation of the two sets did not significantly affect the proportion of the two biases ( $t_1(15) = 1.74, p = n.s.$ ;  $t_2(10) = 0.79, p = n.s.$ ). So it is not the case that listeners chose an interpretation on the basis of prosody and then on hearing the second sentence merely gave the same response as with the first sentence. If it were the case, we'd find an interaction between list presentation and bias type such that when the best set was heard first, people would apply the best set interpretations to the subsequent matched set, but when the matched set was heard first, people would apply the matched set interpretations to the best set. What happens instead is that people choose an interpretation based on their interpretation bias, which in most cases corresponds to the best set interpretation, and then apply that same interpretation to the second hearing.

So both prosody and context can influence the interpretation that listeners assign to a syntactically ambiguous utterance. The contextual influence is particularly strong, and its influence masks the effect of prosody in the whole set. Taking the biased responses out and looking at the 86 cases that were not part of the 87% of interpretively biased pairs -- that is, the 12% of pairs that were correct, the 1% that were incorrect, and all the other items that contained only one useable response -- 70% were paraphrased correctly ( $t_1(16) = 4.04, p < .001$ ;  $t_2(11) = 2.32, p < .025$ ). This means that the effect of prosody over all the items may actually be a more pronounced effect that's been buried by interpretation biases.

The fact that subjects can accurately interpret the prosodic cues in the best set and matched set items shows that the high accuracy for many of these items in Experiment 1 was not a chance occurrence. But perhaps there is still a difference between the best set and the matched set. On average, 66% of the best set items were accurately paraphrased, compared to 44% of the matched set ( $t_1(16) = 5.58, p < .001$ ;  $t_2(11) = .45, p = n.s.$ ), showing a trend towards a best set advantage. But when the bias effect is removed by excluding those items where the same answer was given for both versions, the difference between the best set and the matched set is no longer significant (79% vs 61%;  $t_1(11) = 2.11, p = n.s.$  [five subjects had no unbiased responses];  $t_2(11) = .29, p = n.s.$ ). So the best set advantage over the matched set is primarily driven by bias effects.

### Discussion

When presented in isolation and paraphrased, the 12 most unambiguous items from Experiment 1 and their matched versions are significantly more likely to be paraphrased correctly than incorrectly. The reliability across items shows that the stimuli contained enough disambiguating prosodic information to allow a correct interpretation, although the size of the effect shows that listeners did not always pick up on this information.

This finding is in line with work by Ferreira, Anes, and Horine (1996), who also found that when prosody is the only disambiguating cue, accuracy at interpreting meaning was far from perfect, ranging from chance performance to 84% accuracy depending on the type of ambiguity

heard.

In addition to the effect of prosody, Experiment 2 also revealed that there are strong biases towards particular interpretations of the ambiguous stimuli. When people demonstrated an interpretation bias by providing the same paraphrase for stimuli originally produced with different meanings, they preferred the interpretation of the best set over the matched set. So most of the difference between the items of the best set and those of the matched set in Experiments 1 and 2 can be ascribed to a preference towards choosing a context or coming up with a paraphrase that agrees with the best set interpretation. Of these most unambiguous items, those items where listeners did not show an interpretation bias were paraphrased correctly 70% of the time. So, people can use prosodic cues to interpret ambiguous sentences that were naïvely produced, and they do so more when they are not swayed by interpretation biases.

In Experiment 1, we investigated the role of naïvely produced prosody in choosing between two contexts and found that prosodic cues were too weak to lead listeners to make accurate choices. In Experiment 2, we selected the stimuli most likely to contain strong-enough prosodic cues and investigated the role of prosody in the absence of disambiguating context. We found that prosody can influence listeners' interpretations, though nonprosodic biases also play a major role in decisions about utterance meaning. In Experiment 3, we investigate the relative contribution of prosody and context to the utterance interpretations listeners assign. In both Experiments 1 and 2, there was no conflicting information; listeners could either use or not use prosody in making their choices about contexts or in developing their paraphrases. In Experiment 3, listeners had to choose between using context or prosody in interpreting the ambiguous utterances.

We pitted context against prosody by splicing the ambiguous sentences into appropriate and inappropriate contexts and comparing answers to questions about the resulting passages. In passages that combine the ambiguous sentence of one context with the surrounding sentences of the other context, listeners might become confused and provide less systematic answers than

when the two factors point to the same interpretation. Another possibility is that when faced with conflicting signals, listeners will rely more on prosody than they did in Experiments 1 and 2, and provide answers that correspond more often with prosody over context. Alternatively, prosody might be totally ignored in the presence of disambiguating context. Responses would then not be affected by whether prosody and context match. Yet another possibility is that listeners rely heavily on context but do process the prosodic cues. In that case, differences between the passages might not show up on the content questions but might affect how confident the listeners are about their understanding of the passages.

Because listeners are introduced to the ambiguous sentence by a contextualizing utterance, interpretation biases should be minimized, allowing prosodic effects to make their full contribution. These interpretation biases were not minimized in Experiments 1 and 2 because listeners first heard the sentences in isolation; they could have imagined appropriate contexts, and then maintained their interpretations when making the forced choices or when paraphrasing (see Wales & Toner, 1979, for a similar argument). By restricting the interpretation biases in Experiment 3, we increase the contribution of the prosodic cues that were found to be of a reliable advantage in Experiment 2.

### Experiment 3

#### Method

Subjects. Thirty-six students at UCSC participated in this experiment for course credit, nine students per list.

Materials. The same 24 ambiguous stimuli used in Experiment 2 were used in the current experiment. The original contexts in which the ambiguous sentences were produced as well as their alternative contexts spoken by the same speaker were selected and digitized. An example of a stimuli pair spoken by one speaker is repeated here:

(6) Toni went deep sea diving in the Pacific Ocean. She saw a man-eating fish. It scared her.

(7) Jenny went to the Seafood restaurant. She saw a man having dinner. He seemed to like it.

Recall that each speaker produced each three sentence passage twice, and that only one passage

of a pair contained the ambiguous sentence. The ambiguous sentences were then spliced into one of the recitals of the alternative contexts and also spliced into the recital of the proper context that was not the original recital. For example, if the stimulus occurred in the speaker's first production, then the context of the second recital was selected. The fact that all conditions contained spliced materials is important because it ensures that the naturalness judgements are based on the same smoothness or disruptiveness of the middle sentence relative to the surrounding context across both the match and mismatch conditions. So for each of the 24 stimuli, two auditory passages were created:

Passage 1: Prosody/Context Congruence (PCC). The ambiguous sentence originally produced in a given context is spliced into the other token of that context.

Passage 2: Prosody/Context Incongruence (PCI). The ambiguous sentence originally produced in a given context is spliced into the alternative context.

The resulting 48 stimuli represent 12 pairs in four conditions. Here is an example of the four passages of she saw a man eating fish, assuming that for both Contexts A and B, the first recitation of the critical ambiguous sentence was used (the labels A and B refer to the context, the numbers refer to the speakers who produced the stimuli, and f or s refers to the first or second recitation of the utterances by the speaker, with f/s meaning either recitation could be used):

(8) PCC: (Toni went deep sea diving in the Pacific Ocean.)A1s (She saw a man eating fish.)A1f (It scared her.)A1s

(9) PCI: (Jenny went to the Seafood restaurant.)B1f/s (She saw a man eating fish.)A1f (He seemed to like it.)B1f/s

(10) PCC: (Jenny went to the Seafood restaurant.)B2s (She saw a man eating fish.)B2f (He seemed to like it.)B2s

(11) PCI: (Toni went deep sea diving in the Pacific Ocean.)A2f/s (She saw a man eating fish.)B2f (It scared her.)A2f/s

As the examples show, each passage contained speech from only one speaker.

Design. In total, there were 48 stimuli divided equally over the conditions as well as

three practice trials. Four lists of 12 critical trials were created. Each list contained the 12 ambiguous sentences in one of the four conditions such that each list contained three instances of each condition. The order of the materials was the same in each list.

Procedure. Subjects were tested one at a time. They were seated in front of a computer where they read the instructions. After the three practice trials, they completed the 12 critical trials. A trial had the following structure: There was a warning beep, followed by the passage. The passage was played in three parts. The first part contained everything before the ambiguous sentence, the second part contained the ambiguous sentence, and the third part contained the rest of the passage. After each part had played, a row of dots appeared on the screen. The next part was played after the subject pressed a key. This ensured that subjects had enough time to understand each sentence before moving on to the next.

After each passage was played, a question appeared on the screen. The question was devised to test subjects' understanding of the passage. The questions all followed the format given by the following example, which accompanied the she saw a man eating fish ambiguous sentence: (12) The fish in this passage was probably a (1) cod (2) shark.

Other questions included This passage suggests that it is better to learn (1) less (2) at a faster pace for when you learn gradually you worry more, They got up early in the (1) month (2) morning for They rose early in May, and It is dangerous to (1) fly a plane (2) have a plane fly over you for flying planes can be dangerous. All 12 questions were a forced choice between two responses, one referring to Context A, and one referring to Context B. After answering the question, subjects gave a confidence rating for their answer and also rated how natural the passage sounded on a 7-point scale with 1 being unnatural/not confident and 7 being natural/confident. Naturalness ratings were intended as a collection of listeners' subjective impressions, which we thought might vary across the congruent versus incongruent conditions. They were not intended to be sophisticated measures of prosodic continuity. The naturalness ratings do prime listeners to pay attention to prosody, but as we will see, this bias was negligible. The experiment lasted approximately 10 minutes.

There are at least three possible patterns of responses to the questions, which we'll illustrate with reference to examples (8) - (11). If subjects relied solely on the context and not on the prosody of the ambiguous sentences, they would answer cod for (9) and (10) and shark for (8) and (11). But if subjects relied on prosody, then they would answer cod for (10) and (11), but shark for (8) and (9). Alternatively, both prosody and context might influence the results. One possible pattern might then be that subjects relied mostly on context to answer the content questions (answering cod for (9) and (10) and shark for (8) and (11)), but their confidence ratings might be higher when context and prosody were congruent than when they were incongruent. The naturalness ratings might agree or disagree with the confidence ratings.

### Results

Trials in which subjects failed to press a key on the keyboard for either the question or the ratings were excluded from the analysis. This was the case for only 2 of the 432 trials. Table 2 summarizes the results for the answers to the content questions, the confidence ratings, and the naturalness ratings.

-----  
Insert Table 2 About Here  
-----

Let's first consider the answers to the content questions. When prosody and context were congruent, the right answer was given in 84% of the cases. When prosody and context were incongruent, answers can be scored either relative to prosody or to context. When they were scored relative to prosody, the correct answer was given in 19% of the cases. These scores are significantly different ( $F_1(1,35) = 199.42$ ,  $MSe = .68$ ,  $p < .001$ ;  $F_2(1,11) = 139.64$ ,  $MSe = 2.92$ ,  $p < .001$ ). When they were scored relative to context, the correct answer was given in 81% of the cases, which is not significantly different from the 84% correct in the congruent cases ( $F_1$  &  $F_2 < 1.15$ ). These results show that prosody was not used in deciding what the ambiguous sentences meant. Listeners relied on context to determine their answers. But results also show listeners do not automatically choose interpretations based on context; 16% of the time that the context and

prosody matched, unambiguously guiding listeners towards a particular interpretation, listeners actually chose the alternative interpretation.

Confidence ratings were not affected by whether the context and prosody were congruent or were incongruent. The average rating was 6.1 if context and prosody were congruent and 5.9 if prosody and context were incongruent (only the 355 correctly answered items were analyzed;  $F_1(1,34) = 3.15$ ,  $MSe = .90$ ,  $p = n.s.$ ;  $F_2(1,11) = 1.15$ ,  $MSe = .32$ ,  $p = n.s.$ ). Likewise, naturalness ratings showed no systematic pattern. The average rating was 4.4 when context and prosody were congruent and 4.8 when they were not ( $F_1(1,34) = 3.83$ ,  $MSe = 1.03$ ,  $p = n.s.$ ;  $F_2(1,11) = 2.06$ ,  $MSe = .98$ ,  $p = n.s.$ ).

### Discussion

In this final experiment we tested whether people use prosodic cues when they also have contextual cues available. All results clearly indicate that they do not. When context is available, people rely on the content of the passage in interpreting ambiguous sentences, ignoring prosody. Even listeners' judgements of naturalness were unaffected by incongruent prosody. Passages in which ambiguous sentences were spliced into their opposing contexts were not judged less natural than those containing consistent prosody. In addition, subjects were just as confident about their answers to congruent and incongruent passages.

It may seem obvious that when juggling conflicting information, people choose interpretations based on what makes the most sense semantically, yielding unsurprising context-based decisions. Furthermore, given common distortion in speech caused by accents, disfluencies, and other things like background noise, it might come as no surprise that listeners decide on what sentences mean based on how they are used in context rather than how they are pronounced. But there is no reason to think that semantic information necessarily makes more sense than prosodic information, or that distortion necessarily has no effect on comprehension, or that a lack of an effect at the level of deciding what a sentence means necessarily implies a lack of an effect in naturalness or confidence ratings. Indeed, in Experiment 3 listeners chose a

meaning that did not go with the context almost 20% of the time. In some of these cases, context and prosody matched, yet listeners still chose the incorrect alternative interpretations. So, sometimes something other than context can make more sense to listeners.

### General Discussion

The experiments described here suggest that awareness of ambiguity in everyday speech is rare because conversations provide context, and context can reliably resolve the ambiguities. The experiments further suggest that prosody can be used to disambiguate, but generally it isn't, and especially not when the ambiguity can be resolved contextually.

In three experiments, we tested the extent to which prosodic cues help disambiguate syntactically ambiguous sentences. The speech we tested was produced by nonprofessional speakers who were naïve to the presence of the ambiguities. They memorized and then recited ambiguous sentences which were sandwiched between disambiguating contexts.

In the first experiment, we tested whether the isolated ambiguous sentences contained prosodic signals that hinted towards the context in which they were originally produced. Listeners heard each of the speakers' ambiguous sentences and then judged which of two visually-presented contexts the sentence was originally produced in. Because both possible interpretations were available at the same time, contextual influences were minimized while prosodic influences were maximized. Listeners could not accurately determine which context the ambiguous sentences came from. Whatever prosodic cues were present, they were not strong or plentiful enough to allow for successful disambiguation. This finding replicates that of Allbritton et al. (1996), who also found that untrained speakers do not reliably add disambiguating prosody to syntactically ambiguous sentences.

In the second experiment, we identified some sentences where listeners were more successful at choosing the original context and tested whether new listeners could accurately paraphrase these sentences when presented in isolation. The listeners in Experiment 2 were not informed about the multiple interpretations of the sentences, allowing us to measure whether or not one context was inherently more likely than another. The results of this second experiment

indeed showed strong inherent biases towards certain interpretations of the sentences, but it also showed that prosodic cues were present and could be used in interpreting these ambiguous sentences.

In the third and final experiment, both context and prosody were available to subjects in interpreting the ambiguous sentences. A forced-choice answer task, confidence ratings, and ratings of naturalness all showed that context was used to interpret the passages. There was no measurable effect of prosodic information. Prosodic cues that could be used reliably in the absence of contextual cues were ignored when context was available.

A possible counterargument to the Experiment 3 results is that the tests were not sensitive enough to detect prosodic influences. But prior research has suggested that both judgements of naturalness and confidence ratings should have been sensitive enough.

In a study of the role of prosody in predicting whether an utterance like Jay believe the gossip originally preceded about the neighbors right away or about the neighbors wasn't true, Stirling and Wales (1996) found that people judged cross-spliced sentences as containing less normal intonation than congruent sentences, although the effect was small. Our results may have differed from Stirling and Wales' because the sentences were produced in different ways.

In a study on the role of prosody in disambiguating syntactic ambiguities and also ambiguities of sentence focus, Speer et al. (1991) found that confidence ratings reflected differences in prosodic information. In that study, people first listened to sentences in anticipation of a subsequent memory task. In the memory task, the listeners rated whether or not they had heard the same wording in the earlier learning phase. Sometimes the ambiguous sentences in the learning and test phases contained similar prosodic cues, such as an emphasis on frying in They are FRYING chicken. But other times, the test sentence contained a different prosody from the learning phase, as in They are frying CHICKEN. Though their answers were accurate, people were less confident when the alternative prosody was heard in the test phase than when the original prosody was heard. Our results may have differed from Speer et al.'s

because the tasks were different. Speer et al. investigated how prosody affected memory for speech, not how prosody affected understanding. A memory effect does not necessarily imply a comprehension effect. We can remember the words for songs better when we remember the accompanying tune, but this doesn't mean that singing something helps us to understand it better. In addition, the Speer et al. subjects knew that their memory would be tested and may have listened to the sentence differently. Finally, in the Speer et al. study sentences were presented in isolation, without contexts. The lack of context might have heightened the role of prosodic information, as we found happened in Experiment 2.

In the current paper, we are not addressing the contribution of prosody to on-line disambiguation processes, but rather the relative contribution of prosody and context to listeners' end interpretations of ambiguous utterances. We found that listeners based their ultimate interpretations of ambiguous utterances on the context the utterances appeared in, ignoring prosody, even when potentially disambiguating prosodic information could have been used.

Several studies on the role of prosody in disambiguating speech have found positive contributions of prosodic information. But there are problems extrapolating this research to everyday speech. As we have already mentioned, Beach (1991) and Wales and Toner (1979) studied synthesized and read speech, which may have contained more prosodic cues than natural speech. Price et al. (1991) tested speech produced by professional radio newscasters, which also might have contained exaggerated prosody. Marslen-Wilson et al. (1992) don't mention whether their speaker was naïve to their syntactic manipulation or not, or whether she supplied extra-clear prosody. But even assuming the speech was naïvely produced, the sentences tested were presented in isolation, without disambiguating context. What these studies show is that when speech is carefully prepared or when it is presented in isolation, prosody can play a role in disambiguating syntactic ambiguities. But everyday speech is neither carefully prepared nor uttered in the absence of context.

Our results pertain only to prosody as used in syntactic disambiguation. We are not arguing that prosody is irrelevant to language processing. In contrast, there is evidence that

prosody influences listeners' sense of the continuity of the speech stream (Nooteboom, Brokx, & de Rooij, 1978). There is also evidence that prosody can affect the pre-linguistic process of segmenting the continuous speech stream. Cutler and colleagues (Cutler, Mehler, Norris, & Segui, 1986; Cutler & Norris, 1988) have shown in a number of studies that the rhythmic alternation of heavy and light syllables in English is used as pointers to the beginnings of words. Prosody can also play a role in conveying affect (Cosmides, 1983) and communicative intent, such as approval or prohibition (Fernald, 1989). We recognize that prosodic information is much more than just a set of cues to syntactic boundaries, and that prosodic information can play a vital role in interpretation. What we suggest is that prosody might not play as crucial a role at the level of syntactic disambiguation in everyday speech as some prosodic research done with carefully prepared speech might predict.

Why is it that prosodic cues that can be used in isolation are not taken into account when contextual information is present? The explanation can be deduced from our first experiment where we found that disambiguating prosodic cues are either not present, not strong enough, or not reliable enough to influence listeners' interpretations. These characteristics of disambiguating prosody have also been noted by other researchers (Allbritton et al. 1996; Coulthard and Brazil, 1982; Fernald & McRoberts, 1996; Ferreira, Anes, & Horine, 1996; Pinker, 1995; Stenström, 1988). At best, prosodic cues for syntactic disambiguation may be used reliably and consistently only by professional speakers; they may simply not be consistently present in everyday talk. A comprehension system that would try to detect and interpret these occasional cues would only seldom be successful and would not gain any information on top of what can be inferred from context, which is a more reliable source of information. So in the interest of efficiency, the comprehension system may ignore prosodic cues for syntactic disambiguation in everyday speech or limit its processing of this information to specialized situations in which no other information is available to assist in interpreting what's been said.

## References

- Allbritton, D. W., McKoon, G., & Ratcliff, R. (1996). Reliability of prosodic cues for resolving syntactic ambiguity. Journal of Experimental Psychology: Learning, Memory, and Cognition, *22*(3), 714-735.
- Altenberg, B. (1990). Some functions of the booster. In J. Svartvik (Ed.), The London-Lund Corpus of Spoken English: Description and Research (pp. 193-209). Lund: Lund University Press.
- Altmann, G. (1987). Modularity and interaction in sentence processing. In J. Garfield (Ed.), Modularity in Knowledge-Representation and Natural-Language Understanding (pp. 249-257). Cambridge, MA: The MIT Press.
- Altmann, G., & Steedman, M. (1988). Interaction with context during human sentence processing. Cognition, *30*, 191-238.
- Atkinson, M. (1984). Our master's voices: The language and body language of politics. New York: Methuen.
- Beach, C. (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.
- Chawla, P., & Krauss, R. M. (1994). Gesture and speech in spontaneous and rehearsed narratives. Journal of Experimental Social Psychology, *30*, 580-601.
- Clemmer, E. J., O'Connell, D. C., & Loui, W. (1979). Rhetorical pauses in oral reading. Language and Speech, *22*(4), 397-405.
- Cosmides, L. (1983). Invariances in the acoustic expression of emotion during speech. Journal of Experimental Psychology: Human Perception and Performance, *9*(6), 864-881.
- Coulthard, M., & Brazil, D. (1982). The place of intonation in the description of interaction. In D. Tannen (Ed.), Georgetown University Round Table on Languages and Linguistics 1981 (pp. 94-112). Washington, D. C.: Georgetown University Press.
- Crain, S. & Steedman, M. (1985). On not being led up the garden path: The use of context by the psychological syntax processor. In D. R. Dowty, L. Karttunen, & A. M. Zwicky

(Eds), Natural Language Parsing (pp. 320-358). Cambridge: Cambridge University Press.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1986). The syllable's differing role in the segmentation of French and English. Journal of Memory and Language, *25*, 385-400.

Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. Journal of Experimental Psychology: Human Perception and Performance, *14*, 113-21.

Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: Is the melody the message? Child Development, *60*, 1497-1510.

Fernald, A., & McRoberts, G. (1996). Prosodic bootstrapping: A critical analysis of the argument and the evidence. In J. L. Morgan & K. Demuth (Eds.), Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition (pp. 365-388). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Ferreira, F., Anes, M. D., & Horine, M. D. (1996). Exploring the use of prosody during language comprehension using the auditory moving window technique. Journal of Psycholinguistic Research, *25*(2), 273-290.

Ferreira, F., Henderson, J. M., Anes, M. D., Weeks, P. A., Jr., & McFarlane, D. K. (1996). Effects of lexical frequency and syntactic complexity in spoken-language comprehension: Evidence from the auditory moving-window technique. Journal of Experimental Psychology: Learning, Memory, & Cognition, *22*(2), 324-335.

Gernsbacher, M. A., & Jescheniak, J. D. (1995). Cataphoric devices in spoken discourse. Cognitive Psychology, *29*, 24-58.

Howell, P., & Kadi-Hanifi, K. (1991). Comparison of prosodic properties between read and spontaneous speech material. Speech Communication, *10*, 163-169.

Ladd, D. R., & Cutler, A. (1983). Models and measurements in the study of prosody. In A. Cutler & D. R. Ladd (Eds.), Prosody: Models and Measurements (pp. 1-10). New York: Springer-Verlag.

Levelt, W. J. M. (1989). Speaking: From Intention to Articulation. Cambridge: MIT Press.

- Marslen-Wilson, W. D., Tyler, L. K., Warren, P., Grenier, P., & Lee, C. S. (1992). Prosodic effects in minimal attachment. The Quarterly Journal of Experimental Psychology, *45*, 73-87.
- McDonald, J. L., Bock, K., & Kelly, M. H. (1993). Word and world order: Semantic, phonological, and metrical determinants of serial position. Cognitive Psychology, *25*, 188-230.
- Mehta, G., & Cutler, A. (1988). Detection of target phonemes in spontaneous and read speech. Language and Speech, *31* (Part 2), 135-156.
- Nicol, J. L. & Pickering, M. J. (1993). Processing syntactically ambiguous sentences: Evidence from semantic priming. Journal of Psycholinguistic Research, *22*(2), 207-237.
- Nooteboom, S.G., Brokx, J. P. L., & de Rooij, J.J. (1978). Contributions of prosody to speech perception. In W. J. M. Levelt & G. B. Flores d'Arcais (Eds.), Studies in the perception of language (pp. 75-107). New York: Wiley.
- Pinker, S. (1995). Language acquisition. In L. R. Gleitman & M. Liberman (Eds.), Language: An Invitation to Cognitive Science (2nd ed., Vol. 1, pp. 135-182). Cambridge, MA: MIT Press.
- Price, P., Ostendorf, S., 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*(1/2), 165-191.
- Schafer, A., Carter, J., Clifton, C., & Frazier, L. (1996). Focus in relative clause construal. Language and Cognitive Processes, *11*(1/2), 135-163.
- Scherer, K. R., & Wallbott, H. G. (1985). Analysis of nonverbal behavior. In T. A. van Dijk (Ed.), Handbook of Discourse Analysis: Vol. 2. Dimensions of Discourse (pp. 199-230). San Diego: Academic Press.
- Selkirk, E. (1984). Phonology in Syntax: The Relation Between Sound and Structure. Cambridge: MIT Press.
- Speer, S., Crowder, R., & Thomas, L. (1993). Prosodic structure and sentence

recognition. Journal of Memory and Language, 32, 336-358.

Stenström, A.-B. (1986). A study of pauses as demarcators in discourse and syntax. In J. Aarts & W. Meijs (Eds.), Corpus Linguistics II: New Studies in the Analysis and Exploitation of Computer Corpora (pp. 203-218). Amsterdam: Rodopi.

Stenström, A.-B. (1988). Adverbial commas and prosodic segmentation. In M. Kytö, O. Ihalainen, & M. Rissanen (Eds.), Corpus Linguistics, Hard and Soft: Proceedings of the 8th International Conference on English Language Research on Computerized Corpora (pp. 221-236). Amsterdam: Rodopi.

Stirling, L., & Wales, R. (1996). Does prosody support or direct sentence processing? Language and Cognitive Processes, 11(1/2), 193-212.

Swerts, M., & Geluykens, R. (1994). Prosody as a marker of information flow in spoken discourse. Language and Speech, 37(1), 21-43.

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 (pp. 135-158). Hillsdale, NJ: Lawrence Erlbaum Associates.

Warren, P. (1996). Prosody and parsing: An introduction. Language and Cognitive Processes, 11(1/2), 1-16.

Winner, E., Windmueller, G., Rosenblatt, E., Bosco, L., Best, E., & Gardner, H. (1987). Making sense of literal and nonliteral falsehood. Metaphor and Symbolic Activity, 2(1), 13-32.

## Appendix

Ambiguous sentences are underlined>. The first five are taken from Price et al. (1991) along with their preceding contexts. Asterisks mark which of the sentence versions produced the highest percentage of correct responses in Experiment 1.

Context A	Context B	
He felt he had to read the journal, though it was poorly written and without content. Rollo read the review, literally learning not an iota. He was quite frustrated that there wasn't anything in it he didn't already know.	* Rollo was terribly literal, often missing the forest for the trees. His approach to the satirical journal was no exception. Rollo read the review literally, learning not an iota. Every satirical comment flew over his head.	
When I was a kid, I sneaked into an x-rated movie. As I was eleven only, I knew my dad would be angry. He was always a stickler for following the rules.	* The other children were too young to know they were doing anything wrong. As I was eleven, only I knew my Dad would be angry. He would lecture me about not doing something one knows is wrong because it is sinful.	
* Mike and Tom liked running on a track. When they found out that part of the race went through the woods, they considered not running. Although they did run, in the woods they were uneasy. They missed the smoothness of the track; in the woods they were concerned about tripping on loose sticks.	John thought jogging in the woods would calm anyone down, but my nervous city cousins showed he was wrong. Although they did run in the woods, they were uneasy. They lurched at every sound, thinking some wild animal was about to attack him. In spring there was always more work to do on a farm. May was the hardest month. They rose early in May. This meant their day began at 5 a.m. and ended 12 hours later.	* Bears sleep all winter long, usually coming out of hibernation in late April, but this year they were a little slow. They rose early in May. The rain probably fooled them into thinking it was still April.
* My experience with slow learners has shown one	As you begin to study about nuclear war it	

<p>thing. When you learn gradually, you worry more. I think this is because there's more time to spend worrying.</p>	<p>becomes frightening. When you learn, gradually you worry more. The more you learn, the more you have to worry about.</p>	
<p>The psych. experiment on how lovers behave in public is taking place today. On the wharf you can see a lot of people watching couples. You can tell who the researchers are because they are all wearing funky sunglasses.</p>	<p>* It's amazing how many lovers sit and observe others. On the wharf you can see a lot of people-watching couples. They almost look like they are watching aliens.</p>	
<p>Toni went deep sea diving in the Pacific Ocean. She saw a man-eating fish. It scared her.</p>	<p>* Jenny went to the Seafood restaurant. She saw a man eating fish. He seemed to like it.</p>	
<p>* It's hard to get close to someone because of their mortality. You will always lose someone you love to death. But we love them anyway.</p>	<p>Sometimes we smother those we love, causing them to leave us. You will always lose someone you love-to-death. It's always good to give others their space. Suzie visits her Grandma at the retirement home every Wednesday. At the park yesterday she saw old men and women talking on a bench. They seemed happy which pleased Suzie.</p>	<p>* Jenny always thought of old men as horny bastards who try to get younger women. At the park yesterday she saw old men and women talking on a bench. She sighed as she believed this proved her belief correct.</p>
<p>I really don't like it when my Mother-in-law stays at our house. I hate visiting relatives. She always orders me around in my own house.</p>	<p>* Most of my extended family is weird, they are cold and uninviting. I hate visiting relatives. I wish I could just stay home.</p>	
<p>* What's that smell you ask? That's frying chicken. It'll be ready by 6.</p>	<p>No no no, you bought the wrong kind of chicken. That's frying chicken. I wanted the type for chicken soup.</p>	
<p>Having an airport near a highway is a bad idea. Flying planes can be</p>	<p>* Pilots get paid a lot because they have risky jobs. Flying planes can be</p>	

<p>dangerous. They have been known to crash-land on passing cars.</p>	<p>dangerous. There's always a chance of a major accident, and this chance is multiplied by the amount of time they spend in the air.</p>	
---	---	--

Table 1

Experiment 1: Average Percentages of Correct Responses			
	Overall	Best items	Matched best
All 12 Stimuli	51	74 *	49
Left/Right Attachment	51	71 *	50
Not Left/Right Attachment	50	77 *	49

\*  $p < .05$ .

Table 2

Experiment 3: Average Percentages of Correct Responses			
	Best Set	Matched Set	Both
Prosody/Context Congruence	90 (5.9 / 4.8)	72 (5.9 / 4.7)	81 (5.9 / 4.8)
Prosody/Context Incongruence	81 (5.8 / 4.6)	87 (6.4 / 4.2)	84 (6.1 / 4.4)
Both	86 (5.9 / 4.7)	80 (6.2 / 4.5)	

Note. Cells contain responses to the content questions (confidence ratings/naturalness ratings) with correctness scored with respect to context.

### Author Note

This research was supported by faculty research funds granted by the University of California, Santa Cruz and by a Netherlands Organization for Scientific Research (NWO) Talent grant. We thank Mark Goettsche for technical assistance, Brian Altman, Catherine Quiroz, and James Bartok for help running the experiments, and many colleagues for fruitful discussions of this work. Correspondence concerning this paper should be addressed to Jean E. Fox Tree, Psychology Department, Social Sciences II, University of California, Santa Cruz, Santa Cruz, CA 95064.