# Situationally independent prosodic phrasing

SHARI R. SPEER*, PAUL WARREN**, and AMY J. SCHAFER***

*The Ohio State University
**Victoria University of Wellington
***University of Hawai'i

*Abstract*

*A series of speech production and categorization experiments demonstrates that naïve speakers and listeners reliably use correspondences between prosodic phrasing and syntactic constituent structure to resolve standing and temporary ambiguity. Materials obtained from a co-operative gameboard task show that prosodic phrasing effects (e.g., the location of the strongest break in an utterance) are independent of discourse factors that might be expected to influence the impact of syntactic ambiguity, including the availability of visual referents for the meanings of ambiguous utterances and the use of utterances as instructions versus confirmations of instructions. These effects hold across two dialects of English, spoken in the American Midwest, and New Zealand. Results from PP-attachment and verb transitivity ambiguities indicate clearly that the production of prosody-syntax correspondences is not conditional upon situational disambiguation of syntactic structure, but is rather more directly tied to grammatical constraints on the production of prosodic and syntactic form. Differences between our results and those reported elsewhere are best explained in terms of differences in task demands.*

## 1. Introduction

Prosodic structure influences linguistic analysis at a variety of levels during the production and comprehension of spoken language (see Cutler, Dahan, and van Donselaar 1997; Warren 1999; Speer, Warren, and Schafer 2003; Speer and Blodgett 2006; Wagner and Watson 2010 for reviews). These include the pre-lexical segmentation of the speech stream into words, access of word forms from the mental lexicon, the segmentation of utterances into syntactic constituents, the determination of linguistic and paralinguistic meaning and the establishment and maintenance of discourse functions (Venditti and Hirschberg 2003). In this paper we concentrate on the effect of the discourse situation on the relationship between prosodic and syntactic structures, exploring the effects of contextual constraints.

JLP 2 (2011), 35–98
DOI 10.1515/LABPHON.2011.002
1836-6346/11/01–35

We have reported preliminary results of this research elsewhere (Schafer, Speer, Warren, and White 2000; Warren, Schafer, Speer, and White 2000; Speer et al. 2003; Schafer, Speer, and Warren 2005). Here we expand on those earlier reports to explicitly compare and contrast two types of syntactic ambiguity in a matched set of speakers, present the full range of situational variation we tested, and examine productions from both American and New Zealand varieties of English. This moves research in this domain beyond the initial question of whether one factor can affect prosody to the more interesting question of how prosodic structure is generated and processed across a range of speech situations.

We see the issue of situational dependency as having significance for the types of speech production and perception/comprehension models that we might entertain. Thus, distinctly different models of production are entailed depending on whether speakers regularly adjust their productions for situational ambiguity (as opposed to merely being able to do so when unusual circumstances call for it; see Ferreira, 2007, for specific connections of different types of disambiguation to production models). Similarly, distinctly different models of comprehension are entailed depending on whether the prosodic correlates of syntactic structure are robust or vary significantly according to variation in situational ambiguity. A parallel can be seen in research on speakers' early choice of syntactic structure. Some of this research denies that speakers choose syntactic structures specifically in order to avoid ambiguity (Ferreira and Dell 2000; Arnold, Wasow, Asudeh, and Alrenga 2004), while other work suggests that aspects of lexical and syntactic choice may depend on issues such as whether the listener is known to share knowledge of the situation being described (Lockridge and Brennan 2002). The former set of findings would suggest an encapsulated model of language production, at least as far as these early syntactic choices are concerned, while the latter results imply that speakers adapt their plans to the requirements of the situation. Snedeker and Trueswell (2003) have argued for situationally driven adjustment of the role contextual constraints play in speakers' selection of prosodic structures for the resolution of syntactic ambiguity.

Under a situationally dependent production model, we would expect the factors subsumed under 'audience design' (Bell 1984; Clark and Schober 1992; Schober, Conrad, and Fricker 2000; Lockridge and Brennan 2002) to have a primary influence on speakers' prosody. A corollary of this is that speakers use appropriate prosodic cues to help listeners comprehend, and to avoid being misunderstood. That is, if the production of prosodic patterns is adjusted according to situational constraints, this implies that numerous linguistic and non-linguistic factors regularly interact in the initial construction of output representations, and that speakers continuously evaluate the potential for linguistic ambiguity, as well as the plausibility of alternative interpretations, given the discourse situation. In this case, prosodic cues to syntactic structure will be most likely in situations perceived as ambiguous, and less likely in situations where other sources of disambiguation are available, as with temporary syntactic ambiguities.

A situationally dependent model of comprehension predicts that listeners will use prosodic marking of syntax in the same measure as speakers produce it. If speakers only produce prosodic disambiguation when they perceive syntactic ambiguity for the listener, such cues might be relatively infrequent and inconsistent, and listeners' use of them therefore also relatively uncommon. When other sources of disambiguation are available, prosody might be expected to play a relatively minor role in the comprehension process. Listeners might also behave differently when responding to carefully recorded laboratory speech (i.e., speech from a situation where it can be assumed that there is little or no ambiguity for the speaker) as compared to more spontaneous speech situations. Listeners might "tune" their reliance on prosody, increasing it when they realize a situation is ambiguous.

In contrast, a situationally independent model of prosody in production predicts that grammatical correspondences among linguistic structures determine linguistic output, and thus the prosody produced by speakers. In such a model, multiple grammatical factors including phrase structure, information status, focus, phonological length, and so on could directly affect the prosodic realization of utterances, consistent with models linking syntactic and prosodic constituency in speakers' behavior (e.g., Watson and Gibson 2004, 2005). However, situational ambiguity would only indirectly affect speakers' prosodic choices. This might occur, for instance, when attention is drawn to situational ambiguity (as when speakers are given explicit instructions to disambiguate). Such conscious efforts to disambiguate may or may not be successful, depending on the speakers' awareness of a felicitous prosody to assign to a given intended syntactic structure.

A situationally independent comprehension model would be sensitive to common and predictable grammatical correspondences between incoming prosody and other linguistic structures. This is because if prosodic cues to syntactic structure are common and reliable, they might be expected to play a more central role in comprehension. Much recent work in the language processing literature has indeed shown that listeners' recognition of syntactic structure is predictably conditioned by prosodic phrasing (Carlson, Clifton, and Frazier 2001; Clifton, Carlson, and Frazier 2002; Watson, Breen, and Gibson 2006). Under situational independence, we would expect listeners to behave similarly in laboratory speech and spontaneous speech situations. The prosodic structure produced for a particular utterance, then, would be recognized and processed as a fundamental component of the linguistic meaning available to the listener.

Situational dependence and situational independence are two logical extremes for the production system. It is possible that speakers might show situational dependence for some areas of language and independence for others (Ferreira 2007; Jaeger 2010), or that they might follow a sort of weak dependence: Their prosody would be primarily determined by grammatical factors, but would show sensitivity to situational demands when processing load was low enough to free up resources for tracking situational ambiguity. Our goal here is to explore the extent to which we see evidence for situational dependence, situational independence, or an intermediate position across multiple manipulations of situational ambiguity and in a task that is reasonably reflective of natural speech situations.

Possible support for situational dependency comes from production studies that show reliable disambiguation only in ambiguous situations, suggesting that untrained speakers do not reliably disambiguate syntactic phrase structure with prosody when other information in the discourse can be used to resolve the ambiguity (Allbritton, McKoon, and Ratcliff 1996; Straub 1997; Fox Tree and Meijer 2000; Snedeker and Trueswell 2003). Thus, Allbritton et al. (1996) argued that a speaker's awareness of ambiguity is a primary factor that influences the salience of prosodic contrasts in speakers' production of ambiguous sentence materials. They compared naïve speakers' productions of ambiguous test sentences presented in two types of context. When the sentences were preceded by disambiguating paragraph contexts so that readers were potentially unaware of any ambiguity, there was little prosodic marking of disambiguation. In contrast, when isolated sentences were presented with explicit instructions to pronounce them with one of two meanings, speakers provided prosodic disambiguation. Such evidence suggests that when laboratory speech materials spoken to include disambiguating prosody are used in perception experiments, results may not generalize to other conversational settings, such as everyday dialogue. That is, if accurate, such evidence suggests that prosody may be a less important factor than has been implied by most studies showing prosodic effects on comprehension. We note, though, that many perception and comprehension studies in the phonetic literature routinely base the prosody of their experimental utterances on the results of companion production studies using untrained and naïve speakers (e.g., Warren, Grabe, and Nolan 1995).

Unfortunately, the lack of consistent prosodic disambiguation in the studies mentioned above does not necessarily generalize to typical conversational speech any more than the results from laboratory speech in other experiments. The failure to find consistent disambiguation may be due to the use of production tasks focussed on text, with informants reading aloud, or repeating back a memorized text sentence. Text was presented in short paragraph contexts, or as an instruction to be repeated to a listener who did not give a spoken response. Allbritton et al. acknowledged the significance of this factor, though they were unable to investigate it, since controlled contrasting materials are unlikely to arise in spontaneous speech. There are a number of concerns regarding basing a model of the human language processing mechanism solely on data collected in reading tasks. For one, readers and speakers have different pragmatic goals. Spontaneous speech is generally produced in a contextually appropriate manner in order to achieve a speaker's communicative goals. In contrast, people instructed to repeat or read aloud a provided sentence, even with instructions to speak clearly, are more likely to focus on clearly pronouncing the lexical items than on conveying sentence-level meaning, which can be presumed to be already known to the listener, who is often the one who provided the sentence (i.e., the experimenter). Also, 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 versus spontaneous speech. For example, read speech has fewer and shorter pauses, and fewer prosodic phrases for utterances of comparable word length (Howell and Kadi-Hanifi 1991; Ayers 1994; Blaauw 1994). 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. In addition, disambiguating contexts in some previous studies may not have been strong enough to establish the correct syntactic structure in the "disambiguated" condition, and thus the syntactic attachments may have been inconsistent across tokens.[1]

A further concern with reading tasks is that the disambiguating contexts that are provided for readers may have unintended effects on the prosodic structure of the target sentence. For instance, disambiguating contexts may induce information structures that interfere with prosodic disambiguation of syntax. In a controlled test of this possibility, Schafer and Jun (2001) found that the placement of informational focus at an early point in an utterance resulted in the prosodic reduction of a later part of the utterance, and consequently in a reduction of the durational marking of a syntactic contrast in that part of the utterance (a contrast involving prepositional phrase ambiguities similar to those discussed in our Experiments 1–3 below). Ferreira (1993) showed a related finding: Informational focus can increase the prosodic boundaries in the focus region.

For these reasons, studies of the production of read speech may provide unreliable estimates of the location, size, and frequency of the prosodic disambiguation of syntactic structure that occurs in spontaneous speech. 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. In a study using a task somewhat similar to our own described below, and developed independently of it, Snedeker and Trueswell (2003) had pairs of naïve participants complete a task in which speakers produced structurally ambiguous sentences as instructions to a listener for the movement of objects. Participants could see neither each other nor each other's array of objects. The instructions were read silently from a card shown to the speaker by the experimenter, who also demonstrated the action requested by the instruction, and then spoken aloud by the participants after a short delay and without further reference to the card. Across two experiments, the utterances were made in two types of context. In one context the sentence was potentially ambiguous both for the speaker and for the listener, while in the other the arrangement of objects was such that the sentence was unambiguous for the speaker, but (unknown to the speaker) still ambiguous for the listener. Snedeker and Trueswell present phonetic and phonological analyses of the utterances produced by the speakers, as well as measures of the listeners' responses to the instructions. (See also Snedeker and Casserly, 2010, for further analyses of these materials.) They found good disambiguation of the structures in the first type of context but not in the second; i.e., their naïve speakers failed to provide reliable disambiguation when the utterances they were required to produce were unambiguous in the context of the situational information available to them. We will return to this study in later discussion, since there are clear differences between Snedeker and Trueswell's data and our own findings presented below, as well as important differences between the tasks.

Kraljic and Brennan (2005) also used a task in which speakers instructed listeners to move objects. Speakers were not shown text versions of the instructions, but instead were cued by a schematic drawing indicating which object had to be moved and to where. As in other studies, changes in the object arrays created contexts that made the potential utterances either ambiguous or situationally unambiguous. Kraljic and Brennan's 24 speakers in their Experiment 1 took part in a total of 480 critical trials, and produced 255 utterances that had the required syntactic and lexical structure (i.e., were potential ambiguities of the type being investigated). In contrast to the Snedeker and Trueswell findings, the analysis of both production data and listeners' actions in response to the instructions indicated that the utterances were pronounced in a way that resolved syntactic ambiguity, and that this disambiguation was not modified by the extent of situational ambiguity. This was replicated in their Experiment 3, which provided multiple tests of whether speakers provide more prosodic disambiguation when they are more aware of a syntactic ambiguity or there is arguably more need for disambiguation for the benefit of a listener. Throughout their data, Kraljic and Brennan found situational independence.

There are thus conflicting reports on the issue of how situational constraints on syntactic ambiguity affect the production of potentially disambiguating prosodic structures. In the current paper we present data from a series of studies using the cooperative game task developed in our research program (see for instance Schafer et al. 2000; Warren et al. 2000; Speer et al. 2003; Schafer et al. 2005). Our task design had a number of key goals. One was to collect recordings of syntactic ambiguities in a setting that was as close as possible to conversational speech, i.e., in which speaker/listeners who are engaged in an interactive task make important choices concerning the sequencing and content of utterances for the purpose of a non-linguistic goal, and in which they exchange responses that can implicitly indicate the communicative success of the utterances they have produced. Thus our task differs significantly from those used by Snedeker and Trueswell or by Kraljic and Brennan, who had traditional trial structures. Another was to allow variation in the situational constraints on the ambiguities being expressed, through the natural development of the discourse as participants moved toward a solution of the game, and using a richer range of situational constraints than in previous research. For example, our gameboard configuration contrast was unlike those tested before, and the length of our game task allowed detailed investigation of changes across the timecourse of our speakers' interactive discourse. A third was to assess syntactic disambiguation in contrasting pairs of structures. Such contrasts have not previously been explored from the perspective of a situational need to disambiguate. A fourth was to compare and contrast prosodic disambiguation across language varieties, as this allows us to hold constant the lexical-syntactic structures and need to disambiguate, and yet potentially see the effects of differing grammatical settings.

Our studies therefore involved ambiguous sentences that were carefully controlled for their syntactic properties and lexical content, and yet which were produced in a relatively spontaneous manner and in a situationally appropriate discourse context, by speakers of two distinct varieties of English. Our expectation was that speakers in the task would produce speech that is more spontaneous than that found in typical psycholinguistic laboratory production tasks, where sentences are presented as separate events in a list (with or without text prompts), and any visual referents change with each consecutive trial. Here, we focus on the effect of different types of situational constraints that resulted from the changing contexts of the task and from two different syntactic contrasts. During the course of the game, the availability of certain moves and gamepieces inevitably changed, making certain interpretations of syntactically ambiguous sentences more or less likely. This allowed us to investigate whether prosodic cues to syntactic structure varied as a function of the situational constraints on certain interpretations. We examined the prosodic realization of global prepositional phrase (PP) attachment ambiguities, as in (1), and of a temporary closure ambiguity, shown in (2), produced by naïve speakers using utterances that were situationally motivated.[2] Because global ambiguities like (1) do not contain enough morphosyntactic information within the sentence to fully specify the intended meaning, prosody is fundamental to their disambiguation. In contrast, temporary syntactic ambiguities are resolved by the presence of disambiguating material within the sentence (in (2), the disambiguating word immediately follows the ambiguously attached noun phrase). Thus, we expected the comparison of prosodic disambiguation in (1) vs. (2) to be particularly informative. The inclusion of these two syntactic contrasts allows evaluation of whether prosodic disambiguation is greater in type (1), as a situationally dependent model should predict for a global bracketing ambiguity vs. a temporary ambiguity, or in type (2), as a situationally independent model should predict given current linguistic assumptions for prosody-syntax mappings (e.g., Selkirk 1984, 1994, 2000; Nespor and Vogel 1986; Truckenbrodt 1995, 1999).[3] It also allows comparison of these results with those from previous studies that have compared these two types of structures, primarily with laboratory speech productions (Lehiste 1973; Warren 1985; Price, Ostendorf, Shattuck-Hufnagel, and Fong 1991).

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

(2) When that moves the square {will encounter a cookie / it should land in a good spot}.

Experiment 1 evaluates the effect of contextual constraints on the prosodic realizations of the PP attachment ambiguity in (1). Experiment 2 investigates whether naïve listeners can reliably categorize examples of (1), recorded in Experiment 1 but then isolated from the context of the game. Experiment 3 uses a between-subjects design, with new participants, to address a further issue arising out of differences between our Experiment 1 and Snedeker and Trueswell's (2003) study, namely the influence on prosodic disambiguation of the speaker's awareness that there are two meanings to the utterance in (1). Experiment 4 introduces production data from the temporary ambiguity in (2) (recorded in the same sessions of the gameboard task as the data analyzed in Experiment 1). Experiment 5 presents categorization data for the temporary ambiguities collected in Experiment 4, determining whether a new set of listeners can identify the speakers' intent without recourse to morphosyntactic resolution.

Our initial expectation for global ambiguities in Experiment 1 was that both prosody-syntax correspondences and properties of the developing situational context of the game would be particularly important.

## 2. Experiment 1. Production study of prepositional phrase ambiguities

The PP attachment sentences analyzed in this experiment are given in (1) above, repeated as (3), plus sentences (4)–(6). On one reading of (3) (low, NP attachment of the PP *with the triangle* as a modifier of *square*), the move involves a combined square-and-triangle piece. (Gamepieces and boards are illustrated in Appendix A.[4]) On the other reading (high, VP attachment), a triangle-shaped gamepiece is used as an instrument to move, by pushing, a square piece. This interpretation is compatible with a rule in our game that the square must be moved by another object, which could be a triangle, a cylinder, or a combined square-and-triangle piece. This high attachment is in fact the only reading of another utterance in our materials, given in (4). There is no combined square-and-cylinder piece, so the low attachment is infelicitous in the context of the game. The two players in each game, Driver and Slider, had different roles in the game, explained below. Slider utterances in (5) and (6) have properties that parallel those described for (3) and (4), although in this case the high attachment is to the noun phrase headed by *move* instead of to a verb phase. The complete list of lines available to players is given in Appendix B. Unlike most sentence production experiments, which have multiple items, the game design limited our PP analysis to just the carefully matched sentences given in (3)–(6), but allowed for multiple productions of these sentences from each participant.

Driver utterances:

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

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

<u>Slider utterances:</u>

(5) I am able to confirm the move of the square with the triangle.

(6) I am able to confirm the move of the square with the cylinder.

Previous production studies of PP attachment ambiguities using professional or instructed naïve readers have revealed small but consistent phonetic differences between the two interpretations. These include longer and more frequent silent durations before the preposition in the high attachment cases, as well as longer durations of the immediately preceding word (Cooper and Paccia-Cooper 1980; Price et al. 1991). This pattern reflects a correspondence between the right edge of a syntactic phrase boundary and the right edge of a prosodic boundary immediately preceding the preposition in the high attachment case, versus the lack of a syntactic phrase boundary at that position in low attachment (e.g., Selkirk 2000). In contrast to many (but not all) previous studies, the current experiment uses naïve speakers who are not simply reading sentences, but using PP structures in an interactive and cooperative task. Our initial prediction was that phonological and phonetic analyses of utterances such as those in (3) to (6) would show a difference in the realizations of high and low attachments, with a stronger prosodic break before the PP in high-attached utterances.

The design of the gameboard task included several situational factors that might influence the production of prosodic disambiguation. These manipulations will be outlined briefly here, and expanded in the results section below.

## 2.1. *Design*

### 2.1.1. *Speaker role contrast*

Drivers issued instructions such as (3) and (4), while Sliders followed these instructions and confirmed moves using (5) and (6). Thus disambiguation was more pragmatically important in Driver utterances, where an unclear instruction might lead to an incorrect move. If speakers are influenced by such pragmatic factors, Driver utterances should show a clearer prosodic distinction between high and low attachment than Slider utterances.

### 2.1.2. *Sequential position*

Previous studies of interacting speakers (Clark and Schober 1992) have shown that changes occur in production over the course of a task. In our study, participants played several rounds of the game, and became increasingly familiar with the utterances that made up the instruction and response set. Heightened awareness of the contrast between high and low attachments might influence their realizations of the ambiguous PP utterances as they continue to play. Two situational factors that might contribute to this are that the first production of each attachment occurred in a configuration that contained no situational ambiguity, and that Drivers received immediate feedback from Sliders about whether their utterances had been correctly interpreted.

2.1.3. *English variety* Our gameboard task was completed by participants from two English-speaking regions, the Midwestern United States and New Zealand. Including two geographically distinct varieties allows us to test the generalizability of our findings. Situationally dependent models would be especially supported if we see evidence of ambiguity avoidance across a range of grammatical settings. Situationally independent models would be especially supported if we see variation in prosody-syntax mappings between varieties but not across situations. If the two varieties behave quite similarly, the increased power and control for idiosyncrasies of one lab or speech community lends stronger support to whichever model of situational factors is best supported.

2.1.4. *Gameboard configuration contrast* Multiple sources of information were potentially available to influence the resolution of the PP ambiguities (3) and (5), including preceding linguistic information like that in the studies by Straub (1997) and Allbritton et al. (1996). For example, because players completed moves of particular pieces in series, a player who had just referred to the combined square-and-triangle gamepiece was biased to expect the next utterance to refer to it again, unless this piece had just been placed in its goal. In other situations, the PP utterance was situationally unambiguous, for example when there was no triangle piece in a position to push a square piece.

2.1.5. *Gamepiece contrast* As the examples in (3) to (6) illustrate, our study contrasted *cylinder* sentences with *triangle* sentences. Importantly, the *cylinder* sentences were never ambiguous in the situation of the game, since there was no combined square-and-cylinder piece. Comparing the production of the unambiguous utterances in (4) and (6) with those of the two readings of the ambiguous utterances in (3) and (5) allows a test of whether the presence of visual evidence for two interpretations of an utterance has an impact on the production of prosodic disambiguation. This contrast is similar to the contrast between ambiguous and unambiguous situations in Snedeker and Trueswell (2003), but here it is carried out in a within-subjects design.

The gameboard and gamepiece contrasts involve visual information in the common ground. That is, 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 and Keysar 1996), at least early in the game.

## 2.2. *Method*

2.2.1. *Participants* Our analyses are based on data from 15 native speakers of Midwestern American English (AmE) from the University of Kansas and 15 native speakers of New Zealand English (NZE) from Victoria University of Wellington. Eight pairs of speakers from each variety were initially recorded, but recordings from one AmE participant were not analysed because she produced few well-formed utterances and many lexical and segmental errors, and recordings from one NZE participant were excluded because a post-experiment questionnaire revealed that he was not a native speaker of that variety.[5]

2.2.2. *Design and materials* Driver-Slider participant pairs used partly scripted sentences to negotiate moves of gamepieces from starting positions to goals on a gameboard. 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. Gameboards differed slightly according to participants' roles, with the Driver's board showing locations of goals for the gamepieces, but not of bonuses (cookies) and hazards (ravenous goats). Conversely, the Slider's board showed bonuses and hazards but not goals. All remaining information (gamepieces on boards; locations of gamepieces, etc.) was identical.

The Driver's role was to tell the Slider which piece to move, to inform the Slider of incorrect moves, and to confirm that a gamepiece has reached its goal. The Slider's role was to choose directions to move in and to report moves back to the Driver, but also to ask the Driver for more information when necessary. Neither player could see the other's board. Board design and game rules encouraged negotiation and the strategic use of moves. In addition, board layouts were designed to allow varying levels of situational constraints on possible moves; i.e., at certain points in the game only a subset of the gamepieces could be or needed to be moved. Three pairs of gameboards with differing layouts were used, plus a pair of practice boards and a demonstration board. All gameboards are available for download from the online Appendices.

### 2.2.3. *Procedures*

2.2.3.1. *Production task* Participants sat at tables separated by a divider in a sound-treated room or booth, accompanied by an experimenter who supervised the game, offering advice when necessary and announcing the award or deduction of points. Two additional experimenters monitored the course of game play for each participant, logged the sequence of utterances and moves being made by the participants, and monitored the recordings. The players' conversation was restricted to a set of provided sentence frames and gamepiece names. They were free to choose the sequence of moves to be followed, and thus had to choose which sentences to use and when to use them. Speakers quickly became familiar with the utterances available for use under the rules of the game, and learned to produce them fluently and eventually without recourse to printed lists of frames. In order not to bias the participants' choice of prosodic structure, the experimenters were careful to avoid using the scripted sentences or piece names themselves. The full list of sentence frames is given in Appendix B. Participants were not made explicitly aware of the syntactic ambiguities in the game sentences, and were never told to use disambiguating pronunciations.

Participants wore head-mounted microphones, and utterances were recorded to computer disk. All participant pairs completed a practice board and then played at least two more boards twice each, swapping roles. Players played for two hours, completing as many games as they could. Up to six games were played, involving three different boards. Because the participants had great freedom in choosing which pieces they wanted to move when, and in which direction, the number of critical tokens varied across participants. The constraints of the game nevertheless ensured that the critical sentences were mixed with the other available sentence frames.

2.2.3.2. *Phonological and phonetic analyses* Our analysis of production materials involves both phonological and phonetic measures of prosodic disambiguation. Production studies of prosodic marking of ambiguity resolution frequently report phonetic parameters such as duration and pitch, treating these as separate though related parameters of speakers' disambiguation. However, a more appropriate measure is likely to result from a phonological prosodic analysis, since speakers will differ from one another (and from themselves in separate recordings) in how they use different acoustic cues to prosodic structure (Cutler and Isard 1980; Henderson 1980; Beach 1991; Straub 1997).

2.2.3.3. *Phonological analyses* We assume the intonational theory of Pierrehumbert and Beckman (Pierrehumbert 1980; Beckman and Pierrehumbert 1986) and followed the conventions of the English ToBI annotation system (Silverman et al. 1992; Beckman and Ayers 1997). Analyses presented here focus on prosodic phrasing. In the ToBI system, each utterance has one or more intonation phrases (IPs), the end of which is marked by a boundary tone (which can be H(igh) or L(ow)). Each IP contains one or more intermediate phrases (ip) whose end is marked by a (H or L) phrase accent. The analysis also includes break indices, marking the degree of disjuncture between constituents. These range from 0, indicating that some connected speech process has taken place across a word boundary, through 1 for a 'default' break between two words within a phrase, to 3 and 4 for ip and IP boundaries, respectively. One of our primary measures, the location of the strongest utterance-internal prosodic break, is derived from these break indices. This measure is based on the simple notion that relative break sizes are important in the ascription of syntactic structure (see discussion of the comprehension literature on this in Experiment 2 below, and Snedeker & Casserly, 2010, for discussion of relative versus absolute break sizes).

The phonological analyses for the PP utterances are based on independent transcriptions carried out by at least three expert transcribers (the authors), with some including two further experts. An examination of transcription reliability for our AmE data (following the procedures outlined in Pitrelli, Beckman, and Hirschberg 1994) showed over 90% agreement among the panel of five transcribers on the locations of pitch accents, phrase accents, and boundary tones, with over 94% agreement for the smaller panel of three transcribers. Investigation of the type of tone transcribed at each position by the larger panel shows agreement ranging from 60% for pitch accents to 83% agreement for the type of boundary tone. If the difference between down-stepped and non-down-stepped accents is discounted, then these figures rise to a range from 72% for pitch accents to 84% for boundary tones. These agreement levels are comparable with those given by Pitrelli et al. for a much larger comparison set. Cases of disagreement were resolved through further discussion among the transcribers or by selecting the majority analysis.

Phonological transcriptions of the PP utterances showed a high degree of variation in the prosodic realization of the sequence *the position of the square*, within and across speakers. For instance, for 13 AmE speakers in the Driver role we obtained a total of 62 different patterns on 78 high-attached utterances, and 87 patterns on 101 low-attached utterances. It is clear that the precise prosodic form of an utterance, including accent type and location, cannot be predicted simply on the basis of morphosyntactic structure. However, we expected there to be some constraints on the patterning of prosodic structure, possibly at a more abstract level such as the relative sizes of breaks through the utterance. This would be consistent with comprehension data from Schafer (1997) and Carlson, Clifton, and Frazier (2001) showing that disambiguation of PP attachments was influenced not only by a boundary immediately prior to the PP but also by the pattern of boundary strengths in the preceding material. Thus recordings of (3) were expected to show the strongest boundary at the end of *square* when the speaker intended high attachment. Our phonological analysis accordingly considered the position of the strongest utterance-internal prosodic break (referred to henceforward as Strongest Break Location or SBL). Driver utterances were categorized into those that had the strongest break immediately before the PP (i.e., at *square*), those with the strongest break elsewhere in the utterance, and those where there was a tie for the SBL between *square* and another location. These three SBL categories were labelled ‘square’, ‘notsquare’, and ‘equal’, respectively.[6]

2.2.3.4. *Phonetic analyses* Because of the substantial variation in tonal patterns exhibited by our speakers, we limit our reported phonetic analyses to durational measurements of critical regions. Among other cues, prosodic phrase boundaries are typically marked by lengthening of the final syllable of the prosodic phrase and a following silent interval (Wightman, Shattuck-Hufnagel, Ostendorf, and Price

1992; Ferreira 1993). In general, IP boundaries show more extreme phonetic effects than ip boundaries, corresponding to the higher break index associated with the IP. The phonetic measures for the PP attachment utterances include the duration of each of the words and silent durations (if any) in the sequence *position of the square with the triangle*.

### 2.3. *Predictions*

If there is a syntactic influence on prosodic realization, then the prosodic boundary between *square* and the following PP should generally be greater in the high attachment than in the low attachment utterances. The situationally independent model predicts that this difference should be present regardless of the situational need to disambiguate. In contrast, the situationally dependent model predicts, in its strongest form, that the difference should be significant when there is a pragmatic need to disambiguate but that it should disappear when there is no need to disambiguate. Because of the number of comparisons presented in the results section, more specific predictions are presented with each analysis.

### 2.4. *Statistical analysis*

Utterances were excluded as disfluent if they met one or more of the following criteria: mispronunciation, substitution, addition or deletion of words, word-internal pausing, re-start, presence of a filled pause, and presence of non-speech events (sniffing, coughing, laughing) in the critical region. The number of included utterances varied by participant, partly because of differences in fluency but importantly also because participants followed differing strategies in moving objects to their goals. In the following sections, individual participants are generally excluded if they failed to produce more than two fluent utterances in all of the conditions relevant to that section. In a few analyses, however, this criterion is relaxed because of low utterance counts.

Our statistical analysis uses linear mixed effects regression through the *lmer* function in R (Baayen 2008; Baayen, Davidson, and Bates 2008). This approach allows us to include individual data points and is thus more inclusive than analysis of variance based on participant averages.[7] It also allows inclusion of a number of predictors that would not easily be included in analyses based on participant averages, given the paucity of data points in some condition combinations. Following model comparison, all analyses reported throughout this paper include participants as a random effect, with intercepts varying by participant. Random slopes are also included where these provided a better statistical model. For example, use of random slopes allows the model to fit the relationship between high and low attachment durations for each participant, often (but not always) providing a better fit when attachment was the main factor of interest in an analysis. Where necessary, predictor variables are either centered or contrast-coded to remove collinearity between predictors. Where the dependent variable was a durational measure, the durations were subjected to an inverse transform, since preliminary analysis over the complete dataset showed that this transform produced a more normal distribution of durations (see also Baayen and Milin in press). The transform was −1000/duration. A negative numerator was used so that positive analysis coefficients correspond to increases in duration. Since the t-values output by *lmer* tend to be anti-conservative, our measure of statistical significance is based on Markov chain Monte Carlo (MCMC) sampling (10,000 iterations). In our presentation of significant results we include the mean and lower and upper limits of the 95% interval for the Higher Posterior Density and the probability levels generated by the MCMC sampling. For ease of interpretation, we present bar-charts of means and standard errors in the relevant conditions for our duration results, rather than *lmer* model estimates based on transformed data.

Analyses of phonetic data included the duration of the word *square*, any silence that followed that word and preceded the following PP, and the combined duration of *square*+silence. Since the results for the first two measures showed essentially the same pattern as those for the *square*+silence duration, we present data for the latter only.

For the phonological analyses we constructed two *lmer* models for each set of contrasts, treating SBL as a binomial variable.[8] Phonological analysis 1 followed a conservative approach to the 'equal' SBL tokens, treating these as ambiguous and therefore excluding them from an analysis that included only the 'square' tokens (predicted to align with high-attached PP structures) and the 'not square' tokens (predicted to align with low-attached PP structures). Phonological analysis 2 included the 'equal' tokens, but treated them as equivalent to the 'square' tokens, reflecting the findings of Pynte and Prieur (1996) and Snedeker and Casserly (2010) that the presence of a phrase break immediately preceding a PP can restore high-attachment preferences even when there is an earlier phrase break.

## 2.5. *Results*

A situationally dependent model of the relationship between prosodic and syntactic structure predicts that the strength of the prosodic marking of high vs. low PP attachment will vary according to contextual factors. An Attachment factor will thus enter into significant interactions with such factors, which here include the speaker Role (Driver vs. Slider), the Sequential Position of the utterance in the experimental session, the English Variety spoken by the participant, the Ambiguity level determined by the gameboard configuration, and the Gamepiece contrast. The contrasting prediction of a situationally independent model is that the prosodic marking of the attachment level will not be affected by these factors, and that the simple effect of Attachment will be consistently marked.

Despite the benefits of greater inclusion of data using a mixed effects regression model, gaps in our data distribution meant that we could not reliably test all six fixed effects (Attachment plus the five contextual factors) in a single analysis. We therefore start with the most completely representative dataset, examining the effects of Role, Sequential Position, and Variety, before proceeding to separate analyses for Ambiguity and Gamepiece. This first analysis is based on fluent *triangle* utterances from 25 participants – 13 AmE and 12 NZE – who had at least two fluent utterances in each relevant cell. Our *lmer* models tested for the basic effect of Attachment, for possible interactions of Attachment with Role, Sequential position, and Variety and for simple effects of each of these contextual factors. (See Tables A1–A3 for the full outputs of these models.)

2.5.1. *Attachment* The difference between high and low attachment was a significant factor in the analyses of phonetic and phonological measures. The phonetic data showed longer durations of *square*+silence in high attached utterances, in line with previous findings for PP ambiguities ($MCMC_{mean} = -1.277$, $HPD95_{lower} = -1.451$, $HPD95_{upper} = -1.093$, $p < 0.0001$). Both phonological analyses indicated that an SBL of 'square' was more likely for high attachment (analysis 1: $\beta = -5.225$, $SE = 0.581$, $z = -8.995$, $p < 0.0001$; analysis 2: $\beta = -3.637$, $SE = 0.377$, $z = -9.650$, $p < 0.0001$).

2.5.2. *Participant role* On pragmatic grounds, situational dependence predicts greater differences between high- and low-attached utterances for Drivers, whose role was to convey instructions, than for Sliders, who confirmed that the instruction, now in common ground, had been followed. An interaction of Attachment with Role was found only in the first phonological analysis, excluding the 'equal' cases ($\beta = 1.468$, $SE = 0.766$, $z = 1.918$, $p < 0.06$). This is the only indication that participant Role affected the prosodic marking of syntactic attachment, with Sliders slightly less likely than Drivers to produce major breaks at 'square' for high-attached tokens. This interaction is absent from the phonetic analysis and from phonological analysis 2, when the 'equal' SBL is included as a variant of the 'square' SBL. This suggests that there may be a subtle difference between Sliders and Drivers in the detailed nature of the prosodic patterns used to mark high attachment, i.e., Drivers are more likely than Sliders to use a break at 'square' alone, while Sliders are more likely to allow the combination of such a break with an earlier break. On the whole, though, there is little convincing evidence that Sliders mark the syntactic attachment of the PP less clearly than Drivers.

Interestingly, Role shows a simple main effect in each analysis (significant in the analysis of phonetic data: $MCMC_{mean} = -0.344$, $HPD95_{lower} = -0.446$, $HPD95_{upper} = -0.237$, $p < 0.0001$, and in phonological analysis 2: $\beta = -0.629$, $SE = 0.232$, $z = -2.712$, $p < 0.01$; marginal in phonological analysis 1: $\beta = -0.733$, $SE = 0.436$, $z = -1.681$, $p < 0.10$; analysis 2). *square*+silence durations were shorter for Slider utterances, which also had fewer main breaks transcribed at 'square'. This effect of speaker Role may reflect differences in the care with which speakers articulated utterances as Drivers and Sliders, along the lines of differ-

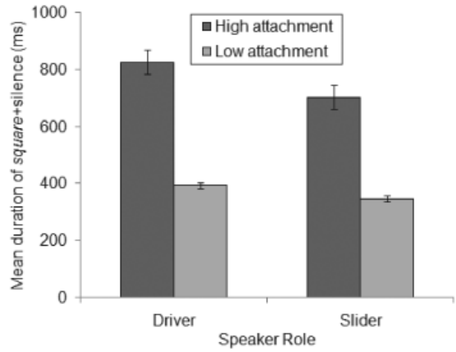


Figure 1. *Mean* square+*silence durations (with standard error bars) for Driver and Slider utterances of high and low PP attachments in Experiment 1; based on 139, 197, 111, and 200 tokens for Driver high and low attachments and Slider high and low attachments respectively, from a total of 25 speakers (13 AmE and 12 NZE with at least 2 fluent utterances in each condition; data in the columns, from left-to-right, represent 90%, 81%, 88%, and 88% of the total count of fluent utterances recorded in these conditions for the complete set of 30 speakers).*

ences in rate and clarity of diction associated with the production of new and given information (Fowler and Housum 1987). The overall averages for *square*+silence durations broken down by Attachment and Role in Figure 1 show the clear main effects and the absence of an interaction.

2.5.3. *Sequential position* We included Sequential Position as a factor to examine whether participants' increasing familiarity with the materials and potential sensitivity to the ambiguity may have affected the extent to which they used prosody to disambiguate. This should result in a clear interaction of Attachment with Sequential Position. However, the absence of any such interaction in the *lmer* analyses for phonetic and phonological data suggests that our participants maintained an equivalent distinction between the two types of attachment even if they did become increasingly aware of the ambiguity.

The simple effect of Sequential Position was significant in the analysis of the phonetic data ($MCMC_{mean} = -0.299$, $HPD95_{lower} = -0.469$, $HPD95_{upper} = -0.116$, $p < 0.01$), and in the more inclusive phonological analysis 2, where 'equal' tokens were included as a variant of 'square' ($\beta = -0.846$, $SE = 0.398$, $z = -2.126$, $p < 0.05$). Later utterances had shorter *square*+silence durations, and were less likely to be analysed phonologically as having the major break at 'square' (i.e., 'square' and 'equal'). The shorter durations may reflect a general speeding up of utterances across the experimental sessions as participants became more familiar with the utterances.

One way in which increasing familiarity may have affected production is through the lexicalization of the phrase *the square with the triangle* as a relatively unanalyzed label for the combined square-and-triangle piece in the low attachment cases.[9] Clearly this might affect the pronunciation of the low attachment reading and therefore the distinction between high and low attachment (for general comments on the effects of lexicalization on prosody see Liberman and Sproat 1992). To investigate this, we carried out a separate analysis of the break index strengths in the low attachment cases, for the AmE data (since more complete prosodic analyses are available for those cases). The average break index at *square* decreased across the repetitions of the low-attached utterances, from 1.85 on the initial utterances to 1.38 on the final utterances, and the average break index across the PP as a whole decreased from 1.20 to 1.09. Both of these are larger changes than the overall reduction across the utterance prior to this phrase, from 1.30 to 1.25. None of these trends are statistically significant, yet the changes at *square*, together with the trend towards fewer strongest breaks at *square*, may indicate an increase in the internal coherence of the phrase *the square with the triangle* across the experimental session, and increasing lexicalization of that phrase. A further consideration, however, is that the low-attached PPs, unlike the high-attached PPs, frequently occurred in sequences of repetitions, as the players moved a combined square-and-triangle piece through a series of positions toward its goal. These repeated sequences tended to occur toward the end of the game sequence and had lower break indices at *square* than non-repeated items. In fact, if repeated items are excluded, then the change in the average break index at *square* in low-attached items from the beginning to the end of experimental sessions is not as dramatic, falling from 1.85 to 1.69. Thus, it appears that changes in boundary strength over the game sequence were primarily cases of reduction for adjacent sequences of low-attached utterances, combined with a general reduction in duration and boundary strength as play continued.

It has been suggested that lexicalization of *the square with the triangle* might also result in a shift in the stress pattern over the phrase (Snedeker and Trueswell 2003: 126). In the case of our NP-PP structure, it is not clear whether lexicalization would result in early stress (as implied by the example of *BLACKbird* vs. *black BIRD* cited by Snedeker and Trueswell) or late stress (which is our prediction for the other example they cite, *The Cat in the Hat*, and might also be predicted for *the square with the triangle* if this is lexicalized with a meaning that distinguishes it from another object in the game, i.e., the plain square with no attached triangle). However, there is no clear evidence from our transcriptions of a change in stress patterns over the course of the experiment; in the instances of *the square with the triangle* that are pronounced as a single intermediate phrase (which we take to be a prerequisite for a single lexical item), there is no clear change in the pattern of accents on *square* and *triangle* across the experiment.

Whether or not there is increasing lexicalization of low-attached *square-with-the-triangle* across the experiment, clear prosodic disambiguation between high- and low-attached items is maintained in our phonetic and phonological analyses across the experimental session. Such findings are compatible with other studies (e.g., Snedeker and Trueswell 2003; Kraljic and Brennan 2005) showing prosodic differences between high and low attached PP structures even with methods that do not include a lot of verbatim repetition, i.e., where increasing lexicalization of one or the other structure is unlikely to be a confounding issue. Further, even if the low-attached utterances are discounted as lexicalized, the size of the prosodic boundaries in the high-attached utterances stands alone as a strong indicator of prosodic disambiguation when compared to phonetic and phonological analyses in other studies. Overall the data suggest at best a limited effect of Sequential Position or changes in awareness of ambiguity over the course of play, versus a robust effect of syntactic phrasing on prosodic phrasing.

2.5.4. *Variety* The Variety of English spoken by the participants affected neither the phonetic nor the phonological distinction between high and low attachment. It also had no simple effect on durational or phonological data.

2.5.5. *Gameboard configuration contrast* Situational context changed over the course of the games as pieces were moved from start to goal positions. This was logged by recording the locations of key gamepieces before and during critical utterances. Utterances were then grouped into three situational categories, reflecting the contexts in which they were produced:

*Ambiguous*. Situational context did not constrain the possible moves; the square-with-triangle piece was available for a legal move and a triangle piece was available to legally move the square. Thus the Driver utterance in (3) could (disregarding possible prosodic disambiguation) equally be interpreted as an instruction to move the square-with-triangle piece or to use the triangle to push the square. Further, the situation was not biased towards either interpretation.

*Biased*. Both interpretations of the utterance corresponded to legal moves, but one was much more likely. For instance, if a triangle has just been moved next to a square, then using the triangle to move the square is an obvious and typical next move, so the situation is coded as a bias for high attachment. Similarly, an instruction to move the square-and-triangle piece that immediately followed a move of that piece was coded as a bias for low attachment.

*Unambiguous*. Only one move describable by the word sequence could be legally completed. For example, the combined square-with-triangle piece could have been boxed in, or the square could have been moved to its goal location, with no triangle in position to move it from its goal.

Speaker sensitivity to situational constraints should be reflected in greater disambiguation between high and low attachments for ambiguous situations than for biased ones and for biased than for unambiguous ones. Note that these constraints differ qualitatively from those used by Straub (1997) and Allbritton et al. (1996), who manipulated the ambiguity of the speech situation through lexical differences in the texts participants read (either in the preceding biasing paragraph or in the material that completed local ambiguities).

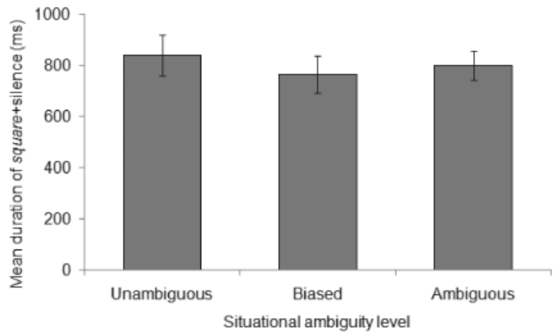


Figure 2. *Mean* square+*silence durations (with standard error bars) for high-attached triangle tokens, by situational ambiguity level, for speakers in the Driver role in Experiment 1; based on 37, 42, and 42 tokens for each of the columns – left to right in the figure – from 21 speakers (11 AmE and 10 NZE with at least one fluent utterance in each condition; data represent 72%, 88%, and 81% of the total count of fluent utterances for the complete set of 30 speakers).*

Our analyses of phonetic and phonological data use these three categories of ambiguity level. We include data from the Driver role only, since this is the role with respect to which the ambiguity categories are defined (recall that the Slider confirms the move already requested by the Driver) and where we would expect to see the strongest effects of situational ambiguity. Also, only *triangle* utterances are included, since the *square with the cylinder* sequence was never ambiguous.

Because of a large number of empty or sparsely populated cells for low attached items, *lmer* analyses including Attachment and Ambiguity as fixed effects proved to be unreliable (strong collinearity persisted between the two factors, even after centering). Therefore our analyses are restricted to speakers who had at least one instance in each ambiguity by attachment condition, and include only high-attached utterances from 21 participants (11 AmE and 10 NZE). The analyses thus included only Ambiguity level as a fixed effect. If the marking of high attachment is dependent on ambiguity level, then we would expect longer *square*+silence durations and more SBLs at 'square' in ambiguous contexts than in biased and unambiguous conditions. However, ambiguity level was not significant in any of the analyses (see Figure 2 for the duration data and Tables A4–A6 for the *lmer* analyses), suggesting that although our PP attachment ambiguities were pronounced with a variety of prosodic patterns, the variability in boundary strengths cannot be explained by the level of situational ambiguity.[10]

2.5.6. *Gamepiece contrast* Participants saw displays that included the combined square-and-triangle piece, simple (unmodified) squares, simple triangles, and cylinders, but no combined square-and-cylinder pieces. Previous work in sentence comprehension suggests that this visual information should have been extremely important in any calculations of the ambiguity of the situation, so that references to "the square with the triangle" would be ambiguous, while those to "the square with the cylinder" would not. 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 (7). 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.

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

This and other studies, including the work of Snedeker and Trueswell (2003) described above, suggest that the players of our game should be strongly biased toward high attachment for the *cylinder* sentence, since that is the only interpretation supported by the display of gamepieces. Situational sensitivity thus predicts that prosodic features indicating high attachment (i.e., a prosodic break before the PP) should not be as clearly marked in the *cylinder* utterances as in the *triangle* utterances, or that such breaks should be inconsistently produced. This would match Straub's (1997) 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.

Because there are considerably fewer cylinder utterances than triangle utterances, a relaxed criterion had to be used for inclusion of utterances in the gamepiece comparison. Our phonetic analysis comparing Driver and Slider high-attachment utterances containing the word *cylinder*, as in (4) and (6), to utterances containing the word *triangle* and intended as either high- or low-attachment, as in (3) and (5), therefore includes all speakers with at least one utterance in each condition. This left a total of 27 participants in the analysis (12 AmE and 15 NZE). The *lmer* analyses included Attachment and Role as fixed effects. Attachment was treatment coded for the three levels low-attached triangle, high-attached triangle, and cylinder, with the latter selected as the reference level. A situationally dependent model of prosodic marking would predict that the attachment marking would be weaker for the unambiguously high-attached cylinder cases than for the high-attached triangle cases, and stronger for Drivers than for Sliders.

Neither the phonetic analysis nor the phonological analyses showed a significant effect of Role (see Tables A7– A9), and Role did not interact with Attachment. The Attachment factor affected phonetic and phonological measures as follows. First, the difference between cylinder tokens and the low-attached tokens was consistently significant (phonetics: $MCMC_{mean} = -1.054$, $HPD95_{lower} = -1.208$, $HPD95_{upper} = -0.901$, $p < 0.0001$; phonology analysis 1: $\beta = -4.680$, $SE = 0.513$,

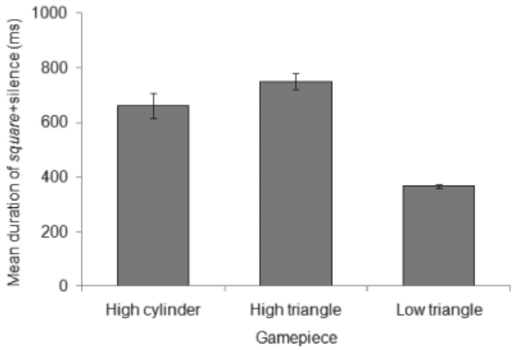


Figure 3. *Mean* square + *silence durations (with standard error bars) for utterances of high-attached cylinder tokens and high- and low-attached triangle tokens, from speakers in both Driver and Slider roles in Experiment 1; based on 97, 257, and 435 tokens for each of the columns – left to right – from a total of 27 speakers (12 AmE and 15 NZE with at least one fluent utterance in each condition; data represent 94%, 92%, and 93% of the total count of fluent utterances for the complete set of 30 speakers).*

$z = -9.124, p < 0.0001$; phonology analysis 2: $\beta = -3.890$, SE = 0.342, $z = -11.387$, $p < 0.0001$). These effects confirm that the high-attached cylinder tokens show distinct prosodic characteristics from those of the low-attached triangles, i.e., that syntactic attachment is prosodically marked even in the absence of ambiguity in the cylinder cases. Second, the smaller difference between cylinder (always high-attached) and high-attached triangle utterances was significant in the phonetic analysis ($MCMC_{mean} = 0.185$, $HPD95_{lower} = 0.022$, $HPD95_{upper} = 0.350$, $p < 0.05$), and marginally significant in the second phonological analysis ($\beta = -0.565$, SE = 0.334, $z = -1.695$, $p < 0.10$). These slight differences between high-attached cylinder utterances and low-attached triangle utterances provide some support for a situationally dependent approach, but only in the weak form that it reduces an extremely strong syntactic effect. Figure 3 gives mean and standard error data for the phonetic data by Attachment.

## 2.6. *Summary*

The phonetic and phonological analyses have shown robust effects of syntactic attachment on prosody. There is only weak and inconsistent evidence that prosodic marking of Attachment is affected by the speakers' Role in the game or of the Sequential Position of the utterance, and what evidence there is may be due to other factors (such as reduction for adjacent repetitions). The Variety of English spoken has no hint of an effect, nor does the level of Ambiguity in the gameboard configuration. The unambiguous high-attached cylinder utterances differ significantly from the low-attached triangle utterances, indicating that attachment is reflected in prosody even when it is not required for disambiguation. Nevertheless, the cylinder utterances did show some marginal differences from the high-attached triangle utterances. Since the two sets of high-attached utterances contained different words, there may be other factors at play here. This possibility will be addressed in Experiment 3 below. Overall, though, the results reported above indicate an overwhelming dominance of syntactic phrasing information over situational ambiguity in determining the prosodic form. Considering the entire shape of the dataset, we view the current results as strongly supporting a situationally independent model of the relationship between syntactic and prosodic structure.

## 3. Experiment 2: Perception study of prepositional phrase ambiguities

Previous comprehension data (Warren 1985; Price et al. 1991; Pynte and Prieur 1996; Schafer 1997; Straub 1997; Snedeker, Trueswell, Gleitman, and Levine 1999; Snedeker and Trueswell 2003; Snedeker and Casserly 2010) show mixed support for prosodic resolution of syntactic PP ambiguity, whether experiments used expert speakers' productions or those from naïve participants, across a range of comprehension tasks. Yet, none of these tasks matched the gameboard task presented here in the degree of interaction between speakers and listeners. The present work is also distinctive in the large number of participants that contributed to the productions, the total number of tokens, and the range of situational constraints on ambiguity. To test listener responses to the disambiguation found in Experiment 1, we conducted a forced-choice experiment, using a subset of our recorded materials. Our goal was to determine whether our naïve speakers' utterances could be accurately categorized by a new set of naïve listeners. In particular, we were interested in whether accuracy would be affected by the situational constraints that were in place at the time of recording. For instance, if utterances produced in ambiguous conditions carry stronger prosodic marking of syntactic structure, then those utterances presented in isolation in the listening experiment should be more accurately identified.

### 3.1. *Method*

3.1.1. *Participants* Nineteen native speakers of Midwestern American English at the University of Kansas and 14 native speakers of New Zealand English at Victoria University of Wellington took part as listeners in separate experiments for AmE and NZE respectively. None of these participants had previously taken part in the production experiments described above.

3.1.2. *Design and materials* Fluent productions of all Driver PP utterances ((3) above) were used in this experiment. Including Slider utterances would have made the experiment unmanageably large. The Driver role is in any case where disambiguation would be predicted to be most crucial (since it involves issuing instructions). One NZE speaker with only one high-attached Driver token was excluded, as were all *cylinder* utterances, since these were only produced in high-attached versions. The resulting materials tested in the categorization experiments were 196 *triangle* utterances from AmE 15 speakers, and 186 *triangle* utterances from 14 NZE speakers. Materials were blocked by speaker to maintain coherence for listeners, and utterances intended as high vs. low attachment were pseudo-randomized within each speaker to distribute attachment type. The order of each speaker's utterances did not match their order in the production experiment. Note that for comparability and representativeness of phonological and phonetic analyses, the individual analyses below include responses only to those utterances that were also included in the corresponding analyses of production data, and thus a slight reduction from the full set of presented tokens.

3.1.3. *Procedure* Participants were seated alone in a quiet room with a computer. Utterances were presented as complete sentences, over headphones. On each trial, participants selected as quickly and accurately as possible one of two visual representations, indicating high attachment (a square pushing a triangle) or low attachment (a square with a triangle on top moving as a single piece). Position of the representations on the screen was counterbalanced across participants.

The design of this experiment allowed us to examine listeners' use of the prosodic marking of high and low attachment utterances to determine the speaker's intended meaning, and whether this was sensitive to situational factors. The dependent variable is whether or not a participant correctly classified a token as high or low attached, and so the *lmer* models were configured for a binomial outcome. Random effects in the models were speakers and listeners, and the fixed effects were a subset of those in Experiment 1, i.e., Attachment, Sequential Position, Variety of English, Ambiguity Level. As indicated above, Role and Gamepiece did not feature in this experiment. An additional factor is the location of the strongest break (SBL) as measured in the phonological analysis for Experiment 1. The first analysis included Attachment, Sequential Position, Variety, and Ambiguity Level and interactions of the last three factors with Attachment. The analysis was based on responses to the Driver data analysed in the first duration analysis in Experiment 1, i.e., to 179 AmE and 157 NZE utterances.

### 3.2. *Results*

3.2.1. *Sequential position* On a situationally dependent view, if increasing sensitivity to the structural contrast results in stronger prosodic contrasts for later utterances, then listeners should more accurately identify tokens taken from later in the production experiment (recall that order of presentation in the listening experiment was randomized within speaker). While this is not true for the dataset as a

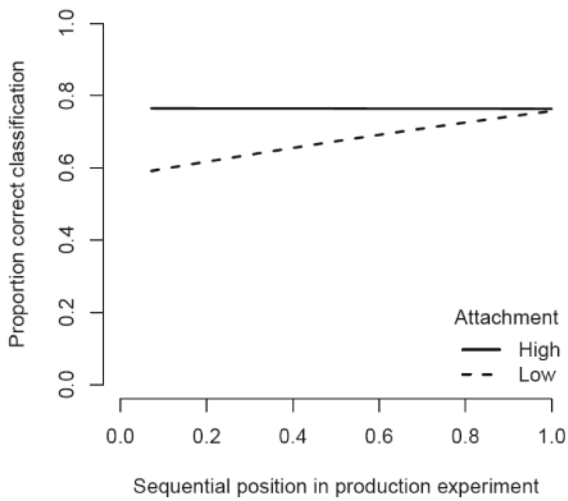


Figure 4. *Proportion (* lmer *model estimates) of correct categorizations of high- and low-attached PP utterances in Experiment 2, by sequential position in each speaker's fluent productions; based on responses from 19 AmE and 14 NZE listeners to tokens from 13 AmE and 12 NZE speakers respectively in the Driver role (i.e., the 139 high- and 197 low-attached Driver tokens for which duration data are shown in Figure 1), with a total of 2336 responses to high-attached and 3263 responses to low-attached items.*

whole, there was a significant interaction of Sequential Position and Attachment ($\beta = 1.006$, SE = 0.254, $z = 3.969$, $p < 0.001$; for the full model see Table A10), reflecting an increasing accuracy of categorization of low-attached utterances the later they occurred in the production experiment, but a stable high level of categorization accuracy for the high-attached tokens (see Figure 4). Recall that the production study suggested greater internal coherence of the low-attached PP as the experiment proceeded, most likely as a result of adjacent repetitions and a general increase in rate of speech across the game. It turns out that low-attached productions that had been adjacent repetitions were identified more accurately (0.69) than other low-attached productions (0.61). Since adjacent repetitions were more likely later in the game, this pattern probably contributed to the upward trend for the low attachment in Figure 4.

3.2.2. *Variety* There was no overall difference between responses to AmE and NZE data, nor was there an interaction of Variety with Attachment.

3.2.3. *Gameboard configuration contrast* If speakers produce prosodic structures that reflect syntactic structure only when the situation requires (using cues that may have escaped our phonetic and phonological analyses), then categorization accuracy should be higher for tokens originally produced in the ambiguous

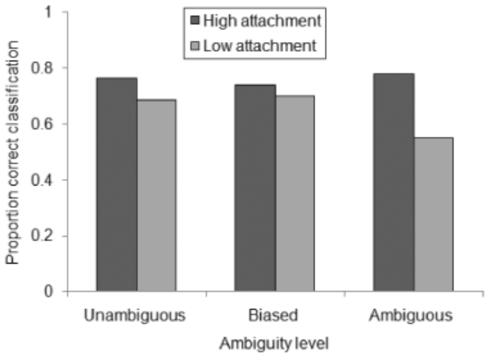


Figure 5. *Proportion (* lmer *model estimates) of correct categorizations of high- and low-attached PP utterances in Experiment 2, by level of situational ambiguity, based on same dataset as Figure 4; responses are to 48 unambiguous, 43 biased, and 48 ambiguous high-attached tokens (for total response sets of 807, 722, and 807, respectively), and to 165 unambiguous, 25 biased, and 7 unambiguous low-attached tokens (2735, 410, and 118 responses).*

gameboard configuration than for those produced in either the biased or the unambiguous gameboard configurations. Our *lmer* model shows no overall Ambiguity effect, but does show interaction of Ambiguity with Attachment, resulting from an effect of Ambiguity on the low-attached items, but not on the high-attached (see Table A10 and Figure 5). The effect for low attachment is the *lower* correct score for ambiguous than for biased or unambiguous items. Note that a situationally dependent model would predict a quite different result, namely that ambiguous utterances would have been more clearly distinguished than unambiguous ones and would therefore be more likely to be correctly identified than unambiguous ones. It is worth noting, however, that while counts of the three ambiguity categories are well-balanced for high-attached utterances, there are far fewer ambiguous than unambiguous low-attached utterances (see caption to Figure 5).

3.2.4. *Attachment* There was a strong effect of Attachment, with high-attached utterances more accurately identified ($\beta = -1.065$, SE = 0.370, $z = -2.878$, $p < 0.005$). This may reflect a bias towards high attachment of the PP, which is compatible with the operation of a default parsing strategy such as minimal attachment (Frazier 1987; Clifton, Speer, and Abney 1991; Frazier and Clifton 1996).

A second *lmer* analysis assessed whether the location of the strongest prosodic boundary was influential in the recovery of the intended syntax, to verify one of our common production measures in Experiment 1 and to further investigate the cues used in prosodic disambiguation. If one of the factors listeners are sensitive to in their recovery of syntactic structure is the location of prosodic boundaries, the location of the strongest break (SBL) in an utterance should be a good predictor of listeners' performance (e.g., Price et al. 1991; Schafer 1997; Carlson et al. 2001). Since the prediction related to SBL is different for high- and low-attached tokens,

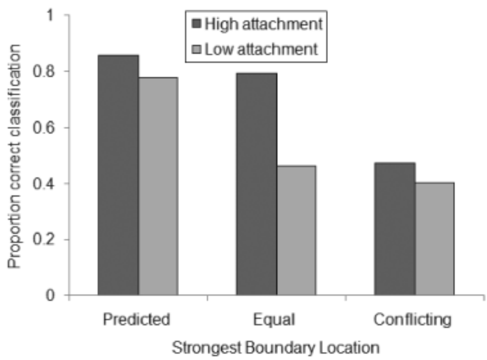


Figure 6. *Proportion (* lmer *model estimates) of correct categorizations of high- and low-attached PP utterances in Experiment 2, by location of strongest prosodic break, based on same dataset as Figure 4; responses are to 78 predicted, 34 equal, and 27 conflicting high-attached tokens (with totals of 1317, 566, and 453 responses), and to 158 predicted, 32 equal, and 7 conflicting low-attached tokens (2597, 543, and 123 responses).*

'square' SBLs were recoded to 'predicted' for high-attached and 'conflicting' for low-attached, and 'not square' SBLs were recoded to 'conflicting' for high-attached and 'predicted' for low-attached. 'Equal' SBLs remained 'equal'. The analysis included this recoded SBL and Attachment. The result showed clear effects of Attachment ($\beta = -0.541$, SE = 0.094, $z = -5.784$, $p < 0.0001$) and of recoded SBL (baseline was set at 'predicted', compared with which the effect for 'conflicting' was $\beta = -1.909$, SE = 0.137, $z = -13.948$, $p < 0.0001$, and for 'equal' was $\beta = -0.463$, SE = 0.133, $z = -3.474$, $p < 0.0001$). SBL also interacted with Attachment, with the attachment difference greater for the 'equal' cases than elsewhere (see Table A11 and Figure 6). Note that 'equal' tokens are more reliably identified when they are intended as high-attachment utterances than when they are intended as low-attachment utterances, consistent with the general bias toward high-attached interpretations mentioned above. This pattern for equal break utterances is consistent with that previously found for French PP attachment ambiguities with two equal strength intonation phrase breaks (Pynte and Prieur 1996) and for English PP attachment ambiguities with two intermediate phrase breaks (Snedeker and Casserly 2010).

### 3.3. *Discussion*

Results from Experiment 2 demonstrate that listeners can use prosodic structure produced by naïve speakers to resolve ambiguous PP attachments. High- and low-attached PPs were accurately classified as such using appropriate prosodic boundaries, although the accuracy of low-attached was not as great as that of high-attached, which possibly reflects a well-attested bias towards high attachment of PPs in structures similar to ours. Situational ambiguity had no effect on high attachments. Low attachments were overall more variable and they exhibited better categorization for tokens from later in the production sequence, likely reflecting the effect of adjacent repetitions with the same referent. On balance we view this as confirming the finding from our production data in Experiment 1 that our speakers made no significant modification to their prosodic realizations of the structures to reflect differing levels of situational ambiguity. In contrast, the relative location and size of prosodic breaks were good predictors of accuracy in the categorization task. This pattern of findings is consistent with situationally independent models of comprehension.

However, the utterances we used in the comprehension study were taken from game situations in which the sequence ‘the square with the triangle’ was consistent with two visually available referents. That is, speakers in our game task could always see both the square-and-triangle piece and the separate square and triangle pieces. It is possible, as argued by Straub (1997) and Snedeker and Trueswell (2003), that prosodic disambiguation of syntax in this situation was due to mere exposure to objects in the discourse environment that could be consistent with either of the two meanings of the critical word sequence, even in situations (such as the unambiguous gameboard configurations) when one of the interpretations is ruled out. This would be consistent with a processing system in which situational dependence is important, but it is calculated at a very coarse grain. For this reason, we conducted an additional production experiment where syntactic structure was manipulated between participants, so that only one meaning of the ambiguous sequence was consistent with the game pieces available to the speakers.

## 4. Experiment 3: Production of prepositional phrase ambiguities in a between-participants design

### 4.1. *Method*

4.1.1. *Participants* Eight pairs of native speakers of Midwestern American English from The Ohio State University, naïve to the purposes of the experiment, participated in partial fulfilment of a course requirement.

4.1.2. *Design and materials* Driver-Slider roles, game rules, and gameboards were identical to those described for Experiment 1 with the following exceptions. We manipulated the syntax used to refer to the critical sequence between subjects, so that four pairs of subjects participated in the low-attached condition, and four in the high-attached. We created two new practice boards and six new pairs of gameboards, so that half of the participant pairs would see only high-attached referents of the phrase *the square with the triangle* (separate square and triangle pieces) and half would see only low-attached referents (a square with a triangle on top). Two driver boards, one each for low-attached and high-attached versions, are given in

Appendix C. Game rules in the high-attached condition were as before, except that no combined square and triangle piece was available on the board (nor was there any other piece described with a modifier). Game rules in the low-attached condition specified that the square must drag behind it an adjacent piece on the opposite side from the direction of its next move. This change preserved the complexity of the original game and its focus on strategic ordering of moves. The Driver's utterance frame for this move was "I want to change the position of the square and the (triangle/cylinder/square with the triangle)," and the Slider's was "I am able to confirm the move of the square and the (triangle/cylinder/square with the triangle)." In the low-attached condition, there were no high-attached, instrumental utterances using the square or any other object.

4.1.3. *Procedures* Participants sat at a table separated by a divider in a quiet room, accompanied by two experimenters, one who supervised the game, offering advice when necessary, and one who supervised the recording of the speech data to computer disk. Participants wore head-mounted microphones. All participant pairs completed a practice game and at least two more games, using separate boards and 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 four games, involving three different boards.

### 4.2. *Results*

Our phonetic and phonological analyses were the same as for Experiment 1. The phonetic data were *square*+silence durations and the phonological data were strongest boundary location (SBL) categories based on the allocation of break index patterns to 'square', 'not square', and 'equal' (see description of that experiment for details). Break indices were based on independent transcriptions from three expert ToBI transcribers (the authors) of the Driver utterances. Each *lmer* analysis included speaker as a random factor and Attachment as a fixed factor. The phonetic data also included Role as a fixed factor, and tested also for the interaction of Attachment and Role. The analysis of phonological data involved two *lmer* models, one excluding 'equal' SBLs, and the other including 'equal' as part of the 'square' set.

The phonetic data showed simple effects of Role ($MCMC_{mean} = -0.310$, $HPD95_{lower} = -0.493$, $HPD95_{upper} = -0.136$, $p < 0.001$) and of Attachment ($MCMC_{mean} = -0.680$, $HPD95_{lower} = -1.010$, $HPD95_{upper} = -0.352$, $p < 0.001$), but no interaction. The full model is in Table A12, and the mean durations by Role and Attachment are shown in Figure 7.

Statistical comparison of the phonetic data from Experiments 1 and 3, including Speaker as a random effect and Experiment, Role, and Attachment as fixed effects, confirms the robust simple effects across the two experiments of Role

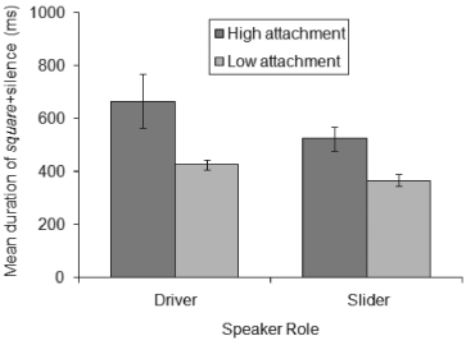


Figure 7. *Mean* square + *silence durations (with standard error bars) for Driver and Slider utterances of high and low PP attachments in Experiment 3; based on 47, 58, 52, and 65 tokens for Driver high and low attachments and Slider high and low attachments respectively, from a total of 16 AmE speakers. (All fluent utterances are included.)*

($MCMC_{mean} = -0.339$, $HPD95_{lower} = -0.426$, $HPD95_{upper} = -0.250$, $p < 0.001$) and Attachment ($MCMC_{mean} = -1.131$, $HPD95_{lower} = -1.284$, $HPD95_{upper} = -0.985$, $p < 0.001$), and the absence of an interaction of Role and Attachment. But the analysis also shows a clear interaction of Experiment and Attachment ($MCMC_{mean} = 0.592$, $HPD95_{lower} = 0.163$, $HPD95_{upper} = 1.007$, $p < 0.005$), with no main effect of Experiment (see Table A13). The model's predicted effect of Attachment in Experiment 1, at 247 ms, is over twice that predicted for Experiment 3, at 119 ms, and is largely due to the much larger durations of the high-attached tokens in Experiment 1, as comparison of Figures 1 and 7 makes clear.[11]

The analyses of the phonological data in Experiment 3 showed clear effects of Attachment, with the high-attached data showing a greater incidence of 'square' SBLs whether the analysis excluded 'equal' altogether ($\beta = -5.006$, SE = 1.557, $z = -3.215$, $p < 0.002$, Table A14), or treated these as 'square' tokens ($\beta = -5.039$, SE = 1.216, $z = -4.145$, $p < 0.0001$, Table A15).

A combined analysis of the Driver data from Experiment 1 with the data from Experiment 3 shows that the greater incidence of 'square' SBLs in high-attached tokens is consistent across the experiments. The Attachment effect is significant in phonological analysis 1, with 'equal' excluded ($\beta = -8.750$, SE = 1.344, $z = -6.412$, $p < 0.0001$) and in phonological analysis 2, with 'equal' grouped with 'square' ($\beta = -4.837$, SE = 0.655, $z = -7.380$, $p < 0.0001$), while neither analysis shows any interaction of Attachment with Experiment (see Tables A16 and A17). However, note that the Experiment 3 data shows a numerically lower tendency for strongest boundaries before the PP in the high attached cases than was found for the equivalent data in Experiment 1 (56.1% in Experiment 1, 34.0% in Experiment 3, climbing to 80.6% and 72.3% respectively if equal-strength cases are included). What is more, when the strongest boundary was at *square* it was much more likely to be a full IP boundary in Experiment 1 (86% versus 32% in Experiment 3).

The robust effect of speaker Role in the phonetic data across the experiments indicates that participants may have spoken more carefully and slowly as Drivers than they did as Sliders. The Attachment effects in both the phonetic and phonological analyses provide further consistent support for the prediction that speakers mark the level of PP attachment. But when we also consider the interaction of Experiment and Attachment in the phonetic data and the numerically weaker tendency to mark high-attached tokens with strong breaks after *square* in Experiment 3, then it would seem that while the fact of prosodic marking of PP attachment in the earlier Experiment is not due to our previous use of a within-subjects design, the extent of this marking may be dependent on the experimental design. Although it is possible that the difference between the two experiments is due to other differences such as the relative fluency levels of the participants or dialect differences between the students from Kansas or Wellington in Experiment 1 and the Ohio State University students in Experiment 3, we consider it most likely that the more ambiguous context of Experiment 1 may have resulted in more "deliberate" pronunciations. What is particularly interesting about this effect is that it occurred on such a broad level. Our speakers did not make fine-grained adjustments to reflect the utterance-by-utterance ambiguity level, but they may have made a global adjustment in their use of strong intonation phrase boundaries for the experiment-wide ambiguity in Experiment 1.

## 5. Discussion of experiments 1 to 3

A situationally dependent model of prosody claims that prosodic cues to syntactic structure are most likely to be produced and interpreted when the situational need to disambiguate is greatest. Previous proposals for situationally dependent models imply that situational need is defined primarily on the basis of whether there is an alternative syntactic parse that is grammatical for the utterance, consistent with the discourse situation, and perhaps, reasonably frequent. If so, sentences such as (3) in Experiment 1 should have been ideal candidates for situationally based disambiguation. Our review of PP productions has shown that the effects of situation should not have been limited by prosodic constraints on the structure, as it is possible for speakers to produce both disambiguating and ambiguous prosodies for PP sentences. However, we found no compelling evidence for situationally dependent prosody in Experiment 1, in which situational ambiguity was manipulated across gamepieces, speaker role, and gameboard configuration. In Experiment 3, we again saw no indication of a fine-grained effect from the contrast in speaker Role, and we saw robust effects of Attachment even in the Slider productions. This constrasts with some previous findings, especially the work of Snedeker and Trueswell (2003), a topic that we take up in the general discussion. We will also return there to the difference in degree of disambiguation between Experiments 1 and 3, which provides the best evidence in any of our comparisons for situational sensitivity (but only at a very coarse level). Most of our findings so far are consistent with situationally independent models of prosodic production. Nevertheless, they do so on the basis of what could be considered null effects: the similar difference between high and low attachments across gameboard configurations, the similar difference between high-attached utterances and low-attached ones regardless of whether the high attachments are *triangle* utterances or *cylinder* utterances, and so forth.

Situationally dependent and independent models can also be contrasted by looking at their predictions for temporarily ambiguous sentences. Unlike sentences with standing ambiguities like (3), temporarily ambiguous sentences should not create strong motivation for prosodic disambiguation, as prosodic structure is not required in order to determine their meaning. As an example, Straub (1997) found reduced disambiguation with PPs that were pragmatically biased for high or low attachment by the object of the PP, compared to PPs that did not contain such cues to attachment. Similarly, Gahl and Garnsey (2004) found greater prosodic disambiguation when SC/DO verbs were followed by their less-probable complements. In Experiments 4 and 5, we examine production and comprehension patterns for the temporary ambiguity caused by an optionally transitive verb in a subordinate clause. The two versions of the structure were produced with several non-prosodic indicators of the intended syntax, some of which occurred before the point of temporary ambiguity. Previous research (e.g., Kjelgaard and Speer 1999) has demonstrated that it is possible for a native speaker to produce both ambiguous and disambiguating pronunciations of this type of ambiguity. Therefore, a strong situationally dependent model predicts that prosodic disambiguation should be weak for this structure, and certainly weaker than the disambiguation found in the PP structure. In contrast, a situationally independent model predicts that the degree of disambiguation depends on the grammatical constraints relating prosodic structure to other linguistic structures, especially syntactic structure. Under standard linguistic assumptions, the stronger syntactic contrast in our temporarily ambiguous structure should result in more robust prosodic distinctions between the two parses than those found for the two parses of the PP sentences (e.g., Nespor and Vogel 1986; Selkirk 2000). Thus, the situationally independent model predicts strong prosodic disambiguation for our temporarily ambiguous closure structure, and stronger disambiguation than in the PP utterances.

Finally, we can consider a weak situationally dependent model, in which prosody-syntax correspondences and situational ambiguity are both important factors in a multiple-constraints or probabilistic system. Such a model can allow for prosodic disambiguation for otherwise disambiguated structures, as long as the prosody-syntax constraints are strong enough. Yet, if we do not know in advance exactly how strong prosody-syntax constraints versus situational ambiguity ones are, we cannot make clear predictions about which will dominate, or whether we will find some degree of intermediate disambiguation. Therefore, it is informative to compare the prosodic disambiguation for distinct constructions, with distinct situational demands, within the same set of speakers and from the same task.

## 6. Experiment 4: Production study of closure ambiguities

Experiment 4 examined whether prosodic correspondences to syntactic structure would occur in sentences with little to no situational need for prosodic disambiguation. It tested sentences containing a temporary ambiguity between Early (8) and Late (9) Closure of a subordinate clause and its verb phrase. As with the PPs, the experimental materials were never pronounced by the experimenters during the testing situation, to avoid modeling particular contours. The materials were presented without commas, to prevent subjects from simply providing a strong prosodic boundary at the location of a comma.

(8) When that moves the square will encounter a {cookie/ravenous goat}.

(9) When that moves the square it should land in a good spot.

Within any particular game situation, there was little question about which way a sentence beginning with "When that moves . . ." would resolve. The Early Closure sentence was part of the Slider's lines while the Late Closure sentence was part of the Driver's lines. Thus, the speaker/listener role alone distinguished the two completions. (Recall that Drivers and Sliders switched roles between games, providing a within-subjects test of the structure.) The Slider's Early Closure sentence was part of a longer utterance, which began with either "Good choice" or "Bad luck" and was part of a conversational turn that confirmed the movement of the square piece. The Driver's Late Closure sentence was also part of a longer utterance, beginning with the phrase "The [brown/blue/red/green/yellow] one." None of these initial phrases were ever used in another context within the game, or in the other role. Obviously, the temporary ambiguity was resolved before the end of the sentence. The disambiguating morphosyntactic material came immediately following the ambiguously attached NP *the square*, and thus just two syllables after the first potential disambiguating prosodic boundary, the boundary between *moves* and *the square*.

Previous production studies using materials such as those in (10), with a standing ambiguity between Early and Late Closure of the subordinate clause (i.e., *gradually* modifies either the following or the preceding verb; commas were not included), have revealed reliable disambiguation by trained speakers (Price et al. 1991), but not by naïve speakers in an oral reading task (Allbritton et al. 1996). Price et al. (1991) found that their speakers disambiguated sentences such as (10) by placing the strongest prosodic boundary at the subordinate clause boundary.

(10) When you learn gradually you worry more.

Naïve readers also provided clear prosodic disambiguation of temporary closure ambiguities like (11), both by marking a prosodic boundary at the clause boundary and by placing stress on either the first or second element of the sequence *Hong Kong*, depending on whether this sequence was in the same intonational phrase as the following word (Warren et al. 1995).

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

Nevertheless, phonetic marking of a clause boundary does not seem to be obligatory. Kjelgaard and Speer (1999) included a "baseline" prosody for temporary ambiguities of clause location such as (12) and (13). The baseline prosody was tonally consistent with the presence of an intermediate phrase boundary somewhere in the ambiguous region (e.g., in the span *leaves the house*), but the rhythmic pattern did not clearly indicate the presence of a boundary. This prosody was shown in an auditory pretest to be highly acceptable for either syntactic form. As mentioned above, Allbritton et al. (1996) found that naïve speakers did not consistently disambiguate sentences such as (10) until they were informed of the ambiguity, nor did they find consistently strong disambiguation effects for uninformed professional speakers. (The expectations for disambiguation in Allbritton et al.'s materials depends on one's assumptions. In a situationally independent model in which prosody-syntax correspondences are important, the prosody for ambiguously associated adverbial phrases might be expected to be less biasing than that for ambiguously associated NP arguments, as in (12), (13), and our stimuli. In a situationally dependent model, prosodic disambiguation for standing ambiguities should generally be stronger than for tempory closure ambiguities like ours.)

(12) When Roger leaves, the house is dark.

(13) When Roger leaves the house, it's dark.

While prosodic marking of a clause boundary is not obligatory, we nevertheless expected it to be a common occurrence when speakers were engaged in a naturalistic task and had a clear meaning in mind. If, contrary to Allbritton et al.'s findings, naïve speakers prosodically disambiguate sentences such as (8) and (9) even when there is non-prosodic disambiguation, then we would expect to find the strongest boundary after *moves* in sentence (8) but after *square* in (9). However, if naïve speakers only provide prosodic differences when the sentence context fails to do so, or only in response to explicit instructions to disambiguate, then we would expect no such prosodic disambiguation, since the ambiguity is resolved through other sources. The prosody in the *moves the square* region should not differ for these two syntactic types, or at least not in a way that can be reliably predicted by the syntax.

### 6.1. *Method*

The closure data were obtained in the same game sessions as the first set of PP attachment data, so details of data collection, participants, and analysis methods are as for Experiment 1.

### 6.2. *Results*

6.2.1. *Excluded participants and utterances* Four participants excluded from Experiment 1 were also excluded from Experiment 4. In addition, one AmE participant and one NZE participant produced no Late Closure utterances and one NZE participant produced no Early Closure utterances, and were excluded from this experiment and the following categorization experiment.

When we came to create balanced sets of Late versus Early Closure utterances for the categorization experiment reported below (Experiment 5), we excluded four Late Closure utterances (all AmE) and five Early Closure utterances (one AmE and four NZE). These exclusions resulted in the removal of another AmE participant who produced only one Early Closure utterance. For ease of comparison of production and categorization data, the same exclusions were applied to the analysis of the current experiment. No other utterances clearly met our criteria for exclusion, described above for Experiment 1. The remaining data comprise 53 fluent Late Closure and 53 fluent Early Closure utterances from 13 AmE speakers and 38 fluent Late Closure and 38 fluent Early Closure utterances from 13 NZE speakers.

6.2.2. *Phonetic and phonological analyses* The key position for analysis in the closure cases is the major syntactic boundary between the two clauses. This is after *moves* in the Early Closure utterances and after *square* in the Late Closure utterances. Situational independence therefore predicts an interaction of Closure category and word region, while situational dependence in its strongest form predicts no interaction. Our phonetic measure for each word region is the duration of the relevant word and any following silence. Our phonological measure is the location of the strongest break (SBL), based on ToBI transcriptions. Two sets of ToBI transcription data were collected for these closure ambiguities. First, a team of experienced transcribers performed a prosodic transcription of each utterance in its original phonetic and syntactic context. In other words, the full utterances (see (14) and (15)) were transcribed. As was found for the PP attachment data, there was much variation in the prosodic structure of the lexically ambiguous portion of the utterances (e.g., for the 53 pairs of AmE utterances there were 30 distinct patterns on the string “moves the square” for Early Closure and 23 for Late Closure). Second, two independent researchers transcribed the ambiguous region without its original phonetic and syntactic context, i.e., they only heard the lexically ambiguous region (as in (16)).

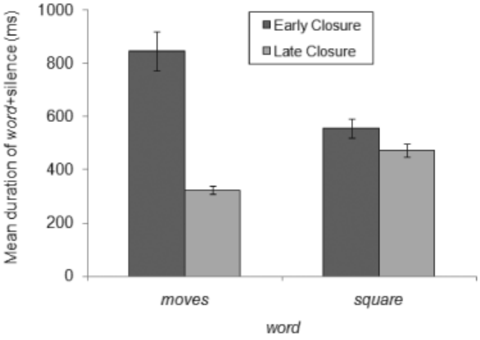


Figure 8. *Mean durations of key word (*moves *or* square*) and following silence (with standard error bars) for Early and Late Closure utterances in Experiment 4; based on a total of 91 Early and 91 Late Closure utterances, from a total of 26 speakers (13 AmE and 13 NZE; these represent 95% and 96% of the total count of fluent Early and Late Closure utterances from 27 participants).*

(14) Good choice. When that moves the square will encounter a {cookie/ravenous goat}.

(15) The tan one. When that moves the square it will land in a good spot.

(16) When that moves the square

Using the break indices from the ToBI transcriptions, utterances were categorised into 'predicted,' 'equal,' and 'conflicting.' 'Predicted' SBLs are at *moves* for the Early Closure utterances and at *square* for Late Closure, and 'conflicting' are at *square* for Early and at *moves* for Late.

The fixed effects in our *lmer* model for the duration data are Variety (AmE vs. NZE), Closure (Early vs. Late), and Region (*moves* vs. *square*). Participants were entered as a random effect, as was Utterance (each utterance had duration data for the two regions). The analysis showed no simple effect or interaction involving Variety (see Table A18). There were however simple effects of Region (*moves*+silence was overall longer than *square*+silence, $MCMC_{mean} = -0.722$, $HPD95_{lower} = -1.014$, $HPD95_{upper} = -0.424$, $p = 0.0001$) and of Closure (overall Early Closure measures were longer than Late Closure, $MCMC_{mean} = -1.700$, $HPD95_{lower} = -2.010$, $HPD95_{upper} = -1.413$, $p = 0.0001$). Importantly, the interaction of Region and Closure was significant ($MCMC_{mean} = 1.605$, $HPD95_{lower} = 1.183$, $HPD95_{upper} = 2.021$, $p = 0.0001$); the difference between Early and Late Closure is present for *moves* but not for *square*, as can be seen in Figure 8.

The finding of no noticeable difference in *square*+silence durations, rather than the longer values in Late Closure utterances (versus Early Closure ones) that would match a stronger boundary there, seems to indicate that the duration data provide only partial support for the prosodic marking of the major syntactic boundary, i.e.,

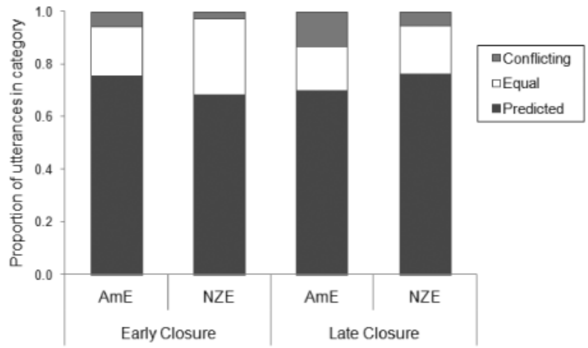


Figure 9. *Proportions of utterances in Experiment 4 where the relative strength of the prosodic breaks (based on phonological transcriptions) at* moves *and* square *is as predicted by the syntactic structure, where the breaks are equal and where the relative strength is in conflict with the syntactic prediction, for 53 AmE and 38 NZE utterances in each of the Early and Late Closure groups (representing 98%, 91%, 93%, and 100% of the total fluent tokens for the four columns in the figure).*

this is marked at *moves* in Early Closure but not at *square* in Late Closure. However, the longer durations overall for the Early Closure utterances in Figure 8 may reflect in part the fact that these utterances provided new information to the discourse (the presence of a cookie or a ravenous goat in the gamepiece path), and might thus have been spoken more clearly, while the Late Closure utterances added no new information to the discourse (see (14) and (15) above). Nevertheless, note from Figure 8 that the *relative* durations of *moves*+silence and *square*+silence within each closure type are in a direction that matches the predicted boundary location. This is confirmed by the phonological data. For example, Figure 9 shows the proportions of 'predicted', 'equal', and 'conflicting' SBLs for each closure type for AmE and NZE data, from the transcriptions carried out on the truncated utterances (see (16) above). These are shown rather than the results from the transcriptions of the complete utterances because it is the truncated utterances that are used in Experiment 5 below. The transcription results from the entire utterances show a very similar pattern, but with fewer 'equal' SBLs and more 'predicted', across the Closure and Variety types. There were also more IP and fewer ip boundaries marked at *square* in the transcriptions of the full Late Closure items. This difference arises because of the truncation process, which aimed to remove any coarticulatory information about the following word, but also led to the loss of segmental material that carried boundary tone information for *square*. In addition, pitch re-set information across the boundary was not available to the transcribers.

The counts of SBL data in the 'conflicting' category are too small for reliable statistical analysis of the distributions in Figure 9, but the pattern is clear enough

– the relative strengths of the prosodic boundaries match the marking of the clause boundary in each case. In addition, note that a prosodic boundary after *square* in the Early Closure sentence is not entirely unexpected because it is the right edge of the subject noun phrase (e.g., Truckenbrodt 1999; Selkirk 2000). Thus, these results are consistent with syntactically based predictions.

Phonetic and phonological analyses of the closure ambiguities reveal prosodic resolution of temporary syntactic ambiguity that is similar or even stronger than that found in the analyses of standing PP ambiguities. Indeed, visual comparison of the effects for PPs in Figures 1 and 7 with those for *moves*+silence in Figure 8 indicates that the closure ambiguities are more clearly marked prosodically (and see also further discussion later in this paper). This supports situationally independent models of prosody, but goes against the predictions of models in which situational dependence dominates. The structure that receives the most prosodic disambiguation is not the one with the most situational ambiguity, but the one with the greatest difference between parses in the location of right edges of syntactic boundaries. This is precisely the pair predicted to differ most by linguistic models of English prosody.

Experiment 5 determines whether the prosodically marked closure structures are successfully discriminated in the same type of perceptual task – forced choice categorization – used to study the perceptual resolution of PP ambiguities.

## 7. Experiment 5: Perceptual study of closure ambiguities

### 7.1. *Method*

Ambiguous fragments of the utterances (see (16) above) were presented over headphones to listeners seated in a sound-attenuated booth for categorisation. On each trial they heard each fragment twice, and were then required to choose between the Early and Late Closure continuations of the fragment, which were displayed in text on either side of a computer screen. Presentation order and location of the continuations on the screen were counter-balanced across participants. Because the original following phonetic context for the two fragments differed, i.e., [ɪ] for *it* in Late Closure and [w] for *will* in Early Closure, the experiment was run in two blocks, so that we could control for any cuing effects of co-articulated material. In one block, the written continuations were the originals for the sentences, as shown in (17). In the other, written continuations were constructed that, if spoken, would have had the initial sound segments of the opposite condition, as shown in (18). Each fragment was tested in both blocks for each listener, and block order was balanced across listeners. Within each block the fragments were grouped by speaker, to maintain coherence for the listeners and match the procedure for the PP experiment.

(17) Original Continuations
When that moves the square
a. . . . it should land in a good spot. [ɪ], Late Closure
b. . . . will encounter a cookie. [w], Early Closure

(18) Segmentally-Crossed Continuations
When that moves the square
a. . . . we'll encounter a problem. [w], Late Closure
b. . . . is shut off from the best path. [ɪ], Early Closure

7.1.1. *Participants* This experiment was run in two parts. AmE Closure tokens were judged by 16 native speakers of Midwestern American English from the University of Kansas and NZE closure tokens by 16 native speakers of New Zealand English from Victoria University of Wellington. No listeners had previously taken part in any of the experiments described above.

7.1.2. *Materials* The materials consisted of a balanced set of 53 Early and 53 Late Closure fragments from 13 AmE speakers, and a balanced set of 38 Late and 38 Early Closure fragments from 13 NZE speakers, i.e., the truncated utterances reported in the production experiment (Experiment 4). The data sets allowed comparison of three different prosodic conditions, defined by the SBL categories from the transcriptions of the truncated utterances in Experiment 4. The largest group of tokens provided a "predicted" boundary strength condition (N = 132, consisting of 77 AmE tokens and 55 NZE tokens), with the strongest prosodic boundary at the clause boundary (*moves* and *square* for Early and Late Closure, respectively). The "equal" condition contained tokens that received equal boundary strengths at *moves* and *square* in the transcription (N = 37, 19 for AmE and 18 for NZE), and the "conflicting" condition contained tokens with weaker prosodic boundaries at the major syntactic boundary than at the other critical location (N = 13, 10 for AmE and 3 for NZE).

## 7.2. *Predictions*

In line with our claim for the marking of sentence structure through prosody, we predicted that the phonetic and phonological contrasts noted in Experiment 4 would disambiguate the structures presented to listeners in this experiment. However, the production data show variability, with a number of phonological analyses revealing 'conflicting' or 'equal' SBL patterns. We predicted for these cases that they would be less well categorised than the 'predicted' SBL patterns. We also predicted that the effect of a 'conflicting' pattern would be stronger for the Late Closure cases, since a (relatively) strong break is less likely after *moves* when this aligns with the V-NP boundary in Late Closure than after *square* when this aligns with the NP-VP boundary in Early Closure (see above). Further, we predicted an effect of Continuation, with the segmentally-crossed items in (18) showing a lower

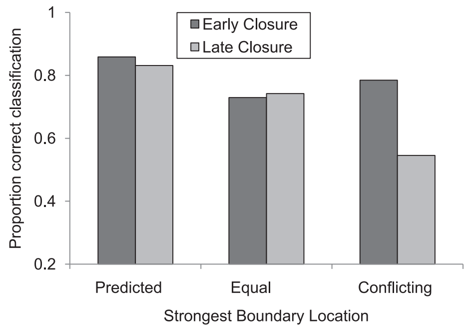


Figure 10. *Proportion (*lmer *model estimates) of correct categorizations of Early and Late Closure utterances in Experiment 5, by location of strongest prosodic break; datapoints are based on totals of 2112, 672, and 128 responses to predicted, early, and conflicting Early Closure tokens, and 2112, 512, and 288 responses to predicted, early, and conflicting Late Closure tokens, respectively. Stimuli were 132 predicted, 42 equal, and 8 conflicting Early Closure tokens, and 132 predicted, 32 equal, and 18 conflicting Late Closure tokens.*

rate of correct categorization than the original uncrossed items (17). This is because segment-level coarticulation cues matched the orthographic continuation for each closure type in the original items but not in the crossed items. Since the coarticulation cues would be weaker for the Late Closure items, because of larger breaks after *square* for this closure type and therefore separation of segmental material at the end of *square* from the beginning of the next word, we predicted an interaction of Continuation with Closure type. We did not predict any effect of Variety, since there was no clear evidence in the production data that AmE and NZE differed in the successful marking of the syntactic structure.

Categorization data were analysed using *lmer*, with speakers and listeners as random effects and the four fixed effects of Closure (Early or Late), Continuation (original or crossed), SBL pattern (‘predicted’, ‘equal’, and ‘conflicting’, as described above) and Variety (AmE or NZE). In addition to the four simple effects, the *lmer* also tested for interactions of Closure with SBL pattern and with Continuation, as motivated in the preceding paragraph.

## 7.3. *Results*

As expected based on the results of Experiment 4, categorization accuracy was very high, with an overall proportion correct of 0.746. Our *lmer* results showed an interaction of SBL pattern and Closure (see Figure 10 and Table A19). As predicted above, the strongest effect is the lower accuracy score for the Late Closure items with conflicting SBL patterns ($\beta = -1.160$, SE = 0.281, $z = -4.127$, $p < 0.0001$).

There is also a simple effect of Completion ($\beta = -0.657$, SE = 0.094, $z = -7.022$, $p < 0.0001$) and an interaction of Completion with Closure ($\beta = 0.405$, SE = 0.130, $z = 3.112$, $p < 0.005$). The interaction is as predicted above – there is a greater ef-

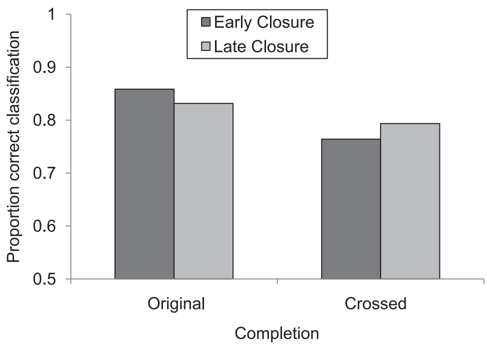


Figure 11. *Proportion (* lmer *model estimates) of correct categorizations of Early and Late Closure utterances in Experiment 5, by Completion; each datapoint is based on a total set of 1456 responses (16 AmE listeners categorizing 53 AmE utterances and 16 NZE listeners categorizing 38 NZE utterances ).*

fect of Completion for the Early Closure than for the Late Closure items. This is shown in Figure 11, which also shows clearly the effect of Completion, with better categorization for original than for segmentally-crossed items.

In addition to these comparisons between conditions, we carried out t-tests on the accuracy data in the whole dataset to determine whether categorization was significantly better than chance. This was the case overall for both Early and Late Closure items (with performance levels of 0.75 and 0.74, respectively), as well as for each of the separate sub-conditions in Figures 10 and 11, with the exception of the Late Closure 'conflicting' boundary case in Figure 10. (All significant t-tests are so at $p < 0.0001$.) The finding that the 'equal' sets performed at a level significantly higher than chance and that the 'conflicting' boundary conditions were either at chance level (Late Closure) or also significantly higher than chance (Early Closure) is somewhat surprising, since we might expect 'equal' boundaries to have been at chance and 'conflicting' boundaries below chance, if prosodic boundary strength alone was determining categorization. Thus, while the categorization results show that for the majority of items the prosodic differences found in the phonetic and phonological analyses are good predictors of disambiguation, they also show that boundary strength is not the only aspect of prosody that can influence syntactic parsing. Indeed, these utterances contain other types of prosodic information that may influence syntactic interpretation, such as pitch accent placement, accent and edge tone types, and differences in pitch range.

## 8. Discussion of Experiments 4 and 5 and comparison with Experiments 1–3

The results of Experiments 4 and 5 show significant patterns of prosodic disambiguation for an ambiguity that does not require prosodic disambiguation within the discourse situation. These results are inconsistent with the strongest versions of situationally dependent models, which should predict no consistent pattern of disambiguation for the closure sentences we tested. Overall, they are also inconsistent with weaker versions that might predict some tendency for prosodic disambiguation in the closure structure, but more disambiguation in the situationally ambiguous PPs. We found some support for this in the duration measurements for the *square*+silence region, which may be argued to show stronger effects for the PPs than for the closure sentences. However, the comparison of duration measurements is limited by the differences between the experiments. In Experiment 1, high versus low attachments are predicted to differ in the prosodic boundary at *square*, and the two attachments contrasted within each speaker role (Driver vs. Slider). In Experiment 4, as discussed above, the expectations for durations are more complex and the Early and Late Closure versions are divided between the two speaker roles.

A comparison between the phonological analyses from Experiment 1 and Experiment 4 is more straightforward, and shows that speakers were more likely to produce clear prosodic disambiguation for Closure than for PP utterances, as predicted by situationally independent models but not by situationally dependent ones. The SBL patterns for the two ambiguity types were analysed using *lmer* for the 22 speakers who are represented in both data sets (12 AmE and 10 NZE). For both types, the transcriptions used were those based on the complete utterances. Two analyses were carried out, reflecting the procedure for Experiment 2. In both, speakers were a random effect, and ‘equal’ categorizations were omitted, while in the second the ‘equal’ tokens were recategorized as ‘predicted’. In both analyses (Tables A20–A21), there was an effect of Type ($\beta = -1.933$, $SE = 0.638$, $z = -3.033$, $p < 0.005$; $\beta = -1.798$, $SE = 0.633$, $z = -2.843$, $p < 0.005$), but no other effects. The Type effect confirms that the Closure ambiguities had more disambiguating boundaries in the predicted location than the PP ambiguities.

The comparison between categorization Experiments 2 and 5 is complicated by the differences in tasks required by the stimuli. Recall that participants heard entire sentences for the PP ambiguity, but only the initial, ambiguous portion of the Closure sentences. For the latter stimuli, participants chose between the original continuations in half of the blocks and segmentally crossed ones in the other half. We carried out two *lmer* analyses of data from the 22 speakers represented in both datasets. Each analysis had speakers and listeners as random effects and Type (PP vs Closure) and Variety (AmE vs NZE) as fixed effects. One included both original and crossed Closure data (Table A22), the other excluded the crossed data (Table A23). Both analyses showed an interaction of Type and Variety, though this was only marginally significant in the more inclusive model (full model: $\beta = 0.517$, $SE = 0.297$, $z = 1.745$, $p < 0.09$; model without crossed data: $\beta = 0.763$, $SE = 0.295$, $z = 2.586$, $p < 0.01$). PPs were categorized more accurately overall in the NZE data, but Closures in the AmE data. Recall, however, that Variety does not even approach significance in the separate analyses of PP categorization and Closure categorization, nor in any of the phonetic or phonological analyses. Overall, NZE listeners performed slightly better at choosing correct paraphrases for the PPs than correct continuations for the Closure sentences, while AmE listeners performed more poorly with low-attached PPs. Given the general lack of Variety effects across the data, we think the safest conclusion is that there is generally quite good categorization for both structures, consistent with a strong role for prosody-syntax correspondences.

Our results contrast with those of Allbritton et al. (1996). There are several factors that may account for this. First, as we have argued above, our speakers were speaking quasi-spontaneously and Allbritton et al.'s speakers were reading text aloud. Second, our speakers had a clear meaning in mind but the other speakers may not have fully interpreted the sentences. Third, our materials contained an ambiguously attached NP that was either the direct object of a preceding verb or the subject of the matrix clause. Allbritton et al.'s materials contained phrases that were adjuncts, which may be more likely to be phrased separately from the other material within their clause (Truckenbrodt 1995, 1999). Finally, our materials allowed deaccenting on the ambiguously attached phrase *the square*, since this provides given information in its context. As each prosodic phrase must contain at least one pitch accent, a fully deaccented phrase is more likely to be grouped prosodically with the other material in its clause. We assume that most sentences are uttered within some type of rich discourse context, as they are here. Thus, we believe that this is an aspect of our materials that helps them to generalize well to spontaneous speech, but is well worth further study. In summary there are a number of factors that distinguish the two studies. We found that our results, like theirs, contain cases of successful disambiguation as well as unsuccessful disambiguation; what differs is the proportion. We cannot say with certainty which factor or factors was most important in shifting the proportions of disambiguating utterances. However, the factors we have laid out work together to suggest that a set of linguistic parameters (clearly established meaning, syntactic phrasing, argument status, given/new status, etc.) are the principal determinates of prosodic form, and that situational need to disambiguate is of lesser importance.

## 9. Effects of length and experiment-level ambiguity

Our PP results in Experiments 1–3 contrast with some previous findings and comport with others. The previous work includes production studies of global PP ambiguities such as (19) (Cooper and Paccia-Cooper 1980), (20) (Price et al. 1991), and (21) (Straub 1997), as well as local ambiguities such as (22) (Warren 1985) and (23) (Straub 1997).

(19) Lieutenant Baker instructed the troop with a handicap.

(20) Raoul murdered the man with a gun.

(21) The chauffeur annoyed the man with the cigar.

(22) Sam climbed the peak with *{snow on top | Pete and Dave}*.

(23) The chauffeur annoyed the man with the *{sunburn | song}*.

Phonetic data from such studies, like our own production data in Experiments 1 and 3 above, have shown high attachments to be marked by longer and more frequent silences before the preposition, as well as by greater durations of the immediately preceding word. We have focused here on the durational aspects of the PP stimuli, since fundamental frequency differences in these utterances are less clear (Warren 1985; Straub 1997), and durational differences alone can be sufficient for disambiguation (Lehiste, Olive, and Streeter 1976).

When such previous studies have placed ambiguous materials in disambiguating contexts, or when the ambiguity is local, as in (22) and (23), then results regarding the phonetic marking of the structural difference are mixed. When PP ambiguities followed disambiguating paragraphs, Price et al. (1991) still found stronger perceived breaks before the PP in the high-attachment cases. On the other hand, Cooper and Paccia-Cooper (1980) and Straub (1997) found a reduction in durational contrasts between high and low PP attachments when these were presented in disambiguating contexts. Straub encapsulated this finding in her Contingent (Prosodic) Cueing Hypothesis, which proposes that the production system only allocates cognitive resources necessary to provide prosodic disambiguation when other sources of disambiguating information are unavailable.

Snedeker and Trueswell (2003), using a task similar to our own, also argued that PP attachment ambiguities are more clearly distinguished under circumstances in which ambiguity is present for the speaker, and that when such ambiguity is absent then so too is prosodic marking of syntactic structure. Our results in Experiments 1 and 3 above appear to be largely at odds with this result, but we believe that there are a number of factors that might account for the difference. Snedeker and Trueswell discuss three properties of our game task that they believe may contribute to the difference. One is the lexicalization of the phrase *the square with the triangle* in the low-attached cases, due to repetition of this phrase. As reported in the presentation of Experiment 1, we do find some evidence for increasing internal coherence of this phrase across the experiment, but this trend was not consistent across speakers, and to a large extent reflects general increasing speech rate and fluency as the speakers get more practiced with the task. In addition, and in agreement with tasks that did not use extensive repetition of the same utterance (Snedeker and Trueswell 2003; Kraljic and Brennan 2005), we found that the distinction between high- and low-attached utterances did not depend on increasing familiarity or on a process of lexicalization, but rather was present from the first critical utterance of the experiment. Snedeker and Trueswell's analysis of utterances produced on the first trial in their Experiment 1 similarly showed that their speakers' prosody was affected by the intended structure from the outset. Finally, the high-attached utterances in our Experiments 1 and 3 show substantial numbers of full IP boundaries at *square*, suggesting prosodic disambiguation of the high attachments independently of the evaluation of the low attachments.

The second difference that Snedeker and Trueswell highlight between their task and our own previously published data (Schafer et al. 2005) is that the contextual cues available to mark situational ambiguity in our Experiment 1 were more subtle than theirs, since they ran experimental conditions in which only one of the interpretations of the PP ambiguity was possible. We believe that the contrast in game-pieces (between triangle and cylinder utterances) provided a cue demonstrated to be effective in other sentence processing work (Tanenhaus et al. 1995; Trueswell, Sekerina, Hill, and Logrip 1999; Spivey, Tanenhaus, Eberhard, and Sedivy 2002). Our Experiment 3 provides an even stronger response to this by similarly presenting the speakers with only one possible referent of the PP ambiguity in a between-subjects design. The results of our Experiment 3 showed a consistent distinction between high and low attachments, even with no contextual ambiguity. The production data presented by Snedeker and Trueswell for their single referent experiment (their Experiment 2) suggests that disambiguation was still being signalled by their speakers, since many of the item analyses demonstrate significant differences, though reduced in comparison with their two referent experiment. We believe that the pattern of results in two-referent and single-referent experiments both in our own study and in Snedeker and Trueswell's is consistent with a situationally independent model, where disambiguation is primarily due to the grammatical constraints on prosodic production, but awareness of ambiguity can increase disambiguation in a very coarse fashion, as discussed further below.

The final property that Snedeker and Trueswell claim distinguishes their experiments and our own concerns the nature of the utterances used. The PPs in our experiment are longer and more complex than their PPs. They claim that their utterances, at 6 to 9 syllables in length, are less difficult to produce without pausing than our constructions, which are 16 and 17 syllables in length (12 and 13 words). Although we generally agree with this expectation (and believe we were the first to suggest a length effect to them), the data do not seem to support a length-dependent need for a medial break. If we take the simple measure of the average break index at the critical point of attachment (i.e., at the end of the noun immediately before the PP), then their published data, with average break indices of 3.59 in the high-attached cases and 1.48 in the low-attached cases in their first experiment (Table 2, p. 111) are very similar to our own, which are 3.38 and 1.65 respectively in our Experiment 1. For their second experiment, which we replicate within the constraints of our own game design and materials in Experiment 3 in this paper, they report average break indices at these positions of 2.81 and 2.34 for high- and low-attached, for which the corresponding values in our Experiment 3 are 2.94 and 1.32. These data do not suggest that our materials were any more likely to produce more major breaks at the crucial sentence positions, despite their greater overall length.

Snedeker and Trueswell also make reference to corpus evidence that the mean length of an utterance in casual conversation is “about 6.8 words or roughly 8.6 syllables” (p. 126), drawing on analyses of the Swichboard corpus. Using definitions by Bell et al. (2003), their use of “utterance” seems to be interpretable as an intonational unit such as an IP or an ip, although this is clearly implausible with some of the utterances in the corpus, such as those discussed below. If we conservatively use the larger IP to define utterances, then the average length of utterances in our recordings of *I want to change the position of the square with the triangle* in Experiment 1, based on the analysis of our team of five transcribers, is 10.1 syllables. If an “utterance” is co-terminous with the ip, then the average would be lower still. Given that our experimental stimuli also included much shorter utterances (such as *Good job!*), it is plausible that the average length of utterances defined as IPs over the experiment as a whole is not very different from the estimate they provide.

In our own analysis of data from the same Switchboard corpus we focused on the typical length of utterances that contain the sequence V-NP-PP, and yielded a longer average length (mean 21.8 words, median 9 words, mode 17 words). We included all utterances containing the sequence V-NP-PP, where the verb occurred with a following noun and a subsequent PP that had been automatically parsed as either a “child of VP” or a “child of (the subsequent) NP.” We hand-checked the text content of the resulting 3,390 utterances from the corpus, to ensure that the critical sequence occurred. We excluded utterances that did not contain the target string and therefore utterances shorter than 4 words (V, N, P, NP) were not counted. In order to be conservative in estimating average utterance length, we hand-checked the utterance content to evaluate sentence boundaries. If an utterance contained adjacent complete sentences with no intervening connectives, each sentence was counted as a separate item (e.g., the sequence *Plant one of those Takes a lot of room* was counted as two items, with 4 and 5 words, respectively). As is typical of spontaneous speech, many utterances from the corpus contained word sequences that did not create syntactically well-formed sentences. When a syntactically complete sentence was adjacent to a semantically related sequence of words that did not itself form a sentence, we kept the two together as an utterance (e.g., the following “utterance” from the corpus was divided in our analysis into two utterances of 42 and 7 words respectively, as indicated by bracketing: *[And if you do kill a person and it is you know you’re found to be guilty I believe God establishes the authority of our court system and those prison people know they know a genuine turnaround in a person’s character] [I have a degree in social work]*). These divisions yielded a total of 4,273 utterances. Of these, 3.14% were six or fewer words in length, and 33.3% were 12 words or fewer in length (2.3% of utterances had exactly 6 words, and the percentage was the same for those that had exactly 12 words). The bulk of utterances (55%, 2,335 utterances) fell between 7 and 19 words in length. Inspection of the corpus revealed that the spontaneous utterances, like our sentences, were often syntactically complex, containing modified direct objects, multiple PPs, embedded verb phrases, and so forth. We conclude that the 12- and 13-word stimuli used in our experiment yield utterances that are satisfactorily representative of those found in spontaneous speech. If the objective is to determine how frequently naïve speakers of a language prosodically disambiguate PP attachment in spontaneous speech, we believe that our stimuli provide the better approximation to ordinary speech, although we would like to see future research use a range of sentence lengths and structural details.

The length difference can have another effect than the expected strength of medial boundaries in an utterance. It can also influence the location of prosodic breaks inserted to satisfy phonological constraints on the absolute or relative size of prosodic phrases (Cooper and Paccia-Cooper 1980; Gee and Grosjean 1983; Ferreira 1988, 1993; Jun 2003; Watson and Gibson 2005). In Snedeker and Trueswell's items, the mid-point of the utterance, by syllable count, was often at the critical juncture between the head noun of the direct object and the ambiguously attached PP: *(Tap the frog) (with the flower)*. Placing a prosodic boundary at the most disambiguating location for low attachment creates two phrases that are unbalanced for length: *(Tap) (the frog with the flower)*. In our materials, the longer length allows many more options for prosodic phrasing, avoiding a bias against disambiguating prosody for low-attached PPs.

Snedeker and Casserly (2010) provide a distribution of prosodic break patterns for Snedeker and Trueswell (2003), which allows us to compare how boundary patterns shifted between Snedeker and Trueswell's Experiments 1 and 2, versus the American speakers in our Experiments 1 and 3. These patterns are graphed in Figure 12, as the proportion of high- or low-attached productions within an experiment realized with a specific boundary pattern. Like Snedeker and Casserly, we excluded cases with the ambiguous break index of 2. The first boundary in a set such as (0, 0) indicates the boundary strength prior to the direct object head noun. For their materials, this was a boundary between the verb and critical noun; for ours it is the strongest boundary anywhere in the pre-noun region. The second boundary in a set is always the boundary between the critical noun and the ambiguously attached PP. Zero indicates a word-level boundary or smaller (break index 0 or 1 in ToBI), ip an intermediate phrase, and IP an intonation phrase.

Inspection of Figure 12 reveals, first, that their Experiment 1, in which participants were told to disambiguate, elicited stronger disambiguation than our Experiment 1, in which no instructions about disambiguation were provided. They show high proportions of tokens containing an intonation boundary at the most disambiguating position and no more than word-level boundaries at any other position (upper left panel). In our materials (lower left panel), speakers made more frequent use of the smaller intermediate phrase boundary in the disambiguating position for low attachments, and more frequent use of an (ip, IP) pattern for high attachments,

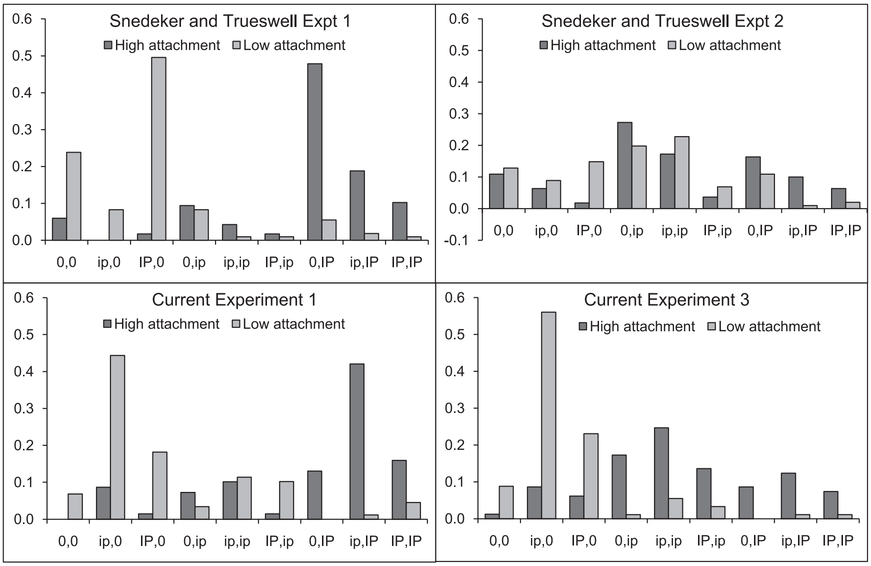


Figure 12. *Distribution of boundary type combinations in PP attachment utterances in two experiments reported by Snedeker and Trueswell (2003) and in the two PP attachment experiments reported in the current paper. See text for details.*

in which the boundary at *square* is the strongest in the utterance, but phrase lengths are better balanced (into three or more prosodic phrases) by having one or more intermediate phrase boundaries earlier in the sentence. Turning to the experiments with between-participants manipulation of attachment, our Experiment 3 maintains a clear difference between high attachments and low attachments, although with generally smaller boundary strengths across the board (lower right panel) than in our Experiment 1. Snedeker and Trueswell's Experiment 2 shows a much weaker distinction between high and low attachments, as seen by the strong degree of overlap in the distribution of boundary patterns. There are also high numbers of tokens using what we might consider phonologically neutral patterns: (0, 0), with no internal breaks, (0, ip) and (0, IP) with one break at the utterance mid-point, and (ip, ip) with breaks at the end of each content word. These are the patterns we would expect if the speakers are not engaged in production processes with a clear syntax in mind – perhaps a form of "good enough" processing (Ferreira, Ferraro, and Bailey 2002; Ferreira and Patson 2007).

We think it is crucial to understanding the differences across PP experiments that in Snedeker and Trueswell's first experiment, where there were two possible referents for the PP construction, instructions were given to the speaker to disambiguate. They were told "to say each sentence in such a way as to get the Listener to perform the same action on the other side of the screen" (Snedeker & Trueswell

2003: 106). In other words, when speakers saw two possible referents of the PP in front of them, watched the experimenter move the objects, and were then handed the card with the sentence they had to use as an instruction, they would have become aware that there was an ambiguity they were supposed to resolve on the basis of their pronunciations (and indeed Snedeker and Trueswell report that in a post-experiment questionnaire 97% of their speakers reported having been aware of the ambiguity). We note also that having the same very simple syntax with a constantly changing array of objects for description on successive trials may well increase the likelihood that speakers will abstract away from individual sentence meaning to note a syntactic pattern in the trials. As a consequence of these factors, it is likely that Snedeker and Trueswell's speakers may have produced deliberate prosodic disambiguation similar to that found in laboratory speech. This would have been more likely in their first experiment, where visual referents were available for both interpretations of the PP ambiguity, than in the second experiment, where only a single referent was available to each speaker (for their Experiment 2, the post-experiment questionnaire showed just 31% of the speakers reported awareness of the ambiguity). Thus, speakers in Snedeker and Trueswell's experiments acted like subjects in other traditional psycholinguistics production experiments, and disambiguated more deliberately when their attention was drawn to the ambiguity, or when they were instructed to disambiguate. It is useful to have such demonstrations of what speakers can (and cannot) do when asked, as well as what exactly they produce. Yet this is different from assessing what speakers automatically and routinely do under varying conditions of situational ambiguity, as we hope to have probed in our study.

A further difference between the procedures of experiments that have shown prosodic disambiguation of syntactic PP attachment (our Experiment 3, and Kraljic and Brennan 2005), and the technique used by Snedeker and Trueswell is that in their task there was no conversation, and no need for pre-linguistic or content-based planning – especially in their second experiment, the speaker made no decision about a communicative strategy for solving a problem, and had no reason to try to convey his/her own ideas and plans to a listener whose correct or incorrect comprehension made a difference in the situation. Thus, both the lack of speaker interaction and the lack of a strong need to communicate a message may have affected the productions, especially in the second experiment. One set of experiments casts doubt on whether speaker interaction can account for differences in disambiguation. Kraljic and Brennan (2005) found comparable prosodic disambiguation in ambiguous and unambiguous situations when conversants sat side-by-side with a partner viewing common display, when a lone speaker addressed a tape recorder, and for speakers who were skilled at detecting ambiguity in the visual scene as well as for those who were below chance at doing so. They found this with stimuli that were slightly longer than those tested by Snedeker and Trueswell (2003) and others, but shorter than ours – sentences such as *Put the dog in the basket on the star*. This suggests that, given materials that afford natural differences in phrasing (Nespor and Vogel 1986), the task of planning a message (at some level higher than mere lexical access and phonetic implementation) may be a key distinguishing factor between disambiguating productions and ambiguous ones.

## 10. General discussion

The experiments presented here have investigated whether the prosodic marking of syntactic structure in production is situationally dependent. A strong situationally dependent model of the prosody-syntax relationship in language processing would entail a specific pattern of variation in correspondence between prosodic and syntactic structures, so that speech situations in which syntax is ambiguous would be exactly those situations in which prosody marks syntactic structure, and unambiguous situations would not result in such prosodic cues. In contrast, we predicted that situational independence of prosody in production would mean that the grammatical correspondences between linguistic structures would regularly determine aspects of the prosodic realization of utterances. We take "grammatical correspondences" to include far more than simply the phrase structure of a sentence, as we describe further below. Nevertheless, we assume that there are significant correspondences between prosodic phrasing and syntactic phrasing, and therefore we predicted that we would see significant – although not invariant – effects of syntactic form on prosodic structure.

Our phonetic and phonological analyses of prosodic boundaries in the PP experiments (Experiments 1 and 3) revealed remarkable consistency in their size and location across different situational contexts, with high attachment pronounced with a stronger break at *square* than low attachment for ambiguous phrases such as *I want to change the position of the square with the triangle*, and each attachment receiving pronunciations expected to be highly biasing (e.g., the strongest prosodic boundary located just prior to the PP for high attachments). These experiments showed that such prosody-syntax correspondences were not dependent on functional load – the differences between high- and low-attachment structures were present both for Drivers who issued instructions and for Sliders who confirmed that those instructions had been followed. Nor was this pattern of results dependent on fluctuating ambiguity within the game – the prosody-syntax correspondences were just as strong in cases of ambiguity as in cases where the configuration of gamepieces ruled out one of the possible referents, and for when there was no competing gamepiece in the game (i.e., no combined square+cylinder piece). Although there were some indications of greater disambiguation for low-attached triangle utterances at later versus earlier points in game play, these effects appear to be due to a general tendency for greater cohesion of the phrase with repetition (especially for adjacent repetitions of the same instruction). The best indication for situational effects comes from the contrast between Experiment 1 and Experiment

3. Yet even here, our speakers showed significant differences between high and low attachments in a between-subjects manipulation of attachment, in all of our production measures. These data indicate a production system in which non-situational factors play a fundamental role in prosodic implementation, supplemented by coarse-grained changes when a speaker becomes aware of a need to disambiguate. We assume that the former effects stem from automatic processes in the grammatical encoding system, while the latter may come from the use of monitoring and more conscious encoding (Ferreira 2007). The conclusions from our production experiments were fully supported by the results of the categorization study (Experiment 2).

The comparison of Experiments 1–3 on the one hand with Experiments 4 and 5 on the other shows that reliable prosody-syntax correspondences are produced for both standing (PP attachment) ambiguities and temporary (clause closure) ambiguities. Our phonetic and phonological analyses in Experiment 4 showed that a stronger prosodic boundary was present at the syntactic clause boundary than at the alternative critical position for closure ambiguities. The categorization results in Experiment 5 showed that listeners could correctly identify the syntactic structure of spoken fragments excised from their syntactic contexts. In other words, prosody disambiguates phrase structure even when the intended structure of an ambiguous word-string is indicated by morphosyntactic information immediately following that ambiguous string and by other information available in the discourse situation. The indications we found of a clearer distinction in the production of prosodic contrasts for the Closures than the PPs are similar to the findings for Closure versus PP attachments in earlier research using less interactive speaker situations (Lehiste 1973; Warren 1985; Price et al. 1991). It thus appears that prosody is at least as strong an indicator of syntactic structure for the temporary clause closure ambiguity as for the standing PP ambiguity, even with the switch to a discourse context that called for effective communication between participants and mimicked many features of natural conversation.

Together these five experiments present compelling evidence that effects of syntactic phrasing on prosodic structure are only weakly or coarsely influenced by situational demands. There are at least four points that we believe are relevant to this conclusion. First, we tested several kinds of situational ambiguity. For each situational variable manipulated, even at the least ambiguous level of the factor, speakers produced prosodic structures that resolved syntactic ambiguity. This suggests that some factor other than situational ambiguity is an important determinant of prosodic form. Second, we saw little evidence of prosodic disambiguation increasing with increasing situational ambiguity. This did not appear to be because of ceiling effects on prosodic disambiguation – the PPs were disambiguated less strongly than the Closure sentences according to our production analyses, and there were many PP tokens in which the speaker could have opted for a more obviously disambiguating pattern than the one chosen (such as the use of an IP boundary at the critical location, with no boundaries stronger than word-level boundaries in other locations). Third, we found in addition to the absence of significant effects of situation within Experiments 1 and 2 a reversal of the predicted effects of situation when we compared transcription patterns for Experiments 1 and 4, looking at PP attachment versus Closure utterances recorded in the same game-playing sessions. Such a reversal counters concerns that the evidence against a situational ambiguity effect depends on null results (Kraljic and Brennan 2005). The prosodic pattern we found for the two syntactic types is explained by a situationally independent model but not a situationally dependent one, as is the relatively weak prosodic disambiguation other studies have found in another type of syntactic ambiguity, the NP-complement versus sentence complement ambiguity (Beach 1991; Anderson and Carlson 2004; but see also Warren 1985; Marslen-Wilson, Tyler, Warren, Grenier, and Lee 1992; Stirling and Wales 1996; Watt and Murray 1996). Fourth, we see logical problems with situationally dependent models of prosody, which these data help illustrate.

According to a strongly situationally dependent model, speakers only produce prosodic disambiguation to remove ambiguity for the listener. Speakers' production of prosodic cues to syntactic structure might then be relatively infrequent, assuming that other sources of disambiguating information are fairly common. Because every sentence has some type of prosodic structure, a situationally dependent model is most plausible if there is a subset of infrequently produced prosody-syntax correspondences which are strongly predictive of certain parsing choices and which are likely to occur in the critical location only under certain circumstances, such as in the absence of further disambiguating information sources. Snedeker and Trueswell (2003) argue that this is likely for IP boundaries at the left edge of a high-attached PP, based on the prosodic boundary patterns in their six-to-nine syllable PP attachment utterances. A thorough evaluation of the frequency and consistency of prosodic cues to syntactic structure requires the examination of a broader range of utterance types and the full set of prosodic events that can be used as cues to syntactic structure. Current evidence for English suggests that each of the following prosodic events may affect the comprehension of syntactic form: pitch accent location and type (e.g., Schafer, Carter, Clifton, and Frazier 1996; Carlson 2002), intermediate phrase boundary location (e.g., Schafer 1997; Kjelgaard and Speer 1999; Clifton et al. 2002; Snedeker and Casserly 2010), and intonation phrase boundary location (see Cutler et al. 1997; Warren 1999; Speer et al. 2003; Speer and Blodgett 2006 for reviews). Other prosodic contrasts exist but have not yet been tested. Studies of the frequency of intermediate phrase boundaries for English suggest that they occur, on average, every 5.3 syllables (Ueyama 1998) or roughly every 4 content words (Ayers 1994). Each intermediate phrase boundary in English entails the presence of at least one pitch accent, but in certain contexts pitch accents may occur on every content word. At least in some constructions intermediate phrase boundaries are as disambiguating as full intonation phrase boundaries (Kjelgaard and Speer 1999; Clifton et al. 2002; Blodgett 2004; Snedeker and Casserly 2010).

Together these findings suggest that prosodic events that are relevant to interpretation are quite frequent, at least in utterances of lengths found in typical conversation (see our corpus analysis, above). Producing an ambiguous prosody, therefore, is not generally a matter of leaving off rare and optional aspects of production. Rather, it is a case of constructing an ambiguous prosodic form instead of some other, less ambiguous form (Beckman 1996). Once the picture is expanded beyond IP boundaries in short tokens to encompass the full range of prosodic cues and a broader class of utterances, and once it is recognized that grammatical correspondences to prosodic form include more than syntactic phrase structure, learning the associations between prosodic events and other grammatical structures seems far less simple for a situationally dependent model. Although the bulk of the evidence for prosodic cues to disambiguation comes from controlled productions by trained speakers, our data support the conclusion that a complex set of prosodic cues influence interpretation, as shown for example by the better than chance performance for Closure sentences with equal strength prosodic boundaries at the clause and non-clause critical positions. Thus, while prosodic disambiguation is sometimes characterized as a simple process of inserting a large break in a disambiguating position, we believe that the production of prosody involves adopting a complex phonological structure that, like syntactic structure, adheres to well-formedness constraints and requires planning (e.g., Beckman and Pierrehumbert 1986; Shattuck-Hufnagel 2000). Research on comprehension indicates that it is the overall pattern of prosodic structure that determines the degree of disambiguation, and not simply the presence or absence of a boundary at one point in the utterance (Schafer 1997; Schafer et al. 2000; Carlson et al. 2001; Clifton et al. 2002; Snedeker and Casserly 2010).

A case of situational dependence has to be one in which it is possible to produce a disambiguating prosody. What would be the advantage of *not* choosing such a prosody in syntactically unambiguous situations, leaving aside other grammatical factors like the indication of informational structure or length constraints on prosodic phrase size? It is certainly plausible that some prosodic structures are easier to produce than others. It strikes us as less plausible that ambiguous prosodies are easier to produce, as a class, than disambiguating ones – and especially implausible that any labor saved would be worth the effort to the production system of tracking the ambiguity of the situation for the listener and the range of prosodic structures that are possible. Disambiguating phrasing has been seen in the psycholinguistic literature as a tendency to produce prosodic boundaries at the initiation or completion of long syntactic phrases that are challenging for the production system (Cooper and Paccia-Cooper 1980; Gee and Grosjean 1983; Ferreira 1988, 1993; Watson and Gibson 2004), suggesting a functional advantage of such phrasing for the production system. In contrast, tracking the ambiguity of planned speech with respect to the discourse situation would require several steps of planning. The production system would need the ability to predict the syntactic and lexical form of upcoming material in enough detail to evaluate its ambiguity, and then would have to carry out an appropriate monitoring procedure to determine that the planned material is in fact linguistically ambiguous, and moreover that this linguistic ambiguity would not be resolved by a number of non-linguistic factors for the given listener. The production system would not be able to make phrasing decisions simply on the basis of approximate plans, such as the knowledge that the next phrase is long or syntactically complex. It would instead have to use fairly detailed ones, such as the knowledge the next phrase is large and will contain a PP attachment ambiguity that is intended with low attachment but could plausibly be interpreted with high attachment by the listener because the scene contains an object that could plausibly be used as an instrument and the particular preposition used in the utterance allows either attachment. This is extensive computation for a system that appears to be highly incremental. The absence of systematic adjustment of prosody for situational ambiguity in our data is in concord with the prosodic findings of Kraljic and Brennan (2005) and fits with other recent findings that suggest that speakers do not systematically avoid ambiguity by adding optional words (Ferreira and Dell 2000; Jaeger 2010; but see also Haywood, Pickering, and Branigan 2005) or changing word order (Arnold, Tanenhaus, Altmann, and Fagnano 2004), two other disambiguating cues that would require fine-grained adjustments and sophisticated planning. Some of these experiments showed evidence of a general strategy to increase clarity under instructions that emphasized successful communication, but not a specific strategy of adding or re-ordering words in just those sentences that were ambiguous.

In our experiments prosodic phrasing regularly corresponded to aspects of syntactic constituent structure. While this finding supports the importance of such relationships, we believe that viewing prosodic phrasing as merely a reflex of syntactic phrasing, or of a need to resolve syntactic ambiguity, fails to account for the range of findings in the literature and leads to an impoverished model of sentence processing. Instead, our results and those of other experiments suggest that prosody-syntax associations are always a factor in producing sentences, a state we have described as situationally independent production of prosody. This does not entail that all sentences are reliably disambiguated by prosody, or that those structures that can be prosodically distinguished are disambiguated in all tokens. Some syntactic ambiguities may not have structures that are readily separated by the prosody-to-phrase-structure mappings in the grammar. Relatively short utterances may tend to occur in a single intonation phrase or even a single intermediate phrase in many discourse conditions, following phonological or performance constraints that lead to a simpler prosodic structure unless the length or complexity of the utterance requires additional phrasing (Gee and Grosjean 1983; Truckenbrodt 1999). Of course, if the speaker does not have a particular syntactic structure or its associated semantic structure in mind during a production task, the prosody will be less likely to reflect the syntax chosen by the experimenter. Moreover, our phonological analyses noted a wide range of suprasegmental realizations for the utterances under investigation. Despite the relative consistency of phrasing over a range of different contextual ambiguities, as measured by the locations and types of breaks between prosodic constituents, we found extensive variability in tonal realizations. This variability undoubtedly stems from the openness of tonal marking to variation in speech style and to differences in the selection of information-bearing elements. As a consequence, it is clear that the prosodic representation of an utterance is not fully predictable from the syntax. Even a disambiguated syntactic structure can be associated with several prosodic representations, which vary in such things as the location of pitch accents and the choice of high versus low pitch accents and phrasal tones.

How can we reconcile the consistency and variability found in prosodic form, both in our data and in the literature? Clearly, the syntactic organization of a sentence is an important determinant of prosodic form, but not the only one. We view prosody as a critical part of sentence and discourse production which reflects a joint conversation space created by the speaker and hearers in response to the larger situation. We pointed out above that every sentence has some type of prosodic structure. More specifically, prosody is a multi-dimensional component of the speech stream that can be analyzed into sub-components and has internal constraints that specify a well-formed structure.[12] This structure, we believe, is a frame for speech that simultaneously performs several functions. At the syntactic level it frequently groups heads of phrases with their arguments, in both head-initial and head-final languages (Selkirk 1986; Truckenbrodt 1999; Jun 2003), and separates distinct phrases. This has obvious advantages for comprehension, and as mentioned above, may be a synchronic or historical reflex of production planning mechanisms. At the discourse level, prosodic structure is used to convey distinctions in foregrounding, topicalization, emphasis and the like, through the choice and location of pitch accents in languages like English and variation in phrasing and other cues (see Venditti and Hirschberg, 2003, for a review), as well as conveying important affective distinctions. It also seems to function as a framework or "rhythmic scaffolding" (Arbisi-Kelm and Beckman 2009) for the reception and specification of phonological and phonetic information at lower levels of perception and production processes (e.g., Shattuck-Hufnagel and Turk 1996; Shattuck-Hufnagel 2000 and references therein). Thus, prosodic structure helps organize material in working memory both in production and in comprehension (Ferreira 1988; Speer, Crowder, and Thomas 1993).

Because prosody reflects multiple levels of linguistic structure simultaneously, one danger in its investigation is to concentrate too narrowly on patterns at just one level. A prosodic form that is particularly effective in conveying one type of contrastive focus may do so by making less distinctive the prosodic marking of following material, leaving the syntactic structure of that material less differentiated; another information state may allow for a particularly clear indication of syntactic form (Schafer and Jun 2001, 2005; Jun 2003). Thus, while tightly controlled laboratory speech situations are extremely useful in identifying some of the effects of prosodic structure, the lack of a rich discourse situation in many experiments may underdetermine the appropriate interpretation of prosodic patterns and underestimate prosody's importance (as well as the number and types of linguistic levels prosody reflects). A sentence may contain extremely useful cues to the intended meaning, but in certain cases those cues may be primarily useful for the information structure, not the syntactic structure, or only useful for the syntactic structure given certain assumptions about the discourse structure.

Models of the speech production process (e.g., Levelt 1989; Ferreira 1993) see the generation of prosodic structure as an early part of the planning of speech output, preceding the development of detailed syntactic structure. This being so, giving subjects a complete utterance to repeat – as in a reading task – will result in an unusual sequence of phases of production in utterance construction, and possibly therefore in atypical outputs. Kraljic and Brennan (2005) avoided this complicating factor by giving their participants a constrained scenario to describe with specified object names, but without constraining the available syntactic choices. Studies such as that of Snedeker and Trueswell (2003) manipulated the situational constraints on ambiguity, but continued to constrain tightly the lexical and syntactic choices available to participants by requiring them to repeat text presented on cards. We believe that our study sits between these two with respect to syntactic planning, in that we did not provide the entire content of the utterances to be used by our participants (since this depended on how they chose to pursue the goals of the task), though we did provide the syntactic frames. Further, our sentences were produced as part of a problem-solving task cooperatively carried out by a pair of interacting speakers. Each utterance was, we assume, produced with an unambiguous message in mind as part of a conversation. The message required not just an appropriate syntactic structure but also appropriate markings of how to connect the utterance to the conversation, which employed varying states of contrast or emphasis for referents, varied syntactic structures, and utterances that conveyed a mix of instructions, questions, commentary, and other discourse functions. As such, we suspect it was easier for our participants to provide prosodic structures typical of spontaneous speech than those in other testing situations. Although we find considerable variability in some aspects of our productions, this is exactly what we would expect if prosody is systematically indicating pragmatic contrasts as well as syntactic ones.

From this perspective, in which prosodic form reflects multiple linguistic levels of analysis, there is a sense in which the production of prosody *must* be dependent on the situation. That is, while prosody is not determined solely by situational ambiguity – what we have described as situationally dependent production – neither is it invariant across situations. Discourse factors such as whether a word presents given or new information affect the likelihood of prominence on that word (e.g., Lieberman 1963; Fowler and Housum 1987; Bard et al. 2000; Dahan, Tanenhaus, and Chambers 2002), and the pattern of pitch accents interacts with the choices for prosodic phrasing, as mentioned above. Speakers can clearly change their speech in response to their audience, e.g., providing the enhanced prosody characteristic of child-directed speech (e.g., Fernald et al. 1989) or matching the tempo of previous speech (Jungers, Palmer, and Speer 2002). Critically, though, we believe there is a tendency to reflect prosody-syntax relationships regardless of the situational ambiguity. Under some discourse settings these emerge strongly, while under others they may be overridden by the need to indicate other aspects of the message. In cases where a speaker is told to prosodically disambiguate, several changes in production may or may not take place. First, a speaker, especially one who is reading aloud, may shift attention from careful articulation at the lexical level to a sentential level, resulting in change in prosodic emphasis. Second, the speaker may become more aware of the contrast in meanings, and thus more capable of delivering a helpful prosody via normal production mechanisms. Third, the speaker may make coarse changes, such as decreasing the rate of speech or expanding the pitch range, that felicitously lead to stronger or more salient utterance-internal prosodic boundaries in disambiguating positions. Finally, the speaker may adopt fine-grained purposeful changes in prosody-syntax correspondences to indicate the desired form. We have not yet seen convincing evidence that the last of these changes accounts for the apparent cases of situational adjustment of prosody reported in the literature. This lack of evidence leaves open the possibility that untrained speakers never make detailed adjustments in their prosody to indicate syntactic structure. Speakers do seem to make adjustments to other aspects of their production for the benefit of their audiences (e.g., Brown and Dell 1987; Schober and Clark 1989; Brennan and Clark 1996; Horton and Keysar 1996; Lockridge and Brennan 2002). However, if speakers provide any sort of audience design in their prosodic disambiguation of syntax, it seems to be a non-primary factor in the production process.

## 11. Conclusion

We have argued on the basis of five experiments and two varieties of English that the production of prosody is not solely dependent on the contextually-based need for syntactic disambiguation and is better explained by grammatical form, perhaps in conjunction with its effects on production difficulty. Our results replicate and extend results by Kraljic and Brennan (2005) but challenge earlier research (Allbritton et al. 1996; Fox Tree and Meijer 2000; Snedeker and Trueswell 2003). Two factors seem to distinguish our results and those of Kraljic and Brennan from the others. The first is the use of more naturalistic tasks that allowed for quasi-spontaneous production of speech by participants who had clear sentence-level meanings to convey. The second is the use of somewhat longer and more complex sentences (for our PP studies, although not the clause closure ones) which may have encouraged the use of internal prosodic breaks and, crucially, which appear to be more typical of everyday speech. Length and complexity are well-known as important factors in prosodic phrasing (Cooper and Paccia-Cooper 1980; Gee and Grosjean 1983; Ferreira 1993; Watson and Gibson 2004, 2005). Since length constraints on prosody differ across languages (Jun 2003), this may be a particularly productive factor to explore cross-linguistically. Crucially, our experiments showed disambiguation that was just as strong, if not stronger, for an ambiguity that was always resolved by the situation (the clause closure ambiguity) than for one which was not (the PP attachment ambiguity), providing positive evidence against situationally determined prosodic form.

Our results suggest that correspondences between structures and other aspects of grammatical form are pervasive in natural speech, yet complexly determined. Prosody conveys information about multiple levels of linguistic analysis, providing redundant and simultaneous conveyance of phonetic, phonological, word-level, syntactic, semantic, pragmatic, sociolinguistic, and affective information. A complete model of language processing must ultimately account for all of these levels, cross-linguistically, as well as the way they interact in the production and comprehension of language of different speech styles, from monologues to conversations at various levels of formality. At least some of the correspondences between prosodic structure and other levels of linguistic structure are readily used by comprehenders; most notably, the relative strength of prosodic boundaries within an utterance provides helpful information about syntactic phrasing, although other prosodic contrasts seem to contribute to syntactic categorization as well. Further research on prosodic comprehension will be needed to explore the relationship between production patterns and their use by listeners, and further research on prosodic production will be needed to evaluate the relationship between grammatical form and production demands. Finally, although we have emphasized the usefulness of linguistic assumptions in describing prosodic form, they are clearly incomplete. Linguistic models have illuminated the basic relationship between syntactic phrase structure and prosodic phrasing. Our data show strong support for common prosody-syntax assumptions, but they also show that other suprasegmental factors affect syntactic categorization, for example, giving better than expected categorization in cases of "neutral" or "conflicting" boundaries. Linguistic theories need to pay further attention to properties such as pitch accent and pitch range that may account for these results. And, both linguistic and psycholinguistic theories have much additional work to do on the interaction of syntax with other factors in the construction of prosodic form.

## Authors' Note

All authors contributed equally to the work reported in this paper. The author order follows our most recent rotation of alphabetical order. Authors can be contacted by email at speer@ling.ohio-state.edu, paul.warren@vuw.ac.nz, and amy.schafer@hawaii.edu. This work was supported by NIH training grant DC-00029 (Schafer), NIH research grant MH-51768 and NSF grant 0088175 (Speer), and by NZ/USA Cooperative Science Programme grant CSP95/01 and Marsden Fund grant VUW604 (Warren). The authors would like to thank the following for the assistance with prosodic transcription and data collection and analysis: Amy Austin, Kelly Barrow, Karen Carmody, Nicola Daly, Amanda Fisch, Christa Hansen, Hyekyung Hwang, Lauren Kling, Jenny Kneale, Jennifer Ludlow, John Mast-Finn, Ben Parrot, Cara Prall, Lisa Rief, Shari Sokol, Aaron Soltz, Jill Story, Patricia Vermillion, S. David White, Gerald Whiteside, and Frank Wilson. Thanks also to Chris Brew and Adam Baker for help with corpus analysis.

Ilse Lehiste, whose pioneering research on prosody and processing laid the groundwork for the studies reviewed and presented here, died on December 25, 2010. This paper is dedicated to her memory.

## Appendix

Appendices are available online at http://dx.doi.org/10.1515/LABPHON.2011.002.

Correspondence e-mail address: speer@ling.ohio-state.edu

## Notes

1. For instance, Allbritton et al. (1996) use the example "When you learn gradually you worry more". Note that no commas were included in the materials given to readers, and that the preceding sentence, "As you begin to study about nuclear war it becomes frightening" did not provide a particularly strong bias against a Late Closure interpretation.
2. Our game task also allowed us to investigate two further types of commonly studied structural ambiguities, which are not reported here.
3. As we will discuss below, we assume that multiple factors affect prosodic realization. For example, sentences (1) and (2) differ in length, and syntactic phrase length influences prosodic phrasing. These additional factors should be kept in mind as potential confounds when comparing and contrasting these structures. Nevertheless, directly comparing these two types of ambiguity in a single study allows us to begin to tease apart the potential factors that affect their prosodic form, and the relative strengths of those factors.
4. Appendices and statistical tables are available at http://dx.doi.org/10.1515/LABPHON.2011.002.
5. One further pair of AmE participants had considerable difficulty planning sufficiently strategic moves in the game, and produced very few utterances, and so their recordings were also excluded.
6. We thank an anonymous reviewer for reminding us that Break Indices (BI) of strength 2, as defined in the original ToBI guidelines used here (Beckman and Ayers 1997), are not always of intermediate strength between BIs of 1 and 3. Our use of BI 2 was consistent with a boundary that was weaker than an unambiguous intermediate phrase (ip) boundary, and we therefore placed the SBL at the location of a BI of 3 in the very few cases where the critical comparison for SBL was between a BI 2 and a BI 3.

7. In earlier analyses of our data we did run such ANOVAs and found substantially the same significant effects as reported for the *lmer* analysis here.
8. Although techniques have been developed for *lmer* analysis of multinomial data (Arppe 2008), the highly uneven distribution of tokens across our three levels of SBL meant that we were not able to build reliable multinomial models.
9. We thank Gary Dell for first mentioning this possibility to us, and Jesse Snedeker, Mike Tanenhaus, and John Trueswell for similar comments.
10. Some breaks before *square* in the high-attached utterances may serve to balance the lengths of prosodic phrases, or to avoid long phrases (Gee and Grosjean 1983; Nespor and Vogel 1986). Such factors may account for some of the remaining variability in break location.
11. A separate analysis comparing the three English Varieties involved, i.e., Kansas and New Zealand (Experiment 1) and Ohio (Experiment 3), showed a significant interaction of Variety and Attachment that was attributable to the differences between Kansas and New Zealand on the one hand and Ohio on the other, i.e., was effectively the experiment difference reported above.
12. Even when the prosodic structure is ill-formed, it can reflect useful information about the production process for the comprehender, such as the likelihood that material following a hesitation pause is new information (e.g., J. E. Arnold et al. 2004).

## References


Allbritton, David W., Gail McKoon, & Roger Ratcliff. 1996. Reliability of prosodic cues for resolving syntactic ambiguity. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22. 714–735.

Anderson, Catherine, & Katy Carlson. 2004. Prosodic phrasing in DO/SC and closure sentences. Paper presented at the 17th Annual CUNY conference on human sentence processing, University of Maryland.

Arbisi-Kelm, Timothy, & Mary E. Beckman. 2009. Prosodic structure and consonant development across languages. In Marina Cláudia Vigário, Sónia Frota, & Maria João Freitas (eds.), *Phonetics and Phonology*, 109–136. Amsterdam: John Benjamins.

Arnold, Jennifer A., Thomas Wasow, Ash Asudeh, & Pete Alrenga. 2004. Avoiding attachment ambiguities: The role of constituent ordering. *Journal of Memory and Language* 51. 55–70.

Arnold, Jennifer E., Michael K. Tanenhaus, Rebecca J. Altmann, & Maria Fagnano. 2004. The old and thee, uh, new: Disfluency and reference resolution. *Psychological Science* 15(9). 578–582.

Arppe, Antti 2008. *Univariate, bivariate, and multivariate methods in corpus-based lexicography – a study of synonymy.* University of Helsinki.

Ayers, Gayle. 1994. Discourse functions of pitch range in spontaneous and read speech. *OSU Working Papers in Linguistics* 44. 1–49.

Baayen, R. Harald. 2008. *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press.

Baayen, R. Harald, Doug J. Davidson, & Douglas M. Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59(4). 390–412.

Baayen, R. Harald, & Petar Milin. In press. Analyzing reaction times. *International Journal of Psychological Research*.

Bard, Ellen Gurman, Anne H. Anderson, Catherine Sotillo, Matthew Aylett, Gwyneth Doherty-Sneddon, & Alison Newlands. 2000. Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language* 42(1). 1–22.

Beach, Cheryl 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, Mary E. 1996. The parsing of prosody. *Language and Cognitive Processes* 11. 17–67.

Beckman, Mary E., & Gayle Ayers. 1997. Guidelines for ToBI labelling. Unpublished manuscript, Ohio State University, Columbus.

Beckman, Mary E., & Janet B. Pierrehumbert. 1986. Intonational structure in Japanese and English. *Phonology* 3. 255–309.

Bell, Allan. 1984. Language style as audience design. *Language in Society* 13(2). 145–204.

Bell, Alan, Daniel Jurafsky, Eric Fosler-Lussier, Cynthia Girand, Michelle Gregory, & Daniel Gildea. 2003. Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *Journal of the Acoustical Society of America* 113(2). 1001–1024.

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

Blodgett, Allison. 2004. The interaction of prosodic phrasing, verb bias, and plausibility during spoken sentence comprehension. Ph.D. dissertation, The Ohio State University.

Brennan, Susan, & Herbert H. Clark. 1996. Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22(6). 1482–1493.

Brown, Paula M., & Gary S. Dell. 1987. Adapting production to comprehension: The explicit mention of instruments. *Cognitive Psychology* 19. 441–472.

Carlson, Katy. 2002. *Parallelism and Prosody in the Processing of Ellipsis Sentences*. London and New York: Routledge.

Carlson, Katy, Charles Clifton, & Lyn Frazier. 2001. Prosodic boundaries in adjunct attachment. *Journal of Memory and Language* 45. 58–81.

Clark, Herbert H., & Michael F. Schober. 1992. Understanding by addressees and overhearers. In Herbert H. Clark (ed.), *Arenas of Language Use*, 176–197. Chicago: University of Chicago Press.

Clifton, Charles, Katy Carlson, & Lyn Frazier. 2002. Informative prosodic boundaries. *Language and Speech* 45. 87–114.

Clifton, Charles, Shari R. Speer, & Steven Abney. 1991. Parsing arguments: Phrase structure and argument structure as determinants of initial parsing decisions. *Journal of Memory and Language* 30. 251–271.

Cooper, William E., & Jeanne Paccia-Cooper. 1980. *Syntax and Speech*. Cambridge, MA: Harvard University Press.

Cutler, Anne, Delphine Dahan, & Wilma van Donselaar. 1997. Prosody in the comprehension of spoken language: A literature review. *Language and Speech* 40. 141–201.

Cutler, Anne, & Stephen D. Isard. 1980. The production of prosody. In B. Butterworth (ed.), *Language Production, Vol. 1: Speech and Talk*. New York: Academic Press.

Dahan, Delphine, Michael K. Tanenhaus, & Craig G. Chambers. 2002. Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language* 47. 292–314.

Fernald, Anne, Traute Taeschner, Judy Dunn, Mechthild Papousek, Bénédictine de Boysson-Bardies, & Ikuko Fukui. 1989. A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language* 16. 477–501.

Ferreira, Fernanda. 1993. Creation of prosody during sentence production. *Psychological Review* 100. 233–253.

Ferreira, Fernanda, Vittoria Ferraro, & Karl G. D. Bailey. 2002. Good-enough representations in language comprehension. *Current Directions in Psychological Science* 11. 11–15.

Ferreira, Fernanda, & Nikole Patson. 2007. The good enough approach to language comprehension. *Language and Linguistics Compass* 1. 71–83.

Ferreira, Maria Fernanda. 1988. Planning and timing in sentence production: The syntax-to-phonology conversion. Ph.D. dissertation, University of Massachusetts, Amherst.

Ferreira, Victor S. 2007. How are speakers' linguistic choices affected by ambiguity? In Antje Meyer, Linda Wheeldon, & Andrea Krott (eds.), *Automaticity and Control in Language Processing*, 63–92. Hove: Psychology Press.

Ferreira, Victor S., & Gary S. Dell. 2000. The effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology* 40. 296–340.

Fowler, Carol, & Jonathan Housum. 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.
Fox Tree, Jean E., & Paul J. A. Meijer. 2000. Untrained speakers' use of prosody in syntactic disambiguation and listeners' interpretations. *Psychological Research* 63. 1–13.
Frazier, Lyn. 1987. Sentence processing: A tutorial review. In Max Coltheart (ed.), *Attention and Performance,* Vol. XII. Hillsdale, NJ: Erlbaum.
Frazier, Lyn, & Charles Clifton. 1996. *Construal*. Cambridge, MA: MIT Press.
Gahl, Susanne, & Susan M. Garnsey. 2004. Knowledge of grammar, knowledge of usage: Syntactic probabilities affect pronunciation variation. *Language* 80. 748–775.
Gee, James, & François Grosjean. 1983. Performance structures: A psycholinguistic and linguistic appraisal. *Cognitive Psychology* 15. 411–458.
Haywood, Sarah L., Martin J. Pickering, & Holly P. Branigan. 2005. Do speakers avoid ambiguity during dialogue? *Psychological Science* 16. 362–366.
Henderson, Alan. 1980. Juncture, pause and intonation fall and the perceptual segmentation of speech. In Hans W. Dechert & Manfred Raupach (eds.), *Temporal Variables in Speech: Studies in Honour of Frieda Goldmann-Eisler*, 199–226. The Hague: Mouton.
Horton, William S., & Boaz Keysar. 1996. When do speakers take into account common ground? *Cognition* 59. 91–117.
Howell, Peter, & Karima Kadi-Hanifi. 1991. Comparison of prosodic properties between read and spontaneous speech. *Speech Communication 10*. 163–169.
Jaeger, T. Florian. 2010. Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology* 61. 23–62.
Jun, Sun-Ah. 2003. Prosodic phrasing and attachment preferences. *Journal of Psycholinguistic Research* 32. 219–249.
Jungers, Melissa K., Caroline Palmer, & Shari R. Speer. 2002. Time after time: The coordinating influence of tempo in music and speech. *Cognitive Processing* 3. 21–35.
Kjelgaard, Margaret M., & Shari R. Speer. 1999. Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language* 40. 153–194.
Kraljic, Tanya, & Suan E. Brennan. 2005. Prosodic disambiguation of syntactic structure: For the speaker or for the addressee? *Cognitive Psychology* 50. 194–231.
Lehiste, Ilse. 1973. Phonetic disambiguation of syntactic ambiguity. *Glossa* 7. 103–122.
Lehiste, Ilse, Joseph P. Olive, & Lynn A. Streeter. 1976. Role of duration in disambiguating syntactically ambiguous sentences. *Journal of the Acoustical Society of America* 60. 1199–1202.
Levelt, Willem J. M. 1989. *Speaking: From Intention to Articulation*. Cambridge, MA: MIT Press.
Liberman, Mark, & Richard Sproat. 1992. The stress and structure of modified noun phrases in English. In I. Sag (ed.), *Lexical Matters*, 131–181. Chicago: University of Chicago Press.
Lieberman, Philip. 1963. Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech* 6. 172–187.
Lockridge, Calion B., & Susan E. Brennan. 2002. Addressee's needs influence speakers' early syntactic choices. *Psychonomic Bulletin and Review* 9(3). 550–557.
Marslen-Wilson, William D., Lorraine K. Tyler, Paul Warren, Paula Grenier, & Catherine S. Lee. 1992. Prosodic effects in minimal attachment. *Quarterly Journal of Experimental Psychology* 45. 73–87.
Nespor, Marina A., & Irene Vogel. 1986. *Prosodic Phonology*. Boston: Kluwer.
Pierrehumbert, Janet B. 1980. The phonology and phonetics of English intonation. Ph.D. dissertation, MIT.
Pitrelli, John F., Mary E. Beckman, & Julia Hirschberg. 1994. Evaluation of prosodic transcription labeling reliability in the ToBI framework. Paper presented at the Third International Conference on Spoken Language Processing, Yokohama, Japan.
Price, Patti, Mari Ostendorf, Stefanie Shattuck-Hufnagel, & Cynthia Fong. 1991. The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America* 90. 2956–2970.

Pynte, Joel, & Benedicte Prieur. 1996. Prosodic breaks and attachment decisions in sentence parsing. *Language and Cognitive Processes* 11. 165–192.
Schafer, Amy Jean. 1997. Prosodic parsing: The role of prosody in sentence comprehension. Ph.D. dissertation, University of Massachusetts.
Schafer, Amy Jean, Juli Carter, Charles Clifton, & Lyn Frazier. 1996. Focus in relative clause construal. *Language and Cognitive Processes* 11. 135–163.
Schafer, Amy Jean, & Sun-Ah Jun. 2001. Effects of focus on prosodic reflections of phrase structure in American English. Paper presented at the Prosody in Processing Workshop, Utrecht University, Utrecht, Netherlands.
Schafer, Amy Jean, & Sun-Ah Jun. 2005. Effects of contrastive focus on prosodic cues to clause boundaries. Poster presented at the Eighteenth annual CUNY conference on human sentence processing, Tucson, AZ.
Schafer, Amy Jean, Shari R. Speer, & Paul Warren. 2005. Prosodic influences on the production and comprehension of syntactic ambiguity in a game-based conversation task. In John C. Trueswell & Michael K. Tanenhaus (eds.), *Approaches to Studying World-Situated Language Use*, 209–225. Cambridge, MA: MIT Press.
Schafer, Amy Jean, Shari R. Speer, Paul Warren, & S. David White. 2000. Intonational disambiguation in sentence production and comprehension. *Journal of Psycholinguistic Research* 29. 169–182.
Schober, Michael F., & Herbert H. Clark. 1989. Understanding by addressees and overhearers. *Cognitive Psychology* 21(2). 211–232.
Schober, Michael F., Frederick G. Conrad, & Scott S. Fricker. 2000. Listeners often don't recognize when their conceptions differ from speakers'. Paper presented at the Annual Meeting of the Psychonomics Society, New Orleans, LA.
Selkirk, Elisabeth O. 1984. *Phonology and Syntax: The Relation Between Sound and Structure*. Cambridge, MA: MIT Press.
Selkirk, Elisabeth O. 1986. On derived domains in sentence phonology. *Phonology Yearbook* 3. 371–405.
Selkirk, Elisabeth O. 1994. Sentence prosody: Intonation, stress and phrasing. In John Goldsmith (ed.), *Handbook of Phonological Theory*, 550–569. Oxford: Blackwell.
Selkirk, Elisabeth O. 2000. The interaction of constraints on prosodic phrasing. In M. Horne (ed.), *Prosody: Theory and Experiment. Studies Presented to Gösta Bruce*, 231–261. Dordrecht: Kluwer Academic Publishers.
Shattuck-Hufnagel, Stefanie. 2000. Phrase-level phonology in speech production planning: Evidence for the role of prosodic structure. In Merle Horne (ed.), *Prosody: Theory and Experiment*, 201–229. Dordrecht: Kluwer.
Shattuck-Hufnagel, Stefanie, & Alice E. Turk. 1996. A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research* 25(2). 193–247.
Silverman, Kim, Mary E. Beckman, John Pitrelli, Mari Ostendorf, Colin Wightman, Patti Price, et al. 1992. ToBI: A standard for labeling English prosody. Paper presented at the 1992 International Conference of Spoken Language Processing.
Snedeker, Jesse, & Elizabeth Casserly. 2010. Is it all relative? Effects of prosodic boundaries on the comprehension and production of attachment ambiguities. *Language and Cognitive Processes* 25(7–9). 1234–1264.
Snedeker, Jesse, & John Trueswell. 2003. Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language* 48. 103–130.
Snedeker, Jesse, John Trueswell, Lila Gleitman, & Lisa Levine. 1999. Prosodic choice: Effects of speaker awareness and referential context. Paper presented at the University of Southern California Speech Production Conference, Los Angeles.
Speer, Shari R., & Allison Blodgett. 2006. Prosody. In Matthew Traxler & Morton Ann Gernsbacher (eds.), *Handbook of Psycholinguistics* (2nd ed.), 505–537. Amsterdam: Elsevier.

Speer, Shari R., Robert G. Crowder, & Lisa Thomas. 1993. Prosodic structure and sentence recognition. *Journal of Memory and Language* 32. 336–358.

Speer, Shari R., Paul Warren, & Amy Schafer. 2003. Intonation and sentence processing. In M. J. Solé, D. Recasens, & J. Romero (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences*, 95–105. Barcelona.

Spivey, Michael J., Michael K. Tanenhaus, Kathleen M. Eberhard, & Julie C. Sedivy. 2002. Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology* 45(4). 447–481.

Stirling, Lesley, & Roger Wales. 1996. Does prosody support or direct sentence processing? *Language and Cognitive Processes* 11. 193–212.

Straub, Kathleen A. 1997. The production of prosodic cues and their role in the comprehension of syntactically ambiguous sentences. Ph.D. dissertation, University of Rochester.

Tanenhaus, Michael K., Michael Spivey-Knowlton, Kathleen Eberhard, & Julie Sedivy. 1995. Integration of visual and linguistic information in spoken language comprehension. *Science* 268. 1632–1634.

Truckenbrodt, Hubert. 1995. Phonological phrases: Their relation to syntax, focus, and prominence. Ph.D. dissertation, MIT, Cambridge, MA.

Truckenbrodt, Hubert. 1999. On the relation between syntactic phrases and phonological phrases. *Linguistic Inquiry* 30. 219–255.

Trueswell, John C., Irina Sekerina, Nicole M. Hill, & Marian L. Logrip. 1999. The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition* 73(2). 89–134.

Ueyama, Motoko. 1998. Speech rate effects on phrasing in English and Japanese. Unpublished manuscript, University of California, Los Angeles.

Venditti, Jennifer J., & Julia Hirschberg. 2003. Intonation and discourse processing. Paper presented at the 15th International Congress of Phonetic Sciences, Barcelona.

Wagner, Michael, & Duane G. Watson. 2010. Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes* 25(7). 905–945.

Warren, Paul. 1985. The temporal organisation and perception of speech. Ph.D. dissertation, University of Cambridge.

Warren, Paul. 1999. Prosody and language processing. In S. Garrod & M. J. Pickering (eds.), *Language Processing*, 155–188. Hove: Psychology Press.

Warren, Paul, Esther Grabe, & Francis J. Nolan. 1995. Prosody, phonology and parsing in closure ambiguities. *Language and Cognitive Processes* 10. 457–486.

Warren, Paul, Amy J. Schafer, Shari R. Speer, & S. David White. 2000. Prosodic resolution of prepositional phrase ambiguity in ambiguous and unambiguous situations. *UCLA Working Papers in Phonetics* 99. 5–33.

Watson, Duane, Mara Breen, & Edward Gibson. 2006. The role of syntactic obligatoriness in the production of intonational boundaries. *Journal of Experimental Psychology: Learning, Memory, & Cognition* 32(5). 1045–1056.

Watson, Duane, & Edward Gibson. 2004. The relationship between intonational phrasing and syntactic structure in language production. *Language & Cognitive Processes* 19. 713–755.

Watson, Duane, & Edward Gibson. 2005. Intonational phrasing and constituency in language production and comprehension. *Studia Linguistica* 59. 279–300.

Watt, Sheila M., & Wayne S. Murray. 1996. Prosodic form and parsing commitments. *Journal of Psycholinguistic Research* 25(2). 291–318.

Wightman, Colin W., Stefanie Shattuck-Hufnagel, Mari Ostendorf, & Patti Price. 1992. Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America* 92. 1707–1717.