UCLA Working Papers in Phonetics, Vol. 99, 5-33.

Prosodic resolution of prepositional phrase ambiguity in ambiguous and unambiguous situations*

Paul Warren, Victoria University of Wellington, New Zealand
Amy J. Schafer, UCLA
Shari R. Speer, Ohio State University
S. David White, Ohio State University

Introduction

There is a growing body of research that demonstrates that listeners are sensitive to prosodic information and are able to utilize such information in their immediate comprehension of spoken sentences involving a wide range of syntactic ambiguity types. Thus, Marslen-Wilson et al (1992) showed that listeners distinguish NP-complement and S-complement versions of sentence fragments such as those in (1) using prosodic information, and Warren et al (1995) have shown similar use of prosody, including stress placement, in disambiguating temporarily ambiguous sentences such as (2). (Note: material in braces shows alternative completions of the preceding fragment.)

(1) The jury believed the testimony of the last witness *{was almost totally misleading / for the prosecution}.*

(2) Whenever parliament discusses Hong Kong problems *{are solved instantly / they are solved instantly}.*

Using on-line reaction time techniques (such as cross-modal naming tasks), these and similar studies have demonstrated the immediacy with which prosodic information is utilized. Note though that the experimental materials used in these comprehension studies have generally been carefully recorded utterances with clear and consistent prosodic contrasts. These materials are frequently recorded by expert speakers using sentence lists, and almost always by speakers who are conscious of the ambiguities involved. It is clearly relevant then that Allbritton et al (1996) have argued that a speaker's awareness of ambiguity is a primary factor that influences the salience of prosodic contrasts in that speaker's production of ambiguous sentence materials. These researchers investigated effects of context and instruction on speakers' prosodic disambiguation. When ambiguous test sentences were presented to speakers in short paragraph contexts, so that readers were arguably not aware of any ambiguity, there was little prosodic marking of disambiguation. However, when speakers were given decontextualized ambiguous sentences, together with explicit instructions about the two meanings and instructions to render the difference with differences in their pronunciations, they did provide prosodic disambiguation. Such evidence suggests that the use of expert and aware readers in producing materials for comprehension experiments will result in test materials that may be quite different from normal conversational speech. As a consequence, the results of such comprehension experiments may not be generalizable to general conversational settings outside of the laboratory.

Some comprehension studies have attempted to use more natural materials by basing their experimental utterances on the results of companion production studies using untrained and naïve speakers (e.g. Warren et al, 1995). In many of these production studies it is also the case that no explicit attention is drawn to the ambiguity, since in contrast to the global ambiguities used by Allbritton et al, these studies have used local ambiguities, i.e. ambiguities that are resolved within the sentence in question (see further below).

A common factor in all of the studies cited above, and one which is a further important issue for naturalness, is that informants are asked to read text rather than to speak spontaneously. Allbritton et al acknowledge the significance of this factor, though they are unable to investigate it, since controlled contrasting materials are unlikely to arise in spontaneous speech. We argue that there are a number of concerns that can be raised regarding the naturalness of read speech. For one, readers and speakers may have different pragmatic goals. Spontaneous speech is generally produced in a contextually appropriate manner in order to achieve some communicative goal, while the communicative goals of reading in language production experiments are somewhat more limited. In addition, reading and speaking have different production demands. For example, readers are provided with a word order and orthographic representations of words, whereas speakers must generate a surface structure from a preverbal message. More generally, the time at which (parts of) semantic, syntactic, and prosodic structures are created may differ for read speech and spontaneously produced speech. Significantly, prosodic structures have been shown to differ for read speech and spontaneous speech (e.g., Howell & Kadi-Hanifi, 1991; Ayers, 1994; Blaauw, 1994). It should be noted that further complications may arise if readers do not fully parse the material before producing it, or parse it with a different structure than the one intended by the experimenter. Though most production studies have readers repeat non-fluent utterances, one can fluently utter a sentence that has been misparsed, particularly with sentences containing global ambiguities.

Despite these concerns, and despite the desirability of investigating the prosodic resolution of ambiguity in spontaneous speech, relevant research using such recordings is relatively scarce. Certain general observations can be made concerning the prosodic patterns used in spontaneous speech to mark particular syntactic structures or sentence types (cf. Kingdon, 1958; Goldman-Eisler, 1968, and others both before and since). However, spontaneous speech is unlikely to provide a rich array of carefully controlled contrasts such as those involved in structural ambiguities, since one consequence of the creativity of grammatical systems is that tightly matched pairs of utterances, differing only in a syntactic contrast of interest, are exceedingly unlikely to occur spontaneously.

One way of constraining the range of utterances used by speakers while still allowing spontaneous speech is to involve them in some form of role play. Again, though, most of the research tasks involving role play are unlikely to provide a large enough set of contrasting structures, largely because they have not been designed with this goal in mind. For instance, map tasks (Anderson et al, 1991) require speakers to give directions to guide a listener from one point on a map to another, and are concerned with eliciting contrasts such as that between given and new information; however, there are no constraints placed on the syntactic forms of expression. Descriptions of networks of colored nodes (e.g. Levelt & Cutler, 1983) have taught us much about the planning, sequencing and repair of utterances, including the use of prosody in these tasks, but, again, these are not designed for eliciting specific syntactic contrasts. And for similar reasons we are unlikely to find relevant data in the tangram task used by Clark and colleagues, which has nevertheless been useful in revealing aspects of how a reference is initiated and accepted as part of the collaborative process of communication (Clark & Wilkes-Gibbs, 1992), and in highlighting differences in successful comprehension for a collaborating addressee and for a non-interacting overhearer (Clark & Schober, 1992).

In the current paper we present data from a series of studies using a cooperative game task. These studies involve syntactic ambiguities that are carefully controlled for their syntactic properties and lexical content, and yet which are produced in a relatively spontaneous manner and in a situationally appropriate discourse context. In our task, speakers use a small set of predetermined expressions to negotiate the movement of gamepieces around a board. Speakers quickly become familiar with the utterances available for use under the rules of the game, and learn to produce them fluently and without recourse to printed sentence lists. Our expectation is that the task will produce a rich source of data for the study of syntactic ambiguity resolution in more spontaneous speech than that collected with traditional sentence reading tasks. Since the utterances are designed to include a number of syntactic ambiguities frequently studied in comprehension studies, we are also able to use our recordings in an examination of the use of prosodic information in ambiguity resolution during comprehension.

In addition, our boards are set out in such a way that the availability of certain movements of gamepieces inevitably changes during the course of a game, making certain interpretations of syntactically ambiguous sentences more or less likely. In this way, we are able to investigate the issue of whether prosodic cues to syntactic structure vary as a function of the situational constraints on certain interpretations.

Prepositional phrase ambiguities

The sentence materials used in this task include ambiguities similar to those exemplified in (1) and (2) above, and also ambiguities such as (3) which involve the attachment of prepositional phrases (PPs), and which provide the data to be discussed in this paper.

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

On one reading of this ambiguity (the low or NP attachment of the PP), the intention is to move a combined square-and-triangle piece (which in fact looks somewhat like a house). On the other reading (the high or VP attachment), the instruction is to use a triangle-shaped gamepiece to move, by shunting, a square piece. Such an interpretation is compatible with a requirement in our game that the square piece cannot move on its own but has to be moved by another object, which can include a triangle piece. This high attachment is in fact the only reading of another possible utterance in our set of instructions, given in (4). There is no combined square-and-cylinder piece, so a low attachment interpretation of (4) is infelicitous in the context of the game.

(4) I want to change the position of the square with the cylinder.

The prepositional phrase ambiguity in (3) is of particular interest here since it is a "global" ambiguity, in which the intended syntactic structure is not clear within the sentence itself. In this respect it contrasts with the "temporary" or "local" ambiguities illustrated in (1) and (2), where the choice of continuation signaled by the lexical information in braces will disambiguate the sentences, even in the absence of prosodic information. A further expectation concerning global ambiguities such as (3) is that a crucial role will be played not only by prosodic information, but also by situational information, i.e. by any properties of the developing situational context of the game that might make one interpretation of the ambiguity more likely than the other. We thus expect prepositional phrase ambiguities like those in (3) to be a productive area for the study of the use of prosody and its interaction with various situational constraints.

Previous PP production studies
Previous production studies of PP materials have revealed small but consistent phonetic differences between the two readings of sentences containing global ambiguities such as (5) (Cooper & Paccia-Cooper, 1980), (6) and (7) (both from Price et al, 1991), and (8) (Straub, 1997), as well as local ambiguities such as (9) (Warren, 1985) and (10) (Straub, 1997).

(5) Lieutenant Baker instructed the troop with a handicap.
(6) I read a review of nasality in German.
(7) Raoul murdered the man with a gun.
(8) The chauffeur annoyed the man with the cigar.
(9) Sam climbed the peak with *{snow on top / Pete and Dave}*.
(10) The chauffeur annoyed the man with the *{sunburn / song}*.

The phonetic differences reside in the incidence and duration of pauses before the preposition and in the duration of the preceding syllable or vowel nucleus (e.g. of *troop* in (5)), with greater durations before the higher or adverbial attachment. Pitch excursion differences over the utterances are less clear (Warren, 1985; Straub, 1997), and durational differences alone can be sufficient for disambiguation (Lehiste et al, 1976).

Studies using disambiguating paragraphs or local ambiguity, i.e. where prosody is not the only potential source of disambiguation, suggest that the structural interpretation is marked prosodically even when perhaps not strictly necessary. For instance, although disambiguating paragraphs preceded their high and low PP attachment ambiguities, Price et al (1991) still found stronger perceived breaks before the PP in the high attachment cases, corresponding to greater pausing and pre-pausal lengthening. On the other hand, Cooper and Paccia-Cooper (1980) found a reduction in durational contrasts when the two versions of sentence (5) were presented in disambiguating contexts compared to when they were read in isolation, a finding replicated for different PP materials by Straub (1997). Straub encapsulates this final finding in her Contingent (Prosodic) Cueing Hypothesis which states that prosodic cues are less marked when alternative sources of disambiguating information (such as preceding context) are available. These results are compatible with those from Allbritton et al's (1996) study of further ambiguity types, as mentioned above.

All of the studies cited above used reading tasks, often of sequences of isolated sentences. While some did involve a broader linguistic context (e.g. paragraph contexts), they still used scripted recordings. They also used either professional speakers (Price et al, 1991) or naïve speakers who were made aware of the ambiguities through instruction, and/or who were given the opportunity to rehearse. Such studies are thus open to the criticism mentioned above that the data they present are not necessarily typical of spontaneous speech. In many cases they also used a small sample of speakers (e.g. four in Straub's and Price et al's studies).

One aim of the current study is to include a larger number of subjects, who produce utterances which are situationally relevant and which are less obviously read from sentence lists or paragraphs.

Prosodic theory and ambiguity resolution
The production studies of PP ambiguities cited above consider the durational and pitch properties of the utterances recorded, treating these as separate though related parameters of speakers' disambiguation. However, a more appropriate measure of the contrast between the structures under investigation is likely to result from a phonological prosodic analysis, since speakers will differ from one another (and from themselves in separate recordings) in the extent to which they use the different acoustic cues to prosodic structure. This was pointed out some time ago by researchers such as Cutler and Isard (1980) and Henderson (1980), and more recently by Beach (1991) and Straub (1997). The latter concludes that the acoustic parameters of prosody have a cumulative effect, though she does not offer an analysis, such as a phonological prosodic one, which might reveal this effect. Recent studies of prosodic disambiguation consider phonological prosodic as well as acoustic parameters (e.g. Ferreira, 1993; Warren et al, 1995; Speer, Kjelgaard & Dobroth, 1996; Schafer, 1997; Kjelgaard & Speer, 1999), and the current study similarly includes the results of phonological analyses of the recorded materials, as well as acoustic data relating to duration and pitch, although in this preliminary report we include only acoustic analyses of duration.

In our prosodic analyses, we assume the intonational theory of Pierrehumbert and Beckman (Pierrehumbert, 1980; Beckman & Pierrehumbert, 1986) and follow the conventions of the ToBI transcription system (Beckman & Ayers, 1997; Silverman et al, 1992). In these systems, it is assumed that each utterance of American English is produced in one or more intonation phrases (IPhs). Each intonation phrase is made up of one or more intermediate phrases (ip), each of which must contain at least one pitch accent. American English employs several distinct pitch accents, including high (H*), low (L*) and bitonal accents (e.g., L+H*). The end of each prosodic phrase is associated with an edge tone. The end of an intonation phrase is marked by a boundary tone, which can be high (H%) or low (L%). The end of an intermediate phrase is marked by a phrase accent, which can be high (H-), low (L-), or a downstepped high (!H-). The phonological analysis also includes break indices, which mark the degree of disjuncture between constituents. The 'default' break between two words within a prosodic phrase is 1, with a break of 0 indicating that some connected speech process has taken place across the word boundary. An intonation phrase boundary generally corresponds to a break index of 4 and an intermediate phrase boundary to a break of 3. The index of 2 is used for a rhythmically salient break that is not also associated with a phrase accent, or for a phrase accent that is not also associated with the expected rhythmically salient break.

Roughly speaking, high edge tones and pitch accents are realized phonetically with a relatively high fundamental frequency (F0), and low edge tones and pitch accents with a low F0, but the exact F0 contour depends on the particular tone sequence, plus such factors as the kind of segments which carry the tune. Prosodic phrase boundaries can also be marked phonetically by lengthening of the final syllable of the prosodic phrase and a following silent interval (Wightman et al, 1992; Ferreira, 1993), segmental variation in the initial or final segments (Pierrehumbert & Talkin, 1992; Fougeron & Keating, 1997), and a new pitch range (Beckman & Pierrehumbert, 1986). In general, intonation phrase boundaries show more extreme effects than intermediate phrase boundaries. For example, an intonation phrase-final syllable tends to have a longer duration than a intermediate phrase-final syllable, corresponding to the higher break index associated with the intonation phrase. Thus, the durational differences preceding high and low attached PPs discussed above are evidence of differences in the phonological prosodic structure between high and low attachment, with high attachments likely having stronger prosodic boundaries preceding the PP than low attachments.

PP comprehension studies

The comprehension literature relating to PP attachments in speech presents a mixed picture, with the disambiguating effects of prosody generally small, but sensitive to situational constraints on the ambiguity. In an off-line forced-choice comprehension study, Warren (1985) found that laboratory recordings of PP ambiguities were correctly identified with only little more than chance accuracy. To an extent, that result could be explained with reference to an *a priori* bias for adverbial interpretations of the PP materials used (as reflected in a clear preference for adverbial continuations in a written task). It was clear, however, that the minor differences observed in the production data (compared with other syntactic contrasts included in the same study) had relatively small effects on subjects' choices.

Similarly, Straub (1997) — using speech materials from her four naïve speakers — found only chance-level correct identification of PP ambiguities in a forced-choice experiment, both for materials recorded in a biasing context and for those recorded with no biasing context. Latencies in a naming task, however, did show sensitivity to differences in the recorded materials. Continuation words (e.g. *sunburn* or *song* for (10) above) were responded to more quickly if they were consistent with the intended interpretation of the original recording than otherwise. Perhaps unsurprisingly given the production data surveyed earlier, this effect was strongest for globally ambiguous materials produced without a biasing context — i.e. where prosodic information is the sole means by which the speaker may potentially disambiguate.

Using the speech data from their four expert speakers, Price et al (1991) asked a panel of naïve subjects to match contexts to ambiguous sentences isolated from their paragraphs. They found that their subjects correctly identified 78% of the high attachments and 63% of the low attachments, though with low confidence compared to other ambiguity types (i.e. confidence levels were 27% and 20% for high and low PP attachments respectively, compared with 52% for all ambiguity types together).

The locus of likely prosodic effects has been considered in an experiment combining prosodic break location and verb argument structure in French. Pynte and Prieur (1996) considered PP attachment sites in sentences such as (11) and (12), each of which can have a PP modifying the verb phrase (VP) or the second noun phrase (NP).

(11) Les espions informent les gardes du *{complot / palais}*.
(The spies inform the guards of the *{conspiracy / palace}*.)

(12) L'étudiant choisit un appartement avec *{soin/balcon}*.
(The student chooses a flat with *{ care / a balcony}*.)

Pynte and Prieur had the sentences read out by a trained native speaker, and then manipulated the position of prosodic breaks, marked by lower F0 and increased segment duration as well as by a pause (which they standardized to 150 msec). Using word monitoring tasks, they found that the presence of a break after the verb resulted in an advantage for NP attachment words (*palais* in (11); *balcon* in (12)), while a second break before the PP neutralized this effect, resulting in faster monitoring times for VP attachments and markedly longer latencies for NP attachments.

Schafer (1997) examined PP attachments in the context of her Prosodic Visibility Hypothesis. This claims that attachments will more readily be made to constituents that are within the intermediate phrase currently being processed, and that attachment to syntactic nodes associated with material in earlier intermediate phrases becomes more difficult the more intervening intermediate phrase boundaries there are.[1] This Hypothesis was tested using sentences like (13) with intermediate phrase boundaries as marked. Schafer's Prosodic Visibility Hypothesis predicted that an intermediate phrase boundary before the PP would lower the visibility of all of the preceding nodes, including both the NP and the VP nodes, resulting in a greater proportion of default VP-attachment preferences for the PP in (13a) and (13d), compared with (13c) and (13b) respectively. It also predicted that there would be fewer high VP attachments when an intermediate phrase boundary falls between the V and the NP, as in (13b) and (13d), compared with (13c) and (13a) respectively, since this makes the high attachment site less visible when the PP is being processed. The results of Schafer's study confirmed these predictions, with the strongest effect resulting from the presence of intermediate phrase boundaries between V and NP. The percentage value given after each example indicates the level of VP attachment choice for that type in Schafer's data.

(13a)	(The bus driver angered the rider)$_{ip}$ (with a mean look).	61.5%
(13b)	(The bus driver angered)$_{ip}$ (the rider with a mean look).	44.3%
(13c)	(The bus driver angered the rider with a mean look).	59.9%
(13d)	(The bus driver)$_{ip}$ (angered)$_{ip}$ (the rider)$_{ip}$ (with a mean look).	52.6%

Finally, Snedeker, Trueswell, Gleitman, and Levine (1999), reporting on work in progress, tested sentences like (14) in two speech situations. In one situation, a speaker and a listener saw matching displays containing an unadorned frog, a flower, and a frog wearing a flower. The speaker uttered instructions like (14) to the listener. These instructions were written on a card and shown to the speaker after the experimenter had demonstrated the appropriate move with the objects to the speaker, providing disambiguation of the globally ambiguous instruction for the speaker. In this first speech situation, Snedeker et al reported significant differences in prosodic disambiguation, as measured by the listener's ability to act out the correct movement of objects. (Phonetic analyses of the prosody were not provided.)

(14) Tap the frog with the flower.

In the second speech situation, with different subjects, the speaker saw objects that matched only one interpretation of the ambiguous sentences, e.g., a plain frog and a flower but not a frog wearing a flower. Unbeknownst to the speaker, the listener saw displays like those in the first speech situation, which supported both interpretations. The two different syntactic structures were tested between groups of subjects, i.e. each listener heard sentences produced for only one of the PP attachments. As before, the speaker read aloud written instructions following a demonstration of the move by the experimenter. In this speech situation, however, when the ambiguity of the sentence was resolved for the speaker by the display of objects (as well as the demonstration), the listener's manipulation of objects showed no evidence of prosodic disambiguation. (Again, phonetic analyses were not provided.)

Without knowing something about the pronunciations the speakers provided in the two situations, there are at least two ways to interpret these results. It is possible that the speakers produced fewer prosodic cues to syntactic structure when they saw an unambiguous display, and that listeners correspondingly found the prosodic structures from the second situation to be more ambiguous than the ones produced in the first situation. It is also possible that the speakers produced similar prosodic structures in the first and second speech situations, but that listeners responded differently when hearing prosodic structures intended for just one of the interpretations available in their display. That is, speakers may have continued to provide prosodic cues to syntactic structure in the second situation, but listeners, hearing only high attached sentences or only low attached sentences, and seeing a display that supported both interpretations, may have shifted their categorization of the prosodic structures to better divide them between the two interpretations available in the situation. Nevertheless, speakers presumably produced prosodic disambiguation in the first situation, and may have adjusted their prosody in response to their perception of the situational ambiguity.

There is thus some support for the role of prosodic information in disambiguating PP ambiguities. This is true both for experiments using speech from expert speakers (e.g. Pynte and Prieur) and for experiments using naïve subjects (e.g. Straub). The current study extends these findings, using naïve subjects as informants and recording utterances which, although from a constrained inventory, are nevertheless used by our subjects to achieve communicative goals in a contextually relevant way. The game contexts are deliberately varied, allowing us to assess the degree of prosodic disambiguation for ambiguous and non-ambiguous versions of the same utterances, without the need for explicit disambiguation.

Situational ambiguity

As noted above, previous research has suggested that the degree of prosodic disambiguation of syntax might directly depend on whether other factors are available to resolve the ambiguity. For example, Straub (1997) proposed that the production system would only allocate the cognitive resources necessary to provide prosodic disambiguation[2] when other sources of disambiguating information were unavailable. Therefore, under her proposal, prosodic disambiguation should not be expected when the context in which a sentence is uttered has made the intended meaning more plausible than alternate grammatical interpretations.

Straub (1997) and Allbritton et al (1996) manipulated the ambiguity of the speech situation through lexical differences in the texts subjects read. Globally ambiguous sentences followed short paragraphs that biased interpretation toward one meaning, or locally ambiguous sentences ended with material that pragmatically biased the interpretation. In our game task, multiple sources of information were potentially available to influence the resolution of the PP ambiguity. One source of information that was occasionally available was preceding linguistic information, as in the studies by Straub and Allbritton et al. For example, if the players had just referred to the combined square-and-triangle gamepiece, we assumed that they might be biased to expect the next utterance to refer to it again. Our speakers also had visual information, from the layout of the gamepieces on the gameboards, which sometimes resolved the attachment ambiguity. There were two categories of disambiguating visual information in our task, one of which was similar to the information given to Snedeker et al's speakers.

As described with examples (3) and (4) above, our study contrasted *cylinder* sentences with *triangle* sentences. Speakers saw displays which included the combined square-and-triangle piece, simple (unmodified) squares, simple triangles, and cylinders, but no combined square-and-cylinder pieces. Previous sentence comprehension work suggests that this visual information should have been extremely important in any calculations of the ambiguity of the situation. For example, Tanenhaus et al (1995) examined eye fixations on objects in a visual display during the comprehension of spoken sentences with temporary PP attachment ambiguities such as (15). They argued that listeners made extremely rapid assessments of whether more than one referent was available in the display to support a spoken description. Listeners showed eye movements consistent with initial attachment of the phrase *on the towel* to the NP *the apple* when there were two apples in the visual display (one of which was on a towel), but evidence of initial VP attachment of the phrase when there was only one apple in the display. This and other studies, including the work of Snedeker et al described above, suggest that the players of our game should be strongly biased toward VP attachment for the *cylinder* sentence, since that is the only interpretation supported by the display of gamepieces, but either biased toward NP attachment or unbiased for the *triangle* sentence, since both interpretations are supported by referents in the discourse situation.

(15) Put the apple on the towel in the box.

A more subtle type of disambiguating information was available from the arrangement of pieces on the gameboard. Certain gameboard configurations made the move associated with one or the other interpretation of our triangle sentence impossible under the rules of the game. For example, a piece could not move if it did not have an open path in some direction, and the square could only be pushed by a triangle if the triangle was adjacent to it. Therefore, the critical PP sentence would most plausibly refer to the combined square-and-triangle piece if that piece had an open path and none of the squares were adjacent to triangles; some of the other configurations should have biased interpretation toward the other meaning.

Both types of disambiguating visual information were in the common ground. The information was available to both players in the game, and both players should have been aware of the other's knowledge of it. Thus, if a speaker's conscious or unconscious evaluation of situational ambiguity depends on whether the speaker believes the ambiguity has been resolved for the listener (and not just for the speaker), the visual information in our game should still be considered a relevant source of information, although it might be expected to be more influential in monitoring and correcting sentence production than in the initial planning of it (Horton & Keysar, 1996).

Experiments

The research we report in this paper uses our gameboard task (introduced briefly above and to be described in more detail below) to investigate the prosodic realization of global PP ambiguities by naïve speakers using utterances that are situationally motivated. Our task allows for variation in contextual constraints on the different possible interpretations of the ambiguity, and Experiment 1 (our production study) includes the evaluation of the effects of such constraints on prosodic realizations. In Experiment 2 we will also investigate the extent to which the utterances recorded in Experiment 1 but then isolated from the context of the game can be reliably categorized by naïve listeners.

Hypotheses

The main issue in these experiments concerns the syntactic and situational determination of the prosodic realization of PP ambiguities by naïve speakers in the game task, using utterances such as (16) to (19). We conducted our experiments with the following hypotheses in mind.

(16) I want to change the position of the square with the triangle.
(17) I am able to confirm the move of the square with the triangle.
(18) I want to change the position of the square with the cylinder.
(19) I am able to confirm the move of the square with the cylinder.

1. *Syntactic determination*. We predict that our phonological and acoustic phonetic analyses will support a difference in the realizations of high (VP) and low (NP) attachments of the PP in structurally ambiguous utterances such as (16) and (17). In each case, the high attachment will be reflected in a stronger prosodic break before the PP than found in low attachment sentences.

2. *Illocutionary force*. In our game task, one speaker (the Driver, for detail see below) issues instructions, as in (16) above, while another (the Slider) follows these instructions and confirms that moves have taken place, using utterances such as that given in (17) above. Disambiguation is potentially more crucial in Driver utterances, since the incorrect move may otherwise result. If prosodic realization is sensitive to such pragmatic factors, then we would predict greater disambiguation for Driver utterances than for Slider utterances.

3. *Level of situational ambiguity 1: gameboard configuration*. In the game, there will be situations in which the Driver's instructions are truly ambiguous (either high or low attachments would be situationally possible), others in which they are unambiguous (only one relevant move is possible given the rules of the game and configuration of the board), and others in which there is ambiguity but recent moves would result in a bias towards one interpretation over the other. If speakers are sensitive to situational constraints, then we

predict that there will be greater disambiguation between high and low attachments for ambiguous situations than for biased or unambiguous ones.

4. *Level of situational ambiguity 2: gamepiece contrast*. In addition to *the square with the triangle*, our game data includes *the square with the cylinder* sequences (see (18) and (19) above), in which the only interpretation in the context of the game is that of a high attachment (since there is no combined square-with-cylinder piece). Again, situational sensitivity predicts that the features that indicate high attachment (i.e. the prosodic break before the PP) will not be as clearly marked in the *cylinder* utterances as in the *triangle* utterances. This would match Straub's (1996) finding of reduced marking of attachment contrasts for local versus global ambiguities. That is, although the attachment of the preposition preceding *triangle* or *cylinder* is ambiguous, the presence of *cylinder* disambiguates the preposition's attachment to a high one. As noted above, several recent studies suggest that speakers and listeners are sensitive to information in visual displays of the sort found in our triangle/cylinder gamepiece contrast.

We also predict that naïve listeners will be sensitive to prosodic differences in our materials. We develop specific hypotheses conerning comprehension in our presentation of Experiment 2 below.

<u>Experiment 1 – speech production task</u>

Production task

Our production study used a cooperative game task in which two players used scripted sentences to negotiate moves of gamepieces from starting positions to goals. The players were restricted to a set of provided sentences, but they had to choose the sequence of moves to be followed, and so they had some freedom in choosing which sentences to use and when to use them. The full list of sentences from which subjects could choose is given in Appendix A. In order not to bias the subjects' choice of prosodic structure, the experimenters were careful to avoid using the scripted sentences themselves. Subjects were not made explicitly aware of the syntactic ambiguities in the game sentences, and they were never told to use disambiguating pronunciations.

The game was non-competitive, and players were encouraged to work together to accumulate points for the successful movement of objects to their goals, while avoiding the deduction of points for false moves or incorrect usage of expressions. Each gameboard had two slightly differing versions. One, used by the player known as the Driver, included the goal locations for the gamepieces, but not the locations of bonuses (cookies) and hazards (ravenous goats). The other board, used by the Slider, did not have the goals marked, but did show the locations of bonuses and hazards. All remaining information (gamepieces on boards; locations of gamepieces, etc.) was identical for the two boards, and players were aware of what was and was not the same for the two boards. The Driver's role was to tell the Slider which piece to move, to inform the Slider when he or she moved incorrectly, and to confirm that a gamepiece had reached its goal position. 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. Neither player could see the board being used by the other, and the design of the boards and of the rules of the game encouraged negotiation and the strategic use of moves. In addition, board layouts were designed to allow varying levels of situational constraints on the moves that were possible, i.e. at certain points in the game only a subset of the gamepieces could be or needed to be moved. Four pairs of gameboards, with differing layouts, were used, plus a pair of practice boards and a demonstration board. The initial configurations of the four gameboards are shown in Appendix B.

Each pair of subjects played multiple rounds of the game, switching roles and gameboards between rounds. They wore head-mounted microphones, and their utterances were recorded simultaneously to both computer disk and cassette tape. Subjects were also required to use a placemarker, which they placed in the position of the gamepiece that they wished to move, and which provided the experimenter with extra information concerning the players' intentions.

Situational ambiguity

Our measure of situational ambiguity for the PP attachments depended on our keeping track of the game situation during the recording session. For this purpose the experimenter used a check-sheet, on which various key pieces of information were noted during the playing of the game. The analysis of situational ambiguity was into three categories, as follows.

Ambiguous. This category included all instances in which the Driver's global ambiguity *I want to change the position of the square with the triangle* could (disregarding disambiguating prosody) with equal likelihood be interpreted as an instruction either to move the combined square-with-triangle piece or to use the triangle to shunt the square. In other words, there was no a priori reason to expect one move rather than the other to have been intended.

Unambiguous. In this category we included all cases where only one piece could be moved legally. Reasons for this state-of-affairs could be, for instance, that the square-with-triangle was boxed in (other pieces had to be moved out of its way first), as in the starting configuration for gameboard 1 (see Appendix B), or that the square had been moved to its goal, and no triangle could be used to move it back out of its goal.

Biased. This category included situations where both interpretations of the utterance were possible, but one was considerably more likely, for instance as a result of the Driver having just given a set of instructions that moved a triangle up next to a square, so that using the triangle to move the square was 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 completed a practice game, with one participant as Driver and one participant as Slider. In addition, each pair played at least two more games, using a separate board. Players played for two hours, and completed as many games as they could within that time, playing each board twice, once in the role of the Slider, and once in the role of the Driver. The maximum number of games played, not including the practice game, was five games, involving three different boards.

Excluded subjects and utterances
Two subjects' data were excluded from the analysis of PP attachment, because each of them produced only one fluent utterance in the high attached syntax condition. The data from a third subject were excluded because the subject had great difficulty producing fluent utterances, primarily because of segmental errors.

For the remaining 13 subjects, 22 non-fluent utterances were excluded: 12 Driver utterances, 4 with high attached syntax (both *triangle* and *cylinder* types) and 8 with low attached, and 10 Slider utterances, 4 with high attached syntax and 6 with low attached. Utterances were classified as non-fluent according to the following criteria: mispronunciation, substitution, addition or deletion of words, word-internal pausing, re-starting the sentence, presence of a filled pause, and presence of non-speech events (sniffing, coughing, laughing) in the critical region. Excluded utterances met one or more of these criteria.

Acoustic phonetic measures
Durations of each of the words and pauses (if any) in the sequence *position of the square with the triangle* were taken from the digitized speech waves. Future analyses will include F0 data at the key positions in the utterances, though the studies reviewed in the *Introduction* lead us to believe that reliable contrasts in local F0 patterns are unlikely.

Results
Hypotheses 1 and 2. Our first two hypotheses were tested by comparing the durations of the word *square* and of any following pause in Analyses of Variance with attachment type of the PP (high or low) and speaker role (Driver or Slider) as factors. Hypothesis 1 predicted that durations would be greater in the high attachment case than in the low attachment case (see *Previous PP production studies* above), resulting in a main effect of attachment type. In accordance with the second hypothesis, any such difference was predicted to be greater in the Driver case than in the Slider case and should be reflected in an interaction of speaker role and attachment type. The overall averages are shown in Figure 1.

Separate analyses were conducted on the average durations of *square* and of the following pause, as well as on the combined *square* + pause durations. Each of these analyses produced a significant main effect of the level of PP attachment (*square*: $F[1,12] = 30.74$, $p<0.001$; pause: $F[1,12] = 17.76$, $p<0.002$; *square* + pause: $F[1,12] = 30.54$, $p<0.001$). These results provide clear support for the first of our hypotheses, i.e. that speakers mark the level of attachment of the PP.

However, none of the analyses produced an interaction of speaker role and attachment type, and so our second hypothesis is not supported directly by these data, i.e. it does not seem to be the case that speakers disambiguated more clearly in the Driver role, where instructions were issued, than in the Slider role, where a requested move was confirmed.[3] There were, however, main effects of speaker role for both the durations of *square* on its own ($F[1,12] = 18.11$, $p<0.002$) and the combined *square* + pause duration ($F[1,12] = 8.67$, $p<0.02$), with Drivers producing greater durations than Sliders. This may reflect differences in the care with which speakers articulated utterances in Driver and Slider roles.

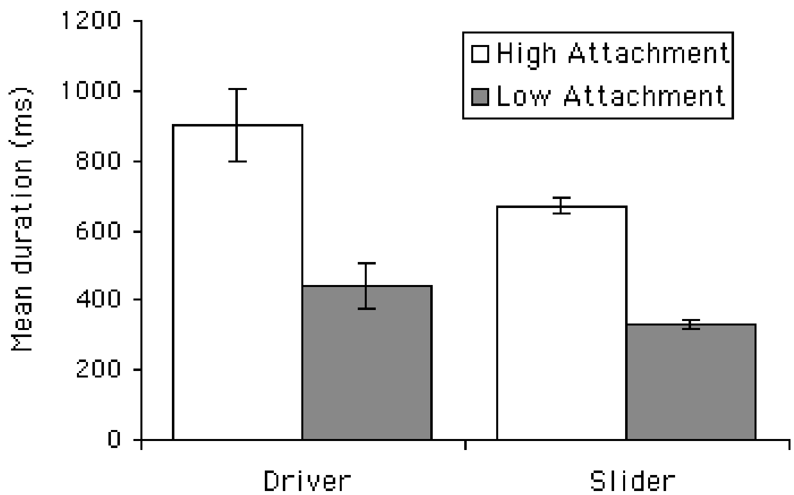


Figure 1. Mean *square* + pause durations (with standard error bars) for Driver and Slider utterances of high and low PP attachments. These means are based on 79, 101, 59 and 100 tokens for Driver high and low attachments and Slider high and low attachments respectively, from 13 speakers.

Hypothesis 3. Our third hypothesis concerned effects on disambiguation in production of the level of situational ambiguity at the point in the game when the utterances were produced. To test this hypothesis, we assigned our durational data to the three categories of ambiguity level determined as above (see *Situational ambiguity* above). Here, we report data from only the Driver role, the role in which we expected to see the strongest effects of the level of situational ambiguity. 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 2.

Clearly, the variable number of tokens making up these data make the comparison of overall means rather unreliable (see the error bars in Figure 2 as an indication of the amount of variability and the Figure caption for the number of tokens in each condition). In particular, the breakdown by ambiguity level left some speakers with very small or empty cells for some conditions, which could have resulted in individual speaker data inflating the duration means in some conditions but not in others. Therefore, we restricted our statistical analysis to those speakers who had at least one instance in each ambiguity x attachment condition. The resulting ANOVAs could be based on data from only 11 of our speakers, and could only cover the high attachment cases. The analyses showed no effect of ambiguity level on the duration of *square* ($F[2,20]<1$), of the following pause ($F[2,20]=1.06$) or of the combined duration of *square* + pause ($F[2,20]<1$).

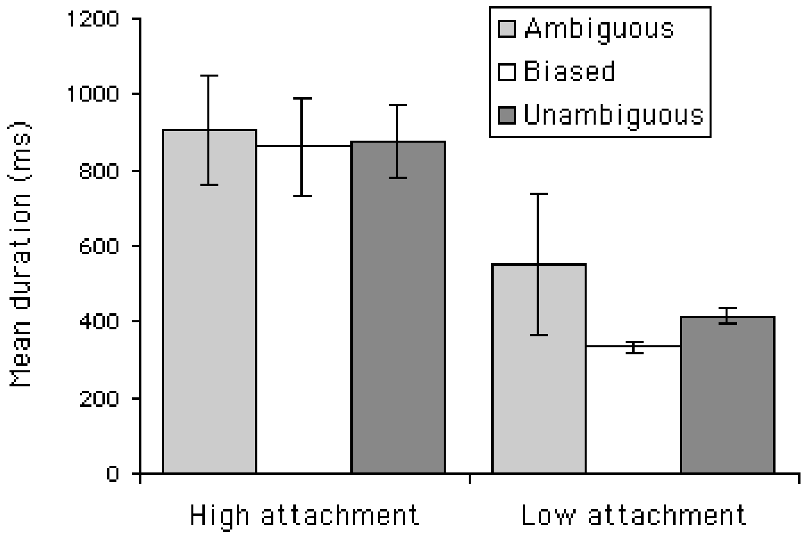


Figure 2. Mean *square* + pause durations (with standard error bars) for high and low attached *triangle* tokens, by situational ambiguity level (see text for details). These means are based on 27, 25, 27, 4, 12 and 85 tokens for each of the columns — left-to-right — in the figure, from 13 speakers.

Hypothesis 4. Our further hypothesis concerning situational ambiguity involved a contrast resulting from the use of different gamepieces. Recall that a parallel pair of sentences to those in (16) and (17) has the word *cylinder* in the place of *triangle*, as in (18) and (19), reflecting the possibility that a cylinder piece could be used to move the square. This provided a further set of high attachment sentences only, the low attachment equivalent not being possible, since there was no combined square-with-cylinder piece. If speakers are sensitive to situational aspects of ambiguity, then the unambiguous high attachments (in (18) for Driver sentences and in (19) for Slider sentences) may not be as clearly differentiated from low attachment utterances as the ambiguous high attachment versions of (16) and (17). To test this prediction, ANOVAs were performed on the duration data from high and low PP attachments with *triangle* and unambiguous high attachments with *cylinder*. The dependent variable was the combined *square* + pause duration. This analysis was performed on data taken from 11 speakers who produced at least one utterance in each condition.

The average results (Figure 3) show that our speakers marked the syntactic properties of the sentences regardless of whether there was a potential ambiguity between high and low attachments. This was confirmed in the ANOVA for this data set, which indicated a significant main effect of utterance type ($F[2,20]=13.13$, $p<0.001$) and a weak effect of speaker role ($F[1,10]=4.68$, $p<0.06$). Planned comparisons showed that the utterance type effect was due to significant differences between the low attachment triangle utterances and each of the high attachment cases, for both Driver and Slider utterances. There was no significant difference between the high attachment conditions. Analyses of other word regions in the sentences confirmed that our speakers used similar rates of speech for the *triangle* and *cylinder* utterances.

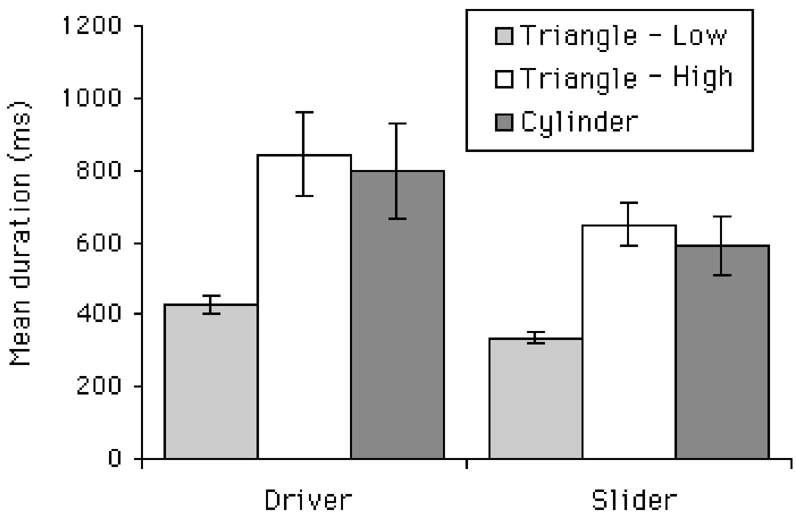


Figure 3. Mean *square* + pause durations (with standard error bars) for Driver and Slider utterances of high and low attached *triangle* tokens and high attached *cylinder* tokens. These means are based on 86, 68, 20, 91, 48, and 14 tokens for each of the columns — left-to-right — in the figure, from 11 speakers.

Summary

Our production data show that speakers in our task marked with some consistency the syntactic difference between high and low attachments of PPs. Lengthening of the word and pause before the PP was found for unambiguous as well as ambiguous high attachments, compared with low attachments. Such lengthening does not appear to be contingent on situational factors such as the differing communicative goals of Drivers and Sliders, nor on the level of ambiguity as determined by the possible movements of gamepieces at the relevant stage of the game. This need not mean, of course, that these situational constraints result in no differences whatsoever in the prosodic marking of the utterances, but simply that they are not reflected in durational differences at the key point. Further phonetic analyses of our materials are in progress, including measurements of fundamental frequency patterns across the utterances, as well as phonological analyses through ToBI-style prosodic transcriptions.

It is particularly striking that the gamepiece manipulation had no apparent effect on the degree of prosodic disambiguation. One-referent versus two-referent contrasts, some very similar to our cylinder versus triangle contrast, have resulted in significant effects on comprehension in more than one laboratory (e.g. Tanenhaus et al, 1995; Trueswell et al, 1999). Such a contrast may have also affected the production of prosody in Snedeker et al's reading task. Nevertheless, the productions in our study show robust differences in duration between high and low attachment sentences, but no significant reduction in the durational markings of high attachment for the unambiguous, one-referent *cylinder* utterances from those found with the ambiguous, two-referent *triangle* utterances.

While perhaps surprising from the perspective of comprehension studies, our findings are in keeping with other work in the production literature. Ferreira and Dell (2000) showed that speakers do not consistently include a disambiguating lexical item, *that*, to avoid temporary syntactic ambiguities in sentence complement structures. They argued that inclusion or omission of the complementizer depends instead on the availability (with respect to sentence production demands) of the material that would be produced next. Together these studies suggest that the calculation of phrase structure ambiguity may not play a significant part in the sentence production process, at least under typical production conditions when speakers are not consciously attempting to avoid syntactic ambiguities. We believe that the production of sentence prosody is controlled primarily by grammatical factors, such as phonosyntactic constraints relating prosodic form to syntactic form and phonological constraints governing the weight of prosodic units. Under this view, it is to be expected that the high attached *cylinder* utterances would have similar prosodic structures to the high attached *triangle* utterances, since they have the same syntactic structure and contain similarly sized phonological units.

<u>Experiment 2 – comprehension task</u>

The second set of data to be reported in this paper come from a comprehension study conducted using a subset of the materials recorded from the game task and used in the production study. The purpose of this comprehension study was to determine whether the patterns indicated in the production study are useful to listeners in making judgments about the utterances (in this case a simple forced-choice categorization of the intended utterance). At the same time, it was felt that a comprehension study might provide some information relevant to the issue of whether other phonetic and/or phonological cues to the structure of the utterances might be present than those that we had planned to measure.

Method

For the comprehension task we selected the Driver's ambiguous PP utterances (i.e. sentences as in (16) above) from our production study. These utterances were presented to listeners as complete sentences in a forced-choice task. The listeners were required to select between two paraphrases, one indicating high attachment, the other low attachment.

Subjects

19 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.

Analysis sets

All *triangle* utterances produced in the Driver role from our production study were presented to each of the listeners, except for the utterances from the speaker with high numbers of non-fluent utterances, for a total of 196 utterances from 15 speakers. This included 16 utterances from the two speakers who produced only one token each in the high attachment condition, which were excluded from analysis. Only *triangle* utterances were used, since the *cylinder* utterances had only high attachment recordings. We carried out two analyses of these comprehension data, analyzing listener's responses to the same two sets of utterances described in conjunction with the test of Hypothesis 3 above. The first analysis included responses to all fluent utterances produced by 13 speakers. The second analysis included responses to only those speakers who had produced at least one fluent token in each of the three ambiguity conditions (ambiguous, biased, unambiguous — see *Situational ambiguity* above). The resulting ANOVA could be based on data from only 11 speakers in the high attachment cases, who together produced 23 ambiguous, 25 biased, and 19 unambiguous tokens. The results for both the entire set and the subset are very similar, and so the presentation below concentrates on the larger data set.

Hypotheses

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

We further predicted that if speakers provide prosodic disambiguation only when the situation requires it, the percentages of 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. Alternatively, if speakers provide prosodic disambiguation regardless of situational ambiguity, then percentages of correct categorization should not differ systematically across the levels of situational ambiguity.

Results

The percentage correct classifications for the first analysis set are given in Figure 4. The overall classification was greater than chance, and therefore supports the first hypothesis above, namely that listeners were able to make use of distinctions in the productions of these utterances that signal high or low attachment. The percentage correct scores for each condition and for each individual participant were subjected to an arcsine transformation, ($2\text{arcsine}\sqrt{p}$), 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, which is compatible with the operation of a default parsing strategy such as minimal attachment (Frazier, 1987; Clifton, Speer & Abney, 1991; Frazier & Clifton, 1996). 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 significant effect for low attachment items resulted from the lower correct score for ambiguous than for biased or unambiguous items. At first blush this is a surprising result, since we would expect contextually ambiguous items to have been better disambiguated by speakers — and to result in better comprehension scores — than unambiguous ones. However, our production results for these items (see Figure 2) might be interpreted as showing that speakers were being more deliberate in their production of ambiguous low attachment sentences, resulting in greater durations for items in this condition. If this is the case, then the strategy would appear to backfire, since the result was more incorrect responses. However, given that listeners heard only 4 utterances in the low attachment-ambiguous situation condition, which were produced by just 3 speakers, the results for this condition must be viewed with caution.

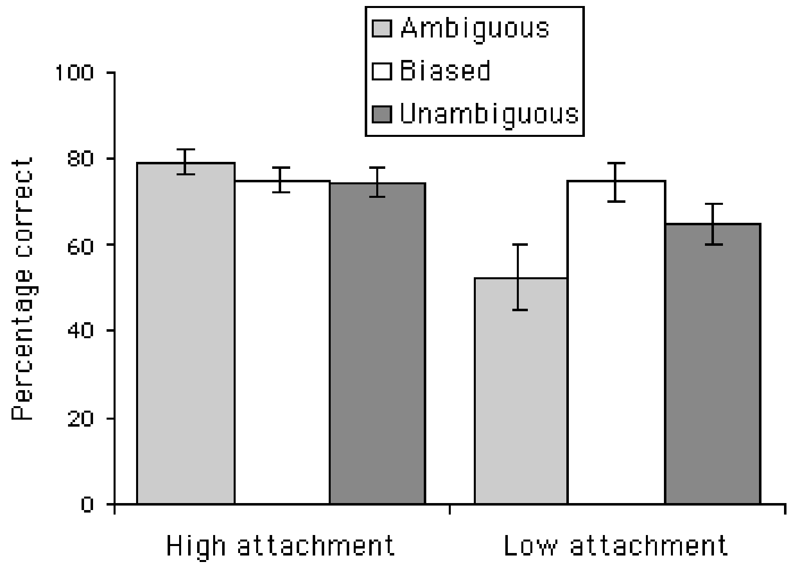


Figure 4. Percentage of correct classifications of tokens as high or low attached sentences, by level of situational ambiguity. The percentages are averages of the values obtained for 19 subjects, and are based on each subject hearing 27, 25, and 27 tokens for high attached ambiguous, biased, and unambiguous conditions and 4, 12, and 85 tokens for low attached ambiguous, biased, and unambiguous conditions respectively.

To test the suggestion of a relationship between duration and the likelihood of high and low attachment responses, we carried out a regression analysis of correct responses and *square* + pause durations for the low attachment sentences. This showed a significant effect, with an R-squared of 0.257 and a correlation of -0.528. In other words, participants in the comprehension experiment appear to have used the duration of the *square* + pause sequence as a cue to the high versus low attachment interpretation of the decontextualized utterances they were presented.

The subset analysis showed no significant effect of level of situational ambiguity on comprehension ($F[2,36]<1$). As described above, this analysis was limited to the comprehension of utterances from 11 speakers who produced at least one token in each of the situations for high attached utterances. The results of this analysis were extremely similar to the results from the complete set of fluent utterances, with mean percentages of correct responses of 73%, 75%, and 74% in the ambiguous, biased, and unambiguous conditions respectively. Reductions of the data set to just the first token or just the last token produced by each of the 11 speakers in each of the situational conditions continued to yield non-significant effects of situation on comprehension.

The lack of a main effect of ambiguity level suggests that there is no strong support for the suggestion that speakers produce different patterns of prosodic marking according to the ambiguity level of the situation. Thus, the results of the comprehension experiment support the conclusions from the production experiment. Our speakers tended to disambiguate the PP sentence, and they did so regardless of the ambiguity of the situation.

General Discussion

Consistent with previous findings from our production game task (e.g., Schafer et al, 2000), the PP attachment sentences were reliably disambiguated through prosody even when disambiguating information was available in the speech situation. Our durational analyses showed prosodic boundaries that were consistent with the syntactic structure, with high attached PPs marked by lengthening and silent intervals immediately preceding the PP. Low attached PPs were less likely to have these indications of a prosodic phrase boundary.

Our comprehension results showed that listeners were able to correctly choose the syntactic structure of spoken fragments, though the accuracy of this categorization was higher for high attachments than for low attachments.

None of our analyses indicated that our speakers modulated their prosodic structure to reflect the level of situational ambiguity. First, the durational difference between high and low attached PPs was present both in the directive utterances produced by speakers in the Driver role and in the confirmatory utterances produced by the same speakers participating as Sliders. There was no interaction of speaker role with the syntactically determined differences.

Second, the durations of the critical regions in the sentences did not differ for sentences produced in ambiguous, biased, and unambiguous game configurations. Sentences produced in these contexts also showed similar percentages of correct categorization in the comprehension task.

Third, the durations of the critical regions in the high attached PP sentences did not differ for sentences that could be mapped to only one set of discourse referents (sentences with *cylinder*, as in (18) and (19) above) and sentences that could be mapped to two sets of discourse referents (sentences with *triangle*, as in (16) and (17)).

These results show that our speakers provided prosodically distinguishable tokens for the two syntactic interpretations of the prepositional phrase attachment utterances being studied. They did this without being told explicitly to disambiguate, and without having their attention drawn to the existence of ambiguity. The reliability of the prosodic disambiguation therefore suggests that prosody is an important source of information for sentence comprehension in a range of discourse situations, and is not limited to situations with trained speakers or speakers who are consciously attempting to disambiguate. While it is clear that speakers may have become aware of the ambiguity of the *triangle* sentences as they participated in the recording sessions, we believe that the situational contexts in which the utterances were used are much more like non-laboratory situations than other controlled recordings previously studied. We also believe that our situational differences are very much like others which have been shown to have contextual effects, and that they are therefore sufficiently salient to be encoded by players. The fact that these situational differences, reflected in ambiguity levels, had no effect on speakers' realizations, indicates that speakers provide the same kinds of prosodic structures for unambiguous situations that they provide for ambiguous ones (and these structures frequently disambiguate the syntax for comprehenders).

Notes

* This work 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, and Gerald Whiteside for assistance with running subjects and measuring data. Authors' email addresses: Paul Warren <paul.warren@vuw.ac.nz>, Amy J. Schafer <schafer@ucla.edu>, Shari R. Speer <speer@ling.ohio-state.edu>, S. David White <dwhite@ling.ohio-state.edu>

1 Note that Schafer (1997) expresses her hypothesis in terms of phonological phrases. These are equivalent to the intermediate phrases here.

2 Straub does not describe how prosodic structures are developed in the course of production, or discuss why a disambiguating prosodic structure should require more cognitive resources than whatever prosodic structure would otherwise be produced.

3 Note though that a pilot study of New Zealand English data collected as part of our collaborative project did show an interaction of attachment level and speaker role, suggesting that speakers in that study may have been sensitive to this aspect of the situational constraints on disambiguation (cf. Speer et al, 1999).

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, 714-135.

Anderson, A.H. et al. (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.

Blaauw, E. (1994). The contribution of prosodic boundary markers to the perceptual difference between read and spontaneous speech. *Speech Communication*, 14, 359-375.

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

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

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

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. & Wilkes-Gibbs, D. (1992). Referring as a collaborative process. In H. Clark (Ed.) *Arenas of Language Use*. Chicago: University of Chicago Press. 107-143.

Clifton, C., Jr., Speer, S., & Abney, S. (1991). Parsing arguments: Phrase structure and argument structure as determinants of initial parsing decisions. *Journal of Memory and Language*, 30, 251-271.

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

Cutler, A. & Isard, S.D. (1980). The production of prosody. In B.Butterworth (Ed.) *Language Production, Vol. 1: Speech and Talk.* New York: Academic Press. 245-269.

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

Ferreira, F. (1993). Creation of prosody during sentence production. *Psychological Review*, 100, 233-253.

Fougeron, C. & Keating, P. (1997). Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America*, 106, 3728-3740.

Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.) *Attention and Performance XII: The Psychology of Reading.* Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in Spontaneous Speech*. London: Academic Press.

Henderson, A.I. (1980). Juncture pause and intonation fall and the perceptual segmentation of speech. In H.W. Dechert & M. Raupach (Eds.) *Temporal Variables in Speech: Studies in Honour of Frieda Goldman-Eisler*. The Hague: Mouton. 199-206.

Horton, W.S. & Keysar, B. (1996). When do speakers take into account common ground? *Cognition*, 59, 91-117.

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

Kingdon, R. (1958). *The Groundwork of English Intonation*. London: Longman.

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

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.

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

Pierrehumbert, J.B. (1980). *The Phonology and Phonetics of English Intonation.* Unpublished doctoral dissertation, MIT.

Pierrehumbert, J.B. & Talkin, D. (1992). Lenition of /h/ and glottal stop. In G. Docherty & D.R. Ladd (Eds.), *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*. Cambridge: Cambridge University Press. 90-117.

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*. Unpublished doctoral dissertation, Amherst, MA: University of Massachusetts.

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.

Silverman, K., Beckman, M.E., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., & Hirschberg, J. (1992). ToBI: A standard for labeling English prosody. *Proceedings of the 1992 International Conference of Spoken Language Processing*, 2, 867-870.

Snedeker, J., Trueswell, J., Gleitman, L., & Levine, L. (1999). Prosodic choice: Effects of speaker awareness and referential context. Paper presented at The University of Southern California Speech Production Conference. November 1999. Los Angeles.

Speer, S.R., Kjelgaard, M.M., & Dobroth, K.M. (1996). The influence of prosodic structure on the resolution of temporary syntactic closure ambiguities. *Journal of Psycholinguistic Research*, 25, 247-268.

Speer, S.R., Warren, P., Schafer, A.J., White, S.D., & Kneale, J. (1999). Situational constraints on the prosodic resolution of syntactic ambiguity. *Proceedings of the Fourteenth International Congress of Phonetic Sciences*. 1-7 August 1999. Berkeley, CA. 1301-1304.

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

Tanenhaus, M.K., Spivey-Knowlton, M.J., Eberhard, K.M., & Sedivy, J.C. (1995). Integration of visual and linguistic information in spoken language processing. *Science*, 268, 1632-1643.

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

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

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

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.

Appendix A: Sentence List

Driver's Set:

To tell the slider to move something:
I want to change the position of the square with the [cylinder / triangle / square with the triangle].
I want to change the position of the [cylinder / triangle / square with the triangle].

To specify which color object to move:
The [tan / blue / red / green] one. When that moves the square it should land in a good spot.
The [tan / blue / red / green] one. When that moves it should land in a good spot.

To say the slider moved correctly:
Good job.
I am able to confirm the move was the final one. The [cylinder / square / triangle / square with the triangle] has now reached its goal.
Congratulations, we have reached the end of the round.

If the slider moved the wrong way:
Whoops, go back, there's another direction I want you to go in.
Whoops, go back, there's another object I want you to move.

Slider's Set:

To ask which object to move:
Which [cylinder / triangle] do you want to change the position of the square?
Which [cylinder / triangle / square] do you want to change the position of this time?

To tell the driver about a feature:
Good choice. When that moves the [cylinder / square / triangle / square with the triangle] will encounter a cookie.
Bad luck. When that moves the [cylinder / square / triangle / square with the triangle] will encounter a ravenous goat. (Fortunately, I have a cookie to feed it.)

To tell the driver about a completed move:
I am able to confirm the move of the square with the [cylinder / triangle / square with the triangle]. It has moved [1 / 2 / 3 / 4 / 5...] space(s) [up / down / left / right].
I am able to confirm the move of the [cylinder / triangle / square with the triangle]. It has moved [1 / 2 / 3 / 4 / 5...] space(s) [up / down / left / right].

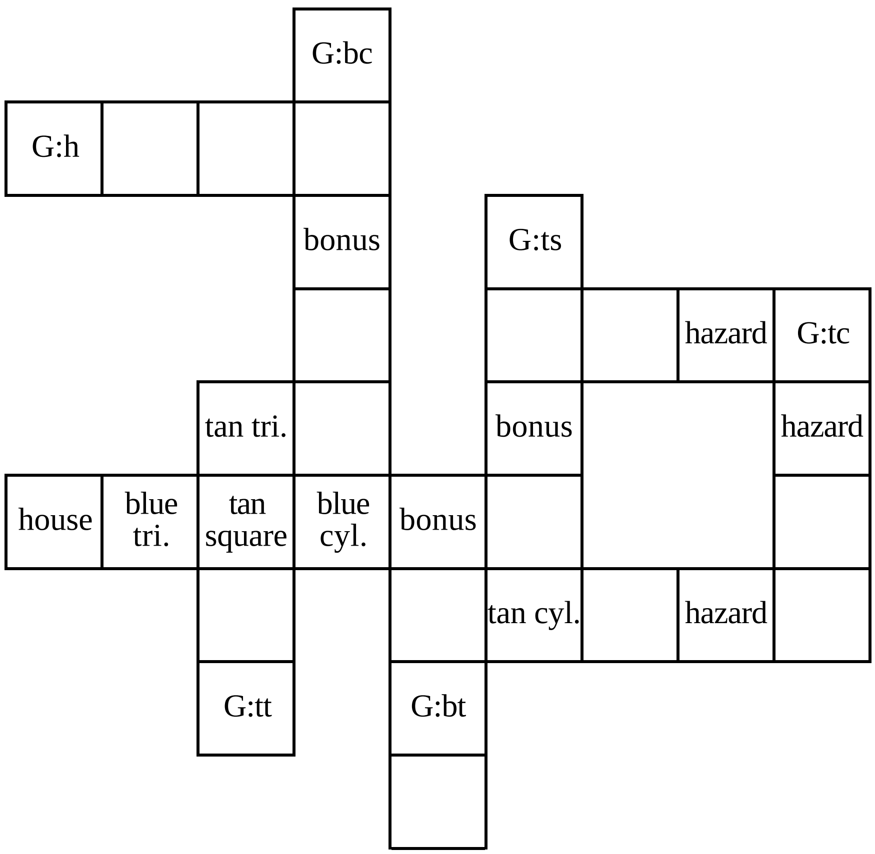
To reply to a correction on a move:
OK, it's back where it was before.

To tell the Driver a move is impossible:
I am unable to complete that move.

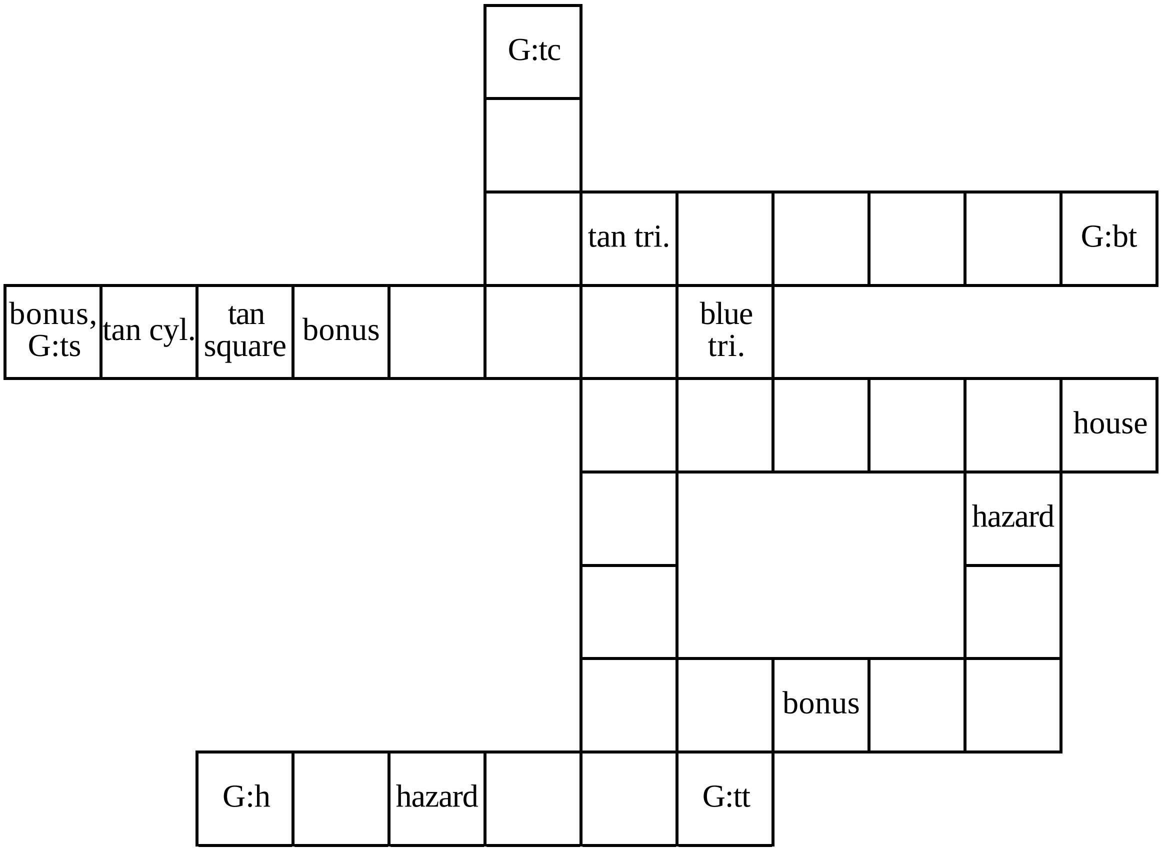
Appendix B: Initial configuration of gameboards

Key:
G:bc - goal for blue cylinder
G:bs - goal for blue square
G:bt - goal for blue triangle
G:gt - goal for green triangle
G:h - goal for combined square-and-triangle piece
G:rt - goal for red triangle
G:tc - goal for tan cylinder
G:ts - goal for tan square
G:tt - goal for tan triangle
house - start position for combined square-and-triangle piece
(name) - start position for gamepiece; constant position for bonus or hazard

Gameboard 1



Gameboard 2


G:tc
tan tri.
G:bt
bonus, G:ts
tan cyl.
tan square
bonus
blue tri.
house
hazard
bonus
G:h
hazard
G:tt

Gameboard 3

	bonus		G:ts	blue cyl.		
hazard				tan cyl.		
G:h				bonus		
tan tri.	tan square				G:tt	G:tc
house	blue tri.				G:bt	G:bc

Gameboard 4

	G:bt				G:h	bonus	
					blue square	red tri.	bonus, G:ts
	bonus, G:tt		hazard		green tri.		G:gt
	G:rt	blue tri.	tan square				
	G:bs	house	tan tri.				