Prosodic influences on the production and comprehension of syntactic ambiguity in a game-based conversation task*

Amy J. Schafer
University of Hawai‘i

Shari R. Speer
Ohio State University

and

Paul Warren
Victoria University of Wellington

Please address correspondence to:

Amy J. Schafer
Department of Linguistics, 569 Moore Hall
University of Hawai‘i at Mānoa
1890 East-West Rd
Honolulu, HI 96822
aschafer@hawaii.edu

There is now considerable evidence that listeners are sensitive to prosodic structure in their syntactic analysis of spoken language (for reviews see Cutler, Dahan & Donselaar, 1997; Warren, 1999). Some recent research suggests that the prosodic contrasts investigated in comprehension research are not produced consistently by speakers, but may directly depend on ambiguity levels in the discourse situation. When naïve untrained speakers produced disambiguating prosody in situations where two or more syntactic parses were plausible, they did so less reliably if the discourse context contained other disambiguating information (Allbritton, McKoon & Ratcliff, 1996; Snedeker & Trueswell, 2003; Straub, 1997). Earlier studies similarly demonstrated stronger prosodic disambiguation when speakers were explicitly instructed to disambiguate (Cooper, Paccia & LaPointe, 1978; Lehiste, 1973), as well as situational dependence, with speakers reducing the length of a description with repeated mention (Clark & Schober, 1992), or producing reduced forms of words on repetition in discourse context (Fowler & Housum, 1987).

However, these studies showing variable prosodic disambiguation may not be representative of typical speech situations. In most of these, the speakers read sentences aloud in paragraph contexts or as (imagined) instructions to listeners who provided no spoken response.[1] Reading tasks--and the prosody produced in them--may differ from spontaneous speech, not least because the pragmatic goals and production constraints of reader-listener pairs differ markedly from those of interacting speaker-listener pairs. Production studies (Ayers, 1994; Butterworth, 1975) have highlighted some of the differences between the prosodic structures of read speech and those of spontaneous speech: Read speech tends to have fewer and shorter pauses, and fewer prosodic phrases. Thus, reading studies might provide a poor guide to the distribution and size of prosodic boundaries in spontaneous speech and therefore to the extent and nature of disambiguation in spontaneous speech (Mazuka, Misono & Kondo, 2001). Studies of conversational language have shown sharp differences in production and comprehension between conversing speaker-hearer pairs and non-interacting speakers or overhearers. Speakers in conversation designed their utterances to reflect the knowledge they had in common with their listeners and to accommodate feedback from listeners about how well they were being understood (Brennan, 1990; Clark & Wilkes-Gibbs, 1986). Overhearers were less accurate than conversing listeners in identifying speakers' intended referents, even when these were visually available objects (Clark & Schober, 1992). Conversational effects such as these are strengthened when speaker-listener pairs are aware of the need to cooperate (Schober, Conrad & Fricker, 2000).

An advantage of using scripted tasks to study correspondences between prosodic and syntactic structure is that they allow experimenter control of lexical and syntactic content, and make it possible to carefully compare alternative resolutions of ambiguous utterances. Some spontaneous speech tasks have been designed to give a certain degree of control over the range of utterances produced, such as map tasks (Anderson et al., 1991), route descriptions (Levelt & Cutler, 1983), or tangram tasks (Clark & Wilkes-Gibbs, 1986). However, even these tasks do not elicit multiple renditions from the same speaker of a targeted syntactic contrast.

The current research employed a cooperative game task, involving a set of predetermined expressions that were used to negotiate the movement of gamepieces around a board. These expressions contained a range of syntactic ambiguities, although not every expression was ambiguous. In this chapter we focus on PP attachments, as in the sentence *I want to change the position of the square with the triangle.* Depending on whether the PP attaches high (to modify the verb) or low (to modify the noun *square*), the utterance might mean "use the triangle to move the square" or "move the combined square+triangle piece", corresponding to two legitimate commands in our games.

Previous small-scale studies of PP ambiguities with read materials have revealed more pausing and pre-pausal lengthening before the PP when it attaches high (Lehiste, Olive & Streeter, 1976; Straub, 1997; Warren, 1985). Of interest to our discussion above, production studies that have included disambiguating contexts have shown evidence of both the maintenance of prosodic contrasts (Price, Ostendorf, Shattuck-Hufnagel & Fong, 1991) and their reduction (Cooper & Paccia-Cooper, 1980; Snedeker & Trueswell, 2003; Straub, 1997), as have listener judgments from these studies. Studies of prosodic boundary location in PP ambiguities have shown that low attachment interpretations were most likely when a prosodic boundary preceded the direct object NP, and high attachments when one followed the NP (Pynte & Prieur, 1996; Schafer, 1997).

Our game task allowed manipulation of the degree of contextual determination of one meaning of an ambiguity over another. Straub (1997), for instance, has proposed that the production system will allocate resources to prosodic disambiguation when other sources of disambiguating information would not be available for the listener in the resulting utterance. This comports with findings that prosodic disambiguation is less marked when utterances are read with disambiguating contexts. In our task, a number of information sources potentially helped disambiguate between the different PP attachments (see Warren, Schafer, Speer & White, 2000). Here, we examine two types of variation of situational ambiguity. One reflects the configuration of gamepieces on the playing boards and the preceding discourse. The other is linked to players' potential awareness of the PP attachment contrast, which might result in increased disambiguation as time spent playing the game increased.

Two experiments provide the relevant data for PP utterances. Experiment

1 presents acoustic and intonational analyses of productions by naïve speakers. Experiment 2 considers the categorization by a second set of naïve listeners of utterances isolated from the game context. The combination allows us to investigate separately the extent to which speakers alter their prosody to reflect syntactic and situational factors, and the extent to which listeners use whatever prosodic cues are present to recover the intended syntactic form. The transcriptions generated in experiment 1, encoding phonological distinctions such as the presence or absence of a prosodic boundary, allow us to relate production patterns to claims made in the comprehension literature. We can also evaluate whether any given token, considered in isolation, is one which we would expect to bias comprehension, on the basis of claims about the prosody-syntax interface (e.g., Schafer, 1997; Selkirk, 1984, Carlson, Clifton & Frazier, 2001).

Our combined analyses allow us to evaluate three aspects of situational effects on prosodic form. As situational ambiguity increases, i) does the proportion of utterances pronounced with disambiguating prosody rise, as determined by categories of prosodic transcriptions; ii) does the strength of the acoustic cues for disambiguation rise, regardless of phonological categorization; iii) do speakers' productions become more effective in helping listeners recover the intended syntactic structure?

Prosodic Assumptions

We assume the analysis of prosodic structure in American English proposed in Beckman & Pierrehumbert (1986) (following Pierrehumbert (1980)). Each utterance is composed of one or more intonation phrases, each of which is made up of one or more intermediate phrases. We collectively refer to intonation phrases and intermediate phrases as prosodic phrases. The ends of prosodic phrases in American English carry edge tones, typically associated with changes in fundamental frequency. They also show *final lengthening*--increased duration for the final syllable of the phrase--and can be followed by a silent interval. These durational effects tend to be more extreme for intonation phrases than for intermediate phrases (Wightman, Shattuck-Hufnagel, Ostendorf & Price, 1992). The edges of prosodic phrases are also associated with changes in segmental articulation (e.g., Keating, Cho, Fougeron & Hsu, to appear), and with resetting of the pitch range.

Experiment 1

Our production study included both phonological and acoustic phonetic analyses of utterances such as (1) to (4), exploring the syntactic and situational determination of the prosodic realization of PP ambiguities by naïve speakers in our game task.

(1) I want to change the position of the square with the triangle.

(2) I am able to confirm the move of the square with the triangle.
(3) I want to change the position of the square with the cylinder.
(4) I am able to confirm the move of the square with the cylinder.

We conducted our experiments with the following hypotheses in mind.

1. *Syntactic determination*. We predicted a difference in the realizations of high (VP) and low (NP) attachments of the PP. The high attachment was predicted to be reflected in a stronger prosodic boundary before the PP than found in the low attachment sentences.

2. *Illocutionary force*. In our game task, one speaker (the Driver) issued instructions, such as (1), while another (the Slider) followed these instructions and confirmed that moves had taken place, using utterances such as (2). Disambiguation was potentially more crucial in Driver utterances, since the incorrect move could otherwise have been chosen. If prosodic realization is sensitive to such pragmatic factors, disambiguation should be greater for Driver than for Slider utterances.

3. *Level of situational ambiguity 1: gamepiece contrast*. Our game included *square with the cylinder* sequences, in which the only interpretation in the context of the game was that of a high attachment, since there was no combined square+cylinder piece. Situational sensitivity predicts that the features that indicate high attachment would not be as clearly marked in the *cylinder* utterances as in the *triangle* utterances.

4. *Level of situational ambiguity 2: gameboard configuration*. In the game there were configurations of the pieces on the board which resulted in the Driver's use of (1) being truly ambiguous, biased toward one interpretation or the other, or unambiguous, as defined below. If speakers are sensitive to situational constraints, then we should expect greater disambiguation for ambiguous situations than for biased or unambiguous ones.

Procedure

In our cooperative game task two players used scripted sentences to negotiate moves of gamepieces from starting positions to goals. By observing gamepiece moves, the experimenter was able to identify each PP utterance as an intended high or low attachment utterance. Neither player could see the board used by the other, although they knew they had identical gamepieces. The design of the boards and the rules of the game encouraged negotiation and the strategic use of moves. The Driver's role was to tell the Slider which piece to move, to inform the Slider when he or she had moved incorrectly, and to indicate when a gamepiece had reached its goal. The Slider's role was to choose directions to move in and to report moves back to the Driver, but the Slider was also required to ask the Driver for more information when necessary. Players were restricted to uttering sentences from a provided list, but chose freely from this list to best match their communicative needs. Through repeated use of the sentences over the course of the experiment, players became increasingly familiar with the sentence forms and less dependent on reading processes. Further information about the methodology is provided in Warren et al. (2000).

Situational ambiguity levels for gameboard configurations

We defined three levels of situational ambiguity for the gameboard configurations. *Ambiguous* – Disregarding prosody and any underlying syntactic or lexical preferences, sentence (1) could with equal likelihood be interpreted with high or low attachment. *Unambiguous* - The global ambiguity could refer to only one legal move. For example, the square was in its goal and no triangle could be used to move it out. *Biased* - Both interpretations of the utterance were possible, but one was more likely. For example, the players had just moved a triangle next to a square, so that using the triangle to move the square would be an obvious next move.

Subjects

Eight pairs of subjects, all native speakers of American English naïve to the purposes of the experiment, were recorded at the University of Kansas. All subject pairs played as many games as they could within two hours, using multiple boards, and exchanging Driver/Slider roles between games. Subjects wore head-mounted microphones, and their utterances were recorded. Further details of the participants, excluded participants, and excluded productions appear in Warren et al. (2000).

Transcription Methods

All PP sentences were excised from the game context, placed in separate audio files, and assigned coded filenames that masked the speaker's intended syntactic structure. The prosody was transcribed by a team of five transcribers, trained to use the English ToBI (Tones and Break Indices) transcription system (Beckman & Ayers, 1997). All were native speakers of English. Each transcriber analyzed a subset of the utterances, using auditory information and visual inspection of waveform displays, F0 tracks, and if desired, spectrograms. Reliability across transcribers was determined on the basis of a subset on which all five overlapped, using the reliability metric of Pitrelli, Beckman & Hirschberg (1994). There was at least 94% agreement on the presence of pitch accents, phrase accents (indicating an intermediate phrase boundary), and boundary tones (indicating an intonational phrase boundary).

Results

As mentioned above, this chapter focuses on the effect of the gameboard configuration manipulation. Therefore, in this section we report only the results for Driver utterances containing the phrase *the square with the triangle*, returning to the Driver versus Slider comparison and triangle versus cylinder comparison in the general discussion.

Transcription results. There was substantial variation in the intonational and durational patterns that were produced for the sequence *the position of the square* in sentence (1), both within and across speakers. In data from 13 speakers, we found 63 distinct patterns on 79 high-attached utterances, and 87 distinct patterns on 101 low-attached utterances. This indicates that the exact prosodic form cannot be predicted solely on the basis of morphosyntactic structure.

We assigned the transcribed utterances to three groups to evaluate the relationships among syntactic structure, situational ambiguity, and prosodic disambiguation. The first group contained all utterances with a stronger prosodic boundary at the end of *square* (i.e., immediately prior to the PP) than at any other location in the sentence. Boundary strength was determined by the phonological category of the boundary (i.e., word, intermediate phrase, or intonation phrase boundary). The second group had been pronounced with the strongest boundary at a location other than at the end of *square*. The third group contained utterances in which the boundary at the end of *square* and at least one other boundary were of equal strength, and these were the strongest boundaries in the sentence.

Previous production results have shown longer duration for the prosodic boundary preceding high PP attachments than low ones (e.g., Warren, 1985). In the comprehension domain, Schafer (1997) and Carlson et al. (2001) have argued that prosodic disambiguation is influenced by the pattern of prosodic boundary strengths in the preceding material. Both proposals predict that pronunciations of (1) should be biased toward high attachment when the strongest prosodic boundary in the sentence is located at the end of *square*.[2, 3]

Our transcription results suggest that the pattern of relative boundary strengths was strongly influenced by the intended syntactic structure. The strongest boundary followed *square* for 57% of the high attachment utterances, versus 7% for low attachment. There is a potential concern that the "low-attached" utterances could have been produced with lexicalization of the phrase *the square with the triangle*.[4] The prosodic evidence concerning lexicalization is complex (Liberman & Sproat, 1992) and beyond the scope of this chapter. However, the existence of lexicalized utterances would not affect the hypotheses for the high-attached sentences, which are our focus for the assessment of effects of situational ambiguity on prosody.[5]

The distribution of transcription patterns for high-attached tokens by level of gameboard ambiguity is given in Table 1. Similar percentages of tokens were pronounced with the strongest boundary following *square* in ambiguous, biased, and unambiguous game situations, with the highest percentage in the unambiguous situation. The results indicate that speakers pronounced the PP sentence with a variety of prosodic structures, which ranged across prosodies expected to be more and less indicative of the syntactic structure. A substantial portion of the variability in boundary strength patterns can be explained by the intended syntax, but none of it seems to be explained by the level of situational ambiguity.[6]

-- insert Table 1 about here --

Duration results. The transcription patterns in Table 1 do not exclude the possibility of significant effects of situational ambiguity on prosodic disambiguation, since matching phonological structures may have systematically differing phonetic realizations. For example, the silent interval of an intonation phrase boundary in a critical position could be reliably longer in utterances produced in ambiguous situations than in unambiguous ones. Using digitized speech waveforms, we compared the durations of the word *square*, of any following pause, and of the combined *square* + pause sequence. Each was significantly longer for high-attached versions of (1) (Warren et al., 2000), providing clear support for the prediction that, in general, speakers would reflect the intended interpretation of the PP sentences in their prosody.

To examine whether the syntactic effect on prosody was modulated by situational ambiguity we looked at durational data in the three ambiguity levels described above. The overall mean durations of *square* + pause for these ambiguity levels for each of the high and low attachment conditions for 13 speakers in the Driver role are shown in Figure 1.

-- insert Figure 1 about here --

The variable number of tokens making up these data (see Figure 1) made the comparison of overall means rather unreliable. In particular, the breakdown by ambiguity level left some speakers with very small or empty cells for some conditions. Therefore, we restricted our statistical analysis to those speakers with at least one instance in each ambiguity x attachment condition. The resulting ANOVAs were consequently for high attachment data only, from just 11 of our speakers. They showed no effect of ambiguity level on the duration of *square*, of the following pause, or of *square* + pause (Warren et al., 2000).

Thus, the duration results, like the transcription results, show that speakers in our task marked the syntactic difference between high and low attachments of PPs with some consistency. Yet the lengthening of the word and pause before the PP in high attachments, compared with low attachments, did not depend on the level of situational ambiguity.

Experiment 2

In the second experiment, game task materials collected in the production study were presented to listeners in a categorization task in order to determine whether the prosodic patterns identified in the production study would be useful to listeners faced with interpreting the utterances. High and low attachment tokens of sentence (1) were presented to listeners as complete sentences in a forced-choice task in which they selected between paraphrases indicating high versus low attachment. Nineteen native speakers of Midwestern American English from the University of Kansas took part in this experiment. None of them had previously taken part in the production experiment described above.

Hypotheses

If speakers produce prosodic structures that reflect syntactic structure, and that are useful to listeners, then percentages of correct classification in the comprehension experiment should be above chance for both high and low attachment sentences.

Further, if speakers increase prosodic disambiguation to reflect situational need, then correct categorization should be higher for tokens produced in the ambiguous condition than in the biased condition, and higher for tokens produced in the biased condition than in the unambiguous condition. Note that this would imply the use of further prosodic cues to disambiguation than just the boundary strength and durational differences measured in Experiment 1, which did not reliably distinguish levels of ambiguity.

Results

The percentages of correct classifications are given in Figure 2. The overall classification was greater than chance, showing that listeners were able to make use of distinctions that reflect syntactic structure. The percent correct scores for each condition and for each individual participant were subjected to an arcsine transformation, (2arcsine÷p), and entered into an ANOVA with attachment and ambiguity level as factors. This revealed a significant main effect of attachment type ($F[1,18] = 5.80$, $p<0.027$), with more correct classifications for high than for low attachments (76% vs. 64% overall). This main effect may reflect a slight overall bias towards high attachments of the PP.

-- insert Figure 2 about here --

There was also a significant interaction of attachment type and ambiguity level ($F[2,36] = 5.133$, $p<0.011$), reflecting the fact that there was no effect of ambiguity for the high attachment condition, but a significant effect for the low attachment condition. This latter effect resulted from the *lower* correct score for ambiguous than for biased or unambiguous items. That is, it was low attachments produced in the ambiguous situation that showed the least evidence of prosodic disambiguation.[7] The absence of a main effect of ambiguity level fails to support the hypothesis that speakers produce different degrees of prosodic disambiguation according to differences in situational ambiguity. It supports the conclusion from the production experiment that our speakers tended to disambiguate the PP structure, and they did so regardless of the ambiguity of the situation.

Production Sequence Analysis

Given the extent to which interacting speakers can alter their productions over the course of some tasks (e.g., Clark & Schober, 1992), we might expect that speakers in the game task would have changed their use of prosody across the experiment. They presumably became more aware of the contrast between high and low PP attachments as play continued, especially since the design of the games elicited the first production of each attachment in an unambiguous configuration. They also received evidence, directly after each PP production, of whether their conversation partner had interpreted the sentence correctly or not. Each of these factors might result in a tendency toward stronger disambiguation at the end of the task than at the beginning. Therefore, we reanalyzed the listener categorization results to examine whether categorization improved across the production sequence of Experiment 1. Since each speaker produced at least 5 utterances for each attachment, the percentage of correct categorizations was determined for the first through fourth and last utterance for each of the attachment sequences.

The results are shown in Figure 3. ANOVAs revealed only a marginal effect of utterance sequence on categorization (nor were there any systematic effects with a breakdown into ambiguity classes). We wish to emphasize that the lack of a significant effect cannot be attributed to a high degree of consistency within each speaker's prosodic productions. Each speaker produced utterances that received high percentages of correct categorizations and ones that received low percentages. The average difference across speakers between the utterance with the highest percentage of correct categorization and that with the lowest was 35% for high attachments and 49% for low attachments. As with the other results, there was considerable variation within each speaker, but this variation does not appear to be explained by situational ambiguity, as determined by either the gameboard configuration or a presumed rising awareness of the PP contrast across the course of the game.

-- insert Figure 3 about here --

General Discussion and Conclusions

Our analyses found strong and consistent evidence that prosodic structure reflected syntactic structure, at least in the majority of productions, but no evidence that prosodic disambiguation was modulated by situational need. Transcription, duration, and listener categorization results all showed syntactic effects, but gave no indication that prosodic disambiguation increased with situational ambiguity. Similarly, the investigation of sequence effects demonstrated that speakers' productions at the beginning of the task were just as biasing as those from the end of the task.

Other analyses from our game have produced similar results (Schafer, Speer, Warren & White, 2000; Warren et al., 2000). Speakers strongly disambiguated an early/late closure contrast in our game, which was produced with quite limited situational ambiguity. The durational pattern for *cylinder* PP sentences, which in our game received referential support for only the high-attached interpretation, matched the durational pattern for high-attached *triangle* PP sentences. There was no apparent reduction of disambiguation for the *cylinder* sentences, even though the intended interpretation was unambiguous throughout the game. We also found equally strong durational effects of syntax in PP utterances by Sliders, who were confirming a move, as in utterances by Drivers, who were introducing a move. Thus, across two syntactic ambiguities and multiple types of analyses our results consistently show prosodic reflections of syntactic attachment, unaffected by situational ambiguity.

These results contrast sharply with those from previously published research on prosodic disambiguation, recent work by Snedeker and Trueswell (2003), and other tests of situational effects on production (but see also Ferreira & Dell, 2000). We believe there are several reasons to be cautious in generalizing from the previous prosody results to spontaneous discourse situations. As noted in the introduction, the previous studies either did not include a conversation partner or allowed very limited interaction, and most relied much more heavily than our task on reading processes. Although our task did not elicit fully spontaneous speech, we believe that the utterances we collected are much more similar to spontaneous speech than those in other studies.[8] In addition, we believe our task was extremely effective in clearly establishing a syntactic interpretation of the ambiguous sentence for the speaker in a manner that did not have unintended consequences on the prosodic structure of the utterance. The use of biasing linguistic contexts in some of the previous work might not have always resulted in the speaker recovering the syntactic structure intended by the experimenter. In our study, the speaker's intended meaning was always unambiguously demonstrated to the experimenter by an associated move of a gamepiece. Further, it is quite likely that certain discourse contexts can induce focal structures that impact the prosody-syntax correspondences. Schafer and Jun (2001) have demonstrated that prosodic reflections of PP attachment in English can be affected by changes in focal structure. We believe that such factors were minimized in our task, but may have had significant effects in some of the previous studies.

In this chapter we have been most concerned with effects of situational ambiguity on prosody. We looked for its effects with an experimental design that we hoped would be quite representative of everyday speech. The levels of situational ambiguity in the game fluctuated because of the preceding discourse and because of actions performed on objects in the discourse context. Some of the experimental materials received referential support for both PP interpretations (the *triangle* utterances), and others received referential support for just one interpretation throughout the game (the *cylinder* utterances). We believe that discourse situations such as these should be highly informative with respect to the relative strengths of grammatical constraints on prosodic form (such as prosody-syntax correspondences) and tendencies in speakers to alter the prosodic disambiguation they provide in response to situational needs in non-experimental contexts. Nevertheless, much research remains to be done in this area, and there is a particular need to analyze the prosody found in truly spontaneous speech produced for a range of sentence forms and a range of discourse contexts.

Although we did not find effects of situational ambiguity on prosodic disambiguation in any of our comparisons, we did see an effect of the discourse situation on utterance form. Speakers tended to have faster rates of speech when playing the game in the Slider role than in the Driver role (Warren et al., 2000), suggesting that they may have been more deliberate when they were directing the course of action than when they were confirming it. (Recall that players switched roles after each game.) However, as mentioned above, this difference did not seem to affect the degree of prosodic disambiguation in the Driver versus Slider role.

There are certainly cases—including some in research cited above—in which speakers employ a disambiguating prosodic structure in an attempt to indicate one interpretation over another. Speakers also make conscious and unconscious choices to be generally clearer in certain speech situations, and may therefore do such things as alter their rate of speech in response to the audience. Such changes may have indirect effects on prosodic disambiguation, e.g., the inclusion of stronger prosodic boundaries in several positions within a sentence when it is uttered in a more deliberate style. Nevertheless, we believe that the production of sentence prosody is primarily controlled by grammatical factors, such as phonosyntactic constraints relating prosodic form to syntactic form, phonological constraints governing the length or weight of prosodic units, and semantic/pragmatic constraints relating information/discourse structure and prosody. Under this view, most prosodic disambiguation of syntax in everyday speech is not disambiguation *per se*, but the regular application of grammatical constraints. In such a model, we should expect that the degree of prosodic disambiguation found in most speech depends very little on the degree of situational ambiguity, but very much on the grammatical structures involved, as found in our game task.

References

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

Anderson, A.H., Bader, M., Boyle, E., Bard, E.G., Doherty, G., Garrod, S., Isard, S.D., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H. S. & Weinert, R. (1991). The HCRC map task corpus. *Language and Speech*, 34, 351-366.

Ayers, G. (1994). Discourse functions of pitch range in spontaneous and read speech. *OSU Working Papers in Linguistics*, 44, 1-49.

Beckman, M.E. & Ayers, G. (1997). Guidelines for ToBI labelling. Ms. Columbus, OH: Ohio State University

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

Brennan, S.E. (1990). *Seeking and Providing Evidence for Mutual Understanding*. Doctoral dissertation, Stanford University.

Butterworth, B. (1975). Hesitation and semantic planning in speech. *Journal of Psycholinguistic Research*, 4, 57-87.

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

Clark, H. & Schober, M.F. (1992). Understanding by addressees and overhearers. In H. Clark (Ed.) *Arenas of Language Use*. Chicago: University of Chicago Press. 176-197.

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

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

Cooper W.E., Paccia, J.M., & LaPointe, S.G. (1978). Hierarchical coding in speech timing. *Cognitive Psychology*, 10, 154-177.

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

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

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

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

Keating, P. Cho, T. Fougeron, C. & Hsu, C.-S. (to appear). Domain-initial articulatory strengthening in four languages. In J. Local, R. Ogden & R. Temple (Eds.) *Papers in Laboratory Phonology VI*. Cambridge: Cambridge

University Press.
Lehiste, I. (1973). Phonetic disambiguation of syntactic ambiguity. *Glossa*, 7, 103-122.
Lehiste, I., Olive, J.P., & Streeter, L.A. (1976). Role of duration in disambiguating syntactically ambiguous sentences. *Journal of the Acoustical Society of America*, 60, 1199-1202.
Levelt, W.J.M. & Cutler, A. (1983). Prosodic marking in speech repair. *Journal of Semantics*, 2, 205-217.
Liberman, M. & Sproat, R. (1992). The stress and structure of modified noun phrases in English. In Sag, I. & Szabolcsi, A. (Eds.) *Lexical Matters*. Stanford: CSLI. 131 – 181.
Mazuka, R., Misono, Y. & Kondo, T. (2001). Differences in levels of informativeness of prosodic cues to resolve syntactic ambiguity. Presented at the Fourteenth Annual CUNY Conference on Human Sentence Processing, University of Pennsylvania, Philadelphia, PA, March 2001.
Nespor, M.A., & Vogel, I. (1986). *Prosodic Phonology*. Boston, MA: Kluwer.
Pierrehumbert, J.B. (1980). *The Phonology and Phonetics of English Intonation.* Doctoral dissertation, MIT.
Pitrelli J., Beckman, M., & Hirschberg, J. (1994). Evaluation of prosodic transcription labeling reliability in the ToBI framework. *Proceedings of the International Conference on Spoken Language Processing*, Yokohama, Japan, September 1994. 123-126.
Price, P., Ostendorf, M., Shattuck-Hufnagel, S., & Fong, C. (1991). The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America*, 90, 2956-2970.
Pynte, J. & Prieur, B. (1996). Prosodic breaks and attachment decisions in sentence parsing. *Language and Cognitive Processes*, 11, 165-192.
Schafer, A.J. (1997). *Prosodic Parsing: The Role of Prosody in Sentence Comprehension*. Doctoral dissertation, University of Massachusetts.
Schafer, A.J. & Jun, S.-A. (2001). Effects of focus on prosodic reflections of phrase structure in American English. Presented at The Prosody in Processing Workshop, Utrecht University, Utrecht, Netherlands. July 2001.
Schafer, A.J., Speer, S.R., Warren, P., & White, S.D. (2000). Intonational disambiguation in sentence production and comprehension. *Journal of Psycholinguistic Research*, 29, 169-182.
Schober, M.F., Conrad, F.G., & Fricker, S.S. (2000). Listeners often don't recognize when their conceptions differ from speakers'. Presented at the Annual Meeting of the Psychonomics Society, New Orleans, LA.
Selkirk, E. O. (1984). *Phonology and Syntax: The Relation between Sound and Structure.* Cambridge, MA: MIT Press.
Snedeker, J., & Trueswell, J. (2003). Using prosody to avoid ambiguity: Effects

of speaker awareness and referential context. *Journal of Memory and Language, 48*, 103–130.

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

Warren, P. (1985). *The Temporal Organisation and Perception of Speech*. Doctoral dissertation, University of Cambridge.

Warren, P. (1999). Prosody and language processing. In S. Garrod & M. Pickering (Eds.), *Language Processing*. Hove: Psychology Press. 155-188.

Warren, P., Schafer, A.J., Speer, S.R., & White, S.D. (2000). Prosodic resolution of prepositional phrase ambiguity in ambiguous and unambiguous situations. *UCLA Working Papers in Phonetics*, 99, 5-33.

Wightman, C.W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P.J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America*, 92, 1707-1717.

Footnotes

* This research was supported by NIH research grants DC-00029 and MH-51768, NZ/USA Cooperative Science Programme grant CSP95/01, and Marsden Fund grant VUW604. We thank Kelly Barrow, Karen Carmody, Amanda Fisch, Christa Hansen, Lauren Kling, Jenny Kneale, Jennifer Ludlow, Cara Prall, Lisa Rief, Shari Sokol, Aaron Soltz, Jill Story, David White, and Gerald Whiteside for assistance with running subjects and measuring data.

1. Recent work by Snedeker and Trueswell (to appear) employed a task in which the speaker uttered a series of commands involving the manipulation of a set of toys to a listener separated by a screen. Interaction between the two participants was limited to the speaker asking if the listener was ready. In this task the experimental materials were presented as printed text and acted-out toy manipulations. The textual stimulus was then removed and the command produced by the speaker from memory.

2. Schafer and Carlson et al. differ in their predictions about several boundary strength patterns, such as a pattern with intonation phrase boundaries at the end of both *position* and *square*. In Schafer's proposal, this pronunciation would bias listeners toward high attachment; in Carlson et al.'s proposal it would not. Since more finely graded analyses are beyond the scope of this paper, we focus on the cases where there is consensus.

3. The location of prosodic boundaries is likely influenced by several other factors than the intended attachment site. For example, there is some tendency to produce a prosodic boundary at the midpoint of an utterance (e.g., Gee & Grosjean, 1983). Utterances with the strongest prosodic boundary at the end of *square* are unlikely to be showing solely the influence of this tendency, given the late location of the boundary.

4. We thank Gary Dell for first mentioning this possibility to us, as well as Jesse Snedeker, Mike Tanenhaus, and John Trueswell.

5. In addition to being unaffected by the possibility of lexicalization, the high-attached tokens were more evenly distributed across the three gameboard configurations and had the widest distribution across our three boundary pattern groups.

6. The high-attached utterances with strong prosodic breaks located prior to the end of *square* may reflect the pressure to balance the lengths of prosodic phrases and to avoid long prosodic phrases (e.g., Gee & Grosjean, 1983; Nespor

& Vogel, 1986). Phonological factors such as these likely account for some of the remaining variability in prosodic boundary location.

7. It is possible that this reflects a choice by the speakers to produce more deliberate pronunciations in the low-attachment ambiguous-situation condition. See Warren et al. (2000) for further discussion of this possibility. We note, though, that very few tokens were produced in this condition in Experiment 1. Therefore, the stimuli tested in this condition in the comprehension study might not accurately reflect the range of prosodic patterns that would be found in a larger sample.

8. Snedeker and Trueswell's task, like ours, produced speech that was less dependent on reading processes than that of previous tasks. However, our task involved greater interaction between participants than theirs, seems to have included a greater range of syntactic structures in the discourse situation, and required more varied interaction with the objects in the discourse situation.

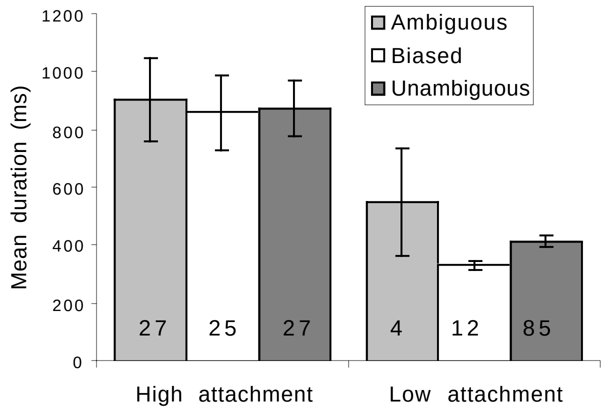
Table 1. Number (and percentages) of high-attached versions of (1) for each gameboard configuration pronounced with the strongest prosodic boundary in the sentence located at the end of *square*, the ends of *square* and at least one other word (two or more boundaries of equal phonological strength), or the end of some word other than *square*.

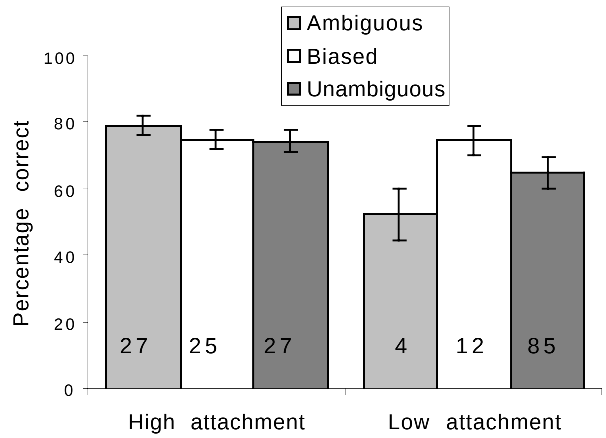
Figure 1. Mean *square* + pause durations (with standard error bars) for high- and low-attached *triangle* tokens, by situational ambiguity level. Number of tokens for each mean are indicated.

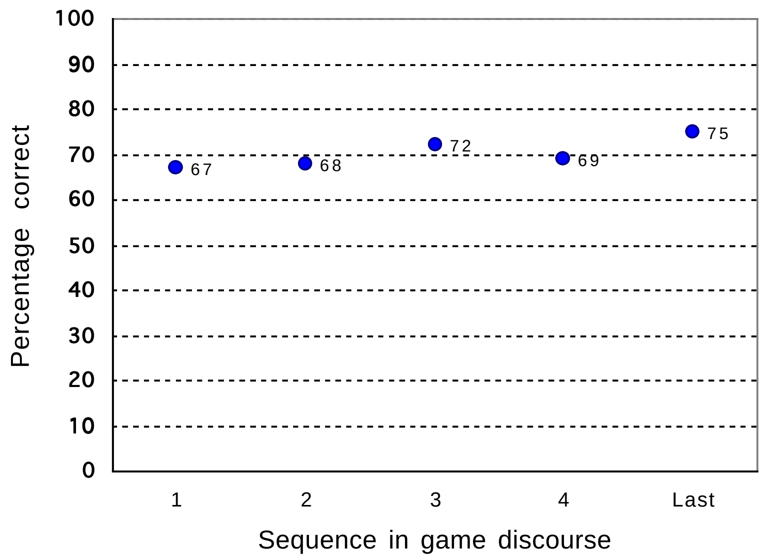
Figure 2. Percentages of correct classifications of tokens as high- or low-attached sentences, by level of situational ambiguity. The percentages are averages of the values obtained for 19 subjects listening to 13 speakers. The numbers of tokens heard in each condition are indicated.

Figure 3. Percentage of correct classifications of tokens as high- or low-attached PP sentences, by sequence within high- or low-attached utterances in the game discourse. The percentages are averages of the values obtained for 19 subjects listening to 13 speakers. Sequence positions were assigned separately for the two attachment conditions; the figure shows the average of the high- and low-attached mean for each position.

	Gameboard configuration		
Strongest boundary at the end of:	Ambiguous	Biased	Unambiguous
square	14 (52%)	12 (48%)	19 (70%)
square and some other word(s)	7 (26%)	8 (32%)	4 (15%)
some other word than *square*	6 (22%)	5 (20%)	4 (15%)
Total number of tokens	27	25	27


1200
1000
800
600
400
200
0
Mean duration (ms)
Ambiguous
Biased
Unambiguous
27
25
27
4
12
85
High attachment
Low attachment


Ambiguous
Biased
Unambiguous
100
80
60
40
20
0
Percentage correct
27
25
27
4
12
85
High attachment
Low attachment



100
90
80
70
60
50
40
30
20
10
0
Percentage correct
67
68
72
69
75
1
2
3
4
Last
Sequence in game discourse