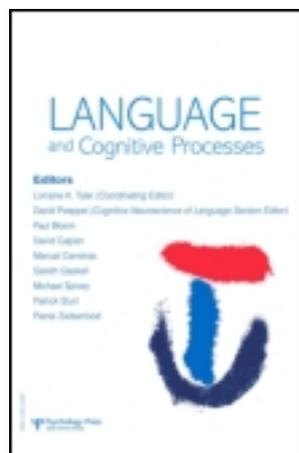


This article was downloaded by: [Harvard College]

On: 04 April 2012, At: 09:01

Publisher: Psychology Press

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Language and Cognitive Processes

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/plcp20>

Is it all relative? Effects of prosodic boundaries on the comprehension and production of attachment ambiguities

Jesse Snedeker^a & Elizabeth Casserly^{a b}

^a Department of Psychology, Harvard University, Cambridge, MA, USA

^b Department of Linguistics, Indiana University, Bloomington, IN, USA

Available online: 20 Jul 2010

To cite this article: Jesse Snedeker & Elizabeth Casserly (2010): Is it all relative? Effects of prosodic boundaries on the comprehension and production of attachment ambiguities, *Language and Cognitive Processes*, 25:7-9, 1234-1264

To link to this article: <http://dx.doi.org/10.1080/01690960903525499>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.tandfonline.com/page/terms-and-conditions>

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Is it all relative? Effects of prosodic boundaries on the comprehension and production of attachment ambiguities

Jesse Snedeker¹ and Elizabeth Casserly^{1,2}

¹*Department of Psychology, Harvard University, Cambridge, MA, USA,*

²*Department of Linguistics, Indiana University, Bloomington, IN, USA*

While there is ample evidence that prosody and syntax mutually constrain each other, there is considerable uncertainty about the nature of this interface. Here, we explore this issue with prepositional phrase attachment ambiguities (*You can feel_A the cat_B with the feather*). Prior research has been motivated by two hypotheses: (1) the absolute boundary hypothesis (ABH) posits that attachment preferences depend on the size of the prosodic boundary before the ambiguous phrase (boundary B) and (2) the relative boundary hypothesis (RBH) links attachment to the relative size of boundary B and any boundary between the high and low attachment site (boundary A). However, few experiments test the unique predictions of either theory. Study 1 examines how syntax influences prosodic production. The results provide modest support for RBH and stronger support for ABH. In Study 2, we systematically vary the size of both boundaries in an offline comprehension task. We find that absolute boundary strength influences interpretation when relative boundary strength is held constant, and relative boundary strength influences interpretation when absolute boundary strength is held constant. Thus, our theory of the prosody–syntax interface must account for effects of both kinds.

Keywords: Parsing; Prosody; Syntactic ambiguity resolution; Intonational boundaries; Comprehension; Prepositional phrase attachment; Production.

Correspondence should be addressed to Jesse Snedeker, Department of Psychology, Harvard University, 1136 WJH, 33 Kirkland St., Cambridge MA, 02138, USA. E-mail: snedeker@wjh.harvard.edu

We are grateful to Jennifer Venditti, who conducted the ToBI analyses on the utterances from Snedeker and Trueswell (2003), and to Amanda Worek, who kept the lab running smoothly. A special thanks to Stefanie Shattuck-Hufnagel for introducing the authors. This material is based upon work supported by the National Science Foundation (0623845).

© 2010 Psychology Press, an imprint of the Taylor & Francis Group, an Informa business

<http://www.psypress.com/lcp>

DOI: 10.1080/01690960903525499

Human languages employ several levels of representation: the syllables in an utterance are organised into a prosodic structure, the words into a syntactic structure, and the concepts into a semantic structure. One of the central questions of linguistics is how these partially independent structures are linked together and constrain one another. One of the central questions of psycholinguistics is how information from one level can be used during language comprehension to draw inferences about another. Some of these interfaces are both systematic and well documented, supporting agreement about the content of the interface even when the mechanisms are the subject of dispute. For example, any theory of the syntax–semantics interface must capture the relations between syntactic positions and thematic roles.

In contrast, the study of the prosody–syntax interface is relatively young. Nevertheless, substantial progress has been made. In the past 20 years, we have clearly established that there is a systematic, albeit imperfect, relation between prosodic structure and syntactic form. Manipulations of prosodic structure influence how listeners interpret syntactically ambiguous utterances (Beach, 1991; Carlson, Clifton, & Frazier, 2001; Cooper & Paccia-Cooper, 1980; Lehiste, 1973; Lehiste, Olive, & Streeter, 1976; Price, Ostendorf, Shattuck-Hufnagel, & Fong, 1991; Schafer, 1997; see Cutler, Dahan, & van Donselaar, 1997, for a review). These effects of prosody emerge quickly during online sentence comprehension, suggesting that this is a robust property of the human parser (Kjelgaard & Speer, 1999; Marslen-Wilson, Tyler, Warren, Grenier, & Lee, 1992; Nagel, Shapiro, Tuller, & Naway, 1996; Pynte & Prieur, 1996; Snedeker & Trueswell, 2003; Steinhauer, Alter, & Friederici, 1999; Warren, Grabe & Nolan, 1995; Weber, Grice, & Crocker, 2006). Naïve speakers systematically vary their prosody depending on the syntactic structure of the utterance and naïve listeners can use this variation to disambiguate the utterance (Kraljic & Brennan, 2005; Schafer, Speer, & Warren, 2005; Snedeker & Trueswell, 2003). Under some circumstances, these prosodic cues disappear when the speaker is unaware of the ambiguity or the context disambiguates the utterance (Allbritton, McKoon, & Ratcliff, 1996; Snedeker & Trueswell, 2003). But in other cases these cues persist, suggesting that they are available to listeners at least some of the time (Kraljic & Brennan, 2005; Schafer et al., 2005). Taken together, these studies indicate that users of a language share implicit knowledge about the relationship between prosody and syntax which they can use during comprehension and production.

However, the precise characterisation of the syntax–prosody mapping is far from clear. Several phenomena suggest that this mapping is unlikely to be simple or deterministic. First, there are major syntactic boundaries which are rarely marked by prosodic boundaries (e.g., the boundary between the subject and the predicate). Second, nonsyntactic factors—such as speech rate, word length, and discourse structure—play a critical role in prosodic

structure (for reviews see Beckman, 1996; Cutler et al., 1997; Fernald & McRoberts, 1996; Shattuck-Hufnagel & Turk, 1996; Warren, 1999). Finally, experimental production studies have demonstrated that the same word string, with the same intended structure, in essentially the same discourse context, can be produced with many different prosodic structures (Schafer et al., 2005; Snedeker & Trueswell, 2003).

Theorists have approached the syntax–prosody interface in two ways. Some take on the whole problem, providing an algorithm for converting any syntactic structure into a prosodic structure (Cooper & Paccia-Cooper, 1980; Ferreira, 1988; Gee & Grosjean, 1983; Watson & Gibson, 2004). Other theories focus on a more limited range of phenomena and characterise the mapping within this domain. One commonly explored phenomenon is the prosodic phrasing of syntactic attachment ambiguities (Carlson et al., 2001; Clifton, Carlson, & Frazier, 2002; Schafer, 1997). There are several reasons why this might be a particularly productive place to begin pinning down the syntax–prosody interface.

First, attachment ambiguities are a diverse but clearly defined set of phenomena. The ambiguously attached constituent can vary in its syntactic category, length, and syntactic complexity, allowing experimenters to examine influence of multiple variables. However, all attachment ambiguities have some features in common: in all cases there is an ambiguous phrase which could be linked to the syntactic tree in more than one location. In English, one of the options typically involves incorporating the phrase into the constituent that immediately preceded it, and one of the options involves attaching the constituent at a higher level in the tree. For example, the sentence in (1) has two alternate interpretations. The prepositional phrase (“with flawed data”) can be a constituent of the noun phrase as in (2), in which case Amanda is the heroine. Or it can be attached directly to the verb phrase as in (3), making Amanda the villain.

1. Amanda attacked the paper with flawed data.
2. Low attachment: Amanda [attacked [the paper [with flawed data]_{PP}]_{NP}]_{VP}.
3. High attachment: Amanda [attacked [the paper]_{NP} [with flawed data]_{PP}]_{VP}.

Second, as this example suggests, attachment ambiguities can often remain globally ambiguous. This allows experimenters to examine the influence of prosody on syntax (or syntax on prosody) while holding the string of words constant. One might also expect that the strength of prosody–syntax correspondences would be greatest for globally ambiguous utterances. From a functional perspective, prosodic cues to structure would be most

helpful when other information is absent. There is no direct evidence that global ambiguity is marked more clearly than local ambiguity. However, there is evidence that under some circumstances speakers who are aware of a global ambiguity produce clearer prosodic cues than those who are not aware of it (Kraljic & Brennan, 2005; Lehiste, 1973, but see Schafer et al., 2005; Snedeker & Trueswell, 2003). Finally, attachment ambiguities are ubiquitous in everyday conversation, ensuring that we are studying constructions that speakers and listeners have ample experience with, thus maximising our chances of finding systematic, replicable patterns of performance.

To date, there have been two primary ways of thinking about the relation between prosodic structure and the interpretation of attachment ambiguities. Many theorists have focused on the boundary immediately before the ambiguously attached phrase (marked as B in 4), noting that the absence of a boundary in this location favours low attachment, while the presence of a boundary favours high attachment (see e.g., Marcus & Hindle, 1990; Price et al., 1991; Pynte & Prieur, 1996; Watson & Gibson, 2005).

4. Amanda attacked_A the paper_B with flawed data.

On the basis of this data one might conclude that the absolute strength of boundary B was the primary consideration in mapping between prosody and syntax.

More recently, however, other theorists have suggested that the boundary before the ambiguous phrase can only be interpreted in light of the global prosodic structure of the utterance (Carlson et al., 2001; Clifton et al., 2002; Schafer, 1997). This claim is made most explicit in the work of Carlson, Clifton, and Frazier who argue that the boundary at location B is always interpreted with respect to any other boundary that occurs before a constituent that contains the lower attachment site but not the higher attachment site (Clifton et al., 2002). In the utterance above, the only boundary that would be relevant is the one marked A. Specifically, Carlson and colleagues argue that effects of prosody depend on the relative size of these two boundaries: prosodic structures in which A is bigger than B favour low attachment, those in which B is bigger than A favour high attachment, and those in which the two boundaries are equivalent favour neither.

As Carlson and colleagues note, most studies demonstrating that the lower boundary influences attachment are compatible with this relative boundary strength hypothesis. In comprehension studies when a boundary is added in location B, it is typically larger than any relevant boundary A and thus changes the interpretation of the utterance on both hypotheses. In a

series of comprehension studies, Carlson and colleagues demonstrate that relative boundary size has an effect on interpretation (Carlson et al., 2001; Clifton et al., 2002). But research to date has not explored whether the absolute size of the final boundary plays a role independent of relative boundary size, and the evidence that the relative boundary size plays a role *independent of absolute boundary size* is limited, coming from a single lab and paradigm (albeit from several constructions). To explore these questions, we introduce two hypotheses which clearly isolate these factors (5 and 6). The absolute boundary hypothesis (ABH) is related to Watson and Gibson's (2005) *anti-attachment hypothesis*, though it is both stronger and more limited in scope. The relative boundary hypothesis (RBH) captures the relevant features of Carlson, Clifton and Frazier's *informative boundary hypothesis* (Clifton et al., 2002).

5. ABH: The absolute size of the prosodic boundary immediately before a constituent (boundary B) predicts syntactic attachment independent of relative boundary size. Larger boundaries are associated with high attachment, smaller boundaries with low attachment.
6. RBH: The relative magnitude of the prosodic boundary immediately before a constituent (boundary B) and any higher relevant boundary (boundary A) predicts syntactic attachment, independent of the absolute size of boundary B. When B is larger than A, high attachment is favoured, when the two are equal there is no preference and when A is larger than B, low attachment is favoured.

These hypotheses are phrased from the perspective of the comprehender, but they could be reframed to make predictions about production as well (in which case prosodic structure would reflect syntactic attachment rather than predicting it).

Notice that both of these hypotheses rely on the notion of boundary strength. Thus to make predictions about the interpretation of particular structures, we need a theory of prosodic boundaries. Like most researchers in this field, we will be describing prosody according to the tones and break indices (ToBI) coding system (Beckman & Hirschberg, 1994), which represents the relative prominence of words in an utterance and their prosodic grouping. According to the prosodic theory underlying ToBI, there are two levels of prosodic structure between the level of the utterance and the prosodic word, the intermediate phrase and the intonational phrase (Beckman & Pierrehumbert, 1986). Each intermediate phrase (or ip) contains at least one pitch accent and ends in a high or low phrase tone. Intermediate phrases are grouped together into intonational phrases (or IP's). An intonational phrase contains at least one intermediate phrase and

ends in a high or low boundary tone (which follows the phrase tone of the final intermediate phrase). While prosodic theories vary in the number of hierarchical levels that they recognise, most include levels that roughly correspond to the intermediate and intonational phrase (see e.g., Nespor & Vogel, 1986; Selkirk, 1986). Most researchers exploring the prosody–syntax interface have explicitly argued or implicitly assumed that these two types of boundaries are discrete and categorical. Syntactic structure can influence how an utterance is divided into ip’s or IP’s, but all boundaries of a given kind are equivalent and thus continuous variation within a category plays no role in syntax–prosody interface. To the best of our knowledge, the experimental evidence for this comes from a single study demonstrating that substantially increasing the salience of an ip boundary does not modify the interpretation of an attachment ambiguity (Experiment 3; Carlson et al., 2001).

By adopting this theory of prosody, we can develop the following predictions about the relation between prosodic structure and syntactic attachment (7 and 8).

7. Predictions about the probability of high attachment under the RBH:

$$\begin{array}{ccc} (\text{ip}, 0) & (0, 0) & (0, \text{ip}) \\ (\text{IP}, 0) & < (\text{ip}, \text{ip}) & < (0, \text{IP}) \\ (\text{IP}, \text{ip}) & (\text{IP}, \text{IP}) & (\text{ip}, \text{IP}) \end{array}$$

8. Predictions about the probability of high attachment under the ABH:

$$\begin{array}{ccc} (0, 0) & (0, \text{ip}) & (0, \text{IP}) \\ (\text{ip}, 0) & < (\text{ip}, \text{ip}) & < (\text{ip}, \text{IP}) \\ (\text{IP}, 0) & (\text{IP}, \text{ip}) & (\text{IP}, \text{IP}) \end{array}$$

Throughout this paper, we will be adopting the convention of describing prosodic structures as ordered pairs in which the first item refers to the boundary at location A and the second to the boundary at location B. Word level breaks are coded as 0, intermediate phrase breaks as ip, and intonational phrase breaks as IP. Notice that the two hypotheses make many of the same predictions. For example on both theories (0, 0) structures result in fewer high attachments than (0, IP) structures. In this paper, we will be focusing our attention on the cases in which one theory predicts that two structures will be equivalent while the other theory predicts a difference. These are the unique predictions of the RBH and ABH, which are listed below (9 and 10).

9. Unique predictions of the relative boundary strength hypothesis:
- 1a. For utterances with no low boundary: (IP, 0) & (ip, 0) < (0, 0)
 - 1b. For utterances with a low ip boundary: (IP, ip) < (ip, ip) < (0, ip)
 - 1c. For utterances with a low IP boundary: (IP, IP) < (0, IP) & (ip, IP)
10. Unique predictions of the absolute boundary strength hypothesis:
- 2a. For utterances with equal boundaries: (0, 0) < (ip, ip) < (IP, IP)
 - 2b. For utterances with a larger high boundary: (IP, 0) & (ip, 0) < (IP, ip)
 - 2c. For utterances with a larger low boundary: (0, ip) < (0, IP) & (ip, IP)

Many of these predictions have never been tested or have been tested only indirectly. To examine whether both factors are characteristics of the prosody–syntax interface, we examined these contrasts both in production and comprehension. Study 1 makes use of data from a referential communication task to explore whether speakers use both approaches to disambiguate syntactic attachments. In Study 2, we parametrically varied the strength of the two boundaries in a comprehension task to examine how relative and absolute boundary strength influenced the ultimate interpretation of globally ambiguous utterances.

STUDY 1: RE-ANALYSIS OF PRODUCTION DATA FROM SNEDEKER AND TRUESWELL

Method and prior findings

To begin exploring this question, we re-examined the production data from Snedeker and Trueswell (2003). This study used a referential communication paradigm to examine the prosody–syntax interface in naïve participants. The speaker and the listener were separated by a screen and each given a set of toys, which they believed to be identical. The speaker was then shown a target action using these toys and given a written sentence to produce. The written sentence was removed and the speaker produced the command. The listener followed the command to the best of his/her ability but was not allowed to ask for clarification.

The critical utterances contained ambiguous prepositional phrase attachments (“Tap the frog with the flower”). In Experiment 1, both participants had a toy set which supported both the low and high attachment (e.g., a frog, a frog carrying a flower, a large flower, a block, and a giraffe wearing a coat) and the intended interpretation of the utterance was manipulated within participants. This was done by varying the demonstration that the speaker saw. On high-attachment/instrument trials, the experimenter used the target instrument (the flower) to carry out the action on the unmodified animal (the frog). On low-attachment/modifier

trials, she used her hand to carry out the action on the modified animal (the frog carrying a flower). Under these circumstances, listeners were able to use the prosody of the speaker's utterance to arrive at the correct interpretation about 70% of the time. However, the speaker was typically aware of the ambiguity.

In Experiment 2, two changes were made to decrease ambiguity awareness. First, the intended interpretation of the ambiguous utterance was manipulated between speakers. Second, each speaker was given a toy set that only supported the intended interpretation of the ambiguous utterance (e.g., for a low attachment, the flower above would be replaced with a leaf). The listener's toy set remained ambiguous. Under these circumstances, most speakers were unaware of the ambiguity and the listeners were unable to disambiguate the utterance and thus performed at chance.

In analysing the prosodic form of the speakers' utterances, Snedeker and Trueswell implicitly adopted the RBH. Utterances were coded using the ToBI labelling system by a highly-trained coder who was blind to experimental condition, then they were classified according to whether the lower boundary was greater than, lesser than, or equal in size to the higher boundary (see also Schafer, Speer, Warren, & White, 2000). The results were consistent with the RBH. For example, in the experiment with aware speakers, 81% of the utterances in which B was greater than A were intended to have a high attachment, while only 6% of the utterances in which A was greater than B were intended to. However, as we noted earlier, results like this do not uniquely support the RBH. Since most of the utterances had only one prosodic boundary, differences in relative boundary strength were typically accompanied by differences in the absolute strength of boundary B. In fact, when we classify the utterances according to absolute boundary strength, the results are equally strong. In aware speakers, 91% of utterances with an IP break in position B were intended to have a high attachment, while only 9% of the utterances with no break in position B were intended to.

Present analyses and discussion

To explore the unique predictions of the relative and absolute boundary strength hypotheses, we went back to the Snedeker and Trueswell data set to determine how often each prosodic form was used with the intention of communicating low or high attachment. All utterances with ambiguous boundary indices (ToBI codes of 2) were eliminated from this analysis.¹ The frequency of each structure and the proportion which were intended to have a high attachment are listed in Table 1. The first thing to note is that the

¹ These comprised 16% of the tokens in Experiment 1 in which speakers were aware of the ambiguity and 4% of the tokens in Experiment 2 in which they were unaware. All the ambiguous boundaries occurred before the prepositional phrase. In Experiment 1, the proportion of these utterances appearing in high-attachment contexts was intermediate between utterances with a 0 boundary in this location and those with an ip boundary there, suggesting that they were truly ambiguous. In Experiment 2, they were too infrequent to characterise.

TABLE 1
 Number of tokens for each prosodic form and the proportion that were produced in the high attachment context

<i>Prosodic form</i>	<i>Production aware speakers</i>		<i>Production unaware speakers</i>	
	<i>Number of tokens</i>	<i>Portion with high attachment</i>	<i>Number of tokens</i>	<i>Portion with high attachment</i>
0, 0	33	0.21	25	0.48
ip, 0	9	0.00	16	0.44
IP, 0	56	0.04	17	0.12
0, ip	20	0.54	50	0.60
ip, ip	6	0.83	42	0.45
IP, ip	3	0.67	11	0.36
0, IP	62	0.90	29	0.62
ip, IP	24	0.92	12	0.92
IP, IP	13	0.92	9	0.78

nine structures are not equally common. Aware speakers tended to produce utterances with a single IP break or with no internal prosodic boundaries. Unaware speakers were more likely to produce the (ip, ip) or (0, ip) structures. Second, the structures appear to be used in systematically different ways, particularly by aware speakers. Those with no break before the prepositional phrase (top three rows) were typically used for low attachments, while those with an IP break before the ambiguous phrase (bottom three rows) were generally used for high attachments.

To test the unique predictions of each hypothesis, we conducted a series of Fisher's exact tests comparing pairs of forms to determine whether one would be used more often to signal high attachment. The results of these analyses appear in Table 2. The rows of the table evaluate the unique predictions of each hypothesis which were described above (see 9 and 10). Several analyses involved structures that appeared less than 20 times in the data set. These are shaded grey. Obviously no strong conclusions can be drawn from null effects in these cells. Only two predictions of the RBH involve structures that are frequent enough to support a robust analysis. One of these predictions is confirmed: speakers are less likely to use the (IP, 0) structure to signal high attachment than the (0, 0) structure. This is true even for the unaware speakers who make less use of the (IP, 0) structure. The second prediction is that for unaware speakers the (0, ip) structure will be more associated with high attachments than the (ip, ip) structure. This contrast includes a total of 92 data points but fails to reach significance.

The tests of the absolute boundary strength hypothesis were more informative. In aware speakers, six of the seven critical contrasts were reliable, many despite small sample sizes. Structures with no boundary

TABLE 2
Tests of RBH and ABH on production data from aware and unaware speakers

Prediction	Productions of aware speakers	Productions of unaware speakers
Relative boundary strength		
1a. (IP, 0) < (0, 0)	$p < .05^*$	$p < .05^*$
1a. (ip, 0) < (0, 0)	$p > .2$	$p > .2$
1b. (IP, ip) < (ip, ip)	$p > .2$	$p > .2$
1b. (ip, ip) < (0, ip)	$p > .2$	$p > .2$
1b. (IP, ip) < (0, ip)	$p > .2$	$p > .2$
1c. (IP, IP) < (0, IP)	$p > .2$	$p > .2$
1c. (IP, IP) < (ip, IP)	$p > .2$	$p > .2$
Absolute boundary strength		
2a. (0, 0) < (ip, ip)	$p < .01^*$	$p > .2$
2a. (ip, ip) < (IP, IP)	$p > .2$	$p = .08$
2a. (0, 0) < (IP, IP)	$p < .005^{**}$	$p > .1$
2b. (IP, 0) < (IP, ip)	$p < .05^*$	$p > .1$
2b. (ip, 0) < (IP, ip)	$p < .05^*$	$p > .2$
2c. (0, ip) < (0, IP)	$p < .005^{**}$	$p > .2$
2c. (0, ip) < (ip, IP)	$p < .05^*$	$p < .05^*$

Cells in grey include a contrast with fewer than 20 tokens. $*p < .05$, $**p < .005$.

before the prepositional phrase were typically used to convey low attachments, those with IP boundaries in this location were used for high attachments, and those with ip boundaries there were used for both. In unaware speakers, most of these contrasts were unreliable, suggesting that the mapping between syntax and boundary size is less robust when speakers are not deliberately marking the syntactic contrast. However, the contrast between (0, ip) structures and (ip, IP) structures continued to be reliable.

In sum, we find modest support for the relative boundary strength hypothesis and more robust support for the absolute boundary strength hypothesis. The interpretation of these findings is limited by the small numbers of utterances in many cells. Nevertheless, these analyses constrain our model of the interface. By demonstrating that at least one unique prediction of each theory is supported—both in aware and unaware speakers—these data indicate that both relative and absolute boundary strength play a role in the mapping from syntax to prosody.

In Study 2, we continued exploring the predictions of the absolute and relative boundary strength hypotheses. However, we switched our focus from language production to language comprehension. We did this for two reasons. First, as we noted in the introduction, RBH was primarily developed in the context of language comprehension. While aspects of the proposal are motivated by assumptions about the prosodic cues that

speakers produce, it is possible that the comprehension system and production system diverge in some respects. Second, our analysis of Study 1 was limited by differences in how often each structure was used. By switching to comprehension, we can ensure that all of the relevant structures are used frequently, thus increasing the power of the study.

STUDY 2: COMPREHENSION OF ATTACHMENT AMBIGUITIES

Methods

Participants

A total of 62 adult native speakers of American English drawn from the Harvard University student body and the greater Cambridge community participated in our experiment. None reported uncorrected hearing or vision problems and all were compensated for their participation with either psychology course credit or five dollars. Two subjects' results were removed from analysis due to equipment malfunction ($n = 1$) or failure to complete the experiment within the allotted time ($n = 1$).

Stimuli

Our critical stimuli consisted of eight base sentences, each containing a prepositional-phrase attachment ambiguity like those in (11) below.² These sentences were based on stimuli used in Snedeker and Trueswell (2004), which had been designed to provide equal support to the modifier and instrument interpretation, which map on to low and high syntactic attachments, respectively.³ The eight verbs had been selected on the basis of a sentence completion study and had given rise to roughly equal numbers of instrument and modifier completions. In each case the prepositional object that was paired with the verb had been rated as a moderately plausible instrument for that particular action.

Each utterance was recorded with both a one syllable noun and a three syllable compound noun (11). The three syllable nouns consisted of redundant

² We used only eight items for two reasons. First, Snedeker and Trueswell (2004) normed potential instruments for only eight equi-biased verbs and we wished to make use of these norms. Second, the present study served as an offline validation study for a visual-world experiment which employed an act out task to test young children, thus placing additional constraints on the verbs that could be used. Each participant heard the same base sentence in four different prosodies, thus the number of critical items was 32 per participant.

³ While the only way to semantically interpret the low attachment is as a modifier, the high attachment of a phrase can have several semantic interpretations (accompaniment, location, instrument, etc.). For these particular verbs and prepositional objects, the instrument interpretation was dominant, as evidenced by participants' responses in an act out task (Snedeker & Trueswell, 2004).

forms of the one syllable nouns, which are common in child-directed speech (e.g., *kitty cat*). The goal of this manipulation was to explore the possibility that length of the direct-object noun phrase might play some role in the interpretation of prosodic structure.⁴ Because the length manipulation had no reliable effect (and did not interact with the other variables), we will not be discussing it further. A complete list of the sentences can be found in the Appendix.

11. a. You can pinch the dog with the barrette.
- b. You can pinch the puppy dog with the barrette.

Each of these critical sentences was recorded in each of the nine prosodic forms listed in Table 1. The second author, a student of phonetics trained in prosodic analysis, produced all the stimuli. The sentences were then transcribed by a naïve native-English-speaking laboratory assistant to check for naturalness and intelligibility. All utterances were fully intelligible, but sentences spoken with two utterance-internal IP breaks were judged to be highly unnatural and were therefore dropped from the experiment. This left us with 128 critical utterances. The ToBI transcriptions for these utterances appear in Table 3.

The speaker had been instructed to produce each utterance in the most natural manner possible and to produce a consistent prosody across items in the same condition. Consequently, the placement and type of pitch accent was free to vary across conditions. All the target sentences had pitch accents on the direct-object noun and the prepositional object. All had pitch accents on the verb except for a subset of the (0, ip) utterances which had an accent on the sentence subject instead. However, the type of pitch accent on each word varied systematically across conditions, an issue that we return to in the discussion.

To ensure that every utterance was produced with the intended prosody, we measured the duration of each word and the pauses that followed them. We would expect that a break in location A would increase the duration of the verb, while a break in location B would increase the duration of the direct

⁴ There are two opposing predictions that might be made about the effects of word length on the interpretation of these utterances. First, on a theory in which listeners evaluate prosodic boundaries according to their informativeness (e.g., Clifton et al., 2002), the use of a longer direct-object noun should make the boundary after it less informative. This would result in more low attachments in the conditions with ip or IP breaks before the prepositional phrase. Second, if we entertain the hypothesis that attachment decisions depend in part on the activation of constituents and that the activation rapidly dissipates (Altmann, 1998), then we might expect that redundant material between the onset of the noun and the ambiguous phrase to decrease the number of low attachments. Thus, the lack of any effect could reflect the opposing effects of these two processes. Or it could suggest that the length manipulation was too weak to exert any effect at all. Across the eight conditions, the long nouns were only 166 ms longer than the short nouns.

TABLE 3
ToBI transcriptions of the stimuli for Study 2

	<i>You</i>	<i>can</i>	<i>verb</i>	<i>the</i>	<i>DO noun</i>	<i>with the</i>	<i>pp-object noun</i>
IP, 0			L+H*	!H-L%	L*		H* L-L%
0, IP			H*		!H*	L-H%	H* L-L%
0, 0			H*		!H*		!H* L-L%
ip, ip			L*	H-	L*	H-	L+H* L-L%
ip, 0			H*	L-	H*		!H* L-L%
0, ip	H* (50%)		L* (50%)		L+H*	L-	L+H* L-L%
IP, ip			H*	!H-L%	L*	H-	L+H* L-L%
ip, IP			L*	H-	L*	L-H%	H* L-L%

object noun. The IP breaks were all accompanied by audible pauses of about 100–300 ms. In contrast most of the ip breaks had no audible pause but many contained short silences (30–100 ms) which were visible when the waveform was visually inspected. Because few of our critical words ended in stop consonants, it was difficult to determine precisely when each word ended and the pause began. For this reason, we calculated the total duration of the critical word and the pause that followed it. These values are given in Table 4. The duration analyses confirm that the IP breaks were produced with more lengthening than the ip breaks which in turn showed more lengthening than the null breaks.

EXPERIMENTAL DESIGN

To avoid fatiguing our participants or eroding their intuitions by requesting many judgments on the same word string, we divided the prosody types into three between-participant conditions (see Table 5) with four prosodic forms appearing in each condition. These conditions were constructed from the perspective of the RBH. We will refer to the between-participant manipulation as Prosody Strength, since the strength of the boundaries was varied across critical cells in these conditions while their relative size was held constant. The within-participant manipulation was termed Prosody Type, because it varied relative boundary strength, and thus, from the perspective of the RBH, the different Prosody Types diverge in the kind of interpretation that they support.

TABLE 4
Durations of critical regions for stimuli for Study 2

<i>Prosodic form</i>	<i>Verb +pause</i>	<i>DO noun +pause</i>
IP, 0	823 ms CI.95 = ± 46	455 ms CI.95 = ± 70
0, IP	327 ms CI.95 = ± 43	856 ms CI.95 = ± 36
0, 0	399 ms CI.95 = ± 25	486 ms CI.95 = ± 58
ip, ip	608 ms CI.95 = ± 49	658 ms CI.95 = ± 43
ip, 0	550 ms CI.95 = ± 38	440 ms CI.95 = ± 43
0, ip	315 ms CI.95 = ± 54	644 ms CI.95 = ± 38
IP, ip	756 ms CI.95 = ± 53	644 CI.95 = ± 54
ip, IP	637 ms CI.95 = ± 44	916 ms CI.95 = ± 51

TABLE 5
 Division of the different prosodic forms into three Prosody Strength conditions

<i>Prosody Type (within participant)</i>	<i>Prosody Strength (between participants)</i>		
	<i>Strong</i>	<i>Weak</i>	<i>Two breaks</i>
Neutral no breaks	∅, ∅	∅, ∅	∅, ∅
Neutral ip breaks	ip, ip	ip, ip	ip, ip
High attachment	∅, IP	∅, ip	ip, IP
Low attachment	IP, ∅	ip, ∅	IP, ip

Specifically, each of the Prosody Strength conditions included both the (0, 0) and (ip, ip) utterances, which were predicted to be neutral according to the RBH. In addition, each condition included one prosodic structure with a larger break in position A than in position B (predicted to promote low attachment under the RBH) and one condition with a larger break in position B than in position A (predicted to promote high attachment under the RBH). The nature of these utterances varied across the Prosody Strength conditions. In the strong prosody condition, the asymmetry was created by pitting an IP break against a null break. In the weak prosody condition an ip break was pitted against a null break. In the two-break prosody condition, an IP break was pitted against an ip break (see Table 5). Note that on the relative boundary strength hypothesis, the three Prosody Strength conditions should be equivalent: they each include two prosodies which are neutral with respect to PP-attachment, a prosody more consistent with high attachment, and a prosody more consistent with low attachment.

In addition, the number of syllables in the direct-object noun was manipulated between subjects, resulting in six lists with 32 critical utterances each. On a single list, each base utterance appeared in four different prosodies. Sixteen filler sentences were created: eight contained a relative clause/complement clause ambiguity (12) and eight contained an ambiguous pronoun (13). These filler items were recorded by the same phonetician, who produced four different prosodic forms of each sentence. For example, in sentences such as (13), stress was alternately placed on the words “hippo”, “kissed”, “horse”, or “bored” in order to create four contrasting prosodies and decrease awareness of the critical manipulation.

12. You can tell the zebra who's mean.

13. The hippo kissed the horse because he was bored.

All the fillers were included on each list. Consequently, each participant heard 32 critical stimuli and 64 filler stimuli. The stimuli were presented to participants in four blocks. Each sentence (eight critical and 16 filler)

appeared once per block in random order and each prosodic structure appeared twice. The order of presentation was reversed for half of the participants.

Thus, the experiment had four independent variables: Prosody Type (high attachment, low attachment, neutral no breaks, neutral ip breaks) was manipulated within participants and Prosody Strength (strong, weak, and two breaks), Syllable Number (1 or 3), and Order (forward, backward) were manipulated between participants.

PROCEDURE

Participants were told that they were going to hear a series of ambiguous sentences that would repeat over the course of the experiment, said with different “intonation” in each instance. They would be offered two possible interpretations of the sentence and asked to choose “which the speaker intended”. Participants were encouraged to respond based on their initial intuitions. Upon agreeing to these instructions, participants were seated at a laptop computer and fitted with noise-cancelling headphones.

The experiment was conducted with PST’s E-Prime software. Each session began with two practice trials—identical to session trials—designed to acclimate participants to the presentation mode using unambiguous sentences. In each trial, participants heard an utterance once, with “Listen carefully” displayed on the screen. Then the written form of the sentence appeared on the screen above the question “What does the speaker mean?” and two possible interpretations, A and B. For critical stimuli such as sentence (11a) (*You can pinch the dog with the barrette*), the two options would be: (A) *use the barrette to pinch the dog*, and (B) *pinch the dog that has the barrette*. With these items still visible, participants had to press a button to indicate that they were ready to listen to the sentence again, listen and then choose an interpretation by pressing A or B. The instrument and modifier interpretations were randomly assigned to either A or B for each trial across the experiment.

RESULTS

The proportion of high-attachment (instrument) responses in each condition was calculated by subject and by item and submitted to subjects and items ANOVA’s with Prosody Strength, Syllable Number, and Order as between-subjects and within-items variables and Prosody Type as a within-subject and within-items variable. Prosody Type distinguished the four prosodic structures within each list and was coded according to the predictions of the relative boundary strength hypothesis (low attachment, high attachment,

neutral no breaks, neutral ip breaks). Prosody Strength reflected how the break asymmetry between the low attachment and high attachment conditions was achieved (through a single IP break for strong prosody, a single ip break for weak prosody and an ip and IP break for the two-break condition).

As Figure 1 illustrates there was a robust effect of Prosody Type, $F(3, 144) = 31.81, p < .001$; $F(3, 21) = 34.95, p < .001$. Participants predominately gave high-attachment interpretations to utterances with high-attachment prosody ($M = 0.84$). Utterances with low-attachment prosody only received high-attachment interpretations about half the time ($M = 0.51$), while the two neutral conditions were intermediate ($M = 0.66$ and $M = 0.77$ for neutral no breaks and neutral ip breaks, respectively). However, there was also a reliable effect of Prosody Strength, $F(2, 48) = 3.59, p < .05$; $F(2, 14) = 6.31, p < .05$, and a Prosody Strength by type interaction, $F(6, 144) = 2.74, p < .05$; $F(6, 42) = 3.91, p < .005$. If the direction of relative boundary strength was the only prosodic factor influencing syntactic analysis, then we would expect no effects or interactions involving Prosodic Strength. Thus, this interaction indicates that the mapping between prosody and syntax cannot be reduced to this single factor.⁵

To explore this interaction in greater detail we compared each of the pairs of Prosodic Strength conditions in separate ANOVA's. When the strong and weak prosody conditions were compared, there was no interaction between Prosody Strength and Prosody Type, $F(3, 96) < 1, p > .5$; $F(3, 21) < 1, p > .5$. However, the critical interaction was present in the comparison between the two-break condition and both the strong and the weak prosody conditions, $F(3, 96) = 3.51, p < .05$; $F(3, 21) = 4.71, p < .05$, and $F(3, 96) = 4.37, p < .01$; $F(3, 21) = 8.08, p < .001$, respectively. As Figure 1 suggests, the difference between the low-attachment and high-attachment prosodies was greater when the sentence contained only a single prosodic break. Nevertheless, there was a reliable effect of Prosody Type in all three Prosody Strength conditions strong: $F(3, 48) = 14.30; p < .001$; $F(3, 21) = 19.43, p < .001$; weak: $F(3, 48) = 17.38; p < .001$; $F(3, 21) = 20.40, p < .001$, and two breaks: $F(3, 48) = 4.37; p < .01$; $F(3, 21) = 6.18, p < .005$.

⁵ In addition, there was a small but reliable interaction between Prosody Strength and Order, $F(2, 48) = 3.19, p = .05$; $F(2, 14) = 8.92, p = .005$. There were no other reliable effects or interactions. Critically, there was no reliable effect of number of syllables in the direct-object noun, $F(1, 48) = 1.12, p > .2$; $F(1, 7) = 4.22, p = .08$, suggesting that our length manipulation did not have a strong influence on interpretation of the ambiguous phrase.

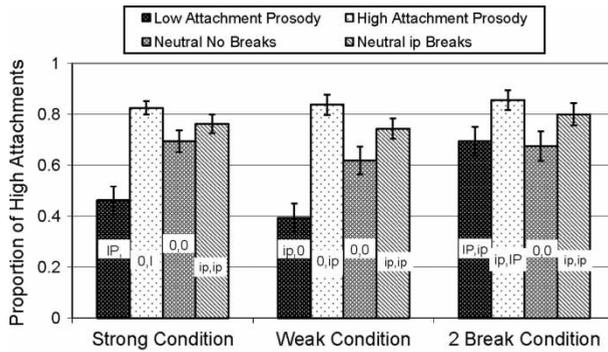


Figure 1. Proportion of high attachment judgements for Study 2.

Next, we examined the unique predictions made by the absolute and relative boundary strength hypotheses in a series of planned comparisons. The results of these tests are given in Table 6.⁶ The first thing to notice is that unique predictions of both the RBH and the ABH receive strong support from this data. In particular, the experiment clearly confirms that relative boundary strength influences interpretation when there is no intonational break before the ambiguous phrase (prediction 1a) and that absolute boundary strength influences the interpretation of utterances that would be predicted to have neutral prosody or low-attachment prosody based on relative boundary strength alone (predictions 2a and 2b). Thus, we conclude that both of these factors play some role in syntactic ambiguity resolution.

The second prediction of the relative boundary strength hypothesis receives more modest support from this experiment: when there is an ip break before the ambiguous phrase, the effect of the higher boundary was reliable in the subjects analyses but only marginal in the items analyses (prediction 1b). This suggests that the effect of differences in relative boundary strength is less robust when there is a boundary before the ambiguous phrase, than when there is no boundary in this position (prediction 1a). Such a difference could reflect a fundamental feature of the algorithm mapping prosody to syntax, or it could be driven by the

⁶ Where possible, these hypotheses were tested in within-subjects comparisons. However, for some hypotheses between-subjects comparisons were necessary (italics in Table 3). This raises the concern that differences between individual cells could reflect the mix of structures in the two lists rather than differences in the particular cells under consideration. To explore this, we conducted one-way ANOVA's to find out whether the interpretation of the neutral no break (0,0) and the neutral ip break (ip, ip) utterances varied across the three Prosody Strength conditions. We found no evidence that they did (all $F_s < 1$, all $p_s > .5$).

TABLE 6
 Tests of the RBH and ABH in comprehension judgment data, Study 2

<i>Prediction</i>	<i>Conditions^a</i>	<i>Result</i>
Relative boundary strength		
1a. (IP, 0) <(0, 0)	Strong condition: low attachment vs. neutral no breaks	<i>t1(19) = 4.23, p < .001**</i> <i>t2(7) = 4.47, p < .005**</i>
1a. (ip, 0) <(0, 0)	Weak condition: low attachment vs. neutral no breaks	<i>t1(19) = 3.67, p < .001**</i> <i>t2(7) = 4.15, p < .005**</i>
1b. (IP, ip) <(ip, ip)	Two-break condition: low attachment vs. neutral ip breaks	<i>t1(19) = 2.43, p < .05*</i> <i>t2(7) = 1.79, p = .058</i>
1b. (ip, ip) <(0, ip)	Weak condition: neutral ip breaks vs. high attachment	<i>t1(19) = 1.86, p < .05*</i> <i>t2(7) = 1.58, p = .079</i>
1b. (IP, ip) <(0, ip)	<i>Low attachment, two-break condition vs. high attachment, weak condition</i>	<i>t1(38) = 2.07, p < .05*</i> <i>t2(7) = 1.87, p = .052</i>
Absolute boundary strength		
2a. (0, 0) <(ip, ip)	All conditions: neutral no breaks vs. neutral ip breaks	<i>t1(59) = 3.29, p < .001**</i> <i>t2(7) = 2.63, p < .05*</i>
2b. (IP, 0) <(IP, ip)	<i>Low attachment, strong condition vs. low attachment, two-break condition</i>	<i>t1(38) = 2.93, p < .005**</i> <i>t2(7) = 3.43, p < .01*</i>
2b. (ip, 0) <(IP, ip)	<i>Low attachment, weak condition vs. low attachment, two-break condition</i>	<i>t1(38) = 3.78, p < .001**</i> <i>t2(7) = 6.77, p < .001**</i>
2c. (0, ip) <(0, IP)	<i>High attachment, weak condition vs. high attachment, strong condition</i>	<i>t1(38) < 1, p > .3</i> <i>t2(7) < 1, p > .3</i>
2c. (0, ip) <(ip, IP)	<i>High attachment, weak condition vs. high attachment, two-break condition</i>	<i>t1(38) < 1, p > .3</i> <i>t2(7) < 1, p > .3</i>

^aPredictions were tested with within subject comparisons wherever possible. All between subjects tests appear in italics. **p* < .05, ***p* < .005.

possibility that our measures have greater sensitivity in conditions near the midpoint, relative to conditions near the floor or ceiling.

Finally, one prediction of the absolute boundary strength hypothesis receives no support from this data: for utterances which are predicted to have high attachment under the relative boundary strength hypothesis, absolute boundary has no measurable effect on interpretation (prediction 2c). Again this could reflect a ceiling effect in the data: because participants interpret even (0, ip) utterances as high attachments 84% of the time, there is little room for improvement. Alternately, the lack of any effect here could suggest that ip and IP boundaries are functionally equivalent for the purposes of ambiguity resolution. The strongest version of this hypothesis can be ruled out by the reliable difference between the (IP, ip) and (ip, IP) utterances in the two-break condition, *t1(19) = 2.39, p < .05; t2(7) = 2.96, p < .05.*

However, this leaves open the possibility that IP and ip breaks are equivalent for the purposes of determining *absolute* boundary strength.

GENERAL DISCUSSION

In two studies, we found evidence for both the relative and absolute boundary strength hypotheses. In Study 1, we analysed production data from a referential communication task (Snedeker & Trueswell, 2003) and found that intended syntactic structure influences relative boundary size (even when absolute boundary size is held constant) and influences the absolute size of the final boundary (even when relative boundary size is constant). Even when speakers were unaware of the ambiguity, one unique prediction of each theory was confirmed. In Study 2, we switched our focus to comprehension, which allowed us to gain greater control over the frequency of each structure. The results provided robust support for unique predictions of both the RBH and ABH.

Below we address four issues: (1) we examine the role that variation in pitch accents may have played in Study 2; (2) we compare our findings to those of previous comprehension studies on the RBH; (3) we examine the parallels and divergences between the comprehension and production data; and (4) we explore whether a third hypothesis, the two absolute boundaries hypothesis, might account for these findings.

Disentangling the effects of pitch accents and prosodic phrasing

In the methods section of Study 2, we noted that the pattern of pitch accents in the critical utterances varied systematically across conditions. The speaker was instructed to produce each prosodic structure with the accent pattern that seemed most natural. Because the length (in syllables) of the prosodic phrases varied considerably across conditions, this resulted in systematic differences in the accent pattern. Thus, we must consider the possibility that the effects we observed here reflect differences in the accent pattern instead of, or in addition to, differences in prosodic phrasing. Presumably these effects would be mediated by the connection between accents and discourse functions.

While the precise mapping between accent types and discourse is controversial, most theorists claim that L + H* accents signal new information (Baumann, 2005; Pierrehumbert & Hirschberg, 1990) or discourse themes (Steedman, 2000). The H* accent is often argued to be functionally similar to the L + H* accent but less marked or salient (Baumann, 2005;

Steedman, 2000). In contrast L* accents are typically associated with information that is either given in the discourse or accessible (Baumann, 2005; Pierrehumbert & Hirschberg, 1990). It is unclear how or whether these differences in discourse function might affect syntactic parsing, particularly in the present study where the discourse context is limited. One simple hypothesis is that accents which suggest that constituent is salient and new (e.g., L + H*) might make attachment to that location more probable. Why say more about what is already given?⁷

On this hypothesis, we might predict little or no effect of the accent pattern in cases in which the verb and noun have the same kind of pitch accent. This is true of five of our eight structures: (0, IP), (0, 0), (ip, ip), (ip, 0), and (ip, IP). The three conditions with asymmetric accenting suggest that accenting alone cannot account for our findings. Our (IP, 0) utterances had a stronger accent on the verb (L + H*) than on the noun (L*) but they had one of the lowest proportions of high attachments, as expected on both the RBH and ABH. The (0, ip) utterances had stronger accents on the noun (L + H*) than the verb (L* or no pitch accent at all) but they were consistently interpreted as high attachments. Our (IP, ip) utterances had stronger accents on the verb (H*) than the noun (L*), which might nudge them towards high attachment. These utterances were interpreted as high attachments about 69% of the time, landing in the middle of the structures tested.

None of the critical conclusions in Table 6 can be attributed to differences in accent patterns. For all five tests of the RBH, the predicted effect of accenting is either neutral or goes in the *opposite direction* of the observed and predicted effect of boundary strength. For example, we observed that (0, ip) utterances received more high attachments than (ip, ip) utterances, despite the fact that the (0, ip) utterances had a stronger accent on the noun (L + H*) than the verb (L* or none) which would be predicted to promote low attachment, while the (ip, ip) utterances had L* accents in both positions. Similarly, for the ABH, there are two critical comparisons in which the predicted effect of accenting is either neutral or goes against the observed effect of boundary strength. For example, in support of prediction 2b, we found that (IP, 0) utterances were less likely to be interpreted as high attachments than (IP, ip) utterances even though they had a stronger accent on the verb (L + H*) than (IP, ip) utterances (H*) and the same type of accent on the noun (both L*). Finally, our null findings for prediction 2c of the ABH cannot be attributed to competing effects of boundary strength and accenting. The ABH hypothesis predicts that (0, ip) utterances should receive fewer high attachments than (0, IP) and (ip, IP) utterances. Given the hypothesis sketched above, the accent pattern on these utterances would

⁷ This is clearly inspired by Schafer and colleagues' *Focus Attraction Hypothesis* (Schafer et al., 1996), though they cannot be blamed for our particular implementation.

be expected to reinforce this prediction: the (0, ip) utterances have an L + H* accent on the noun and a weak accent (L* or none) on the verb, promoting low attachment, while the other two utterances have the same type of accent in both positions (H* and L*, respectively) creating a more neutral prosody.

In fact, of the 10 comparisons in Table 6, there is only one case in which the variation in accent type offers a competing explanation to hypotheses based on boundary strength. In support of prediction 2b, we found that (ip, 0) utterances received fewer high attachments than (IP, ip) utterances. This is consistent not only with the difference in absolute boundary size, but also with a difference in accenting: the (IP, ip) utterances have a weaker accent on the noun (L*) than on the verb (H*) while the (ip, 0) utterances have H* accents in both positions. But this does not alter our conclusions, as we noted above the other comparison testing prediction 2b cannot be explained by accenting.

In sum, accenting fails to account for the observed data pattern, predicting a reversal of several observed effects, as well as effects that are not present. While accenting may influence syntactic attachment under some circumstances (contrast Lee & Watson, 2008; Schafer, Carter, Clifton, & Frazier, 1996) in the present study, the effects of prosodic phrasing appear to dominate. The question of how accenting, discourse structure, and prosodic phrasing influence syntactic parsing clearly warrants more research.

Comparisons with previous comprehension experiments

To those who have followed the literature on prosody and ambiguity resolution, these results may be somewhat surprising. Over the past 6 years many researchers have adopted some version of the RBH and have found support for it in data from both comprehension (Carlson et al., 2001; Clifton et al., 2002) and production (Kraljic & Brennan, 2005; Schafer et al., 2005, Snedeker & Trueswell, 2003), suggesting that there is ample evidence for this position. But in fact few studies test predictions that are unique to the RBH, and no study to date has focused on exploring the predictions of ABH, while holding relative boundary strength constant. Instead, researchers have conceived of the problem as one of “global” vs. “local” prosody, in which any evidence for global structure having an influence would be evidence against the hypothesis that the interface is purely local.

To determine whether these findings were consistent with prior experiments, we looked at prior experimental studies on the effects of prosody on the resolution of attachment ambiguities. We could find only three papers which tested a unique prediction of either the RBH or ABH. These papers are summarised in Table 7.

TABLE 7
 A comparison of the findings of Study 2 and prior experiments testing the relative and absolute boundary hypotheses

<i>Prediction</i>	<i>Prior study</i>	<i>Effect in prior experiments?</i>	<i>Effect in study 2?</i>
Relative boundary strength			
1a. (IP, 0) < (0, 0)		Not tested	✓
1a. (ip, 0) < (0, 0)	Schafer (1997), Exp. 1, $p < .05$	✓	✓
1b. (IP, ip) < (ip, ip)	Carlson et al. (2001), Exp. 3, unreported but likely Clifton et al. (2002), Exp. 2, $p < .05^a$	✓	✓
1b. (ip, ip) < (0, ip)	Carlson et al. (2001), Exp. 4, unreported but likely Clifton et al. (2002), Exp. 1, <i>ns</i> ; Exp. 2, $p < .05^a$	✓	✓
1b. (IP, ip) < (0, ip)	Clifton et al. (2002), Exp. 2, $p < .05$	✓	✓
1c. (IP, IP) < (0, IP)	Clifton et al. (2002), Exp. 1, $p < .05$; Exp. 3, $p < .05$	✓	Not tested
1c. (IP, IP) < (ip, IP)		Not tested	Not tested
Absolute boundary strength			
2a. (0, 0) < (ip, ip)		Not tested	✓
2a. (ip, ip) < (IP, IP)	Carlson et al. (2001), Exp. 3, unlikely Clifton et al. (2002), Exp. 1, unlikely	X	Not tested
2a. (0, 0) < (IP, IP)		Not tested	Not tested
2b. (IP, 0) < (IP, ip)	Carlson et al. (2001), Exp. 3, unlikely	X	✓
2b. (ip, 0) < (IP, ip)		Not tested	✓
2c. (0, ip) < (0, IP)	Carlson et al. (2001), Exp. 4, unlikely Clifton et al. (2002), Exp. 1, unlikely	X	X
2c. (0, ip) < (ip, IP)	Carlson et al. (2001), Exp. 4, unlikely	X	X

Note: X, no effect; ✓, prediction confirmed.

^aClifton et al. (2002), Experiment 2, included four different kinds of attachment ambiguities. This contrast was significant in the overall analysis but not for all ambiguity types.

In compiling this table, we entered all cases in which the comparison of interest was analysed with a direct statistical test. However, since the ABH was rarely tested explicitly, we also entered the results of studies in which no planned comparison of the cells was conducted but the outcome was inferable (from the means and the pattern of main effects). In the table, these findings are labelled “likely” and “unlikely”.

For the most part the results of Study 2 converge with the prior findings. Of the 10 predictions that we tested, seven had been previously explored. In four cases, the prediction was confirmed in all studies and in two cases it was not confirmed by any of the studies. In only one case was there an actual difference in the findings. We found more high attachments for (IP, ip) utterances than for (IP, 0) ones, while Carlson et al. (2001) did not. Two details of their experiment may help to explain this discrepancy. First, as the authors point out, the utterances included in the (IP, 0) condition varied in their prosodic structure. About half appeared to contain an intermediate phrase break in position B. Thus the two conditions actually overlapped in structure. Second, both structures were interpreted as low attachments on about 85% of the trials, raising the possibility that ceiling effects limited the sensitivity of the experiment.

Despite these similarities, the present study diverges from the others in finding substantial support for unique predictions of the ABH. Table 7 suggests that this reflects the particular contrasts that were studied. In Study 2, we found reliable evidence for the influence of absolute boundary size when we contrasted ip or IP boundaries with null boundaries. However, we failed to find effects when we contrasted ip boundaries with IP boundaries. The prior work has focused almost exclusively on the latter contrast.

Integrating the findings from production and comprehension

It is difficult to directly compare the results of the comprehension and production studies. The power of the production study varied across the cells but was generally reduced relative to the comprehension study. In the comprehension study, each prosodic structure appeared equally often but participants had a bias towards making high attachments. In the production study, high and low attachments appeared equally often but the participants produced some prosodic structures more than others. Thus, there is a different profile of sensitivity across the cells in each experiment. Nevertheless a few observations can be made.

First, several predictions were confirmed for both producers and comprehenders, suggesting that these are stable features of the prosody–syntax interface. This included the strong contrast between the (IP, 0) and (0, 0) structures, supporting the RBH, which was present in both aware and unaware speakers. In the case of the ABH, all contrasts which pitted an ip or IP boundary in position B against a null boundary were confirmed for the aware speakers and the comprehenders.

Where the two data sets diverged, the interpretation is less clear. This occurred in two places. First, most of the unique predictions of the RBH were confirmed in comprehension but not in production. Here no strong conclusions are possible. The production data are sparse in most of these cells.

It is tempting to conclude that an effect in comprehension implies that the distinction must be reflected in production: Why use a cue which isn't valid? But it is certainly conceivable that two systems could have different operating principles. For example, some relative boundary strength contrasts could be correlated with syntactic attachment during production because each variable is correlated with some third factor (e.g., absolute boundary strength or the length in syllables of the respective constituents). When these confounds are removed (as in the present studies), we would expect to find no effect of syntax on the relevant contrast during production. The comprehension system, however, could have acquired a mapping between relative boundary strength and attachment on the basis of this correlation and might continue to show sensitivity to the contrast under these circumstances.

The second discrepancy is in the final prediction of the ABH. This contrast pits utterances with the (0, ip) structure against those with the (0, IP) or (ip, IP) structure. Producers are more likely to use the latter for high attachments but comprehenders categorically select high attachments for all three utterance types. This difference is intriguing because the comprehension study has considerably more power than the production study. The discrepancy could reflect ceiling effects in the comprehension data. However, our examination of prior studies suggests that this is a fairly robust finding in judgment tasks (see Carlson et al., 2001; Clifton et al., 2002). Alternately, it could reflect a difference in how the effects of absolute boundary strength arise during comprehension and production. Further research on these structures using more parallel materials and tasks would be informative.

Two absolute boundaries?

These data call out for a third hypothesis that can explain the observed effects of both relative and absolute boundary strength. One tempting proposal is that attachment depends on both the absolute strength of the boundary at both position A and the absolute strength of the boundary at position B. We will refer to this hypothesis as the two absolute boundaries hypothesis (2ABH).⁸ In contrast with the RBH and ABH (see 7 and 8), 2ABH does not provide a relative ordering of all the possible structures, unless additional assumptions are added about how the sizes of the two boundaries interact. However, it does make the following predictions.

⁸ Full development of 2ABH would have to specify how the relevant location A is defined. The ABH (as we defined it) grants no role to an earlier boundary. RBH proposals typically argue that multiple boundaries between the high-attachment site and the low-attachment site may be relevant (see e.g., Schafer et al., 2000). In the short simple sentences used in the present study, we can bypass this issue, since there are no likely break locations between position A and B.

14. The size of the boundary at location A should influence attachment

$$\begin{aligned} &(\text{IP}, 0) < (\text{ip}, 0) < (0, 0) \\ &(\text{IP}, \text{ip}) < (\text{ip}, \text{ip}) < (0, \text{ip}) \\ &(\text{IP}, \text{IP}) < (\text{ip}, \text{IP}) < (0, \text{IP}) \end{aligned}$$

15. The size of the boundary at location B should influence attachment

$$\begin{aligned} &(\text{IP}, 0) < (\text{IP}, \text{ip}) < (\text{IP}, \text{IP}) \\ &(\text{ip}, 0) < (\text{ip}, \text{ip}) < (\text{ip}, \text{IP}) \\ &(0, 0) < (0, \text{ip}) < (0, \text{IP}) \end{aligned}$$

All the predictions in (15) involve varying the size of the boundary at B and thus are also made by the ABH. Most of the predictions in (14) are also made by the RBH, however, there are two predictions which are unique to the 2ABH:

16. Unique predictions of 2ABH (relative to ABH and RBH)

- a. $(\text{IP}, 0) < (\text{ip}, 0)$
- b. $(\text{ip}, \text{IP}) < (0, \text{IP})$

Our data provide scant support for either prediction (see Table 1 and Figure 1). Both of the prosodic structures in 16a are strongly associated with low-attachment contexts in aware speakers and were generally interpreted as low attachments in the comprehension study, while both of the structures in 16b were strongly associated with the high-attachment contexts in aware speakers and interpreted as high attachments in the comprehension study (all $ps > .5$). The data pattern for unaware speakers was murkier. There is a marginal effect suggesting that the (ip, 0) structure appeared more in high-attachment contexts than the (IP, 0) structure ($p = .057$), which would support the prediction in 16a. However, there is also a marginal effect that runs counter to the prediction in 16b: the (ip, IP) structure was more associated with high-attachment contexts than the (0, IP) structure ($p = .073$).

We might also ask whether our data provides any unique support for the RBH or ABH relative to the 2ABH. In the case of the RBH, there are no unique predictions since all predictions which are not shared by the ABH necessarily involve varying the size of the boundary at location A, and thus are also predictions of the 2ABH. However, for the ABH there are several unique predictions relative to both of the other hypotheses.

17. Unique predictions of ABH (relative to 2ABH and RBH)

- c. $(0, 0) < (\text{ip}, \text{ip}) < (\text{IP}, \text{IP})$
- d. $(\text{ip}, 0) < (\text{IP}, \text{ip})$
- e. $(0, \text{ip}) < (\text{ip}, \text{IP})$

Our results provide robust support for these predictions. All these predictions were tested and confirmed for aware speakers in Study 1 (see Table 2). Study 2 only tested some of these predictions because (IP, IP) utterances were not included, but two of the three predictions that were tested were confirmed (see Table 6). To account for these findings the 2ABH would have to be supplemented to grant a privileged role to the lower boundary. In sum, our results suggest that the 2ABH does not provide an adequate explanation for the observed attachment patterns.

FINAL WORDS

These studies support unique predictions of both the relative boundary strength hypothesis and the absolute boundary strength hypothesis. Thus our findings imply that neither hypothesis alone is sufficient to account for the relation between prosodic phrasing and the attachment of an ambiguous phrase. In an ideal world, we would now propose an alternate model of the prosody–syntax interface which could account for both sets of findings. Lacking this insight, we can only point to the questions that must be resolved before such a theory can be constructed.

First, greater clarity is needed about the contexts in which ip and IP boundaries have distinct effects and the contexts in which they are treated as equivalent. In our production data, we find robust differences between the two boundary types in comparisons where ABH is in question. However, our comprehension data (and that of Carlson et al., 2001) suggest that the two boundary types are distinct for tests of the RBH but equivalent in tests of the ABH.

Second, it is critical that we know if the effects of relative and absolute boundary strength stem from the same mechanism or different mechanisms. Resolving this question will require on-line experiments that address the unique predictions of each hypothesis and provide information not only about the pattern of interpretation, but also about the processes by which these patterns arise. Our comprehension data hint at the possibility that the two processes may be distinct. Effects of relative boundary size clearly depend on a prosodic representation that distinguishes between ip and IP boundaries. In contrast, effects of absolute boundary size may depend on a coarser

representation of prosody in which fails to capture this distinction. This raises the intriguing possibility that the effects of absolute and relative boundary size arise at different points in processing. For example, absolute boundary effects might result from lower-level processes that chunk input for analysis and treat all breaks as identical, while relative boundary effects could result from higher-level processes that attempt to align syntactic structure with a rich representation of prosodic structure.

Finally, in the current paper we have followed the dominant theory of prosodic structure and the dominant practice in psycholinguistics by treating prosodic boundaries as discrete categories and assuming that all variation in interpretation is linked to the categorical status of these boundaries. To the best of our knowledge only one comprehension study has ever tested this assumption (in the context of the prosody–syntax interface). This experiment found the predicted null effect (Experiment 2; Carlson et al., 2001) but additional evidence for this position is critical.

REFERENCES

- Allbritton, D., McKoon, G., & Ratcliff, R. (1996). Reliability of prosodic cues for resolving syntactic ambiguity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 714–735. doi:10.111.122.4
- Altmann, G. T. M. (1998). Ambiguity in sentence processing. *Trends in Cognitive Sciences*, 2(4), 146–152. doi:10.1016/S1364-6613(98)01153-X
- Baumann, S. (2005, September). *Degrees of givenness and their prosodic marking*. Paper presented at the international symposium on Discourse and prosody as a complex interface, Aix-en-Provence.
- 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. doi:10.1016/0749-596X(91)90030-N
- Beckman, M. E. (1996). The parsing of prosody. *Language & Cognitive Processes*, 11, 17–67. doi:10.1080/016909696387213
- Beckman, M. E., & Hirschberg, J. (1994). *The ToBI annotation conventions*. Columbus, OH: Ohio State University.
- Beckman, M. E., & Pierrehumbert, J. B. (1986). Intonational structure in Japanese and English. *Phonology Yearbook*, 3, 255–309.
- Carlson, K., Clifton, C., & Frazier, L. (2001). Prosodic boundaries in adjunct attachment. *Journal of Memory & Language*, 45(1), 58–81. doi:10.1006/jmla.2000.2762
- Clifton, C. Jr., Carlson, K., & Frazier, L. (2002). Informative prosodic boundaries. *Language and Speech*, 45, 87–114.
- Cooper, W. E., & Paccia-Cooper, J. (1980). *Syntax and speech*. Cambridge, MA: Harvard University Press.
- Cutler, A., Dahan, D., & van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language & Speech*, 40(2), 141–201.

- Fernald, A., & McRoberts, G. (1996). Prosodic bootstrapping: A critical analysis of the argument and the evidence. In J. Morgan & K. Demuth (Eds.), *Signal to syntax* (pp. 365–388). Mahwah, NJ: Erlbaum.
- Ferreira, F. (1988). *Planning and timing in sentence production: The syntax-to-phonology conversion*. Unpublished dissertation, University of Massachusetts, Amherst, MA.
- Gee, J. P., & Grosjean, F. (1983). Performance structures: A psycholinguistic and linguistic appraisal. *Cognitive Psychology*, *15*, 411–458.
- Kjelgaard, M. M., & Speer, S. R. (1999). Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language*, *40*, 153–194. doi: 10.1006/jmla.1998.2620
- Kraljic, T., & Brennan, S. E. (2005). Prosodic disambiguation of syntactic structure: For the speaker or for the addressee? *Cognitive Psychology*, *50*(2), 194–231.
- Lee, E.-K., & Watson, D. (2008, March). *Effects of focal accents on relative clause attachment: Accent effects or memory effects?* Poster presented at the 21st annual CUNY conference on human sentence processing, Chapel Hill, NC.
- 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.
- Marcus, M., & Hindle, D. (1990). Description theory and intonation boundaries. In G. Altmann (Ed.), *Cognitive models of speech processing: Computational and psycholinguistic perspectives* (pp. 483–512). 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*, *45A*, 73–87. doi:10.1080/14640749208401316
- Nagel, H. N., 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*(2), 319–344.
- Nespor, M. A., & Vogel, I. (1986). *Prosodic phonology*. Boston: Kluwer.
- Pierrehumbert, J. B., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. R. Cohen, J. Morgan, & M. E. Pollack (Eds.), *Intentions in communication* (pp. 271–311). Cambridge, MA: MIT Press.
- Price, P., Ostendorf, M., Shattuck-Hufnagel, S., & Fong, C. (1991). The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America*, *90*, 2956–2970.
- Pynte, J., & Prieur, B. (1996). Prosodic breaks and attachment decisions in sentence parsing. *Language and Cognitive Processes*, *11*, 165–192. doi:10.1080/016909696387259
- Schafer, A., Carter, J., Clifton, C., & Frazier, L. (1996). Focus in relative clause construal. *Language and Cognitive Processes*, *11*(1/2), 135–163. doi:10.1080/016909696387240
- Schafer, A., Speer, S., & Warren, P. (2005). Prosodic influences on the production and comprehension of syntactic ambiguity in a game-based conversation task. In J. C. Trueswell & M. K. Tanenhaus (Eds.), *Approaches to studying world situated language use* (pp. 209–226). Cambridge, MA: MIT Press.
- Schafer, A., Speer, S., Warren, P., & White, S. (2000). Intonational disambiguation in sentence production and comprehension. *Journal of Psycholinguistic Research*, *29*, 169–182.
- Schafer, A. J. (1997). *Prosodic parsing: The role of prosody in sentence comprehension*. Unpublished doctoral dissertation, University of Massachusetts, Amherst, MA
- Selkirk, E. O. (1986). On derived domains in sentence phonology. *Phonology Yearbook*, *3*, 371–405.
- Shattuck-Hufnagel, S., & Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research*, *25*(2), 193–247.
- 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.

- Snedeker, J., & Trueswell, J. C. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology*, *49*(3), 238–299.
- Steedman, M. (2000). Information structure and syntax-phonology interface. *Linguistic Inquiry*, *31*(4), 649–689.
- Steinhauer, K., Alter, K., & Friederici, A. D. (1999). Brain responses indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience*, *2*, 191–196.
- Warren, P. (1999). Prosody and sentence processing. In S. Garrod & M. Pickering (Eds.), *Language processing* (pp. 155–188). Hove, UK: Psychology Press.
- Warren, P., Grabe, E., & Nolan, F. (1995). Prosody, phonology and parsing in closure ambiguities. *Language and Cognitive Processes*, *10*, 457–486. doi:10.1080/01690969508407112
- Watson, D., & Gibson, E. (2004). The relationship between intonational phrasing and syntactic structure in language production. *Language and Cognitive Processes*, *19*, 713–755. doi:10.1080/01690960444000070
- Watson, D., & Gibson, E. (2005). Intonational phrasing and constituency in language production and comprehension. *Studia Linguistica*, *59*, 279–300.
- Weber, A., Grice, M., & Crocker, M. W. (2006). The role of prosody in the interpretation of structural ambiguities: A study of anticipatory eye movements. *Cognition*, *99*, B63–B72. doi:10.1016/j.cognition.2005.07.001

APPENDIX: SENTENCES FOR STUDY 2

Critical sentences

1. You can turn over the whale/killer whale with the stick.
2. You can point at the dog/puppy dog with the flower.
3. You can drag the cat/kitty cat with the pipe cleaner.
4. You can throw the bear/panda bear with the cup.
5. You can scratch the whale/killer whale with the paper.
6. You can pinch the dog/puppy dog with the barrette.
7. You can feel the cat/kitty cat with the feather.
8. You can blow on the bear/panda bear with the fan.

Pronoun ambiguities (implicit causality)

1. The cow envied the zebra because he is mean.
2. The elephant ignored the monkey because he won the race.
3. The giraffe kicked the lion because he lost his rock.
4. The hippo kissed the horse because he was bored.
5. The horse laughed at the giraffe because she had a new toy.
6. The lion liked the frog because she is happy.
7. The lobster looked at the dolphin because she had eaten too much.
8. The pig whispered to the cow because she made new friends easily.

Relative clause/complement clause ambiguities

1. You can whisper to the little bear who's brown.
2. You can tell the giraffe who's naughty.
3. You can inform the hippo who's hungry.
4. You can show the horse who's new.
5. You can teach the cow who's good.
6. You can ask the lobster who's tired.
7. You can whisper to the lion who yawned.
8. You can tell the zebra who's nice.