Available online at www.sciencedirect.com
SCIENCE DIRECT®
Cognitive Psychology 50 (2005) 194–231
Cognitive Psychology
www.elsevier.com/locate/cogpsych

Prosodic disambiguation of syntactic structure: For the speaker or for the addressee?☆

Tanya Kraljic*, Susan E. Brennan

Department of Psychology, State University of New York at Stony Brook, Stony Brook, NY 11794-2500, United States

Accepted 20 August 2004
Available online 2 December 2004

Abstract

Evidence has been mixed on whether speakers spontaneously and reliably produce prosodic cues that resolve syntactic ambiguities. And when speakers do produce such cues, it is unclear whether they do so "for" their addressees (the *audience design* hypothesis) or "for" themselves, as a by-product of planning and articulating utterances. Three experiments addressed these issues. In Experiments 1 and 3, speakers followed pictorial guides to spontaneously instruct addressees to move objects. Critical instructions (e.g., "Put the dog in the basket on the star") were syntactically ambiguous, and the referential situation supported either one or both interpretations. Speakers reliably produced disambiguating cues to syntactic ambiguity whether the situation was ambiguous or not. However, Experiment 2 suggested that most speakers were not yet aware of whether the situation was ambiguous by the time they began to speak, and so adapting to addressees' particular needs may not have been feasible in Experiment 1. Experiment 3 examined individual speakers' awareness of situational ambiguity and the extent to which they signaled structure, with or without addressees present. Speakers tended to produce prosodic cues to syntactic boundaries regardless of their addressees' needs in particular situations. Such cues did prove helpful to addressees, who correctly interpreted speakers' instructions virtually all the time. In fact, even when speakers produced syntactically ambiguous utterances in situations that supported both interpretations, eye-tracking data

☆ This material is based upon work supported by a National Science Foundation Graduate Fellowship and by NSF Grant Nos. 0325188 and 0082602.

* Corresponding author.
E-mail addresses: tkraljic@hotmail.com (T. Kraljic), susan.brennan@sunysb.edu (S.E. Brennan).

doi:10.1016/j.cogpsych.2004.08.002

showed that 40% of the time addressees did not even consider the non-intended objects. We discuss the standards needed for a convincing test of the audience design hypothesis.

Keywords: Prosody; Ambiguity; Optional words; Audience design

1. Introduction

Suppose Danielle is outside watering her plants when she remembers she has not yet fed the dog. She calls out to her friend Ann, who is in the kitchen, "Can you do me a favor and put the dog food in the bowl on the floor?" Danielle's request to Ann is structurally ambiguous: the first prepositional phrase (PP) *in the bowl* can be interpreted either as a modifier (attaching to and modifying the noun *food*) or as a goal (as an object attaching to the verb *put*). In particular, this is a global ambiguity (as opposed to a temporary ambiguity) because it is not resolved by any lexical or syntactic information in the utterance (Cutler, Dahan, & van Donselaar, 1997). Because there is more than one possible structure, other cues are needed to interpret the speaker's intended meaning.

This example illustrates an important point about language—utterances very often have the potential to be ambiguous. But it is not hard to believe that Ann (as well as readers familiar with typical eating and feeding habits of dogs and owners) knew exactly what Danielle was requesting and immediately interpreted the first PP as a goal, without even noticing the potential for ambiguity. Imagine, on the other hand, that the sentence immediately following Danielle's request read: *"This is ridiculous" Ann thought as she arranged some of the dog food from the bowl directly onto the kitchen floor, "I can't believe how much Danielle caters to this dog."* Again, readers would most likely have no trouble imagining this less typical situation, where the first PP needs to be interpreted as a modifier instead. The point is that for the most part, our experience with language is with utterances that can be disambiguated in more than one possible way. Most likely, Ann would not be puzzled by Danielle's request because she could rely on the situation itself for clarification—whether the bowl was on the floor and whether it was empty or full. She might also rely on any disambiguating cues in the utterance to discover Danielle's intended meaning, such as the length of words and pauses, and their pitch. If all else failed, she could ask Danielle for clarification. Danielle, for her own part, might produce disambiguating cues either automatically or deliberately. To signal the modifier attachment, for example, she could prosodically mark a structural boundary after *bowl* by lengthening the noun and the pause immediately following it relative to the first noun *food* and its pause (Ferreira, 2002; Lehiste, 1973). Such prosodic boundary marking perceptually groups the structural elements of the utterance as follows:

Modifier attachment: *Put the dog food in the bow:l.........on the floor.*
Goal attachment: *Put the dog foo:d.........in the bowl on the floor.*

Alternatively, or in addition, she might use an optional function word such as a relative pronoun that would disambiguate her meaning, as in *put the dog food* *that's* *in the bowl on the floor*.

This paper has two main themes. The first focuses on whether prosodic cues, in combination with optional function words and cues from the situational context, are used by speakers to signal the intended interpretation of a structurally ambiguous utterance. This is relevant not only to basic research in human language use, but also for applying such research to the design of human-computer spoken dialogue systems. Some research programs for spoken dialog systems have adopted the goal of enabling prosodic disambiguation of users' utterances (e.g., Hirose, 1997; Marcus & Hindle, 1990; Shriberg & Stolcke, 2004). However, the potential for machine (as well as human) comprehension systems to use prosody for disambiguation depends heavily on how reliable speakers are in actually producing such cues. If speakers produce prosodic cues automatically based on the intended syntactic structure of utterances, then prosody may constitute a reliable cue. If on the other hand speakers produce prosodic cues only when an utterance is ambiguous in a particular situation (and, moreover, when they are aware that their addressees may need help), then prosodic analysis may provide little payoff unless it can be integrated with additional cues from world knowledge and the referential context.

The second, parallel theme is to pursue the broader goal of determining whether certain actions by speakers (e.g., to mark a syntactic boundary prosodically) are typically done because they are easiest for the speakers themselves (emerging automatically from speech planning processes), or whether they are done for addressees. To the extent that a cue is produced because of, or as an adjustment to, an addressee's perceived needs, it represents what has been referred to as *audience design* (Bell, 1984; Clark & Murphy, 1982; Fussell & Krauss, 1992; Lockridge & Brennan, 2002; see also Schober & Brennan, 2003) and would be considered to be *non-natural* in the Gricean sense—that is, an intentionally communicative signal, as opposed to an automatically emerging symptom (Grice, 1975).

The difficulty with investigating hypotheses about audience design is that what looks like a choice governed by audience design is often not (Brown & Dell, 1987; Dell & Brown, 1991; Keysar, 1997; Schober & Brennan, 2003). That is, what is easiest for a speaker to say may turn out to be what is easiest for an addressee to understand because speakers and addressees happen to share the same environment and because there is parallelism between the human production and comprehension systems (leading to what Pickering & Garrod, 2004, call *representational parity*). Discovering *if, when*, and *how* speakers take their addressees' needs into account during speaking carries implications for the architecture of the language processing system, its flexibility, and its adaptability to context.

Our agenda, then, is to evaluate the communicative potential of prosodic cues in spontaneous spoken dialog. We adopt three criteria (suggested by Brennan & Williams, 1995) to test whether such cues are responsive to the needs of an addressee (the audience design hypothesis): First, the cue must be produced reliably and spontaneously by speakers in dialog. Second, it must be interpretable by addressees. Third, it must vary depending on speakers' intentions in the situation or toward addressees. Although prosodic cues would seem intuitively to be useful, previous research (notably Allbritton, McKoon, & Ratcliff, 1996) has left the first of these criteria in doubt; our experiments help resolve this uncertainty by using a task that places speakers in a relatively natural communicative setting. As for the second criterion, previous studies have found that listeners can use prosody to disambiguate otherwise-ambiguous sentences (e.g, Allbritton et al., 1996; Beach, 1991; Price, Ostendorf, Shattuck-Hufnagel, & Fong, 1991; Schafer, Speer, Warren, & White, 2000), and our first experiment confirms it in a controlled situation in which addressees interact freely with speakers. For the third criterion, we aim to provide a more rigorous test of the audience design hypothesis than has been attempted previously (for discussion, see Schober & Brennan, 2003). Generally, the conclusion that speakers do *not* adapt to addressees' needs in the early stages of utterance planning, but only later, as an adjustment or a repair, is based (whether rightly or wrongly) on null audience design results. It is difficult to know whether such conclusions are justified, or whether the enterprise has been hampered by design flaws, insufficient power, or lack of a good-faith effort to find an effect (for discussion of these issues, see Frick, 1995). To conclude that a partner-specific variable has no effect on speaking, it is necessary to make every effort to find the effect (Frick, 1995). *Production of disambiguating prosodic cues.*

Prosody marks syntactic boundaries in several ways. The final word of a phrase and the pause immediately following it may be lengthened; such durational differences (and the phonetic variations that accompany them) are correlated with suprasegmental markers such as pitch accents and boundary tones (see, e.g., Beach, 1991; Beckman, Hirschberg, & Shattuck-Hufnagel, 2004; Ferreira, 1993; Lehiste, 1973; Schafer et al., 2000). Syntactic boundaries can be signaled via pitch contour variation such as a pre-boundary fall-rise or rise (see Beach, 1991) or by a decrease in amplitude on the pre-boundary syllable (Streeter, 1978; see also Scott, 1982). In short, a variety of phonological and intonational cues may be used in combination to mark syntactic structure (Beach, 1991; Lehiste, 1973; Lehiste, Olive, & Streeter, 1976; Schafer et al., 2000; Scott, 1982; Streeter, 1978).

We focus on the prosodic cue of durational differences for several reasons. Durational differences between words or pauses are relatively straightforward and reliable to measure, compared to a multidimensional approach to categorizing boundary strengths such as ToBI (Beckman et al., 2004). Together with observations of the comprehender's behavior, durational differences afford an objective analysis of whether a speaker has successfully produced the intended syntactic structure. In addition, the association between duration and syntactic structure is strong for both speakers and addressees (see Cooper & Paccia-Cooper, 1980; Klatt, 1975; Snedeker & Trueswell, 2003). Durational differences signal syntactic structure even in the absence of other prosodic markers (Lehiste, 1973), and are correlated with other prosodic cues; for example, stronger prosodic boundaries are marked by greater word lengthening and longer pauses (Ferreira, 1993, 2002). Moreover, comprehension studies find that listeners can use durational cues to correctly interpret an ambiguous syntactic structure even in the absence of a disambiguating pitch contour (Lehiste et al., 1976; Scott, 1982). Some have concluded that duration is used more reliably than other phonological cues in listeners' interpretation of ambiguous utterances (Beach, Katz, & Skowronski, 1996; Streeter, 1978), as well as produced more reliably by speakers (see Lehiste, 1973). Several recent studies suggest the duration of a boundary has its effect not as an absolute, but relative to other durations within the utterance (Carlson, Clifton, & Frazier, 2001; Kjelgaard & Speer, 1999; Schafer et al., 2000). Accordingly, if relative boundary durations within an utterance differ significantly and if listeners' behavior confirms that they have interpreted the utterance correctly, we can be confident that the syntactic boundary was signaled successfully (although this alone does not confirm whether audience design was at work).

Until recently, studies investigating the production of prosodic cues have typically examined scripted, tape-recorded utterances produced by speakers who were instructed to contrast two meanings of a sentence (Allbritton et al., 1996; Lehiste, 1973), to read a sentence aloud within different disambiguating contexts (Price et al., 1991), or to memorize and then say aloud a sentence within a single disambiguating context (Fox Tree & Meijer, 2000). In these cases, speakers had neither a clear communicative goal nor a copresent addressee who needed to understand the utterances. Such studies have found that trained, informed speakers are able to prosodically mark the syntactic structure of a scripted utterance (Allbritton et al., 1996; Cooper & Paccia-Cooper, 1980; Lehiste, 1972; Klatt, 1975; Price et al., 1991), and that they can alter the prosodic groupings to signal different meanings of the same ambiguous sentence (Lehiste, 1973; Price et al., 1991; Snedeker & Trueswell, 2003). But the generalizability of these results to untrained, naive speakers engaged in spontaneous communication has been questioned (Allbritton et al., 1996; Fox Tree & Meijer, 2000).

Allbritton et al.'s experiments were the first to explicitly compare the use of disambiguating prosody by non-trained, naive speakers to that by trained speakers who were either aware or unaware of the potential for ambiguity. The speakers, either amateur and professional actors or introductory psychology students, read sentences into a microphone, including nine sentences that were syntactically ambiguous. Half of the actors were informed of the purpose of the experiment, saw paraphrases of the two possible meanings for each critical sentence, and rehearsed pronouncing each sentence once for each meaning before recording the critical sentence. The other half, as well as all of the students, were naïve to the purposes of the experiment and were not shown a paraphrase of a possible alternative meaning. Each speaker read the sentences embedded in a paragraph that provided disambiguating context. Under these circumstances, the students did *not* reliably produce disambiguating prosodic cues to the intended syntactic structure. The actors produced prosodic cues reliably, but only when they were explicitly asked to do so (and not when they were naive). This project's conclusion raised serious doubts about whether the human comprehension system could rely much on prosodic cues in spontaneous conversation, since it appeared that speakers would produce such cues rarely or in attenuated form (Allbritton et al., 1996).

However, the lack of a communicative goal and the use of a reading-aloud method in Allbritton et al.'s study raises a concern about their conclusion that naïve speakers do not reliably and spontaneously signal syntactic structure prosodically.

Many studies across a wide range of contexts indicate that spontaneous utterances are very different from speech read aloud, and utterances in conversation (with a partner) may be different as well (see e.g., Brennan & Schober, 2001; Ferreira, Bailey, & Ferraro, 2002; Krauss, 1987; Pickering & Garrod, 2004; Samuel & Troicki, 1998; Schober & Brennan, 2003).

Several recent studies (e.g., Schafer et al., 2000; Snedeker & Trueswell, 2003) have addressed this concern by having speakers produce previously memorized syntactically ambiguous instructions in the presence of an addressee whose task was to follow the instructions. These studies avoided using read-aloud speech in isolation, but varied in the degree of naturalness and interaction they permitted: Schafer et al.'s task was more interactive and had immediate consequences for misunderstanding, while Snedeker and Trueswell's did not permit spontaneous interaction or feedback regarding misunderstanding. Perhaps because of this difference, those two studies reached different conclusions. Snedeker and Trueswell's (2003) conclusions were consistent with Allbritton et al.'s: they found that speakers produce prosodic cues to syntactic structure only when they are aware of the ambiguity and when other information does not disambiguate the utterance. Schafer et al. (2000), on the other hand, found that speakers do use prosody to disambiguate syntactic structure and that they do so all the time, regardless of other available disambiguating information. Their study supports a much wider role for prosody in spontaneous production and, by extension, in comprehension.

There were other differences in the tasks that these two studies used, making it difficult to know how to reconcile their conclusions. For example, in Schafer et al.'s game-playing task, assessing the potential for ambiguity was a bit difficult; participants had to keep track of a number of things in order to assess whether a particular utterance could be disambiguated given the current state of the game. In communicative contexts, speakers may mark all syntactic boundaries automatically, or if they are generally aware of a potential for ambiguity but are unable to detect specific cases, they may mark all boundaries in order to be as clear as possible (a coarse sort of audience design). Snedeker's task, on the other hand, employed a much simpler referential context that, in their Experiment 2, always clearly disambiguated the speaker's utterance (for the speaker, and, as the speaker was led to believe, for the addressee as well); in addition, the task was blocked with respect to syntactic structure (each speaker experienced only one of two possible kinds of interpretations throughout the experiment). Other potentially important differences discussed by Snedeker and Trueswell include utterance length (Schafer et al.'s utterances were nearly twice as long, on average, as Snedeker & Trueswell's), or the possible lexicalization of the memorized expressions in Schafer et al.'s experiment, which might have led to an exaggerated prosodic grouping of the words in those expressions.

Clearly, then, just when and how prosody is used to mark syntactic structure in conversation is still an open topic. Our experiments attempt to address this topic by getting speakers to both plan and articulate utterances that are syntactically ambiguous, and sometimes contextually ambiguous, in a task with a communicative goal shared by interacting addressees.

1.1. Interpretation of disambiguating prosodic cues

Many studies have concluded that prosody can provide useful cues to listeners in processing potentially ambiguous utterances (e.g., Allbritton et al., 1996; Beach, 1991; Carlson et al., 2001; Kjelgaard & Speer, 1999; Lehiste, 1973; Price et al., 1991; Schafer et al., 2000; Scott, 1982; Shattuck-Hufnagel & Turk, 1996; Snedeker & Trueswell, 2003; Streeter, 1978). As discussed previously, most of the evidence for this conclusion comes from read-aloud utterances in which the prosody was manipulated artificially by trained speakers or speech synthesizers (e.g., Allbritton et al., 1996; Beach, 1991; Price et al., 1991). So while prosodic cues can be helpful to listeners, their disambiguating potential in spontaneous communication remains to be demonstrated. As Allbritton et al. (1996) put it, knowing that listeners *can* make use of prosodic cues when available does not tell us much if speakers rarely produce such cues spontaneously.

In addition, there is a question as to how *quickly* comprehension can make use of prosody. Some researchers suggest that prosodic information contributes to the final structuring of an initially syntactically determined parse, but that it is not available during initial syntactic processing (e.g., Marcus & Hindle, 1990; Pynte & Prieur, 1996). Others have suggested that an abstract prosodic representation is maintained in memory and that this information is available early on to actually guide the syntactic parsing process (Kjelgaard & Speer, 1999; Marslen-Wilson, Tyler, Warren, Grenier, & Lee, 1992; Nagel, Shapiro, Tuller, & Nawy, 1996). These ideas have been difficult to tease apart, because the paradigms used in the experiments (e.g., cross-modal naming and speeded listener judgments) do not afford very specific temporal measurements, and often require interrupting the utterance as it is being heard (see Carlson et al., 2001; Snedeker & Trueswell, 2003, for discussions). Snedeker and Trueswell's (2003) eye-tracking study avoids these problems and supports a very rapid, early role for prosody in comprehension; they found that when an utterance is syntactically ambiguous, prosodic information is used as early as lexical and phonological information.

1.2. Disambiguating prosodic cues: for the speaker or for the addressee?

Finally, if prosodic cues are reliably produced by speakers and used by addressees, this does not tell us *why* such cues emerge during conversation. Typically, experimental studies of prosodic cues have focused on either speakers *or* listeners, in non-interactive contexts (with the exception of Schafer et al., 2000), rather than on people engaged in spontaneous communication. This is not surprising, as it is difficult to control what individual speakers say spontaneously in order to obtain enough tokens of comparable utterances (Bock, 1986), let alone what paired speakers say in conversations. In fact, Pickering and Garrod (2004) argue that language processing in a monologue context (without interacting speakers or addressees present) is qualitatively different from language processing in a dialogue context.

Schafer et al. (2000) devised a controlled way to study the spontaneous articulation of syntactically ambiguous utterances to interacting addressees. They had pairs of speakers play a game in which they produced utterances from a memorized set of sentences in order to get their partners to maneuver game pieces around a board. The utterances were temporarily ambiguous (*When that moves the square will*... vs. *When that moves the square it*...). Although these utterances were pre-set and memorized rather than planned on the fly, their use was not scripted; speakers produced (recalled and articulated) each utterance as it was needed. This game-playing task offered some advantages over Allbritton et al.'s task. Because participants had to respond to each other, each utterance was associated with a communicative goal and consequences for misunderstanding: speakers got immediate feedback from their partners as to whether the meaning of their utterances was clear. This task more closely approximates spontaneous conversation. Consistently, Schafer et al.'s naive speakers produced disambiguating prosodic cues to the intended syntactic structures. Schafer et al. attributed the divergence of their findings from Allbritton et al.'s (1996) to the effects of communicative intentions in an interactive context.

While this study demonstrated convincingly that speakers do spontaneously produce prosodic cues to syntactic structure, it left open the question of why they do so. In Schafer et al. (2000), speakers did not appear to tailor their prosody to addressees' actual needs at a given moment; most if not all of the utterances could have had only one possible meaning at any point, given the state of the game. This suggests that if audience design is to be implicated as a force behind prosodic disambiguation, it would not be as a precise adjustment to the addressee's current needs in a particular situation, but as a coarser ('brute force') adjustment. That is, when speakers share communicative goals with co-present addressees, they may try harder to be clear in general than when they have no communicative purpose.

Evidence about audience design has been collected concerning another cue in speech production: disambiguating but optional function words. Ferreira and Dell (2000) examined the circumstances under which speakers inserted such words into reproduced, memorized utterances that were either temporarily ambiguous or unambiguous (e.g., the complementizer *that* in *The coach knew* _____ *you*..., which is temporarily ambiguous, as opposed to *The coach knew* _____ *I*..., which is disambiguated early by case-marking). They found that disambiguating function words were likely to be produced when upcoming content from the speech plan was *not yet available* to the speaker. These words were no more likely to be produced when there would be temporary ambiguity without them than when there would not, leading to the conclusion that using these words is driven by what is easy for the speaker rather than by design for the addressee. However, when naive addressees (who were going to rate the utterances for clarity later) were physically co-present with speakers in Ferreira and Dell's Experiment 6, the speakers did use more optional complementizers overall than when there were no addressees. This kind of adjustment might be less costly from a processing standpoint than it would be to maintain and update a model of the addressee's needs moment-by-moment, which would presumably be required if speakers produced optional words or prosodic cues only when addressees needed them. On this view, when speakers detect any significant potential for ambiguity, they should not expend the resources to make targeted adjustments, but simply opt to be as clear as feasible—an adjustment to addressees' needs that, while coarse, nevertheless constitutes audience design.

Another issue is whether the potential disambiguating cues of prosodic lengthening and function words are produced redundantly, reciprocally, or independently of one another. Beach (1991) has proposed an analysis of the cognitive effort involved in speaking, claiming that speakers should avoid redundant cues in the interest of saving effort. On that view, a speaker would disambiguate a particular parse by either lengthening the boundary or producing a function word, but not both. However, we expect that the monitoring needed to avoid redundant cues (like the monitoring needed to monitor addressees' disambiguation needs in a fine-grained way) would be more costly than producing prosodic cues and optional words more liberally. A third possibility is that if optional function words emerge simply as placeholders in a speech plan, as Ferreira and Dell (2000) concluded, they may be produced independently of prosodic boundary marking.

Some have argued that certain aspects of utterance planning are too automatic to be influenced by audience design (Bard et al., 2000; Bard & Aylett, 2000). Prosody may emerge simply as a by-product of planning and articulating a particular syntactic structure (Cooper & Danly, 1981; Cooper & Paccia-Cooper, 1980; Klatt, 1975; Selkirk, 1984), regardless of other available disambiguating information. In this case, lengthened prosodic boundaries would be a reliable and potentially informative cue to syntactic structure, independent of referential context.

2. Experiment 1

Experiment 1 examined whether people use prosody to reliably disambiguate syntactic structure during spontaneous speaking and whether they are less likely to do so when the situation disambiguates the utterance for them (as Snedeker & Trueswell, 2003 found), or whether they do this regardless of the referential situation (as Schafer et al., 2000 found). The aim was to get multiple speakers to spontaneously plan and produce comparable, syntactically ambiguous sentences in the same controlled referential circumstances during a task that required interactive communication with an addressee. We had pairs of naive participants do a referential communication task during which they viewed, referred to, and manipulated a set of objects. In each trial, a Director instructed a Matcher to move objects as in Metzing and Brennan (2003) and Tanenhaus, Spivey-Knowlton, Eberhard, and Sedivy (1995) (but neither participant in the present studies was an experimental confederate).

We elicited utterances of the desired form by providing Directors with a picture that they viewed privately before issuing each instruction. Each picture showed a schematic drawing of the current display of objects and indicated by markings which objects the Director was to instruct the Matcher to move, where to move them, and any other objects that needed to be mentioned as part of the instruction to the Matcher. The point was to elicit syntactically ambiguous utterances in which the first prepositional phrase (PP) could be interpreted either as a modifier or as a goal, and where the situation (in this case, the set of objects on the display) was ambiguous (supporting both the modifier and the goal interpretations) or unambiguous (supporting only one interpretation). For example, in the utterance *Put the dog in the basket on the star*, the first PP (*in the basket*) could be interpreted as a modifier (i.e., specifying which dog) or as a goal (i.e., as an argument to *put*, specifying where the dog is to be placed). Thus, for communication to succeed in an ambiguous situation, it would be necessary for Directors to signal the correct meaning using prosody or optional function words. On the other hand, in unambiguous situations, where interpretations were constrained, Directors would not need to disambiguate structurally ambiguous utterances for communication to succeed. In each case, the Matcher's spontaneous response to the Director's instruction (e.g., to move the object or to ask for clarification) confirmed whether or not disambiguation was successful.

Experiment 1 addressed the following questions: (1) Do speakers spontaneously produce disambiguating prosodic cues in utterances with ambiguous syntactic structures? (2) Do they do so even when the referential situation supports only one interpretation of the utterance? (3) Are speakers (Directors) more likely to disambiguate prosodically after they have already experienced the potential for ambiguity in the situation as addressees (Matchers)? And (4) under what circumstances do addressees actually experience utterances as ambiguous?

2.1. Predictions

We predicted that speakers would disambiguate syntactic ambiguity using prosody, consistent with the only other study of prosodic disambiguation to give naive participants both a communicative task and the opportunity to interact freely (Schafer et al., 2000). Specifically, we predicted that speakers would use durational cues to signal syntactic boundaries for the intended interpretation of an utterance. For a goal attachment, speakers should lengthen the first boundary in the utterance (i.e., the first noun and/and the pause immediately following); for a modifier attachment, speakers should lengthen the second boundary (the second noun and/or the pause immediately following).

As for testing the impact of audience design on prosodic disambiguation, we expected that if speakers were sensitive to addressees' precise needs, there would be stronger differences between boundary 1 and boundary 2 when situations support both interpretations than when they support only one. If speakers are more likely to disambiguate prosodically when an utterance is truly ambiguous in its context, this would be consistent with Snedeker and Trueswell's (2003) interpretation of their results. Such a finding would not necessarily be inconsistent with Schafer et al.'s finding (that speakers disambiguate even when a situation is *not* ambiguous), because it may be a matter of degree; speakers may mark syntactic boundaries routinely, but do so even more in the face of potential ambiguity (especially if ambiguity cannot be easily or precisely assessed, as was the case in Schafer et al.). Alternatively, if prosodic boundary marking emerges automatically from syntactic planning, then there may be no difference in the relative strengths of boundary 1 and boundary 2 in ambiguous versus unambiguous situations.

It may be that speakers become aware of potentially ambiguous situations only when they have already experienced the task as addressees. If this is the case, then the stronger prosodic boundary difference that we predicted for ambiguous situations may occur only for those who serve as Matcher for the first half and Director for the second half of the task.

2.2. Method

2.2.1. Participants

Forty eight undergraduate volunteers (17 men, 31 women) from the State University of New York at Stony Brook received research credit in a psychology course for their participation. All participants were 18 years of age or older, and all identified themselves as native English speakers. Participants were scheduled to come to the lab in pairs, and members of a pair were randomly assigned to the role of Director or Matcher.

2.2.2. Design

Directors instructed Matchers to move objects in a physical display. Each of 20 trials included a critical target instruction with a potentially ambiguous prepositional phrase; for half, the PP was intended as a modifier of the object to be moved (modifier attachment), and for half, as the destination of the object (goal attachment). For each kind of attachment, half of the time the target instruction was ambiguous given the set of objects in the display (ambiguous) and half of the time it was not (unambiguous). Each critical instruction appeared in all 4 cells of this design; four versions of the experiment were created, with each critical instruction appearing once on each version and each pair of participants randomly assigned to one version. In addition, each of the 20 trials included 1–2 additional unambiguous filler instructions in which the pair manipulated non-target objects. Trials (displays with a given set of objects) were presented in a single random order.

To look for effects of speakers' awareness of the potential for ambiguity, we had members of each pair switch task roles halfway through the task, so that the first Director was considered to be inexperienced with taking the Matcher's perspective, and the second one, experienced.

2.2.3. Materials

Each display consisted of a set of four manipulable objects; some of these objects rested upon or were next to reference objects. For the trials that contained ambiguous instructions, each set consisted of: (1) the target object to be moved, (2) the destination object, (3) an object which could be the destination if the PP was interpreted incorrectly by the Matcher, and (4) a distracter object. Sets for the unambiguous conditions included the target, the destination, a competitor for the target (for modifier attachments) or for the destination (for goal attachments), and a distracter object. A sample of the four versions of a display set is shown in Fig. 1 for the target utterance: "Put the dog in the basket on the star."

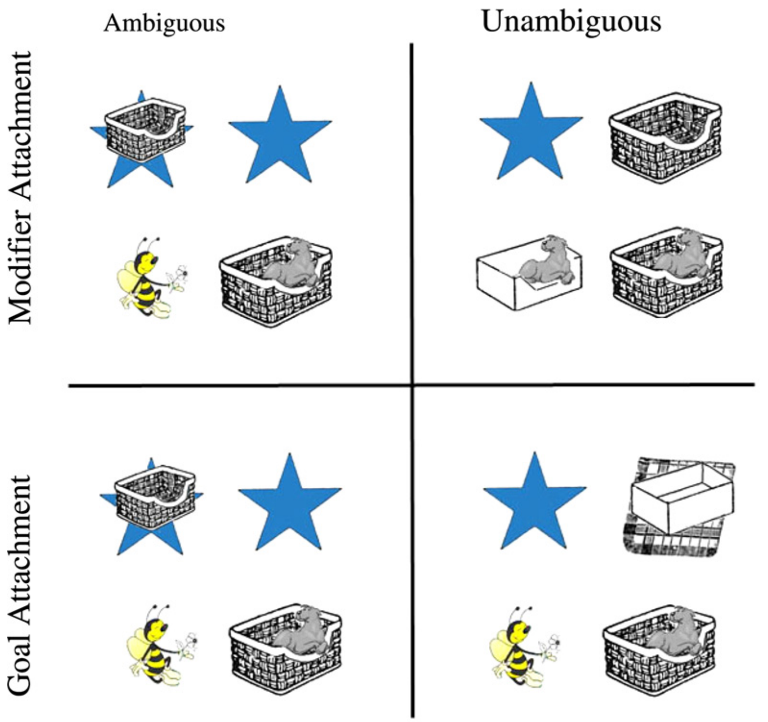
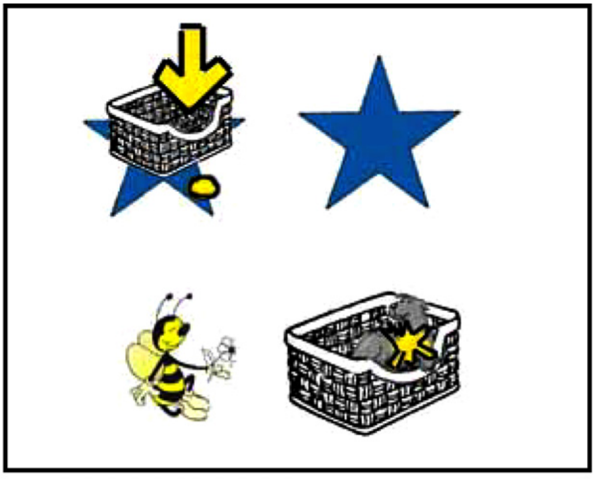


Fig. 1. Schematic sample set of the four versions of a display in Experiment 1.

The objects for each trial were placed on a wooden stair-step display that was positioned on a table directly in front of the Matcher. The display consisted of three shelves, each of which were further divided into 3 8 in. × 8 in. spaces with black tape. At the start of each trial, the objects for that display were placed by the Experimenter on the outermost top and bottom corner spaces of the display, although once the trial began any of the spaces could be used.

To get speakers to produce utterances with structurally ambiguous PP-attachment, we did the following. At the outset of each trial, while the objects in the display were being positioned, speakers were given a card with line drawings of all the objects in the display (Fig. 2). These cards were marked to show which object was to be moved (by a red star on the drawing) and where it was to be moved (by a red arrow). In addition, any item on the sheet that the speaker needed to mention (the potentially ambiguous prepositional phrase) was marked with a red dot. In this way we hoped to influence the speaker to produce instructions like "Put the dog in the basket on the star".

We discovered in piloting that it would be necessary to prevent speakers from falling into a repetitive pattern and producing all the instructions with the same mechanical-sounding prosody. Therefore, each trial consisted of two parts: first, the critical instruction (indicated on the card with red markings), and then one or two filler



Asterisk: Target Object; *Arrow*: Destination; *Dot*: Must Mention

Fig. 2. Example of the card from which Directors spontaneously instructed Matchers in Experiments 1 and 3. Directors instructed Matchers to move the target object (with asterisk) to the destination (with arrow) while also mentioning another object (with dot).

instructions that provided structural variety (indicated by additional markings in black). The filler instructions had Matchers move objects onto spaces in the display that first had to be emptied of other objects.

The Directors' and Matchers' speech was recorded using an omnidirectional microphone.

2.2.4. Eyetracking apparatus

Matchers wore a headband-mounted visor fitted with an ISCAN RK-726PCI pupil/corneal reflection eyetracker and Polhemus head tracker. The visor supported a small video camera that output the Matcher's view of the referent array to the eyetracking computer. Another small video camera tracked the Matcher's eyegaze, which was overlaid as a cursor over the image of the referent array (after being corrected by a head tracker for any head motion). The apparatus weighed 6.5 oz and enabled participants to move naturally while conversing in a seated position. The resulting video image was recorded on digital videotape so that it could be easily synchronized with the speech and coded frame by frame.

2.2.5. Procedure

The Matcher was seated directly in front of the stair-step display; the Director was seated slightly behind and to the side of the Matcher, so that she could not use eye gaze or pointing to inform the Matcher of the correct interpretation. After a brief calibration procedure for the eyetracker, the partners were told that the purpose of the experiment was for the Director to get the Matcher to perform the intended actions by using as few instructions as possible. Directors were also told to begin every utterance with "Put" or "Place". Within these constraints, they were free to instruct Matchers in whatever way they wanted.

At the beginning of each trial, the Experimenter arranged the objects, in view of the participants, onto the display in a predetermined order (counterbalanced for the position of the target item to be moved and the target destination). After the objects were in place, the Experimenter named each of the items (e.g., "Here's a dog in a basket, a basket on a star, a star, and a bee.") and the Director and Matcher were encouraged to check the display to ensure that the Experimenter had placed the correct items in the correct order. In this way, each of the participants could become familiar with the objects in each particular situation, be encouraged to entrain on the terms we wished them to use, and become aware of any potential for ambiguity in the referential situation.

After arranging and naming the objects in a given set, the Experimenter handed the schematic card to the Director while saying "Begin." The Director could then take as long as necessary before speaking to the Matcher. Upon hearing the instruction, the Matcher's task was to carry it out or ask for clarification if unsure of the instruction. The Director was told to observe the Matcher's action to makesure it was correct, and if it was not, to inform the Matcher and to give the instruction again, clarifying in whatever way she felt would help.

At the outset, participants performed eight practice trials. In piloting, the first Directors kept asking for guidelines for how they might say the instructions schematized on the cards; because we wanted to avoid biasing their prosody with spoken examples, we added text versions of target sentences on the eight schematic cards for the practice trials only. Half of these sentences followed the structurally ambiguous form *Put the________ in the ________ on the ________* half included disambiguating optional words (*Put the ________ that's in the ________ on/onto the ________*).

After the Director had instructed the Matcher for 10 trials, they switched roles and performed another 10 trials, with the (former) Matcher now acting as Director. We hoped that second-half (if not first-half) Directors would have sufficient opportunity to be aware of the potential for ambiguity and the Matchers' need for disambiguation, and perhaps be more likely to tailor cues to Matchers' needs. During critical instructions, we recorded Matchers' eyegaze, where they moved the target object, and both partners' speech.

2.3. Coding and analyses

2.3.1. Durational measures

Because utterances were produced spontaneously rather than read aloud, Directors often inserted optional function words (e.g., "Put the dog <*that's*> on the basket <*onto*> the star") or syntactic variations (e.g., "Put the candle in the ashtray with the star in it") that made the utterances structurally *un*ambiguous. We digitized all utterances that were either structurally ambiguous or for which optional function words disambiguated the target PP-attachment. Using SoundEdit for Macintosh, we made five measurements per utterance: the duration of the first noun in the utterance, that of the pause immediately following the first noun, that of the second noun, that of the pause immediately following the second noun, and the length of the entire utterance. For the durational analyses of the structural boundaries, we combined the duration of each NP with that of its subsequent pause because whether a structural boundary is indicated by lengthening a word or by pausing right after it depends on the phonetic characteristics of the word (Ferreira, 2002). For the utterance *Put the* *dog*$_1$ *in the* *basket*$_2$ *on the star*, the first boundary (underlined) was expected to occupy a larger percentage of the utterance than the second boundary whenever *in the basket* was intended to have a goal interpretation, with the reverse being true for a modifier interpretation. For each critical utterance we computed a difference score (boundary 1 – boundary 2), divided by the length of the utterance in which it appeared in order to normalize for speaking rate. We then conducted ANOVAs by-subjects (F1) and by-items (F2) on the normalized difference scores for three factors: attachment (goal or modifier), ambiguity type (ambiguous or unambiguous), and speaker's experience (first or second Director).

2.3.2. *Optional function words*

All speech produced by the Director or the Matcher during the critical instructions (up until the Matcher moved the object correctly) was transcribed, including any restarts, disfluencies, or clarifications. Critical utterances were coded (a) as having structurally ambiguous PP-attachment of the sort we were trying to elicit (e.g., *Place the dog in the basket on the star*), (b) as including disambiguating optional function words (*that, into*, or *onto*, as in *Place the dog that's in the basket onto the star* or *Place the dog into the basket that's on the star*), or (c) as *other* (unanalyzable due to omitted or incorrect information, disfluencies, restarts, or constituents produced in a way that did not follow the task instructions). Unintelligible utterances were treated as missing. To detect whether speakers use prosodic boundary marking and optional function words to disambiguate redundantly, reciprocally, or independently, we then conducted ANOVAs of the strengths of boundary differences conditional on whether optional function words were used.

2.3.3. *Eyegaze measures*

Because we were interested in discovering whether disambiguating cues were immediately useful to Matchers in interpreting structurally ambiguous utterances, we coded Matchers' looks and touches to movable targets, destinations, and distracters during trials in which the Director produced structurally ambiguous instructions. Beginning at the moment when Directors began to utter the first NP, we coded Matchers' looks toward the four relevant corner sections of the display, ending with their touching the target object just prior to moving it to the intended destination. Our assumptions (following Tanenhaus et al., 1995) were that Matchers who did not yet have sufficient disambiguating cues would gaze at any object(s) under consideration as potential referent(s), and that they might look back and forth between two candidate destinations if they were undecided. We predicted that if naive speakers naturally provide good disambiguating prosodic cues to structural boundaries, then addressees would consider only targets and ignore distracter objects, even when the utterance included no disambiguating optional function words and when the context supported both parses.

2.4. Results

The method for eliciting spontaneous utterances was fairly successful: Directors spontaneously produced utterances in analyzable form about half the time, in 248 out of 480 critical trials. The remaining 232 utterances were unanalyzable because they contained events such as coughs or sneezes, disfluencies, misinformation, or other departures from the desired form. Utterances followed the forms suggested by the schematic card and labels given by the Experimenter more often for modifier than for goal attachments, 65% of the time to 39%, $F1(1,47) = 24.14$, $p < .001$; $F2(1,19) = 23.44$, $p < .001$, and marginally more often for situationally unambiguous than situationally ambiguous utterances, 55–48%, $F1(1,47) = 3.26$, $p = .077$; $F2(1,19) = 3.51$, $p = .077$. That utterances with goal interpretations departed more often from predictable forms suggests that they were more difficult to plan than modifier utterances. Of the 248 analyzable utterances, 37.5% were structurally ambiguous and the rest contained optional function words. Numbers and types of analyzable utterances are shown by condition in Table 1.

2.4.1. Speakers' production of disambiguating cues

2.4.1.1. Prosodic marking of syntactic boundaries. Directors indeed disambiguated the syntactic boundaries of their utterances prosodically. The first noun phrase boundary was longer than the second for goal interpretations (by 17.3% of the utterance) and shorter than the second for modifier interpretations (by 10.4%), as shown in Table 2, $F1(1,17) = 131.62$, $p < .001$; $F2(1,5) = 64.44$;[1] $p < .001$. This finding is in stark contrast with Allbritton et al.'s (1996), whose speakers did not disambiguate (read-aloud) sentences prosodically. It is consistent with other studies in which speakers had communicative goals (Schafer et al., 2000; Snedeker & Trueswell, 2003). Our task differed from the tasks used in those studies by eliciting controlled but spontaneously planned utterances. The evidence is strong that prosody can serve as a reliable cue for structural disambiguation in conversation.

Although Directors *did* differentiate structural boundaries depending on whether the critical phrase referred to a goal or a modifier, they did *not* do so any more strongly when the utterance was ambiguous in its referential context than when it was unambiguous, $F1(1,17) = 1.12$, n.s.; $F2(1,6) = .076$, n.s. Moreover, prior experience in the Matcher role did not lead Directors to mark boundaries more strongly; if anything, second Directors marked them *less* strongly than did first Directors (this

[1] In this between-items analysis, the denominator degrees of freedom are low due to missing data for the analysis that includes speaker order. In the ANOVA without speaker order, $F2(1,13) = 276.13$, $p < .001$.

Table 1
Experiment 1: Frequencies with which speakers' spontaneous instructions matched the desired forms, with and without optional words, out of 480 trials

	Syntactically ambiguous		Syntactically unambiguous (due to disambiguating function words)		Total
Spontaneous utterance:	*Put the dog in the basket on the star*		*Put the dog that's in the basket on the star*		
	Situationally ambiguous	Situationally unambiguous	Situationally ambiguous	Situationally unambiguous	
Goal attachment	13	20	30	30	93
Modifier attachment	28	32	44	51	155
Total	41	52	74	81	248

Table 2
Experiment 1: Strength of Directors' prosodic boundary differences[a] as a function of their prior experience in the Matcher role (with percentage of utterances containing optional words in parentheses)

	Director 1 No prior experience	Director 2 Prior experience	Total
Goal attachments			
Unambiguous situation	+20.2% (61.5%)	+13.4% (58.3%)	+16.8% (59.9%)
Ambiguous situation	+15.1% (60.0%)	+20.2% (78.3%)	+17.7% (69.15%)
Total (goals)	+17.9% (60.9%)	+16.7% (68.1%)	+17.3% (64.5%)
Modifier attachments			
Unambiguous situation	−12.8% (64.3%)	−9.8% (58.5%)	−11.3% (61.4%)
Ambiguous situation	−12.6% (60.0%)	−6.5% (62.2%)	−9.6% (61.1%)
Total (modifiers)	−12.7% (62.3%)	−8.2% (60.3%)	−10.45% (61.3%)

[a] The strength of the prosodic boundary difference is boundary 1 − boundary 2, normalized as a percentage of the utterance length. This was predicted to be positive for Goal attachments and negative for Modifier attachments. A larger magnitude difference (in the predicted direction) indicates more disambiguation.

was reliable by-subjects but not by-items), $F1(1,17) = 5.31$, $p = .03$; $F2(1,6) = .00$, n.s. (see Table 2). There was also no interaction of speakers' experience by situational ambiguity; that is, second Directors did not differentiate prosodically between situationally ambiguous vs. unambiguous utterances any more than did first Directors, $F(1,17) = .01$, n.s.; $F(1,5) = 3.14$, n.s. In sum, there was no evidence of audience design driving prosodic boundary marking.

2.4.1.2. Optional function words. Directors used optional disambiguation function words 62.5% of the time. They did this whether the utterance was ambiguous in its referential situation or not, $F1(1,17) = .45$, n.s., $F2(1,5) = .14$, n.s. Experience acting in the Matcher role was not followed by a reliably greater use of optional function words, although the pattern of means in Table 2 is compatible with audience design, particularly for goal interpretations.[2] Optional function word use also showed no interaction between experience and situational ambiguity, $F1(1,17) = 1.21$, $p = .29$; $F2(1,5) = .12$, n.s.

Directors were just as likely to insert optional function words into their utterances when they intended goal interpretations (64.5% of the time) as when they intended modifier interpretations (61.3% of the time), $F1(1,17) = 1.48$, n.s.; $F2(1,5) = .41$, n.s. An optional function word can be used to clarify an utterance either early, before an ambiguous phrase (e.g., *Put the dog* *into* *the box on the coaster* or *Put the dog* *that's* *in the box on the coaster* for goals and modifiers, respectively) or later, after an ambiguous phrase (*Put the dog in the box* *that's* *on the coaster* or *Put the dog in the box* *onto* *the coaster* for goals and modifiers, respectively). When they did use optional function words, Directors disambiguated early 64.5% of the time, a rate that was marginally greater than chance, $t1(34) = 1.76$, $p < .10$; $t2(19) = 2.47$, $p < .03$.

Finally, it is of interest to know whether the cues provided by prosodic boundary marking and optional disambiguating function words are distributed redundantly, reciprocally, or independently of each other, as this may shed light on how such cues emerge. We re-analyzed the normalized boundary differences for utterances that were produced with and without optional function words separately, finding that speakers marked boundary differences strongly whether they inserted disambiguating function words or not (for utterances *with* optional function words, $F1(1,9) = 53.09$, $p < .001$; $F2(1,12) = 103.36$, $p < .001$; for utterances *without* optional function words, $F1(1,1) = 812.13$, $p = .02$; $F2(1,5) = 54.81$, $p = .001$). Although there were too many missing data to compute individual correlations for these two cues for each speaker's target utterances, we computed correlations over all the target utterances for goal and modifier attachments (with the rationale that although this violates the independence assumption, if there is little or no correlation given the inflated degrees of freedom, it is probably safe to conclude that these cues do not occur either redundantly or reciprocally). There was no correlation between these cues for the 93 target utterances with goal attachment ($r = .051$, n.s.) and a very small one for the 155 target utterances with modifier attachment ($r = .158$, $p < .05$), so there was little evidence that optional function words and the size of boundary differences are anything other than independent of one another.

2.4.2. Matchers' interpretation of disambiguating cues

2.4.2.1. Performance. Upon hearing Directors' instructions, Matchers responded by immediately moving the correct object to the correct target location 93.8% of the time in the 480 critical trials; 1.7% of the time they did so after requesting clarification and 4.4% of the time they did so after Directors initiated a repair. Out of 93 utterances that were syntactically ambiguous (that is, with no disambiguating

[2] The repeated measures design resulted in missing data and relatively low power in the analyses for optional function words, such that only 40% of speakers contributed to the by-speakers analysis ($F1(1,17) = 2.90$, $p = .107$) and 30% of the items contributed to the by-items analysis ($F2(1,5) = 3.82$, $p = .108$). The means by-speakers were consistent with an audience design explanation, whereas the means by-items were not.

function words), 78 (84%) were followed immediately by Matchers' correct responses, and the rest, by correct responses after brief clarifications. This strong performance by Matchers demonstrates that on the whole, Directors produced sufficient disambiguating cues for their intended meanings to be recognized, even in syntactically ambiguous utterances. There was only one uncorrected error in all 480 trials; in this case, the Director's instruction signaled a modifier interpretation rather than the targeted goal interpretation, with a strong boundary difference of −20.6% (that is, in the wrong direction). As Table 3 shows, the mean boundary differences were either somewhat weaker or indicated the wrong structural interpretation altogether in cases where Matchers did not immediately place objects in the target locations (although there were too few instances of misunderstandings for a statistical comparison to be made). This pattern supports the prediction that prosodic cues help addressees in disambiguation. Addressees in this task did better than Snedeker and Trueswell's (2003) addressees, who performed the correct actions 70% of the time, and Price et al.'s (1991) listeners, who selected the correct context for ambiguous utterances in a 2-item forced choice task 71% of the time. Speakers and addressees in our task were visually co-present to the same display (unlike in Snedeker and Trueswell's task), and speakers were naive and untrained (unlike in Price et al.'s study where speakers were professional radio announcers). Unlike both of these other tasks, ours allowed speakers and addressees to interact freely.

2.4.2.2. Eyegaze. Of the 78 syntactically ambiguous utterances (lacking optional disambiguating function words) that were immediately interpreted correctly, 70 yielded usable eyetracking data (the other 8 involved events such as blinks or faulty tracking). Often, Matchers did not consider distracters at all; in 40% of these trials, Matchers picked up the correct target object and moved it to the correct destination

Table 3
Experiment 1: Strength of Directors' prosodic boundary differences for trials in which responses were immediately correct vs. clarified or repaired responses (percentage of utterances containing optional words are in parentheses)

	Addressee behavior	
	Immediate correct responses	Clarification requests or repairs
Goal attachments		
Unambiguous situation	+16.9% (59.2%)	+14.6% (100%)[a]
Ambiguous situation	+19.3% (80.0%)	+11.1% (25.0%)
Total (goals)	+17.9% (67.9%)	+11.5% (33.3%)
Modifier attachments		
Unambiguous situation	−12.1% (62.0%)	+3.9% (50.0%)
Ambiguous situation	−9.6% (66.7%)	−8.5% (22.2%)
Total (modifiers)	−10.7% (64.1%)	−6.7% (30.8%)

Note that the number of trials on which a clarification or repair was required was very low: Directors had to clarify/repair listeners' responses on only 22 (5 unambiguous and 17 ambiguous) of the 480 critical trials.

[a] There was only one trial in this cell.

without any looks at all to distracter destinations. In fact, whether the display supported only one or both interpretations did not seem to matter; Matchers ignored distracter destinations 46% of the time even when the syntactically ambiguous utterance was truly ambiguous in its referential situation (57% for goals and 41% for modifiers).

Because goal PP-attachment for our target utterances is marked with a lengthened boundary at the end of the first NP and modifier PP-attachment is marked with a lengthened boundary at the end of the second NP, we expected that addressees would disambiguate goal interpretations earlier than modifier interpretations if they rely on prosodic cues. For the cases in which utterances were syntactically ambiguous and both interpretations were supported (the two left-hand cells in Fig. 1), prosody is presumably the sole disambiguating cue as to whether the Director intends the second PP to be interpreted as goal or modifier. It is worth considering exactly when the prosodic cue occurs in these cases. If addressees actually compare the first and second boundaries in an utterance like "*Put the $\underline{dog}_1$ in the $\underline{basket}_2$ on the $\underline{star}_3$*", they would not have sufficient disambiguating information until after *basket*. If on the other hand they judged boundaries on the basis of a saliently elongated NP, then they might be able to recognize goal interpretations earlier than modifier interpretations. This second possibility appeared to be borne out. Upon hearing syntactically ambiguous instructions, Matchers touched the intended target object sooner for goals (1295 ms after the onset of the third NP) than for modifiers (2545 ms), $F1(1,\ 13) = 21.17$, $p < .001$; $F2(1,12) = 4.64$, $p = .05$.[3] This was also supported by eyegaze behavior; Matchers also looked at the correct destination sooner after the onset of the third NP for goals (304 ms) than for modifiers (1314). Table 4 shows latencies to looks and touches.

2.5. Discussion

It is clear that speakers spontaneously mark the structure of syntactically ambiguous utterances using prosody, even when the referential context supports only one interpretation of the utterance. This result suggests that, at least in the sorts of situations we tested, speakers produce prosodic cues regardless of the specific needs of addressees in a referential context. It is particularly striking that addressees so frequently considered only the interpretations that speakers intended them to, even when utterances were both syntactically and situationally ambiguous. This evidence that prosody plays a strong and reliable role in disambiguation is good news for researchers working on machine processing of spoken dialog, because it means that

[3] This was true even when we compared the points of disambiguation for destinations in situations where the set of objects supported both interpretations; the points of disambiguation were at N2 for goals and N3 for modifiers. For goal attachment in unambiguous situations, the point of disambiguation is at the onset of N3 ("*star*" in Fig. 1, lower right); note that prosody cannot be of any help here since both baskets are plausible destinations. For modifier attachments in unambiguous situations (Fig. 1, upper right), there are two points of disambiguation; for the object to be moved, it is N2 ("*basket*"), and for the destination, it could be at the onset of the final PP ("*on the star*") or else N3 ("*star*") itself.

Table 4
Experiment 1: Latencies (ms) to first looks to targets, distracters, and destinations from first and third noun onsets and from the point of disambiguation[a] (POD) to the touch

N1 N2 N3
"Put the dog in the basket on the star"

	N1-Onset to target look	N3-Onset to dest. look	POD noun to dest. look	N3-Onset to touch	POD noun to touch
Ambiguous					
Goals	1133	304	1333	1295	2324
Modifiers	1962	1314	1314	2545	2545
Unambiguous					
Goals	1633	512	512	2500	2500
Modifiers	1390	1219	1219	2264	2264

[a] Note that the POD is N3 for all cases except for ambiguous goal interpretations, where it is earlier (N2).

prosodic cues are likely to reflect syntactic structure in a consistent way, and so prosody is well worth pursuing as a cue for machine parsers of human speech. However, this result does not provide support for the audience design hypothesis, which predicts that prosodic variation is driven by listeners' needs.

The second way in which speakers failed to adjust to their addressees' particular needs was in their use of optional disambiguating function words such as *that* and *into*; these were produced equally often in syntactically ambiguous and unambiguous sentences. This finding is consistent with those of Ferreira and Dell (2000), whose speakers tended to insert an optional function word when it was available in the planning process rather than when the sentence would be ambiguous to an addressee without it. At best, in those studies optional function words might have represented a coarse adjustment, as they were produced at a slightly higher rate with addressees than without (Ferreira & Dell, 2000). In the current experiment, when speakers *did* use optional function words, the word or words inserted into the utterance tended (albeit marginally) to allow disambiguation sooner rather than later, which may represent a somewhat coarse sort of sensitivity to addressees' needs.

It would be premature to conclude at this point that audience design is not a factor in the production of these two kinds of disambiguating cues. Horton and Gerrig (in press) pointed out two prerequisites for successful audience design: (1) speakers must be aware of or sensitive to addressees' needs, and (2) they must be capable of responding in ways that address these needs. If these prerequisites are satisfied and speakers *still* do not adapt their speech for addressees' needs, then that would constitute evidence that speakers are behaving egocentrically. It is unclear whether Snedeker and Trueswell's subjects were aware of specific ambiguities at the time they were speaking, as they were not interviewed about their awareness of ambiguity until the experiment was over. By having subjects switch Director-Matcher roles in Experiment 1, we tried to maximize the likelihood that second Directors would be aware of the potential for ambiguity, having already experienced it themselves as Matchers. But we failed to find any evidence whatsoever for a difference between first and second Directors in prosodic boundary marking, and there was no conclusive evidence that experience mattered in the use of optional function words.

Our null findings do not settle the audience design question, but lead to an important point about awareness: experience with a general potential for ambiguity may not bring speakers any closer to being able to adjust their production of disambiguating cues to reflect what addressees may need in a specific situation. To engage in audience design with any precision, speakers would need to be able to detect the ambiguity (or lack thereof) in a specific display by the time they began speaking. But it is possible that our speakers were not able to do this. In fact, when questioned during debriefing, many of them indicated that they thought that the situations that we intended to be *unambiguous* were actually *ambiguous*. It seems that this was because our unambiguous displays, while supporting only one interpretation, nevertheless were *temporarily* ambiguous: in order to be as complex as the ambiguous displays, they always contained duplicate objects (see Fig. 1 for an example). Such ambiguity, although temporary, may have been particularly salient to the Directors, who might have assessed ambiguity based on the amount of redundancy in the displays. If this is the case, we cannot rule out the possibility that Directors *were* taking the needs of their listeners into account (at least as they understood those needs) by consistently providing disambiguating prosody.

To understand the extent to which Directors could have been aware of Matchers' needs by the time they began speaking, Experiment 2 evaluates people's speed and accuracy in assessing the potential for ambiguity in target utterances with respect to the complex displays used in Experiment 1 and some simpler displays to be used in Experiment 3.

3. Experiment 2

Experiment 2 aimed to clarify whether the speakers in Experiment 1 were likely to have noticed ambiguity by the time they began to speak, as well as whether they may have relied on the presence of duplicate objects in the display as a heuristic for assessing ambiguity (rather than analyzing exactly how the particular utterance being planned mapped onto the display). A different group of participants individually performed a simple ambiguity judgment task.

3.1. Method

3.1.1. Participants

Forty-four undergraduates (23 men and 21 women) from the State University of New York at Stony Brook participated for course credit. All participants were 18 years of age or older, and all were native English speakers. None had participated in Experiment 1.

3.1.2. Materials and design

The materials consisted of sentences displayed on a screen, each accompanied by an image consisting of a photograph of a set of movable objects. The photographs were taken using a digital camera, digitized into Adobe Photoshop, and edited so that each image was 1440×1080 pixels. The sentences consisted of the 20 critical instructions from Experiment 1 (e.g., *Put the dog in the basket on the star*) plus 10 new instructions and 10 fillers.

The design of Experiment 2 is illustrated in Fig. 3. Each of the 20 critical instructions from Experiment 1 was paired with images of a (globally) ambiguous display (from Experiment 1), a temporarily ambiguous display (that had been considered ‘‘unambiguous’’ in Experiment 1), and a new, simpler unambiguous display (containing no duplicate objects). The 10 new instructions (Fig. 3, lower right) were unambiguous, but referred to *unique* objects in displays containing some duplicate objects (these were the same displays that were paired with ambiguous instructions from Experiment 1, Fig. 3). The globally and temporarily ambiguous displays reproduced from Experiment 1 contained equal numbers of duplicate objects and supported both modifier and goal interpretations of the (text) instructions they were paired with. The unambiguous displays supported only one interpretation (modifier or goal). The 10 filler instructions were paired with (new) ambiguous displays so that ‘Clear’ and ‘Ambiguous’ responses would be equally likely.

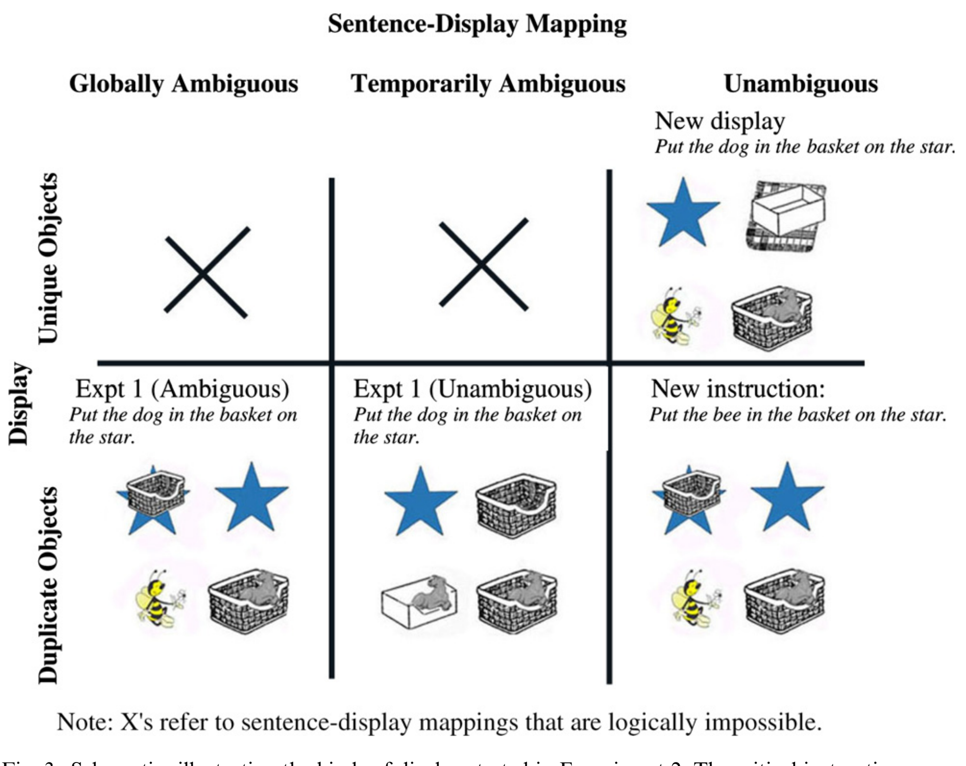


Fig. 3. Schematics illustrating the kinds of displays tested in Experiment 2. The critical instructions were the same for each display except for the panel in the lower right, which had a new instruction paired with the same display as in the lower left.

Critically, all of these sentences are syntactically ambiguous between goal and modifier attachments when considered out of context. When identical sentences are paired with different displays, we can tease apart the effect of object redundancy from actual referential ambiguity. This allows us to test the hypothesis that perhaps people tend to rely on a heuristic of using duplicate objects in a display to estimate the potential for ambiguity of a sentence.

There were four versions of the experiment, each containing 60 trials (30 ambiguous and 30 unambiguous). Each critical instruction appeared three times within a version, but each time was paired with a different display. Trials were presented in a single, pseudo-random order, with the constraints that no two identical images be presented consecutively, that trials with the same instruction be presented at least eight trials apart, and that no more than four ambiguous or four unambiguous trials be presented consecutively.

3.1.3. Procedure

Participants were told to keep their two index fingers over two keys labeled 'Clear' and 'Ambiguous' (the 'f' and the 'j' keys, respectively), and to respond as quickly as possible. Trials were self paced. A photograph of a set of objects would appear, with labels next to each object. These labels were given first so that participants would already be familiar with the names of each object once the instruction appeared. This was done to avoid difficulties in recognizing the objects from their labels and to make the procedure as similar as possible to that of Experiment 1, where the experimenter named each object as she positioned it. The labels remained on the screen for 5000 ms. Once the labels disappeared, participants could take as long as they wanted to familiarize themselves with the display, as in Experiment 1. When they were ready, they pressed a key and the instruction appeared at the top of the screen. After reading the instruction, they judged it as clear or ambiguous, given the objects in the display. Reaction times and response accuracy were recorded.

3.2. Results and discussion

Reaction times over three standard deviations above the mean were discarded (1.7% of trials). ANOVAs were used to compare the 4 types of sentence-display pairings in Fig. 3 by type of sentence attachment (goal or modifier). Planned comparisons were used to test: (1) whether people analyzed the specific ambiguity in a sentence-display mapping, contrasting ambiguous to unambiguous situations, or (2) whether they relied on a duplicate object heuristic, contrasting trials with duplicate objects (globally and temporarily ambiguous sentence-display mappings from Experiment 1 as well as new sentences referring unambiguously to unique objects in displays that also contained duplicate objects) to trials without duplicate objects (sentences from Experiment 1 paired unambiguously with new displays containing only unique objects).

Overall, people took an average of 4.49 s to judge ambiguity correctly, with correct judgment times for unambiguous and ambiguous responses about equal (see Table 5). When a display contained some duplicate objects, responses were more

Table 5
Experiment 2: Mean accuracy and reaction times for all display types

	Globally ambiguous	Temporarily ambiguous	Unambiguous
Unique object displays			
			(*New display*)
% Correct			88.8% (.32)
RT			
Correct items	×	×	4141 ms (1941)
Incorrect items			4902 ms (2579)
Total RT			4227 ms (2033)
Duplicate object displays			
			(*New instruction*)
% Correct	53.9% (.49)	74.7% (.43)	79.2% (.41)
RT			
Correct items	4506 ms (2065)	4451 ms (2107)	4839 ms (2392)
Incorrect items	3430 ms (1868)	4600 ms (2407)	5230 ms (2557)
Total RT	4010 ms (2047)	4489 ms (2185)	4920 ms (2429)

Standard deviations appear in parentheses.

likely to be correct when the sentence contained an instruction to move a unique object (Fig. 3, lower right panel) vs. one of the duplicate objects (lower left panel), 79.2% correct as opposed to 53.9%, $F1(1,43) = 19.07$, $p < .001$; $F2(1,19) = 84.66$, $p < .001$. This higher accuracy was achieved at the expense of speed (about 910 ms slower), $F1(1,43) = 43.86$, $p < .001$; $F2(1,19) = 62.30$, $p < .001$, suggesting that people were more likely to take the time to fully analyze the sentence-display mapping when the object to be moved was unique. Next, we compared cases where the instructions were identical, but the displays were different (containing either duplicate objects as in Fig. 3, lower left and center panels, or all-unique objects, as in the upper right panel). People were more accurate at judging ambiguity for a given instruction when it was paired with a display of all unique objects than with a display containing duplicate objects, 88.8% to 60.8%, $F1(1,43) = 98.52$, $p < .001$; $F2(1,19) = 113.23$, $p < .001$ and equally fast, $F(1,43) = .1$, n.s.; $F(1,19) = .05$, n.s. (see Table 5).

Finally, we compared situations in which the sentence-display mapping consisted of an unambiguous instruction to move a unique object, but within a display context of either unique or duplicate objects (the upper vs. lower right-hand panels of Fig. 3). People were 9.6% more accurate in their responses in situations with all-unique objects ($F1\ (1,43) = 7.93$, $p < .01$; $F2(1,19) = 16.44$, $p = .001$), as well as 698 ms faster to respond correctly, $F1(1,43) = 5.6$, $p < .03$; $F2(1,19) = 4.22$, $p = .05$. Together, these results provide clear evidence that redundancy in a display is used as a cue to potential ambiguity. This is consistent with what participants reported in Experiment 1, that some of the "unambiguous" displays seemed ambiguous to them (all of the displays in Experiment 1 had at least one pair of duplicate objects).

There was an overall speed–accuracy tradeoff among the 44 participants ($r = .51$, $p < .001$). This is particularly salient for the globally ambiguous displays: people were very quick to judge these displays incorrectly (performing at chance), suggesting that they were often guessing. Although there did appear to be a wide range of abilities when it came to judging ambiguity quickly and accurately (those individuals who were correct at least 60% of the time averaged from 2528 to 7864 ms to judge ambiguity, and those who were accurate at least 90% of the time or more averaged from 3903 to 6115 ms), it also appears that judging ambiguity in this situation is not an easy task. It seems that with the exception of the unique-object displays (which were both the fastest and the most accurately judged), when people tried to judge ambiguity, they had to spend considerable time evaluating the situation with respect to the accompanying instruction in order to be accurate. Overall, the present results strongly suggest that speakers in our first experiment were often unaware of situational ambiguity as they began to speak.

4. Experiment 3

Can the marking of prosodic boundaries and the production of optional function words be attributed to audience design? Although we found no evidence in Experiment 1 that speakers adapted specifically to addressees' needs, the mean time it took in Experiment 2 to correctly detect ambiguity (4.49 s) suggests that Experiment 1's speakers may not have been able to detect specific ambiguities by the time they began to speak. Even if they did not take the time to be precise, speakers still may have engaged in audience design of a coarser sort; they may have used heuristics such as estimating the potential for ambiguity based on the redundancy of objects in the display, as suggested by Experiment 2. In that case, audience design may be manifested as more (or stronger) disambiguating cues when speakers refer to non-unique objects in a display than to unique objects. It would seem less costly for speakers to produce prosodic cues and optional words liberally in the face of any potential ambiguity, rather than to produce them only after monitoring addressees' disambiguation needs in a fine-grained way.

In an attempt to leave no stone unturned in testing for audience design, we conducted Experiment 3. Speakers followed schematic drawings on cards to produce instructions that referred to objects in the same ambiguous and temporarily ambiguous situations as in Experiments 1 and 2, as well as to unique objects in the same simple displays as in Experiment 2. We also reasoned that speakers may be more likely to engage in audience design–whether coarse or fine–with an interacting audience present, and so we had 12 Directors produce the instructions without any Matcher to carry them out, and 24 participants produce the instructions in 12 interacting Director-Matcher pairs, as in Experiment 1. Finally, in case there are individual differences in speakers' abilities to detect specific ambiguities in the mapping of utterance to display, we had each participant individually perform a briefer version of the screen-based ambiguity judgment task from Experiment 2 after completing the instruction task with physical objects. This enabled us to look for audience design behavior by speakers who were more and less skilled at detecting ambiguity.

This design enables a more tenacious test of the audience design hypothesis than did our first experiment. In the instruction-giving task, speakers can tailor their utterances to addressees' needs only when they recognize those needs in the context of an instruction-display mapping—whether by computing ambiguity precisely or by using cues provided by duplicate objects to estimate a *potential* for ambiguity. If speakers do engage in audience design for this level of speech planning, then they should disambiguate goal from modifier structures more strongly when instructing addressees to move non-unique objects in redundant displays (that are temporarily or globally ambiguous) than unique objects in simple (unambiguous) displays. Or else audience design may guide these aspects of the utterance only coarsely—speakers may produce more of one or another disambiguating cue when interacting with a co-present addressee than when speaking to an absent, generic addressee.

4.1. Method

4.1.1. Participants

Thirty-six undergraduates (22 women and 14 men) from the State University of New York at Stony Brook volunteered to participate in exchange for research credit in a psychology course. All participants were 18 years of age or older, all were native English speakers, and none had participated in Experiments 1 or 2. Twenty-four of them participated as 12 pairs, switching Director-Matcher roles halfway through the instruction task as in Experiment 1, and 12 participated singly as Directors with no Matcher present.

4.1.2. Design and materials

As before, the referential situations crossed attachment (modifier vs. goal) with ambiguity type. There were 3 types of ambiguity: globally ambiguous, temporarily ambiguous (both of these referred to a non-unique object in a redundant display from Experiment 1), and unambiguous (referring to a unique object in a simple display from Experiment 2). Two critical items from Experiment 1 were removed for counterbalancing purposes. These two were chosen to be removed because the speakers in Experiment 1 produced only one usable token (that is, with the targeted prepositional phrases in the targeted order) for each of these items in the goal condition. This left 18 critical items. The 2×3 design was fully balanced; all 18 critical items appeared in all six conditions, and all participants experienced all conditions. Six versions of the experiment were created, with each item appearing once on each version. Each version contained three items in each of the six conditions, for a total of 18 critical items. Items were presented in a single random order, and participants were randomly assigned to one of the six versions.

4.1.3. Procedure

The task was identical to that in Experiment 1, except that in the pairs, Matchers did not wear an eye-tracker, and the 12 lone Directors produced instructions without any interacting Matcher. In both conditions, the experimenter placed each object on the display in front of the Director and named it before handing over the schematic guide from which to plan the instruction. Within the pairs, people were randomly assigned to the role of Director or Matcher; after completing one version of the experiment, they switched roles and completed a second version. The task and instructions for lone Directors were as similar as possible to those for pairs (which were the same as in Experiment 1), except that lone Directors were told that the purpose of the experiment was to collect recorded instructions for use in later experiments in which other people would follow their instructions. This is consistent with previous production studies in which speakers were asked to read aloud into a microphone with no addressee present (e.g., Allbritton et al., 1996).

After participants completed the task, they were seated individually in front of a computer and performed an ambiguity judgment post-test consisting of 20 trials from Experiment 2. Ten of these trials were unambiguous and 10 were ambiguous. The displays were drawn from those in Experiment 1 and all of them contained non-unique objects. The purpose of this post-test was to evaluate whether speakers who perform relatively poorly at assessing ambiguity might produce fewer or differently distributed cues than speakers who perform relatively well.

4.1.4. Coding and analysis

As in Experiment 1, we digitized and transcribed the critical utterances, coded the duration of the first and second noun phrases and their following pauses, and computed difference scores (boundary 1 − boundary 2) that were normalized for utterance length. Optional function words were categorized as in Experiment 1. In addition to the factors of attachment and ambiguity type, the three factors of particular interest in Experiment 3 were the speaker's experience (as first or second Director), the presence of an interacting addressee (Director-Matcher pairs or lone Director), and speakers' ability to detect ambiguity, as measured in the post-test. Due to the spontaneous nature of the production task (in which speakers did not always produce the targeted forms) and the within-subjects and within-items design, missing observations prevented conducting ANOVAs by-subjects and by-items with all of these factors simultaneously. So separate by-subjects and by-items ANOVAs were conducted, each with one of the three factors of interest, along with attachment (modifier or goal) × ambiguity type (ambiguous, temporarily ambiguous, and unambiguous). In each ANOVA, two planned contrasts tested for the effects of (1) display redundancy, or unique unambiguous objects (upper right-hand panel of Fig. 3) vs. non-unique temporarily or globally ambiguous objects (lower middle and lower left panels) and (2) the type of ambiguity (temporary, or the lower middle panel, vs. global, or the lower left panel).

4.2. Results and discussion

Speakers produced utterances in the target forms somewhat more often than in Experiment 1 (about 68% to 51% ; see Table 6). Directors working in pairs produced analyzable utterances 63% of the time, and lone Directors did so 77% of the time. Of the 440 analyzable utterances, 58% were structurally ambiguous (compared to 38% in Experiment 1) and the rest had optional disambiguating function words. As in Experiment 1, utterances with modifier attachment contributed more analyzable

Table 6
Experiment 3: Frequencies with which speakers' spontaneous instructions matched the desired forms (for Dyads and Solo Directors), out of 648 trials

Spontaneous utterance:	Syntactically ambiguous *Put the dog in the basket on the star*			Disambiguating function words *Put the dog that's in the basket on the star*			Total
Display:	Duplicate objects		Unique objects	Duplicate objects		Unique objects	
Situational ambiguity:	Globally ambig.	Temp. ambig.	Unambiguous	Globally ambig.	Temp. ambig.	Unambiguous	
Goal attachment							
Dyads[a]	22	26	17	16	7	24	112
Solo	17	17	15	9	9	11	78
Total	39	43	32	25	16	35	190
Modifier attachment							
Dyads	22	49	33	20	40	30	194
Solo	14	0	22	10	0	10	56
Total	36	49	55	30	40	40	250
Total	75	92	87	55	56	75	440

[a] Cells in the Dyad rows offered twice as many opportunities for speakers' instructions to match the desired forms.

tokens than utterances with goal attachment, 250–190, $F1(1,35) = 19.75$, $p < .001$; $F2(1,17) = 8.47$, $p = .01$, and those that were temporarily ambiguous (the "unambiguous" ones from Experiment 1) contributed marginally more tokens than the ambiguous ones, $F1(1,35) = 6.22$, $p < .02$; $F2(1,17) = 2.78$, $p = .11$. Utterances in situations where the objects were all unique were somewhat more likely than utterances in the other types of situations to follow the predicted forms, by-subjects but not by-items, $F1(1,35) = 12.37$, $p = .001$; $F2(1,17) = 1.33$, n.s.

4.2.1. Prosodic marking of syntactic boundaries

Speakers robustly marked syntactic boundaries by lengthening them, just as in Experiment 1; in a 2×3 ANOVA with attachment (modifier or goal) vs. ambiguity (ambiguous, temporarily ambiguous, and unambiguous) as factors, first noun phrase boundaries were longer than second by 14.7% of the utterance for goals and shorter than the second for modifiers by 8.8% of the utterance, $F1(1,20) = 103.19$, $p < .001$; $F2(1,15) = 155.20$, $p < .001$. Whether a Director addressed an actual Matcher or not, prosodic boundary marking occurred to the same degree, $F1(1,20) = 0.05$, n.s.[4] Nor did the redundancy of objects (as a potential proxy for situational ambiguity) matter; structural boundaries were marked just as strongly when there were duplicate objects as when the objects were unique, $F1(1,20) = .07$, n.s.; $F2(1,15) = .31$, n.s. As in Experiment 1, boundaries were just as strongly marked

[4] By-items ANOVAs (F2) are not provided for those comparisons where there were too many missing cells for repeated-measures comparisons to be carried out.

in temporarily ambiguous situations as in globally ambiguous situations, $F1(1,20) = 1.19$, n.s; $F2(1.25)$, n.s. Also as in Experiment 1, whether a speaker served as the first Director or the second within a dyad (having already experienced the task as Matcher) made no difference, $F1(1,10) = .299$, n.s. Table 7 shows the mean boundary differences produced by Directors for trials in which Matchers were immediately correct versus correct after a clarification or repair.

4.2.2. Optional function words

Speakers produced optional disambiguating function words in 42.3% of the 440 analyzable utterances, at similar rates for goals and modifiers (40–44%), different by-subjects but not by-items ($F1(1,20) = 8.17$, $p = .01$; $F2(1,15) = 1.07$, $p = .32$). The breakdown by goal/modifier and display is in Table 7. Optional function words did not appear to be responsive to addressees' needs: It did not matter whether a Matcher was present ($F1(1,21) = .03$, n.s.), whether speakers had experience with the Matcher's task ($F1(1,11) = .40$, n.s.), whether the utterance-display mapping was temporarily ambiguous or globally ambiguous (the two kinds of situations in Experiment 1), $F1(1,20) = .07$, n.s.; $F2(1,15) = 2.90$, $p = .11$, nor whether the display contained redundant or only unique objects ($F1(1,20) = .19$, n.s.; $F2(1,15) = 4.92$, $p < .05$) (in fact, the means for this last by-items analysis of display redundancy went in the direction opposite to that predicted for audience design; that is, speakers provided optional words 46% of the time when addressees did not need them, 42.3% when there was global ambiguity, and 37.8% when there was temporary ambiguity). When Directors used optional function words, these words disambiguated early 64% of the time, replicating, Experiment 1.

Table 7
Experiment 3: Strength of Directors' prosodic boundary differences[a] for trials in which responses were immediately correct vs. clarified or repaired (percentage of utterances containing optional words are in parentheses)

	Immediate correct responses	Clarification requests or repairs
Goal attachment		
Displays with duplicate objects:		
Temporarily ambiguous	+15.4% (27.1%)	(No cases)
Globally ambiguous	+14.6% (47.2%)	+14.3% (0%)
Displays with unique objects:		
Unambiguous	+14.3% (52.2%)	(No cases)
Mean (goals)	+14.7% (42.2%)	+14.3% (0%)
Modifier attachment		
Displays with duplicate objects:		
Temporarily ambiguous	−10.4% (44.9%)	(No cases)
Globally ambiguous	−6.5% (46.0%)	−12.4% (33.3%)
Display with unique objects:		
Unambiguous	−8.7% (42.1%)	(No cases)
Mean (Modifiers)	−8.7% (44.3%)	−12.4% (33.3%)

[a] The strength of prosodic boundary differences was calculated as in Experiment 1.

4.2.3. Independence of cues

As in Experiment 1, the degree of boundary lengthening and the presence of optional function words appeared to be independent cures. For the 190 analyzable utterances with goal attachment and the 250 with modifier attachment, there were no correlations of these cues ($r = .044$, n.s. for goals and $r = -.074$ for modifiers).

4.2.4. Matchers' performance

As in Experiment 1, Matchers performed extremely well, moving the correct object to the correct location in immediate response to the Director's instruction 96% of the time in the 648 trials. Five times (about 1% of the time) they requested clarification before performing the correct action, and 18 times (about 3% of the time) they made an incorrect move and then repaired it at the Director's prompting. There were no uncorrected errors. Out of 75 utterances that were both syntactically ambiguous (that is, with no disambiguating function words) and globally ambiguous (the referential context supported both possible parses), 62 (83%) were followed immediately by Matchers' correct responses, and the rest, by correct responses after brief clarifications. As in Experiment 1, this performance exceeds the 70–71% reported for listeners in Snedeker and Trueswell's (2003) and Price et al.'s (1991) experiments. While there were too few clarifications and repairs to test for an effect of experience, it is interesting that 13 of 14 instances of clarifications and repairs occurred when the first Director was speaking with the first Matcher.

4.2.5. Directors' skill at judging ambiguity (post-test)

As Horton and Gerrig (in press) pointed out, speakers must be aware of addressees' needs early enough in speech planning if they are to engage in audience design. As in Experiment 2, there was a wide range of skills at detecting ambiguity in the post-test. Based on the accuracy scores on the post-test, we divided speakers into halves who were more versus less skilled at judging accuracy on the post-test and compared the disambiguating cues they provided by attachment and ambiguity type. Speakers who were more skilled appeared to contrast the normalized boundary differences for goal vs. modifier attachments just as strongly as those who were less skilled, although those speakers who were more skilled tended to produce slightly more optional words overall, $F1(1,\ 21) = 5.82$, $p < .03$.

For the 24 speakers who interacted in pairs during the instruction task, there was an overall speed/accuracy tradeoff in the post-test afterward ($r = .52$, $p = .009$), just as there was for the participants in Experiment 2's ambiguity judgment task. There was no speed/accuracy tradeoff evident in the post-test for the 12 speakers who had done the instruction-giving task alone ($r = -.37$, n.s.).

5. General discussion

In three experiments, we examined whether a particular kind of variation in spontaneous language use, the lengthening of words and pauses at syntactic boundaries, indeed represents one partner's adaptation to the needs of another. Demonstrating audience design for a particular cue requires satisfying three criteria (Brennan & Williams, 1995): (a) that speakers spontaneously produce the cue, (b) that addressees successfully use it in disambiguation, and (c) that speakers produce it more strongly when addressees actually need it (as opposed to simply as a by-product of speakers' planning and articulation, whether addressees need it or not).

Our strongest and clearest conclusion is that speakers *do* reliably lengthen syntactic boundaries. Moreover, addressees are highly accurate in the presence of this cue, even when the utterance is syntactically ambiguous and the referential context supports both interpretations—in fact, addressees often appear to consider only the interpretation the speaker intended. So by the first two criteria, prosodic boundary marking certainly appears to count as an informative cue. But it is *not* one driven by audience design: the cue persists regardless of whether the addressee actually needs it for disambiguation or whether the speaker actually recognizes the addressee's need for it.

We found that speakers spontaneously and reliably produced disambiguating prosody regardless of all of the following: the utterance's actual ambiguity in its referential context, including whether optional disambiguating words were used (Experiments 1 and 3), the redundancy of objects in the display (Experiment 3), their own experience with the task (Experiments 1 and 3), their individual ability to assess referential ambiguity in a timely fashion (Experiments 2 and 3) and even whether they were addressing an interacting addressee or speaking alone (Experiment 3). Taken together, these findings support the conclusion that in spontaneous speaking, prosodic marking of syntactic boundaries emerges from planning and articulating syntactic structure rather than from awareness of an addressee's need for disambiguation.

Of course, null findings must be interpreted cautiously, but we are inclined in this case to conclude that audience design does not drive prosodic boundary marking. Experiment 3 searched for audience design in four different ways: by attempting to increase speaker awareness of the need for disambiguation via prior experience (and replicating the null effect from Experiment 1), by loosening the criterion for ambiguity to include any adaptations that may be driven by *potential* ambiguity (comparing displays with and without duplicate objects), by ascertaining whether individuals who are good at quickly assessing ambiguity produce stronger boundary cues than other individuals, and by looking for the coarse form of audience design that could be said to occur if there were stronger cues overall with addressees present than without. Not only did we fail to find any effects of these manipulations, but in a few cases the means (some of them based on small numbers) tended to pattern in the opposite direction than predicted by the audience design hypothesis.

We also tested for a related effect of secondary interest, that of addressees' need for disambiguation on speakers' production of optional disambiguating function words. While we found no clear effects of audience design with respect to this cue, these findings are not conclusive. As Frick (1995) noted, issues of insufficient power arise when p values are non-significant yet low, as they were in several of our key comparisons for optional function words. Our spontaneous planning and production task allowed speakers considerable freedom, resulting in missing data in Experiments 1 and 3 when speakers departed from our expectations, and so low power is certainly a concern. Note that Ferreira and Dell's (2000) studies suggest a limited role for audience design in the production of optional function words; in their studies, utterances that would be syntactically ambiguous without such words contained no more complementizers than did utterances that were unambiguous, but utterances directed at addressees who needed to understand them contained marginally more complementizers, whether they were necessary for disambiguation or not (a coarse audience design effect). Optional function words were of secondary interest in the current project; more study is necessary to determine whether (or under what circumstances) this cue is responsive to audience design.

6. Conclusions

It is clear from our experiments that speakers disambiguate syntactic boundaries prosodically when speaking spontaneously, and they do so regardless of addressees' needs. Findings elsewhere that speakers fail to disambiguate prosodically when reading aloud (e.g., Allbritton et al., 1996 and Keysar & Henly, 2002) underscore the need for studying language processing in interactive discourse settings and with communicative tasks. Given that speaking is affected by interaction, purpose, and the moment-by-moment demands of planning in the real world, we used a method that elicited syntactically ambiguous utterances in the desired forms, while still having speakers spontaneously plan what they were to say and interact freely with their addressees. Evidence culled from situations that accommodate these forces on speaking and that represent a good faith effort to find audience design can, we propose, be used in support of conclusions that accept the null hypothesis.

Although audience design may not drive prosodic marking of syntactic boundaries, it does influence production at other levels, such as with the marking of referring expressions as definite (Bard & Aylett, 2000; Clark & Wilkes-Gibbs, 1986; Wilkes-Gibbs & Clark, 1992; Lockridge & Brennan, 2002) and the syntactic packaging of information in response to addressees' needs (Lockridge & Brennan, 2002). Partner-specific effects can influence comprehension from its earliest moments as well, demonstrated by addressees' differing interpretations of the same referring expression spoken by different speakers (Metzing & Brennan, 2003), as well as by the finding that addressees use their knowledge of speakers' goals to immediately restrict candidate referents for ambiguous expressions (Hanna & Tanenhaus, 2004). These effects of partner-specific processing and audience design at other levels are relevant to the present studies because they were measured in settings where participants had communicative goals and needs, and because they appear to occur within the early moments of processing, rather than as late adjustments or repairs (as was argued by Horton & Keysar, 1996, to be the necessary source of pragmatic effects). These findings of audience design effects converge with findings that visual context can have an immediate impact on the production (Brennan, 1990; Brennan, in press) and interpretation (Spivey, Tanenhaus, Eberhard, & Sedivy, 2002; Tanenhaus et al., 1995) of linguistic expressions, in order to strengthen the conclusion that pragmatic information can guide linguistic processing that was once proposed to be encapsulated from such influence (see Brown & Dell, 1987).

Even when audience design can guide linguistic processing in production or in comprehension, it does not always do so; speakers can be distracted, unmotivated, inept, biased, uncooperative, or oblivious (see Schober & Brennan, 2003). And even when audience design effects are detected, their effect sizes may be rather small (e.g., Brennan & Clark, 1996; Lockridge & Brennan, 2002; Metzing & Brennan, 2003). This is not surprising, as there are so many contextual factors that influence language use in conversation. When audience design appears to have no impact, as here with prosodic boundary marking of syntactic structures, this need not mean that processing is necessarily encapsulated from pragmatic influences or could *never* be guided by audience design; it may be that the partner-specific information is *typically* not available early enough to be useful in the situation tested. The potential impact of audience design is constrained by the speaker's awareness of the addressee's needs and the timing with which information about the partner is available.

Our main finding, that speakers disambiguate equally well regardless of addressees' needs, is consistent with Schafer et al.'s findings (2000), in which naive speakers consistently produced prosodic cues even when the utterance was disambiguated by the context of the game. However, it contrasts with Snedeker and Trueswell's conclusion (2003) that speakers disambiguate prosodically only when the speaker is aware that there is ambiguity in the situation. Our task differed in several ways from Snedeker and Trueswell's, as discussed previously: their task was not interactive (speakers did not have any feedback from their addressees and therefore experienced no consequence for misunderstandings), and the utterances in their experiments were shorter than ours, involving two noun phrases rather than three (e.g., *tap the frog with the flower*). The potential boundary contrast in their utterances involved a verb and a noun, and as a result there may have been less opportunity for contrasting prosodic boundaries. Snedeker and Trueswell's speakers did mark syntactic boundaries in experiments in which there was syntactic ambiguity (e.g., *with the flower* could attach to either *tap* as an instrument or *frog* as a modifier), but not in their Experiment 2, where the type of attachment was manipulated between speakers so that each speaker was recalling and articulating sentences that had a single, very repetitive structure. Snedeker and Trueswell (2003) concluded that awareness of ambiguity was the critical factor in their production of informative prosody; however, under the conditions in their experiment, it is also possible that speakers may have simply fallen into using routine, uninformative prosody. We have observed this behavior elsewhere, particularly when communication is not interactive (Brennan & Schober, 2001). The current experiments addressed this by eliciting as many utterances with variable forms as with target forms and by allowing speakers and addressees to interact naturally. Even when our speakers were probably unaware of ambiguity, they produced informative prosody (reinforcing the possibility that the result that Snedeker and Trueswell attributed to awareness may have been due to other aspects of their design).

The role that awareness plays is still an open question, both in speakers' use of disambiguating cues and in audience design more generally. In the last part of our Experiment 3, participants judged sentences in referential contexts to determine how well and quickly they could assess specific ambiguities, displaying a wide range in individual ability. Individual ability turned out not to matter for prosodic boundary marking and to matter only marginally for optional function words. Awareness in Snedeker and Trueswell's experiment was assessed at the end of their task via a questionnaire; this method may have caused speakers to over-report awareness. Neither our task nor Snedeker and Trueswell's enabled probing awareness at specific points during speaking. It may be possible to obtain a measure of awareness by having Directors (rather than Matchers as in our Experiment 1) wear an eye-tracker, to enable observing how thoroughly they search the display as well as whether they look at both the target destination and distracter destination (or repeatedly between them) before speaking.

Our findings support a close relationship between syntactic structure and prosodic lengthening. However, we do not argue that this relationship is a direct one, with prosody simply "read off" a syntactic representation. Ferreira (1993) has argued that the relationship between a syntactic representation and its articulation is mediated by a prosodic representation that can also be influenced by semantic factors (see Ferreira, 1993, for a review and for methods and data that tease apart syntactic from prosodic structure). Presumably, semantic or pragmatic information could influence prosodic lengthening if that information were available before articulation; but this information is unlikely to be available if it requires more time and computation than the system ordinarily expends during ordinary conversation. Speakers can use less effortful means, such as over-producing a cue; in this case, what is easy for speakers also happens to be easy for addressees, but is not tailored to them.

Despite our conclusion from the present experiments that in the case of syntactic boundary marking, words are lengthened "for" the speaker and not "for" the addressee, we see no reason to propose any architectural obstacle to audience design in prosodic variation. Prosodic lengthening or shortening also occurs when words that are *new* in a discourse are pronounced more clearly and in longer forms than *given* words that are repeated and more predictable (Bard et al., 2000; Fowler & Housum, 1987; Lieberman, 1963; Samuel & Troicki, 1998); in those cases, prosodic variations *do* seem to be driven (at least partially) by addressees' needs. Specifically, Gregory, Healy, and Jurafsky (2001, 2003) had the same speaker talk to different addressees in order to tease apart whether the shortening of repeated words reflects the words' information status relative to the speaker or relative to an addressee. Although speakers who used a word initially with one addressee shortened it somewhat when speaking to a second addressee (indicating that shortening reflects a word's information status "for" the speaker to some degree), they shortened it significantly *less* than when the repeated word was directed at the same addressee (reflecting a significant influence of audience design, or adaptation "for" the addressee).[5] It makes sense that an audience design effect would emerge on this level, since

[5] Although Bard et al. (2000) concluded that given/new lengthening was *not* driven by audience design, their corpus compared the length of a word's first production to one addressee and its second production to a second addressee, without the necessary control (second production to the original addressee) that Gregory et al.'s study included.

the information status of a word is a direct result of the speaker's prior linguistic experience in the dialogue (either in having produced the word or having comprehended it), and it is likely to be salient to the speaker when she is addressing a different person. In such a situation, the addressee's needs should not require as much time or effort to assess as did the mappings of utterances to referential contexts in our experiments.

In sum, knowledge about a partner's need for disambiguation can serve as just one of many pragmatic constraints that can potentially affect variation in spoken language. However, partner-specific information will not have a measurable effect unless it is available early enough in planning and articulation. The variation resulting from prosodic lengthening of syntactic boundaries appears to be driven by speakers' processing of syntactic structure rather than by addressees' needs for disambiguation; so while prosodic boundary lengthening turns out to be an informative cue for addressees, it does not seem to arise from audience design.

References

Allbritton, D. W., McKoon, G., & Ratcliff, R. (1996). Reliability of prosodic cues for resolving syntactic ambiguity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(3), 714–735.

Bard, E. A., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., & Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language, 42*, 1–22.

Bard, E. A., & Aylett, M. P. (2000). Accessibility, duration, and modeling the listener in spoken dialogue. *Gtelog 2000, Fourth workshop on the semantics and pragmatics of dialogue*. Gtalog, Sweden: Gothenburg University.

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

Beach, C. M., Katz, W. F., & Skowronski, A. (1996). Children's processing of prosodic cues for phrasal interpretation. *Journal of the Acoustical Society of America, 99*(2), 1148–1160.

Beckman, M. E., Hirschberg, J., & Shattuck-Hufnagel, S. (2004). The original ToBI system and the evolution of the ToBI framework. In Sun-Ah Jun (Ed.), *Prosodic typology—The phonology of intonation and phrasing*. Oxford: Oxford University Press.

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

Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology, 18*, 355–387.

Brennan, S. E. (1990). Seeking and providing evidence for mutual understanding. Unpublished doctoral dissertation, Stanford University.

Brennan, S.E. (in press). How conversation is shaped by visual and spoken evidence. In J. Trueswell & M. Tanenhaus (Eds.), World Situated Language Use: Psycholinguistic, Linguistic and Computational Perspectives on Bridging the Product and Action Traditions. Cambridge, MA: MIT Press.

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

Brennan, S. E., & Schober, M. (2001). How listeners compensate for disfluencies in spontaneous speech. *Journal of Memory and Langugage, 44*(2), 274–296.

Brennan, S. E., & Williams, M. (1995). The feeling of another's knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *Journal of Memory and Language, 34*, 383–398.

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

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

Clark, H. H., & Murphy, G. L. (1982). Audience design in meaning and reference. In J. F. Le Ny & W. Kintsch (Eds.), *Language and comprehension* (pp. 287–299). Amsterdam: North-Holland.

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

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

Cooper, W. E., & Danly, M. (1981). Segmental and temporal aspects of utterance-final lengthening. *Phonetica, 38*(1–3), 106–115.

Cutler, A., Dahan, D., & van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language & Speech, 40*(2), 141–201.

Dell, G. S., & Brown, P. M. (1991). Mechanisms for listener-adaptation in language production: Limiting the role of the "model of the listener". In D. J. Napoli & J. A. Kegl (Eds.), *Bridges between psychology and linguistics: A Swarthmore Festschrift for Lila Gleitman* (pp. 105–129). Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Ferreira, F. (2002). Prosody. *Encyclopedia of cognitive science*. London, UK: Macmillan Reference Ltd.

Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science, 11*(1), 11–15.

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

Fowler, C., & Housum, J. (1987). Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language, 37*, 24–40.

Fox Tree, J. E., & Meijer, P. J. A. (2000). Untrained speakers' use of prosody in syntactic disambiguation and listeners' interpretations. *Psychological Research, 63*, 1–13.

Frick, R. W. (1995). Accepting the null hypothesis. *Memory & Cognition, 23*, 132–138.

Fussell, S. R., & Krauss, R. M. (1992). Coordination of knowledge in communication: Effects of speakers' assumptions about what others know. *Journal of personality and Social Psychology, 62*(3), 378–391.

Gregory, Healy, & Jurafsky, (2001). The role of the hearer in durational shortening. In *14th annual CUNY conference on human sentence processing*, University of Pennsylvania, Philadelphia, PA.

Gregory, Healy, & Jurafsky, (2003). Common ground in production: Effects of mutual knowledge on word duration. Unpublished manuscript.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.). *Syntax and semantics, Vol.3: Speech acts* (pp. 225–242). New York: Seminar Press.

Hanna, J. E., & Tanenhaus, M. K. (2004). Pragmatic effects on reference resolution in a collaborative task: Evidence from eye movements. *Cognitive Science, 28*, 105–115.

Hirose, K. (1997). Disambiguating recognition results by prosodic features. In Y. Sagisaka, N. Campbell, & N. Higuchi (Eds.), *Computing prosody: Computational models for processing spontaneous speech* (pp. 327–342). Berlin: Springer.

Horton, W. S., & Gerrig, R. J. (in press). Conversational common ground and memory processes in language production. *Discourse Processes*.

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

Keysar, B. (1997). Unconfounding common ground. *Discourse Processes, 24*, 253–270.

Keysar, B., & Henly, A. S. (2002). Speaker' overestimation of their effectiveness. *Psychological Science, 13*, 207–212.

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

Klatt, D. H. (1975). Vowel lengthening is syntactically determined in connected discourse. *Journal of Phonetics, 3*, 129–140.

Krauss, R. M. (1987). The role of the listener: Addressee influences on message formulation. *Journal of Language and Social Psychology, 6*, 81–98.

Lehiste, I. (1972). The timing of utterances and linguistic boundaries. *Journal of the Acoustical Society of America, 51*, 2018–2024.

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

Lehiste, I., Olive, J. P., & Streeter, L. A. (1976). Role of duration in disambiguating syntactically ambiguous sentences. *Journal of the Acoustical Society of America, 60*, 1199–1202.

Lieberman, P. (1963). Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech, 6*, 172–187.

Lockridge, C., & Brennan, S. E. (2002). Addressees' needs influence speakers' early syntactic choices. *Psychonomic Bulletin & Review, 9*(3), 550–557.

Marcus, M., & Hindle, D. (1990). Description theory and intonation boundaries. In G. Altmann (Ed.), *Cognitive models of speech processing: Computational and psycholinguistic perspectives*. Cambridge, MA: MIT Press.

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

Metzing, C., & Brennan, S. E. (2003). When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions. *Journal of Memory and Language, 49*, 201–213.

Nagel, N. H., Shapiro, L. P., Tuller, B., & Nawy, R. (1996). Prosodic influences on the resolution of temporary ambiguity during on-line sentence processing. *Journal of Psycholinguistic Research, 25*, 319–344.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27*(2).

Price, P. J., Ostendorf, M., Shattuck-Hufnagel, S., & Fong, C. (1991). The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America, 90*(6), 2956–2970.

Pynte, J., & Prieur, B. (1996). Prosodic breaks and attachment decisions in sentence parsing. *Language and Cognitive Processes, 11*, 165–192.

Schafer, A. J., Speer, S., Warren, P., & White, S. D. (2000). Intonational disambiguation in sentence production and comprehension. *Journal of Psycholinguistic Research, 29*(2), 169–182.

Samuel, S. G., & Troicki, M. (1998). Articulation quality is inversely related to redundancy when children or adults have verbal control. *Journal of Memory and Language, 39*, 175–194.

Schober, M. F., & Brennan, S. E. (2003). Processes of interactive spoken discourse: The role of the partner. In A. C. Graesser, M. A. Gernsbacher, & S. R. Goldman (Eds.), *Handbook of discourse processes*. Hillsdale, NJ: Lawrence Erlbaum.

Scott, D. R. (1982). Duration as a cue to the perception of a phrase boundary. *Journal of the Acoustical Society of America, 71*(4), 996–1007.

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

Shattuck-Hufnagel, S., & Turk, A. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research, 25*(2), 193–247.

Shriberg, E., & Stolcke, A. (2004). Prosody modeling for automatic speech recognition and understanding. In M. Johnson, S. Khudanpur, M. Ostendorf, & R. Rosenfeld (Eds.), *Mathematical foundations of speech and language processing. IMA volumes in Mathematics and its applications* (Vol. 138, pp. 105–114). New York: Springer.

Snedeker, J., & Trueswell, J. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language, 48*, 103–130.

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

Streeter, L. A. (1978). Acoustic determinants of phrase boundary perception. *Journal of the Acoustical Society of America, 64*(6), 1582–1592.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*, 1632–1634.

Wilkes-Gibbs, D., & Clark, H. H. (1992). Coordinating beliefs in conversation. *Journal of Memory and Langugage, 31*, 183–194.