

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

## **Prosodic boundaries in adjunct attachment**

Katy Carlson

Charles Clifton, Jr.

and

Lyn Frazier

University of Massachusetts/Amherst

Corresponding author  
Charles Clifton  
Department of Psychology  
University of Massachusetts  
Amherst, MA 01003  
[cec@psych.umass.edu](mailto:cec@psych.umass.edu)  
(413) 545-2653

running head: Prosodic Boundaries

Authors' note: This research was supported in part by research grant HD-18708 and training grant HD-07327 from the National Institutes of Health to the University of Massachusetts. The authors would like to thank Jennifer Boltuch, Maureen Hutchins, Shawn Johnson, Kathleen McSheehy, Danielle Michaud, Michael Parker, Yanira Vasquez, and Elisabeth Villalta for assistance in conducting the experiments reported here, and Elizabeth Selkirk and Amy Schafer for very helpful discussions of previous versions of this paper and the prosodic analyses it contains. The order of authors is alphabetical. ToBI-transcribed materials can be found at <http://www-unix.oit.umass.edu/~cec>. Authors' addresses: Carlson and Frazier, Department of

Prosodic boundaries

August 30, 2000

Linguistics, University of Massachusetts, Amherst, MA 01003; Clifton, Department of  
Psychology, University of Massachusetts, Amherst, MA 01003, [cec@psych.umass.edu](mailto:cec@psych.umass.edu)

**Abstract**

Five studies explored the processing of ambiguous sentences like Martin maintained that the CEO lied when the investigation started/at the start of the investigation. The central question was why particular prosodic boundaries have the effects they do. A written questionnaire provided baseline preferences, and suggested that clausal adjuncts (when the investigation started) receive more high attachments than nonclausal adjuncts (at the start of the investigation). Four auditory studies manipulated the prosodic boundary before the adjunct clause and the prosodic boundary between the matrix clause and its complement. They disconfirm every version of an account where only the local boundary before the adjunct is important, whether the account is based on the acoustic magnitude of the boundary or its phonological type (an intermediate boundary characterized by the presence of a phrase accent vs. an intonational phrase boundary characterized by both a phrase accent and a boundary tone). Instead the results support use of the global prosodic context, crucially the relative size of the local boundary and the distant boundary.

keywords: prosody, sentence comprehension, syntactic ambiguity, auditory language

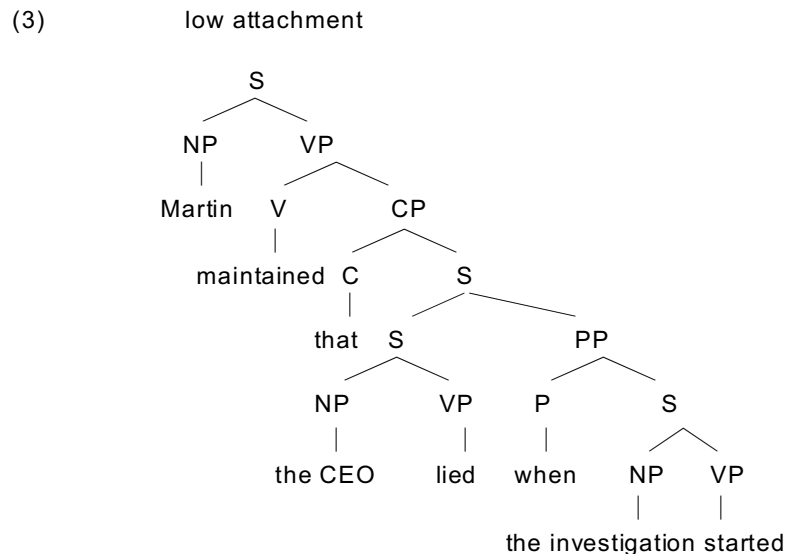
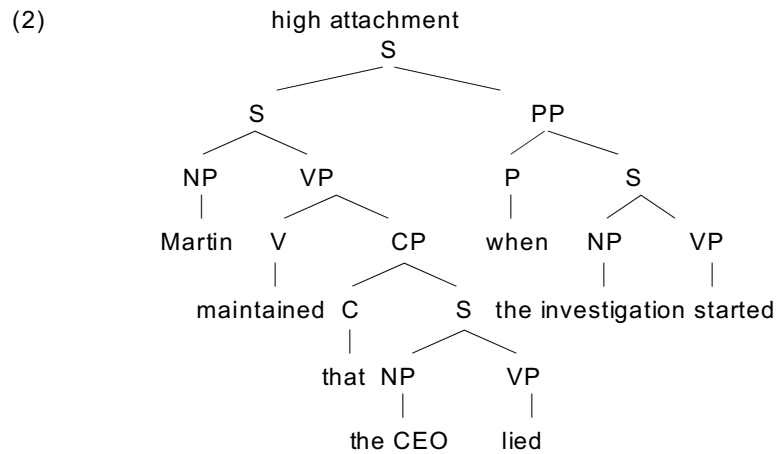
It is clear that prosodic boundaries can bias or disambiguate the interpretations of many constituent structure ambiguities in spoken language (Carroll & Slowiaczek, 1987; Kjelgaard & Speer, 1999; Lehiste, 1973; Price, Ostendorf, Shattuck-Hufnagel & Fong, 1991; Warren, Grabe, & Nolan, 1995, for example). Consider the ambiguity we will focus on in the present paper, the ambiguity between high and low attachment of the final prepositional phrase (PP) in sentences like (1):

(1) Martin maintained that the CEO lied when the investigation started.

The PP when the investigation started can modify either maintained (high attachment) or lied (low attachment). We will demonstrate that the placement of prosodic boundaries in an utterance of (1) can affect its interpretation, and attempt to determine why particular prosodic boundaries have the effects they do. Does the acoustic magnitude of a boundary determine its effectiveness? Or does the phonological category to which it belongs govern its role? Do nonsyntactic factors, such as the length or syntactic complexity of phrases, shape the effect of a prosodic boundary? Finally, is the relation among the prosodic boundaries in an utterance (arguably even the full prosodic description) the critical factor?

In order to be explicit in our arguments, we assume the syntactic structures shown in (2) and (3). In this structure, we treat the adjunct PP as adjoined to a clause. An alternative analysis of this type of adjunct attaches it lower, into the verb phrase (VP), but what matters to the predictions of the prosodic theories we discuss below is whether the PP is treated as part of the

embedded complement sentence as in (3) or only as part of the higher S as in (2).



Several interesting proposals about the prosodic disambiguation of syntactic structure may be found in the processing literature. Marcus and Hindle (1990) analyzed prosody from the perspective of a deterministic D-theory parser in which phrases initially treated as separate can easily be combined later but a string of words treated as a single phrase cannot (without disruption) later be split apart. They claimed that a major intonation boundary is introduced as an "unknown" lexical item into the lexical stream, causing the parser to terminate the current

constituent and treat it as complete. In the absence of such a boundary, new phrases are syntactically combined with preceding phrases into "a single chunk that...may not be broken up later" (p. 483). Applied to our example (1), Martin maintained that the CEO lied when the investigation started, Marcus and Hindle's analysis predicts that the absence of a major prosodic boundary before the PP would encourage a low attachment interpretation. The parser is likely to combine the PP as part of the current S and interpret it as modifying the lowest verb, lied. Conversely, the presence of a major prosodic boundary before the PP would at least delay a low attachment interpretation, allowing a high attachment interpretation to be chosen on the basis of other factors.

Price et al. (1991) studied the effects of prosodic boundaries in a variety of sentence types, finding that the presence or absence of intonational phrase boundaries, which are relatively large, affected the interpretation of several types of ambiguous sentences, while smaller phrase boundaries (called intermediate phrase boundaries) did not usually carry such an impact. Specifically, they stated that "relatively larger break indices tended to mean that syntactic attachment was higher rather than lower" (page 2963). Applying their analysis to our (1), the presence of an intonational phrase boundary before the adjunct in (1) is clearly predicted to favor its high attachment. Predictions are less clear concerning the effect of an intermediate phrase boundary. Given the conclusion of Price et al. that there is a "tendency to associate relatively larger prosodic phrase breaks with larger syntactic breaks" (p. 2966) the prediction seems to be that an intermediate phrase boundary should result in fewer high attachments than an intonational phrase boundary.

An alternative to this suggestion is that the sheer acoustic size of a prosodic boundary is actually what matters, not its phonological type. Existing accounts of the effects of prosodic boundaries interpret differences between prosodic boundaries in terms of differences in phonological category rather than sheer acoustic size. But the sheer acoustic size hypothesis is consistent with at least some of the evidence that Price et al. present and is an alternative we examine below.

Kjelgaard and Speer (1999) presented evidence that the placement of either an intonational or an intermediate prosodic boundary affects the comprehension of temporarily ambiguous "late closure" sentences (cf. Frazier & Rayner, 1982). They demonstrated that appropriate occurrence of either type of prosodic boundary eliminated the usual difficulty with sentences that violated the normal late closure preference. In addition, a baseline condition that lacked all relevant boundaries exhibited a late closure advantage, and cross-splicing sentences to reverse the prosodic phrasing and make it inappropriate inhibited comprehension. Their data thus suggest that an intermediate boundary is sufficient to bias interpretation in sentences like (1) that we are investigating.

Important early work by Nespor and Vogel (1983, 1986) approached the question of the relation between prosody and syntax quite differently, asking what sentences of the language can be disambiguated prosodically. They essentially proposed that a string of words including a lexical head (a noun, verb, or adjective) together with the words on its nonrecursive side (the left side if the head's complements occur on the right, the right side if its complements are on its left) is organized into a phonological phrase (which we take to be equivalent to what we have called

an intermediate phrase). Intermediate phrases are in turn organized into intonational phrases by combining as many intermediate phrases as possible into a single intonational phrase up to the end of the maximal projection of a noun or to the beginning of another clause.<sup>1</sup> They suggested that only structures in which the placement of prosodic phrase boundaries differs can be disambiguated prosodically, and further hypothesized that ambiguities involving analyses that differ in the placement of intonational phrase boundaries will be disambiguated more reliably than ambiguities differing only in the placement of intermediate phrase boundaries. They demonstrated above-chance success in determining a speaker's intended utterance when sentences differed prosodically only in intermediate phrasing, and near-100% success when sentences differed in intonational phrasing.

In Nespor and Vogel's analysis, our example (1), Martin maintained that the CEO lied when the investigation started, could have an intermediate phrase boundary before the final PP under either high or low attachment. Thus, the presence or absence of such a boundary should not affect interpretation. Whether their proposal permits the presence of an intonational phrase boundary before the final PP to disambiguate the sentence seems to depend on details of their stated principles. In Nespor & Vogel (1983), an intonational phrase is continued until another clause (S') begins. If "another clause" is interpreted to mean a new clause that is not a constituent of the clause currently being processed, then a boundary is appropriate for the high attachment case (2) but not for the low attachment case (3). In this case, an intonational phrase boundary should disambiguate the sentences. However, a different understanding of this proposal or use of their 1986 statement (see footnote 1) would permit an intonational phrase boundary in either (2)



or (3), eliminating the possibility of disambiguation.

Nespor and Vogel's account allows speakers to combine intermediate phrases into larger ones, and combine intonational phrases or split them apart, for various reasons (though the effects of this optionality on comprehension are not fully explored). Kjelgaard and Speer (1999) also recognized this optionality, using a baseline condition to demonstrate that relevant prosodic boundaries are not actually necessary to the interpretation of the sentences, just helpful. It has long been known that speakers have options about the location and identity of the prosodic boundaries they use, which complicates both the grammar (the syntax-prosody mapping rules, cf. Selkirk, 1984) and the task of a listener. A listener cannot simply assume that every prosodic boundary is a reflex of syntactic structure, or should be used as a signal for some type of attachment.

For example, the length of phrases can affect prosodic phrasing. In producing a simple sentence with a long subject, speakers are likely to prosodically package the verb with its object. But with a short subject and long object, they are likely to package the verb with the subject (Martin, 1970). Thus in (1), a boundary before the clausal adjunct could be due to the clausal (and therefore weighty) status of the adjunct and thus irrelevant to its attachment, or it could be due to a syntactic boundary at the end of the embedded clause and thus signal high attachment; these two possibilities alone suggest very different uses of the prosodic boundary. Numerous other factors may also affect prosodic choices, such as rate of speech, style, prominence marking, and so on.

An approach to prosodic boundaries that acknowledges optionality and goes beyond

examining the effects of the presence vs. absence of a single local boundary was developed by Schafer (1997; Schafer & Speer, 1997). She suggested that listeners may form a “prosodic package” of material within a single prosodic phrase (either an intermediate phrase or an intonational phrase). Material within the current prosodic phrase is more visible than material outside the current prosodic phrase, making attachment into the current phrase easy. Schafer specifically proposed the prosodic visibility hypothesis in (4):

- (4) Prosodic visibility
- a. The phonological phrasing of an utterance determines the visibility of syntactic nodes.
  - b. Nodes within the [intermediate] phrase currently being processed are more visible than nodes outside that [intermediate] phrase; visibility is gradient across multiple [intermediate] phrases.
  - c. In first analysis and reanalysis, attachment to a node with high visibility is less costly in terms of processing/attentional resources than attachment to a node with low visibility.

Schafer (1997) tested this hypothesis using sentences like those in (5) that ended in a PP (with a mean look) that could either attach into the immediately preceding NP (the rider) or into the VP headed by the verb angered).

- (5)
- a. [The bus driver angered the rider] [with a mean look].
  - b. [The bus driver angered] [the rider with a mean look].
  - c. [The bus driver angered the rider with a mean look].
  - d. [The bus driver] [angered] [the rider] [with a mean look].

She predicted that NP attachment would be preferred in (5b) relative to the other conditions,

because in (5b) the NP is inside the current prosodic package with the PP while the verb is outside that phrase (brackets indicate intermediate phrases). In condition (5c), by contrast, both the NP and the verb are within the same intermediate phrase as the PP, and in (5a), both are separated by one phrase, so attachment should be more evenly split between the two sites. In (5d), the verb is separated by two phrases while the NP is separated by only one, possibly making the frequency of NP modification choices intermediate between (5b) and (5a,c). Data from a listening study confirmed these predictions, which clearly do not depend on an obligatory 1:1 syntax-prosody mapping. Pynte and Prieur (1996) reported a similar finding in French. A prosodic boundary following the verb in French counterparts to (5) favored NP attachment of the PP, but the effect was eliminated by the presence of a second boundary preceding the PP. They discussed two possible loci for the effects, one focusing on the initial grouping of arguments and one acting during revision processes.

Various authors have suggested that global prosodic structures influence processing (Kjelgaard & Speer, 1999; Pynte & Prieur, 1996; Speer, Kjelgaard, & Dobroth, 1996; Slowiaczek, 1981). However, Schafer's (1997) work most explicitly denied the "local cue" logic implicit in much of the discussion of prosodic disambiguation and she developed specific testable hypotheses about the use of nonlocal prosodic structure. In particular, she argued that the interpretation of a prosodic boundary depends on the other boundaries in the sentence. Turning to our example (1), Martin maintained that the CEO lied when the investigation started, Schafer's proposal predicts that low attachment of the adjunct should be preferred when the adjunct is not preceded by a prosodic boundary. When the adjunct is preceded by a prosodic boundary (either

an intermediate phrase or an intonational phrase), the low attachment site should become less visible. How this visibility distinction should affect attachment of the adjunct will depend on further articulation of the proposal. For instance, it is necessary to specify whether listeners always choose the most visible site or probabilistically attach to the lower or higher site based on the size of the visibility differences among alternative attachment sites. Further, it is necessary to work out the interaction between presumed differences in visibility and baseline syntactic attachment preferences (low in the structure we are testing, (1), but high in the structure Schafer tested, (5)).

To summarize, a variety of interesting, different, and empirically-motivated proposals exist concerning prosodic effects in parsing. Most of them clearly predict that sentences like our example (1) can be disambiguated prosodically. Some (Marcus & Hindle, 1990; Price et al., 1991; possibly Nespor & Vogel, 1983) suggest that only an intonational phrase boundary before the final PP will affect interpretation. Others (Kjelgaard & Speer, 1999; Schafer, 1997) are consistent with the possibility that either an intonational or an intermediate boundary will suffice. Schafer's approach holds open the possibility that the global prosodic analysis, in particular the full pattern of prosodic boundaries, can affect interpretation. As we present data testing these predictions below, we will develop a hypothesis in the spirit of Schafer's approach, in which we propose that the relative size (in terms of prosodic categories) of the prosodic boundaries in an utterance is what affects interpretation.<sup>2</sup>

In order to test predictions like these, we must explicitly specify how we represent and identify different categories of prosodic boundaries. We turn to that now.

**Prosody: how to represent it and how to verify representations**

Whenever an acoustic signal is naturally produced, many detailed differences distinguish one token from the next, raising the question of what counts as two instances of the “same” linguistic structure. We will adopt the intonational theory of Pierrehumbert and Beckman (1988; Beckman & Pierrehumbert, 1986; cf. Beckman, 1996; Shattuck-Huffnagel & Turk, 1996), but acknowledge that theories of prosody are still evolving. Without a final, correct, and universally accepted prosodic theory, the problem of categorizing utterances has no perfect solution. The approach we will follow here is to use an explicit system to transcribe the perceptual analyses of sentences (the ToBI labelling system: Beckman & Elam, 1997; Beckman & Hirschberg, 1994; Silverman, Beckman, Pitrelli, Ostendorf, Wightman, Price, Pierrehumbert, & Hirschberg, 1992), but also present acoustic measurements of critical properties as a backup for these transcriptions.

An utterance has many prosodic properties, not all of which are considered important in the intonational theory adopted here or represented in a ToBI transcription. What these systems do represent are the relative prominence and prosodic grouping of words in an utterance, and the tonal events (changes in fundamental frequency, or F0) which typically go along with these factors. Prominent words, which bear relatively more stress than other words in an utterance, are marked with pitch accents in speech (indicated by \* in a prosodic transcription). Pitch accents may be relatively high or low (H\* or L\*), and may also be associated with more complicated tonal movements (such as a low target followed immediately by a rise, represented by L\*+H, and so on).

Two levels of prosodic grouping are postulated in this theory. An intermediate phrase (ip)

is the smaller prosodic group. Each ip contains one or more pitch accents (e.g., H\*, L\*) on prominent words, and must end with a high or low phrase accent (H-, L-). The phrase accents represent tonal events that are less localized than the pitch accents. They are transcribed at the end of the last word in an ip, but tend to be realized starting just after the last pitch accent in an ip, which may result in a fundamental frequency plateau (among other possibilities). Intermediate phrases are themselves grouped into the larger Intonational Phrases. Each Intonational Phrase (IPh) consists of one or more ips, and each IPh ends with the phrase accent marking the final ip boundary plus another tone, a boundary tone. Boundary tones can be high or low (H%, L%), and are generally located on the final syllable or word of a phrase. All of the high and low labels discussed here represent F0 targets rather than absolute tonal levels. Intonational contexts within or between phrases may produce frequencies marked as low that are relatively higher than areas marked as high in another phrase or context. In addition, issues such as the pitch range of the speaker and the use of pitch range to mark discourse (Hirschberg and Pierrehumbert, 1986; Pierrehumbert & Hirschberg, 1990) may result in similar relative judgments.

Several studies (Silverman et al., 1992; Pitrelli, Beckman, & Hirschberg, 1994) have shown that transcribers trained using the ToBI training materials (Beckman & Elam, 1997) show high levels of inter-transcriber agreement: over 80% agreement that a particular category of tonal element (pitch accent, phrase accent, or boundary tone) is present, and agreement of almost 80% on the label within that category. In many cases, a prosodic representation gives almost unambiguous signals as to its transcription. For example, a high F0 peak on the stressed syllable of a word which is prominent in amplitude and duration should clearly be marked with a H\*

pitch accent; an F0 fall to the bottom of a speaker's range at the end of an utterance should clearly be marked with the tones appropriate to the end of an IPh: L-L%. But the training materials acknowledge that there are places where the correct transcription can depend on subtle and sometimes subjective criteria.

One such place is the marking of sentence-internal prosodic boundaries, which will be crucial in the experiments presented here. Thus we will present a short description of the criteria used in our transcriptions. Durational and rhythmic differences, such as lengthening of the word immediately prior to a boundary and the durations of pauses between words, were considered in conjunction with tonal events shown in the F0 tracks. A fall-rise F0 contour, in which a low phrase accent target on or before the pre-boundary word is followed by a rise in pitch within that word to a high boundary tone, coupled with a fairly long duration for the pre-boundary word, was considered to be evidence for an IPh boundary (the L-H% continuation rise). A fall in pitch that reached the bottom of the speaker's range would be considered a L-L% IPh boundary, as in the sentence- and utterance-final boundaries. A fall in pitch that failed to reach the bottom of the speaker's range (as shown by the comparison to the L-L% target at the end of the sentence, for example) was considered to be an ip boundary (L-), even when durational measurements were long. Thus, when tonal and durational events suggested conflicting analyses, the tones were given priority. We feel that prosodic patterns such as fall-rise and fall are more consistently and easily detected at various pitch levels than absolute measures such as word durations. Because all of these judgments are important to our theoretical claims, as well as being difficult to determine, we are including measurements of F0 targets and durations around the regions of interest (Tables

1 and 2) and we provide transcriptions of all experimental sentences on the website mentioned in the Authors' Notes. In this way we hope to allow readers access to the information behind general transcriptions, so that our claims may be evaluated in an informed manner.<sup>3</sup>

### Experiment 1

Experiment 1 was a written questionnaire study. Sixteen sentences like (6) were constructed, with two versions of each, the first with a clausal adjunct and the second with a nominal adjunct. The sentences appear in Appendix 1.

- (6) a. Susie learned that Bill telephoned after John visited.  
b. Susie learned that Bill telephoned after John's visit.

Each sentence contained a matrix clause consisting of a proper name and a verb taking a sentential complement (that Bill telephoned... in (6)). The sentence ended with a temporal adjunct that could be attached low into the embedded clause or high into the matrix clause. The adjunct started with a preposition or an adverb, typically before, after, during, or when. In the clausal version (6a), the object of the preposition was a simple clause consisting of a proper name and a tensed verb. In the nonclausal version (6b), the object of the preposition was a nominal (John's visit). It was generally constructed by adding a possessive marker 's to the subject (John) followed by a deverbal noun (visit) based on the verb in the (a) version or by using a head noun based on the verb together with an argument.

The purpose of conducting a written questionnaire was to obtain a baseline preference for interpreting the ambiguous adjunct. Both clausal and nonclausal adjuncts were tested to determine whether the syntactic structure (as well as the phonological length) of a constituent



contributes to its weight (see earlier discussion and Martin, 1970). If so, then clausal examples (6a) may show more high attachment than nonclausal ones (6b) if readers provide default prosodies which tend to separate clausal elements by assigning them to their own intonational phrase (see Bader, 1998, for evidence of such assignment).

### **Method**

The sixteen sentences were presented in a questionnaire, each followed by the question “which one is true” and two paraphrases. One paraphrase stated the interpretation appropriate if the adjunct had been taken to modify the matrix verb (Susie learned something after John visited) while the other stated the embedded clause modification interpretation (Bill telephoned somebody after John visited). Experimental participants were instructed to read each sentence carefully and to indicate what they first thought the sentence meant. Two forms of the questionnaire were used, one containing the clausal adjunct version of half the sentences and the nominal adjunct version of the rest, with the other versions appearing on the other questionnaire. The questionnaire contained no other items. Forty-five participants completed the questionnaire, 23 with one form and 22 with the other.

### **Results and Discussion**

Forty-eight percent of the responses to sentences with clausal adjuncts (like (6a)) received a matrix modification (high attachment, “early closure”) interpretation. A significantly smaller 39% of the responses to sentences with nominal adjuncts (like (6b)) received a matrix interpretation ( $t_1(44) = 2.71, p < .01$ ;  $t_2(15) = 2.50, p < .02$ ).

The primary purpose of Experiment 1 was to establish a baseline for the auditory studies

that are presented next. The percentage of matrix interpretations of the adjunct seems rather high, given the widely-shared assumption of a preference for recency or local attachment (Frazier, 1978; Gibson, Pearlmutter, Canseco-Gonzalez, & Hickok, 1996; Philips & Gibson, 1998; Stevenson, 1994). However, it does indicate that our sentences are not excessively biased, and thus that they may be expected to respond to prosodic manipulations.

The relative preference for high (matrix) interpretations of adjunct phrases containing clauses (compared to nominal adjuncts) is of potential interest. It may reflect Fodor's (1998) claim that long prosodic constituents are preferentially analyzed as sisters of equally long prosodic constituents (her "balanced sisters" hypothesis, to which we return in Experiment 5). Alternatively, or in addition, it may reflect a greater tendency for a reader to impose a separate prosodic phrase on a syntactic phrase containing a clause than on a phrase containing only a nominal. Bader (1998) proposes that readers assign an intonational/prosodic structure as they read a sentence, using various strategies for assigning accents and prosodic boundaries, including the strategy of beginning a new intonational phrase whenever there is evidence of a clause boundary. Bader provides evidence that if readers must revise their initial syntactic analysis of a sentence and if the new syntactic analysis is inconsistent with the already-assigned intonational structure, reanalysis is difficult. The question of why readers assign an intonational structure during silent reading is an interesting one. Possibly readers must always assign an intonational structure because it is part of the grammar's well-formedness requirements for a sentence. Another possibility is that the intonational/prosodic structure aids maintaining the sentence in memory. We return to further discussion after Experiment 5.

## Experiment 2

As an idealization, prosodic theory treats prosodic boundaries as categorical (so that all ip's, for example, are equivalent). However, an ip or an IPh boundary can be realized in different ways. Experiment 2 tested an idea mentioned in the Introduction, namely that the acoustic prominence of a prosodic boundary, not its categorical identity, affects parsing decisions. The experiment compared sentences with different productions of an ip boundary that were intended to differ in prominence. The prosodic constituents were the same tonally in the different conditions, but differed in the duration of the word before the boundary and the presence of a pause before the boundary. Half of the 16 sentences contained clausal adjuncts (7a, c) and half contained nonclausal adjuncts (7b, d). The sentences appear in Appendix 2 and ToBI transcriptions can be found on the web page cited in the Authors' Notes.

(7) [ip long-ip]

a. Susie learned that Bill telephoned after John visited.

b. Sally learned that Pat telephoned after Tim's visit.

H\*    !H\*    L-                    H\*    L\*            L-                    H\*            L-    L%

[ip ip]

c. Susie learned that Bill telephoned after John visited.

d. Sally learned that Pat telephoned after Tim's visit.

H\*    !H\*    L-                    H\*    L\*            L-                    H\*            L-    L%

### Method

Materials. The experimental sentences were recorded by a phonetically-trained linguist, and phonologically and acoustically analyzed. Each experimental sentence was recorded with

two prosodies, each matching the ToBI transcriptions in (7). These prosodies differed only in the duration of the prosodic boundary preceding the adjunct, which was longer in the (a, b) cases. Eight items (the odd numbered items in Appendix 2) contained clausal adjuncts, and the remaining eight contained non-clausal adjuncts, but the prosodic representations were the same, as shown in (7). As can be seen in the Appendix, each clausal adjunct shared its verbs with one non-clausal item, but this pairing was ignored in conducting and analyzing the experiment.

\*\*\* TABLES 1 AND 2 ABOUT HERE \*\*\*

Selected duration and F0 measurements appear in Tables 1 and 2 (together with measurements from Experiments 3-5). After the first verb (learned), both prosodic conditions had ip boundaries of roughly equivalent size. Durations of the first verb averaged 537 and 516 ms for the (a, b) and (c, d) cases respectively, which was a slight but statistically significant variation ( $t(15) = 3.99$ ,  $p < .001$ ). In most sentences there was a pause following this verb, averaging 61 and 55 ms, which did not significantly differ between conditions. A low F0 target was reached at the end of the first verb or on the following word (that) in all sentences. This target averaged 190 Hz in all conditions, which was low but not at the bottom of the speaker's range. Comparison of the V1 low averages to the final L-L% F0 averages of 169-170 Hz shows significant differences (a, b, V1 vs. final:  $t(15) = 6.28$ ,  $p < .001$ ; c, d, V1 vs. final:  $t(15) = 5.52$ ,  $p < .001$ ); this supports the claim that the low on the first verb only indicated an ip boundary.

At the second verb (telephoned), the durational differences between the prosodic conditions were more marked. The (a, b) sentences had an average verb duration of 493 ms, compared to 411 ms for the (c, d) sentences, which was a significant difference ( $t(15) = 7.00$ ,  $p <$

.001). The duration of the following pause also differed, with the (a, b) sentences averaging a 286 ms break while the (c, d) sentences averaged only a 9 ms break ( $t(15) = 12.91, p < .001$ ). Tonally, however, both prosodic conditions were clearly consistent with an ip boundary. A low F0 target at the end of the second verb averaged 181 and 185 Hz for the (a, b) and (c, d) sentences, which was still well above the bottom of the speaker's range in the majority of sentences. This was confirmed by a significant difference from the low F0 target at the end of the sentences ((a, b), V2 vs. final:  $t(15) = 6.76, p < .001$ ; (c, d), V2 vs. final:  $t(15) = 9.01, p < .001$ ). The high F0 target on the noun (John) in the following adjunct clause showed a significant difference in the amount of pitch resetting in the two prosodic conditions. One might expect more pitch resetting after a larger or stronger boundary; consistent with that, the H\* average was 269 Hz in the (a, b) conditions, while the (c, d) conditions averaged only 230 Hz ( $t(15) = 11.71, p < .001$ ).

Participants and procedure. These 16 sentences were combined with 38 other sentences (including 16 described in Schafer, Carlson, Clifton, & Frazier, 2000, Experiment 2) and presented to 54 students at the University of Massachusetts, who participated in individual half-hour sessions for course credit. Sentences were played to participants individually in a sound-deadened chamber, using Radio Shack Minimus 7 speakers. After hearing seven practice sentences, participants heard the 54 sentences in the experiment presented by a computer in an individually-randomized order. They were instructed to pull a trigger as soon as they had understood the sentence. After they responded, a question appeared on a video screen. Participants were given two choices of what the sentence meant and were required to choose one. For instance, they might see What happened after John visited? together with two alternative

answers very similar to those used in Experiment 1 (e.g., Susie learned something after John visited and Bill telephoned somebody after John visited; the former paraphrases a matrix modification/high attachment analysis and the latter an embedded modification/low attachment analysis). The participant had to pull a trigger under the answer he or she considered correct. Half the matrix interpretation answers occurred on the left, and half on the right. The time taken to pull the trigger was measured from the onset of the question and the alternative answers. However, since the only reaction time effect that approached significance in any experiment was an effect of sentence length in Experiment 5 (simply reflecting the fact that longer sentences had longer questions, which took longer to read), reaction times will not be reported here.

### **Results.**

Listeners assigned a high attachment an average of 35% of the time, ranging from 34.7 to 35.0% for the four types of sentences illustrated in (7). No differences approached significance. The 95% confidence interval for a difference between any two means was 3.5% (treating items as the random variable), a difference less than half the size of any difference we report as significant in later experiments. The data strongly indicate that a phonetically longer ip boundary did not encourage high attachment of the adjunct. Further, the difference between clausal and nonclausal adjuncts observed in the questionnaire Experiment 1 was not observed here.

### **Discussion**

The failure to obtain any difference in interpretation among the sentences used in Experiment 2 means that substantial variation in the physical signal is not sufficient to affect interpretation in the absence of variation in the phonological category of prosodic boundaries.

We note, in anticipation of Experiments 3-5, that the magnitude of the difference in duration of the boundaries in Experiment 2 (82 ms in duration of the second verb, 275 ms in duration of the following pause) is closely comparable to durational differences that did affect interpretation in the later studies when these durational differences were associated with tonal differences (see Table 1). Since the proportions of high attachments observed in Experiment 2 was not large in comparison to what was observed in Experiment 1, Experiment 2 does suggest that the presence of a phonetically very clear ip break before an adjunct is not sufficient to make listeners attach it as a modifier of the matrix verb. However, it implies nothing about whether differences in the phonological category of prosodic boundaries can affect interpretation.

Experiment 2 failed to observe the greater preference for high attachment of clausal adjuncts compared to nonclausal adjuncts observed in Experiment 1. The clausal/nonclausal adjunct manipulation was not a primary focus of Experiment 2, so we manipulated it as a between-items effect. This could have reduced power, but the small size of the confidence interval of a difference between any two means (3.5%) indicates that the experiment had adequate power to detect small effects. An alternative, more interesting, possibility is that readers in the questionnaire study Experiment 1 imposed their own implicit prosodies on what they read and were more likely to place a clausal adjunct in its own prosodic phrase than a non-clausal adjunct (see Bader, 1998, for evidence that prosody constructed during silent reading can affect parsing). This could result in a relative preference for high attachment. In Experiment 2, listeners heard explicit prosodies, decreasing (but not necessarily eliminating, as we will argue after Experiment 5) the likelihood that they will provide their own preferred prosody.

Experiment 3 addressed the question of whether the categorical identity of prosodic boundaries could affect parsing. It contrasted prosodic boundaries that differed in their linguistic category (ip vs IPh), not simply in terms of acoustic salience.

### Experiment 3.

Experiment 3 examined sentences like those in (8), below. Examples (8a) through (8c) all have an IPh before the embedded clause, and differ in the category of the prosodic break before the adjunct: IPh, ip, or none. Example (8d) used a smaller ip boundary before both the embedded clause and the adjunct. Theories that claim that an IPh boundary most strongly disambiguates attachment (Nespor and Vogel, 1983; Price et al., 1991) predict a greater number of high attachments for (8a), with an IPh boundary before the adjunct, than the other conditions. Schafer (1997) did not discuss cases with multiple IPh boundaries or cases with mixed ip and IPh boundaries, as in (8a-c), though she claimed that IPh boundaries lead to further interpretive processing.<sup>4</sup> We will consider a hypothesis that is consistent with Schafer's proposal (cf. Schafer, Speer, Warren, & White, 2000) that the full prosodic analysis matters, not simply a local prosodic boundary. The hypothesis is that the size of a prosodic boundary relative to other relevant boundaries (in a sense to be made more precise later) determines its effectiveness. This hypothesis predicts that low attachment interpretations will be more frequent in (8b) and (8c) than in (8a) and (8d), because in the former conditions the final prosodic boundary is smaller than the earlier boundary, which is not the case in the latter conditions.

(8) a. [IPh IPh]

H\*                    L-H%                    H\*                    L-H%                    H\*                    L-L%



Susie learned that Bill telephoned after John visited.

b. [IPh ip]

H\*            L-H%            H\*    L\*            L-            H\*            L-L%

Susie learned that Bill telephoned after John visited.

c. [IPh 0/ip]

H\*            L-H%            H\*    !H\*            H\*            L-L%

Susie learned that Bill telephoned after John visited.

d. [ip ip]

H\*            !H\*!H-            H\*    L\*            L-            H\*            L-L%

Susie learned that Bill telephoned after John visited.

Experiment 3 used a different technique than the other experiments, an auditory unacceptability judgement task (see Clifton, Frazier, & Deevy, 1999, for a further description). Listeners were instructed to pull a trigger as soon as a sentence became unacceptable. When they found a sentence acceptable, a question was asked about it as in Experiment 2. We used this technique in the belief that it might encourage more careful and attentive listening to the sentences (and because we were searching for the detection of anomalies in some of the other sentences studied in the experiment).

### Method

Sixteen experimental sentences, all with clausal adjuncts (see Appendix 3) were recorded with the 4 prosodies exemplified in the ToBI transcriptions in (8), above. The relevant F0 and duration measurements appear in Tables 1 and 2. At the first verb, the (a-c) sentences all had IPh boundaries, while the (d) sentences had an ip boundary. First verb durations averaged 609 ms for

the (a-c) sentences, but only 469 ms for the (d) sentences (a significant contrast:  $F(1, 15) = 176.64, p < .001$ ). Following this verb was a long pause in the (a-c) sentences, averaging 285 ms, while the (d) sentences averaged only a 9 ms pause (another significant contrast:  $F(1, 15) = 736.29, p < .001$ ). A low F0 target (L-) was reached within the first verb in the (a-c) cases, averaging 171 Hz, followed by a continuation rise (H%) to an average of 199 Hz. The (d) sentences showed a downstepped high target on the verb followed by a slight dip on or after the verb, to an average of 236 Hz. This was analyzed as a !H\* pitch accent followed by a further downstepped phrase accent, !H-, though a case could be made for a relatively high L- phrase accent here. Given these tonal contours and the durational measurements noted above, it is reasonable to conclude that the (a-c) sentences had full IPh boundaries after the first verb while the (d) sentence had only an ip boundary.

At the second verb, there was some variability in the realization of the (c) sentences, but the tonal properties of the (a) sentences showed that they clearly had a larger boundary than the other three conditions. Durational measurements on the verb showed similar averages in the 500 ms range for all sentences but (c), which averaged only 394 ms, reflecting the fact that boundaries were consistently present in the other three conditions (a significant contrast:  $F(1, 15) = 50.49, p < .001$ ). Following this verb was a pause of 370 ms in the (a) case, 269 ms in (b), 7 ms in (c), and 315 ms in (d), with (c) again being the outlier. Thus the durational measurements showed that boundaries of some type were robustly realized in conditions (a, b, d), though durations were equivalent for the ip vs. IPh contrast, and that (c) showed minimal durational cues to a boundary. F0 measurements more clearly distinguished the four conditions. The (a) sentences reached a low

(L-) target of 155 Hz, followed by a rise to an average of 194 Hz, clearly signaling a L-H% continuation rise. The (b) and (d) sentences reached low targets of 185 and 175 Hz, respectively. These targets were significantly higher than the F0 averages for the final IPh boundary (L-L%) in both sentences, which averaged 152 and 151 Hz (b:  $t(15) = 21.97$ ,  $p < .001$ ; d:  $t(15) = 13.96$ ,  $p < .001$ ). This comparison suggests that the V2 measurements in (b) and (d) reflect L- phrase accents marking ip boundaries, rather than IPh boundaries. Finally, in the (c) sentences, the F0 dipped between H\* pitch accents on the preceding and following nouns (Bill and John). In some cases, a prominence was heard on V2, prompting an analysis including a downstepped !H\* pitch accent. Other sentences lacked a prominence at this point, and these were analyzed with !H- phrase accents (alternatively the F0 contours in these latter sentences might just involve downward pitch drifting, but in order to be cautious in our conclusions we raise the possibility that ip boundaries were present rather than no boundaries). F0 measurements averaged 214 Hz in this condition. Tonally, therefore, the (a) sentences are most consistent with an IPh boundary, and the (b, d) sentences are consistent with an ip boundary, while the (c) sentences vary between an ip boundary and no boundary.

These sixteen experimental sentences were embedded in a list of 102 items total, of which 28 were clearly ungrammatical (e.g., subject-verb number disagreement, missing particles or arguments, extra arguments, etc.) or semantically anomalous and another 20 which were marginally acceptable. Experimental participants were first presented with an acceptability pretest in which they indicated the acceptability of each of 10 sentences. The pretest contained a variety of ungrammatical and anomalous, as well as grammatical, items (but none similar to

those in (8)). The very rare disagreements between a participant's judgment and the experimenter's judgment were discussed with the participant. The participant, after a short practice session, then participated in an auditory acceptability judgment experiment. The participant was instructed to listen to a sentence played by the computer as in Experiment 2 and to press a right-hand key if the sentence ended acceptably, but to press a left-hand key if it became unacceptable. Pressing the left-hand key terminated presentation of the sentence. Pressing the right-hand key resulted in the presentation of a question about the meaning of the preceding sentence, as in Experiment 2.

Forty-eight University of Massachusetts undergraduates were tested in individual sessions, receiving individually-randomized lists. Four different counterbalancing conditions were used so that each participant saw four experimental sentences in each version, and so that over all participants, each sentence was tested equally often in each version.

### **Results.**

Between 96 and 98% of the sentences in each prosodic condition were classified as "acceptable" and questions were asked about them. The percentages of high attachment (matrix verb modification) answers to these questions appear in Table 3. A one-way analysis of variance indicated that the conditions differed significantly ( $F(3,141) = 4.56, p < .005$ ;  $F(3,45) = 3.41, p < .03$ ). Examination of the data provide no support for the prediction (Nespor and Vogel, 1983; Price et al., 1991) that condition (a) (IPh IPh) would stand out from the rest. However, there was clear evidence for the proposal that the relative size of the prosodic boundaries affects interpretation. Conditions (b) and (c), in which the boundary before the adjunct was smaller than

the boundary before the embedded sentence, showed the smallest frequency of matrix modification interpretations (averaging 14.4%). Conditions (a) and (d), in which the boundary before the adjunct was the same in size as the boundary before the embedded clause, showed a significantly higher (but still very low) frequency of matrix modification interpretations (averaging 22.8%;  $F1(1,47) = 15.26, p < .01$ ;  $F2(1,15) = 6.24, p < .03$ ).

\*\*\*TABLE 3 ABOUT HERE\*\*\*

## Discussion

The results encourage the idea that it is the relative size of breaks that matters. The frequency of matrix interpretations was higher when the boundary before the embedded clause was of the same phonological category as the boundary before the adjunct than when the earlier boundary was phonologically a more prominent category than the later one. The identity of the boundary before the adjunct did not matter except in relation to the earlier boundary. In particular, the presence of neither an IPh nor an ip boundary before the adjunct was sufficient by itself to favor a high attachment interpretation.

We propose an “informative break” hypothesis to account for the findings of Experiment 3. Given that an ip or even an IPh boundary can occur before the adjunct even when it unambiguously modifies the embedded verb, the presence of this boundary can be justified whether the intended attachment of the adjunct is high or low. If this is correct, then it may not be the phonological type of the boundary that matters so much as whether the boundary before the adjunct is an informative boundary -- one larger, or smaller, than the boundary assigned by the speaker before the embedded clause. In other words, given an early break of size X, the later

(pre-adjunct) break may be prosodically equivalent to X (uninformative), larger than X (favoring high attachment), or smaller than X (favoring low attachment). The breaks before the adjunct in (8a) and (8d) are relatively uninformative, since they are no larger than the earlier break.

However, the pre-adjunct breaks in (8b) and (8c) are informative, since they are smaller than the earlier break. They would signal the absence of an informative prosodic boundary before the adjunct, encouraging low attachment.

As a reviewer pointed out, the term informative has two senses. In the most general sense, a boundary is informative if it is larger or smaller than expected given the other information available to the perceiver, e.g., the other choices the speaker has made at choice points. In a more specific sense, appropriate to the conditions tested here, a boundary is "informative" if it is larger than or smaller than a boundary at another position (the beginning of the embedded clause in our examples). Although the present experiments focus on the specific sense of informative involving the relative size of two boundaries, we fully intend the more general sense and suspect that relative boundary size is just a special case of "informativity."

Experiments 4 and 5 tested the informative break hypothesis. Experiment 4 investigated the interpretation of the sentences used in Experiment 3 when they are prosodically structured with either an IPh or an ip immediately preceding the ambiguous adjunct. This manipulation was crossed with the presence vs. absence of an ip boundary between the matrix and the embedded clause. Thus, Experiment 4 could determine whether the absolute size of the boundary before the adjunct or its size relative to a possible earlier boundary guided interpretation of the adjunct. Experiments 4 and 5 returned to the simpler procedure used in Experiment 2. The Experiment 3

procedure yielded proportions of matrix modification interpretations that were so low that one could be concerned about a floor effect in which no manipulation could drive the proportion any lower. We do not think that this is the case in the Experiment 3 data since we did find a significant effect that could be attributed to the relative size of prosodic boundaries, but the concern could limit one's confidence that the Experiment 3 procedure is able to detect small prosodic effects.

#### Experiment 4

Following the logic described in Experiment 3, the Nespor and Vogel (1983) and Price et al. (1991) systems predict an essentially categorical distinction between examples where an IPh precedes the ambiguity (9a,c) which should strongly encourage high attachment, and examples where only an ip precedes the adjunct (9b,d). Kjelgaard and Speer's (1999) findings suggest that the presence of a prosodic boundary (ip or greater) before the adjunct will permit a high attachment in all cases and therefore no strong prosodic differences should emerge, although presumably the syntactic bias for recent attachments will come into play. The informative break hypothesis predicts a lower frequency of high attachments in (9b) than in (9a, c, d), since the break before the adjunct was larger than the earlier break in these latter three cases.

(9) a. [ip IPh]

Susie learned that Bill telephoned after John visited.

H\*                    L-                    H\*                    L-H%                    H\*                    L-L%

b. [ip ip]

Susie learned that Bill telephoned after John visited.

H\*    L\*    L-                    H\*                    L-                    H\*                    L-L%

c. [0 IPh]

Susie learned that Bill telephoned after John visited.

L\*      H\*                      H\*    L\*      L-H%              H\*              L-L%

d. [0 ip]

Susie learned that Bill telephoned after John visited.

L\*      H\*                      H\*    L\*              L-              H\*              L-L%

## Method

Materials. Four versions of the sixteen experimental sentences used in Experiment 3 were recorded together with 38 filler sentences and seven practice sentences. The experimental sentences appear in Appendix 4. The experimental sentences were recorded by a phonetically-trained linguist, and phonologically and acoustically analyzed. The four different prosodies used are shown in the ToBI transcriptions of (9). The sentences within any one condition differed mainly in the presence or absence of L\* pitch accents in the first or second verb phrase.

Relevant F0 and durational measurements appear in Tables 1 and 2. At the first verb, conditions (a-b) had ip boundaries, while the (c-d) sentences were constructed to avoid a boundary at this position. Specifically, the (c-d) conditions were purposely produced with an F0 contour that began with a fairly low prominence on the first noun, rising to a higher accent on the first verb and an even higher accent on the second noun, avoiding a dip in pitch that could be analyzed as a L-. In the (a-b) sentences, the duration of the first verb averaged 598 ms, while the (c-d) sentences averaged only 419 ms (a significant contrast,  $F(1,15) = 374.12$ ,  $p < .001$ ). Following the verb there was a noticeable pause in the (a-b) sentences, averaging 340 ms, while there was almost no pause in the (c-d) cases, averaging 7 ms (another significant contrast,



$F(1,15) = 1188.32, p < .001$ ). At the end of the first verb, conditions (a-b) reached a low F0 target (L-) averaging 194 Hz, while the (c-d) verbs had H\* pitch accents averaging 227 Hz. The (a-b) sentences showed lowering that failed to reach the bottom of the speaker's pitch range, as suggested by the comparison to the sentence-final low F0s averaging 160 Hz (a:  $t(15) = 19.39, p < .001$ ; b:  $t(15) = 23.94, p < .001$ ). Based on these tonal contours and durational measurements, it is reasonable to conclude that the (a-b) sentences had ip boundaries at this position while the (c-d) sentences avoided such boundaries.

At the second verb, F0 movements clearly supported the boundaries suggested in the ToBI transcriptions, but the durational measurements for the second verb did not show any clear patterns and thus did not reflect the boundary differences. The average duration of the second verb ranged from 512 to 549 ms, with the only significant contrast being the (a-b) sentences averaging slightly longer durations than the (c-d) sentences ( $F(1,15) = 12.20, p = .003$ ). This probably reflects the slightly slower pace of the (a-b) sentences relative to the (c-d) sentences, since the latter were hurried in order to avoid an early prosodic break. The (a-b) sentences showed significantly longer average durations at N1, V1, break 1, N2, and V2; the difference was most distinct at the first verb and pause, where the (a-b) sentences had a boundary and the (c-d) sentences did not, but the averages support an overall rate difference. Pauses after the verb averaged from 362-402 ms, with no significant differences again. These durational measurements are consistent with a prosodic boundary being present in all conditions. The F0 targets at the end of the second verb did differ between the conditions. The (a, c) sentences reached a L- target of 174 Hz and then rose to an average H% of 203 Hz, showing a continuation

rise. The (b, d) sentences reached a L- target averaging 191 Hz with no rise before the pause. Comparison of these low targets to the sentence-final low F0s shows that in all cases the lowering failed to reach the bottom of the speaker's range (a:  $t(15) = 14.54$ ,  $p < .001$ ; b:  $t(15) = 21.88$ ,  $p < .001$ ; c:  $t(15) = 9.02$ ,  $p < .001$ ; d:  $t(15) = 17.78$ ,  $p < .001$ ). The continuation rises in the (a, c) sentences support an analysis with IPh boundaries at this position, while the simple lowering in the (b, d) sentences supports an analysis with ip boundaries at this position. The high F0 target on the noun in the following adjunct clause (John) showed slight differences in the amount of pitch resetting in the different conditions. The (a,c) conditions averaged 250 Hz, while the (b,d) conditions averaged 235 Hz (a significant contrast:  $F(1, 15) = 39.10$ ,  $p < .001$ ), with the higher pitches following the larger boundaries as expected.

Four counterbalanced lists were constructed, so that each participant heard four sentences in each of the four prosodic conditions, and across lists, each sentence was tested once in each prosodic version.

Participants and Procedure. Forty-eight students at the University of Massachusetts participated in individual half-hour sessions for course credit. After hearing the seven practice sentences, they heard the 54 sentences in the experiment presented by a computer in an individually-randomized order. The procedures used were those described in Experiment 2.

## **Results.**

The mean proportions of response choices consistent with modifying the matrix appear in Table 4. The variation among the means is rather small, with an overall preference for low attachment. The numerical pattern of results is consistent only with the informative break

hypothesis: Only sentences like (9b), which had ip boundaries before both the embedded clause and the adjunct phrase, seemed to have a lower preference for high (matrix) attachment than the other conditions. While the interaction between presence vs. absence of the ip boundary before the embedded clause and the presence of an IPh vs an ip boundary before the adjunct phrase was nonsignificant ( $p > .30$  in 2 X 2 participants and items analyses), a contrast testing the informative break hypothesis (by comparing the ip-ip condition to the other three conditions) was significant ( $t_1(47) = 2.43, p < .02$ ;  $t_2(15) = 2.21, p < .05$ ).

\*\*\*TABLE 4 ABOUT HERE\*\*\*

## **Discussion**

Our data are most consistent with the informative break hypothesis. The pattern of differences is clearly inconsistent with any theory that claims an IPh boundary is required to strongly bias attachment. Any difference between the effects of an ip and an IPh boundary present in (9a) vs (9b) should also have appeared in (9c) vs (9d). The pattern of data is, however, expected under the informative break hypothesis: as long as the break before the adjunct is larger than the earlier break, the frequency of choosing the high attachment interpretation is (moderately) high. It becomes lower only when the break before the adjunct is no larger than the earlier break.

## **Experiment 5**

To further test the informative break hypothesis, we conducted one final study using sentences like those in (10) with six versions of each. Three versions (10a-c) had only an early IPh break and no break before the adjunct. Three versions (10d-f) had only an IPh break before

the adjunct but no early break. The (a) and (d) versions contained short adjunct phrases consisting of at most two words. Typically these were bare adverbs like today, yesterday, or this week. In two cases they contained prepositions: in June and on Friday. The (b) and (e) versions contained longer examples of the same adjunct. The (c) and (f) versions were similar to the 'clausal' adjuncts tested in the earlier experiments. The informative break hypothesis predicts more high attachments in (10d-f) than in (a-c).

(10)

a-c. [IPh 0]

Susie learned that Bill telephoned last night.

Susie learned that Bill telephoned extremely late last night.

Susie learned that Bill telephoned after John visited.

H\*            L-H%            L\*            H\*            H\*    H\* L-L%

d-f. [0 IPh]

Susie learned that Bill telephoned last night.

Susie learned that Bill telephoned extremely late last night.

Susie learned that Bill telephoned after John visited.

L\*            H\*            H\*            L-H%            H\*    H\* L-L%

These three types of adjunct were tested in each of the two prosodic conditions in order to return to the issue of whether the clausal status and/or the length of the adjunct influences attachment preferences and the interpretation of prosodic information.

Fodor's (1998) balanced sister hypothesis predicts that the length of an adjunct should be crucial to its analysis. This hypothesis claims that a listener (or reader) prefers to attach a phrase

as a sister to a phrase of similar prosodic weight. Assuming that lengthening the adjunct phonologically (or lengthening it and increasing its syntactic complexity) makes the adjunct heavier and noting that a higher attachment necessarily results in attachment to a heavier sister, there should be fewer high attachments in the phonologically and syntactically light (10a) and (10d) than in the other forms.

### **Method**

Eighteen sentences like (10) were constructed (see Appendix 5) and recorded with the two general prosodic contours exemplified in the ToBI transcriptions in (10). The detailed transcriptions available in the webpage cited in Footnote 4 reveal how these contours were adapted to the sentences of different lengths.

Relevant F0 and durational measurements appear in Tables 1 and 2. At the first verb, conditions (a-c) had IPh boundaries, while the (d-f) sentences were produced with a steadily rising contour up through the second noun to avoid boundaries at this position. Durational and tonal measurements support this distinction. The first verb of the sentence averaged 604 ms for the (a-c) conditions, but only 391 ms for the (d-f) conditions (a significant main effect of prosody:  $F(1, 17) = 487.36, p < .001$ ). Following this verb was a lengthy pause in the (a-c) cases, averaging 451 ms, while no such pause was found in the (d-f) sentences. In addition, F0 measurements on the first verb showed a robust L-H% contour in the (a-c) conditions, reaching an average L- of 187 Hz and rising to a H% of 243 Hz. The first verb in the (d-f) sentences showed no lowering, just a rise from the L\* (or relatively low H\*) pitch accent on the preceding noun to a H\* on the verb averaging 253 Hz. These measurements are most consistent with an IPh

boundary at the end of the first clause in the (a-c) conditions but not the (d-f) conditions.

At the second verb, conditions (d-f) now had IPh boundaries, while conditions (a-c) began a rising contour to avoid inadvertent L- boundaries. Durations of the second verb averaged 339 ms in the (a-c) sentences, but 517 ms in the (d-f) sentences (a significant main effect:  $F(1, 17) = 388.16, p < .001$ ). Following the second verb was a lengthy pause in the (d-f) cases, averaging 460 ms, while no pause was found in the (a-c) sentences. F0 measurements on the second verb showed a clear L-H% contour in the (d-f) sentences, reaching an average L- target of 184 Hz and rising to a H% of 230 Hz. The (a-c) sentences showed a rise from the L\* accent on N2 (Bill) averaging 200 Hz to a H\* on the second verb averaging 225 Hz. This pattern is most consistent with a lack of boundary at this position in (a-c).

F0 measurements from the adjunct clause showed further differences between the prosodic conditions. The (a-c) sentences continued the rising contour begun at the second noun of the sentences, which culminated in an extreme peak on the final word of the sentence (night or visited in this example). The average F0 maximum in conditions (a-c) was 297 Hz. The (d-f) conditions, on the other hand, contained rising contours which covered only the adjunct clause material, following the boundary at the second verb. These cases reached an average F0 peak of 277 Hz, which was significantly lower than the (a-c) value ( $F(1, 17) = 91.03, p < .001$ ). All conditions then fell to a L-L% final boundary, averaging 173 Hz. In some earlier experiments, the height of the main F0 peak in the adjunct clause was used as a measure of pitch resetting following a boundary (the larger the boundary, the higher the expected resetting). In this experiment, however, a rising F0 template was applied to group the adjunct clause with the

preceding material in (a-c) and applied only to the adjunct clause in (d-f). This template replaced the simple H\* accent following a boundary which was found in the earlier experiments, so in fact the higher peak was found here in the conditions with a more distant boundary (a-c).

The 18 experimental sentences were combined with 74 other sentences of a variety of forms (including sentences concerned with the interpretation of pronouns and of negation). Each sentence was followed by a question, tailored to the particular condition a sentence appeared in (e.g., the question would be What happened last night, What happened extremely late last night, or What happened after John visited for the three different adjunct conditions illustrated in (10)). Participants were to choose one of two answers corresponding to matrix or embedded verb interpretation. Half the matrix answers appeared on the left side of the screen, and half on the right, and the participant was to pull a trigger under the correct answer.

Six list forms were created, each containing three sentences in each of the six experimental conditions illustrated in (10). Lists were counterbalanced so that each of the 18 experimental sentences was tested in each condition in one list. Each of 48 University of Massachusetts undergraduates heard one list, in individually-randomized order, after listening to a short practice list. The procedure was as described in Experiment 2.

## Results

The mean proportions of matrix verb modification interpretations appear in Table 5. A 2 X 2 analysis of variance, with type of adjunct (short, long, clausal) and location of prosodic boundaries ([I<sub>Ph</sub> 0] [0 I<sub>Ph</sub>]) as factors indicated highly significant effects of each factor (adjunct type:  $F(2,94) = 5.42$ ,  $p < .01$ ;  $F(2,34) = 6.83$ ,  $p < .01$ ; prosody:  $F(1,47) = 21.41$ ,  $p < .001$ ;

$F(1,17) = 7.85, p < .01$ ), but no interaction ( $F < 1$ ). Substantially more matrix interpretations were given to sentences with [0 IPh] than with [IPh 0] boundary patterns, 42 vs 27%. Further, more matrix interpretations were given to sentences with clausal adjuncts (42%) than with nonclausal phrases, which did not differ between short and long versions (31 vs 30%).

\*\*\*TABLE 5 ABOUT HERE\*\*\*

## Discussion

As predicted by the informative break hypothesis, the (10d-f) examples (with [0 IPh] prosody) showed more high attachment (matrix verb modification interpretation) than (10a-c) ([IPh 0]). The manipulation used in Experiment 5, the maximum contrast possible, succeeded in providing clear evidence that the position of the largest prosodic break in a sentence strongly affects the interpretation of an adjunct phrase.

One might object that manipulating the position of a single prosodic boundary, as in Experiment 5, cannot provide strong support for the informative boundary hypothesis. Indeed, the results of Experiment 5, taken by themselves, would not uniquely support this hypothesis. However, the results of earlier experiments disconfirm any hypothesis based on just the presence, absence, or size of the local boundary preceding the adjunct. Thus, assuming that a valid hypothesis must be consistent with all the results, only the informative break hypothesis will suffice.

Experiment 5 failed to support the version of the balanced sisters hypothesis (Fodor, 1998) that claims that the phonological weight of the adjunct would affect interpretation. Conditions (10b) and (10e) did not result in an increased frequency of high attachments. However, the



experiment did show that the syntactic weight of the adjunct phrase affected interpretation.

There were more high attachments for the syntactically complex clausal adjuncts (10c) and (10f) than for their nonclausal counterparts. In the written questionnaire (Experiment 1), readers also distinguished between clausal and nonclausal adjuncts. They were more likely to attach “after John visited” high than “after John’s visit.” However, highly similar clausal vs. nonclausal adjuncts were not distinguished in the first auditory experiment, Experiment 2. This is perhaps not surprising. In the written study, readers must supply their own prosody and syntactic weight may influence how they prosodically phrased the sentence. In the auditory study, however, the listener was supplied with an actual prosodic/intonational structure and thus syntactic weight differences may have been minimized.

What is surprising, from this perspective, is the outcome of Experiment 5, an auditory study where syntactic weight did matter. In Experiment 5, clausal adjuncts received more high attachments than nonclausal ones regardless of whether the adjunct was preceded by an intonational phrase boundary. This may be because in Experiment 5 the contrast between clausal and nonclausal adjuncts was distinctly greater than in Experiments 1 and 2. In Experiments 1 and 2 both clausal and nonclausal adjuncts were PPs and in both cases the object of the preposition was generally syntactically and thematically complex (multiphrasal). In Experiment 5 the nonclausal adjuncts were typically bare NPs with no internal theta-roles (and the long nonclausal adjuncts generally simply added modifiers) whereas the clausal adjuncts were PPs with a multiphrasal (clausal) object.

Syntactic weight effects may arise because the perceiver organizes auditory input into

processing units which distinguish between bare NPs on the one hand, and clauses and thematically-complex phrases, on the other (see Carroll & Tanenhaus, 1975, for relevant evidence from click mislocation). This, together with the assumption that perceivers impose boundaries on the speech input even in cases where such boundaries are not present acoustically (e.g., Martin, 1967) would explain how a syntactic weight effect could emerge in Experiment 5. Regardless of whether this speculation ultimately proves to be correct, the present results clearly highlight the complexity of the weight effects observed in perceptual processing of speech.

### **General Discussion**

In the Introduction, we raised the question of why particular prosodic boundaries have the specific effects they do on the interpretation of sentences. The results of Experiment 2 suggest that variation in the sheer acoustic duration of a boundary in the absence of tonal differences does not guide interpretation.

However, the current results also argue against the view that particular phonological boundary types (IPh vs. ip) have invariant effects, and against the view that boundaries have a strictly local effect independent of prior context. Instead the results suggest that the interpretation of one prosodic boundary depends on the existence and relative size of other prosodic boundaries in the sentence.

The hypothesis we proposed may well characterize immediate, on-line effects of prosodic boundaries. However, our experiments were designed simply to explore how the absolute or relative sizes of prosodic boundaries constrain the interpretation of utterances, not to address the question of the real-time nature of the use of prosodic information. In fact, few techniques have

been developed that allow researchers to answer questions about the on-line use of prosody. Pynte and Prieur (1996) used a word-detection technique to study early effects of prosody on ambiguity resolution, but explicitly argued that their data on the effects of prosody on word-detection reaction time could equally well be interpreted in terms of on-line guidance effects or effects involving quick revision of an initial analysis. Marslen-Wilson, Tyler, Warren, Grenier, and Lee (1992), Warren, Grabe, & Nolan (1995) and Kjelgaard and Speer (1999, Experiments 3 and 4) used a cross-modal naming time task to avoid effects of off-line deliberation about the resolution of an ambiguity, and the Kjelgaard and Speer obtained closely-comparable results using end-of-sentence judgement tasks and cross-modal naming times. However, just as in the case of Pynte and Prieur (1996), the naming task could not resolve the locus of the effect of prosody any more finely than to indicate that it must have operated by shortly after the end of the ambiguously-attached phrase. Techniques with finer temporal resolution are needed before we can explore the time course of assigning a prosodic analysis to an utterance and using its consequences in interpreting the utterance.

In a study of natural speech production, Schafer, Speer, Warren, and White (2000) argued for a picture of the role of prosody in parsing that, like ours, emphasizes the importance of relative boundary size. They showed that speakers in a constrained game situation (move the X) mark clause boundaries in interestingly consistent ways. For example, given the utterances in (11), 91% of the early closure utterances (like 11b) had a stronger prosodic boundary following “moves” than following “square”.

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

- b. When that moves the square will encounter a cookie.

Ninety-six percent of the late closure utterances (11a) had a stronger prosodic boundary after “square” than after “moves.” However, there was also interesting variability within and across speakers. A variety of pitch accent, phrase accent and boundary tone combinations were used for the same syntactic structure. Schafer et al. indicate that (considering phrases like moves the square in (11)) there were 25 distinct intonational patterns for 35 early closure utterances and 22 distinct patterns for 48 late closure utterances. Nevertheless, the biggest prosodic boundary in the speaker's utterance generally corresponded to the largest syntactic break.

The variability in production precludes an absolute interpretation of any given type of prosodic boundary on the part of the perceiver. But it does not prevent the perceiver from using the prosodic structure the speaker encodes in the utterance. Schafer et al. presented the ambiguous portion of the naturally produced utterances to listeners in a forced-choice task, removing the disambiguating syntactic continuation. For utterances in which the largest prosodic break corresponded to the largest syntactic break, listeners were 85% correct in identifying the structure of the fragment, suggesting that listeners could use the relative size of breaks to disambiguate the utterance. Even in ambiguous prosody conditions, where the size of the breaks before and after square were comparable, listeners performed above chance (at roughly 65% correct). This result shows clearly that listeners were also attending to information beyond relative break size, though at present it remains unclear what that information was.

The Schafer et al. results and the results reported here reinforce each other. Even a single syntactic structure receives highly variable prosodic implementation. This variability precludes

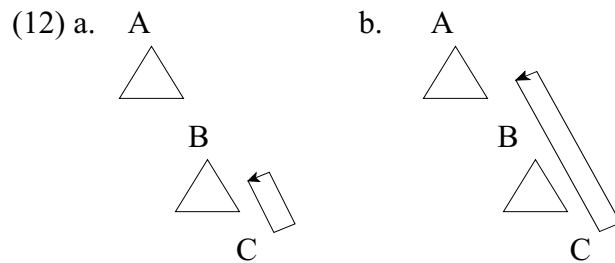
constant interpretation of some particular boundary type and precludes strictly local use of prosodic information. Instead it encourages the use of the full prosodic representation, as argued by Schafer (1997). With respect to prosodic boundaries, the focus of the present experiments, it entails that a given prosodic boundary may be more or less informative depending on the other prosodic choices the speaker has made.

Let us consider in more detail why particular prosodic boundaries have the effect they do in our experiments as well as in Schafer et al. (2000). On our view, part of the answer resides in the grammar and part in the way the grammar is exploited by speakers to convey their intentions. We assume that listeners use their grammar, including prosodic constraints, to assign a structure to a sentence. Therefore, if Structure A of an ambiguous input requires a prosodic boundary after the n'th word and Structure B does not allow a prosodic boundary after that word, prosodic disambiguation should occur, as argued by Nespor and Vogel (1986). However, in natural speech, there is considerable variety in the detailed prosodic structure observed for a sentence, and we suspect that the circumstances in which strict grammatical disambiguation occurs may be more limited than current phonological theory envisions.

Throughout this paper, we have also assumed that the prosodic hierarchy is part of the grammar. Therefore, an intonational phrase boundary, IPh, is larger than an intermediate phrase boundary, ip, which in turn is larger than the absence of a prosodic boundary. This too follows from the grammar, in this case, directly from the grammatical definition of prosodic constituents.

Turning to how speakers and their hearers exploit grammatical options, we have presented experimental evidence that optional boundaries are interpreted in a context-dependent fashion.

As illustrated in (12), given an ambiguous phrase C that may be analyzed as part of the preceding phrase B, as in (12a), or not, as in (12b), the informativity of the boundary preceding C depends on the size (phonological type) of the boundary before B.



If the pre-C boundary is larger than the pre-B boundary, this will encourage high attachment of C. If it is smaller, it will encourage low attachment. Boundaries of equal size will not be informative to the listener concerning the intended attachment site for C.

We will eventually have to address the question of the best way to specify size of boundary. To account for the present data, we may simply consider the phonological type of the boundaries surrounding B in (12). However, one can easily imagine that a more subtle articulation of the informativity hypothesis will ultimately be needed, one where the size of the boundary is a matter of how big it is given how big the listener would expect it to be based on the local syntax of the input and the 'emphaticness' of the speaker, e.g., how the speaker is instantiating unambiguous boundaries. What we think is behind the informativity hypothesis is whether a boundary is bigger/smaller than expected given the speaker's other decisions in formulating and producing an utterance.

Finally, we wish to emphasize that this view in no way presupposes that the speaker is explicitly attempting to help the listener. The speaker need only be self-consistent

in choosing among grammatically-permissible options.

We are now in a position to address a nagging question that we have left dangling so far. Why are the effects established in the present studies so small? We think the answer comes in two parts. First, we have investigated the effect of optional boundaries. By definition, the listener is choosing between two grammatical analyses of a sentence, because on either structure a prosodic boundary may but need not occur before the ambiguous adjunct clause. The listener is responding to the pattern of choices the speaker has made at choice points where the grammar does not dictate a particular decision, not a prosodic structure that renders one analysis of a sentence ungrammatical. Secondly, in the sentences investigated here (apart from Experiment 5), the adjuncts were always "heavy," either clausal or complete functional categories. Long or syntactically heavy constituents will increase the probability that the prosodic boundaries surrounding the heavy phrase are triggered by the internal properties of the phrase rather than by the relation of the phrase to the larger syntactic structure. This allows the listener to justify the existence of an optional prosodic boundary without assuming the prosodic boundary signals closure of the preceding phrase.

If our claims about relative boundary size are descriptively correct, the question arises as to why perceivers would exploit the informative break hypothesis. We think the answer is that listeners assume that speakers are rational, in the sense of being self-consistent: they will not capriciously mark one break as larger than another if they intend a structure/interpretation where the smaller prosodic break corresponds to the larger constituent structure boundary. To reflect this we could rename the informative break hypothesis the "rational speaker hypothesis."

Viewed this way, the hypothesis is potentially explanatory, though clearly in need of further development.

Is the rational speaker hypothesis incompatible with the other hypotheses considered above?

We think not. It is incompatible with strictly local use of isolated "prosodic cues" or context-independent use of boundary type. But it does not vitiate the insights behind the other systems discussed in the introduction. It does suggest that none of them is sufficient on its own as a characterization of the role of prosody in sentence processing.



## References

- Bader, M. (1998). Prosodic influences on reading syntactically ambiguous sentences. In J. Fodor & F. Ferreira (Eds.), *Reanalysis in sentence processing* (pp. 1-46). Dordrecht: Kluwer.
- Beckman, M. (1996). The parsing of prosody. *Language and Cognitive Processes*, 11, 17-67.
- Beckman, M. E., & Elam, G. A. (1997). Guidelines for ToBI labelling, version 3.0. Manuscript and accompanying speech materials. Columbus, OH: Ohio State University.
- 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*, 3, 255-309.
- Carroll, P. J., & Slowiaczek, M. L. (1987). Modes and modules: Multiple pathways to the language processor. In J. L. Garfield (Ed), *Modularity in sentence comprehension: Knowledge representation and natural language understanding* (pp. 221-248). Cambridge, MA: MIT Press.
- Carroll, J. M., & Tanenhaus, M., K. (1975). Functional clauses are the primary units of sentence segmentation. Bloomington, Indiana: Indiana University Linguistics Club.
- Clifton, C., Jr., Frazier, L., & Deevy, P. (1999). Feature manipulation in sentence comprehension. *Rivista di Linguistica*, 11, 11-39.
- Fodor, J. D. (1998). Learning to parse? *Journal of Psycholinguistic Research*, 27, 285-319.
- Frazier, L. (1978). On comprehending sentences: Syntactic parsing strategies. Unpublished doctoral dissertation, University of Connecticut.

- Frazier, L., Rayner, K. 1982. Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology* 14:178-210
- Gibson, E., Pearlmutter, N., Canseco-Gonzalez, E., & Hickok, G. (1996). Recency preference in the human sentence processing mechanism. *Cognition*, 59, 23-59.
- Hirschberg, J., & Pierrehumbert, J. (1986). The intonational structuring of discourse. Paper presented at the 24th Association of Computational Linguistics.
- Kjelgaard, M. M., & Speer, S. R. (1999). Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language*, 40, 153-194.
- Lehiste, I. (1973). Phonetic disambiguation of syntactic ambiguity. *Glossa*, 7, 107-122.
- Marcus, M., & Hindle, D. (1990). Description theory and intonation boundaries. In G. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational 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.
- Martin, J. G. (1967). Hesitations in the speaker's production and listeners' reproduction of utterances. Journal of Verbal Learning and Verbal Behavior, 6, 903-909.
- Martin, E. (1970). Toward an analysis of subjective phrase structure. *Psychological Bulletin*, 74, 153-166.

Nespor, M., & Vogel, I. 1983. Prosodic structure above the word. In A. Cutler & R. Ladd (Eds)

*Prosody: Models and Measurements* (pp 123-140). Berlin: Springer-Verlag.

Nespor, M. & Vogel, I. (1986). *Prosodic Phonology*. Foris: Dordrecht

Pierrehumbert, J., & Beckman, M. 1988. *Japanese tone structure* Cambridge, MA: MIT Press.

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.

Phillips, C., & Gibson, E. (1997). On the strength of the local attachment preference. *Journal of*

*psycholinguistic research*, 26(3), 323.

Pitrelli, J. F., Beckman, M. E., & Hirschberg, H. (1994). Evaluation of prosodic transcription

labeling reliability in the ToBI framework. *Proceedings of the 1994 International*

*Conference on Spoken Language Processing* (pp 123-126). Yokohama, Japan

Price, P. J., Ostendorf, M., Shattuck-Huffnagel, 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 processing.

Language and Cognitive Processes, 11, 165-192.

Schafer, A. (1997). *Prosodic parsing: The role of prosody in sentence comprehension*. PhD

Dissertation, University of Massachusetts, Amherst, MA: GLSA.

Schafer, A., Carlson, K., Clifton Jr., C., & Frazier, L. (2000). Focus and the interpretation of

pitch accent: Disambiguating embedded questions. Language and Speech, 43, 75-105.

Schafer, A. J., & Speer, S. R. (1997). The role of prosodic phrasing in sentence comprehension.

Paper presented at the CUNY Conference on Human Sentence Processing, Santa Monica, CA., March, 1997.

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

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

Shattuck-Hufnagel, S., & Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research*, 25, 193-248.

Silverman, K., Beckman, M. E., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., & Hirschberg, J. (1992). *ToBI: A standard for labeling English prosody*. Paper presented at the Second International Conference on Spoken Language Processing, Banff, Canada.

Slowiaczek, M. L. (1981). Prosodic units as language processing units. Unpublished doctoral dissertation, University of Massachusetts at Amherst, Amherst, Ma.

Speer, S. R., Kjelgaard, M. M., & Dobroth, K. M. (1996). The influence of prosodic structure on the resolution of temporary syntactic closure ambiguities. *Journal of Psycholinguistic Research*, 25, 249-272.

Stevenson, S. (1994). Competition and recency in a hybrid network model of syntactic dependencies. *Journal of Psycholinguistic Research*, 23, 295-322.

Warren, P., Grabe, E., & Nolan, F. (1995). Prosody, phonology and parsing in closure

Prosodic boundaries

August 30, 2000

ambiguities. *Language and Cognitive Processes*, **10**, 457-486.

## Appendix 1: Sentences Used in Questionnaire Experiment 1

1. Susie learned that Bill telephoned after John visited.  
Susie learned that Bill telephoned after John's visit.
2. Pat recorded whatever people said when Sally walked in the garden.  
Pat recorded whatever people said during Sally's walk in the garden.
3. Brian reported that Sam arrived before Tina returned.  
Brian reported that Sam arrived before Tina's return.
4. Emmon announced whatever the decision was immediately after the Dean arrived.  
Emmon announced whatever the decision was right after the Dean's arrival.
5. Ellen mentioned whatever was bothering her when Sarah visited.  
Ellen mentioned whatever was bothering her during Sarah's visit.
6. Tom discovered that Patty got sick when Lucy went to the Bahamas.  
Tom discovered that Patty got sick during Lucy's trip to the Bahamas.
7. Freida remembered whatever she said when Tom was around.  
Freida remembered whatever she said in Tom's presence.
8. Felix recorded that the mayor made a commitment when the Governor was around.  
Felix recorded that the mayor made a commitment in the Governor's presence.
9. Sally expected that Sam would leave when Martha returned.  
Sally expected that Sam would leave upon Martha's return.
10. Lucy voted for whatever Max proposed when the department met.  
Lucy voted for whatever Max proposed at a department meeting.
11. Melinda praised whatever Tony did when the grandparents were present.  
Melinda praised whatever Tony did in the grandparent's presence.
12. Martin maintained that the CEO lied when the investigation started.  
Martin maintained that the CEO lied at the start of the investigation.
13. Patricia found out that Lawrence disappeared when the war broke out.  
Patricia found out that Lawrence disappeared at the outbreak of the war.
14. Tom supported whatever Barbara recommended when a crisis was brewing.  
Tom supported whatever Barbara recommended during any potential crisis.
15. Karl approved whatever Rodney suggested when long-term research was planned.  
Karl approved whatever Rodney suggested during long-term research planning sessions.
16. Linda complained that the secretary was sick whenever an outside review was scheduled.  
Linda complained that the secretary was sick during every scheduled outside review.

## Appendix 2: Sentences Used in Experiment 2

1. Susie learned that Bill telephoned after John visited.
2. Sally learned that Pat telephoned after Tim's visit.
3. Emmon reported that Sue arrived before Amy returned.

4. Brian reported that Sam arrived before Tina's return.
5. Tom discovered that Patty got sick when Lucy went to the Bahamas.
6. Ann discovered that Ellen got sick during Sarah's trip to the Bahamas.
7. Frieda recorded that the major made a commitment when the President was around.
8. Felix recorded that the mayor made a commitment in the Governor's presence
9. Sally expected that Sam would leave when Martha returned.
10. Lucy expected that Max would leave upon Tony's return.
11. Melinda maintained that the Chairman lied when the questioning started.
12. Martin maintained that the CEO lied at the start of the investigation.
13. Patricia found out that Lawrence disappeared when the war broke out.
14. Alicia found out that Terrence disappeared at the outbreak of the war.
15. Rodney complained that the bookkeeper was sick whenever an outside audit was scheduled.
16. Linda complained that the secretary was sick during every scheduled outside review.

#### Appendix 3: Sentences Used in Experiment 3

1. Susie learned that Bill telephoned after John visited.
2. Sally learned that Pat telephoned after Tim visited.
3. Emmon reported that Sue arrived before Amy returned.
4. Brian reported that Sam arrived before Tina returned.
5. Tom discovered that Patty got sick when Lucy went to the Bahamas.
6. Ann discovered that Ellen got sick when Sarah went to the Bahamas.
7. Frieda recorded that the major made a commitment when the President was around.
8. Felix recorded that the mayor made a commitment when the Governor was around
9. Sally expected that Sam would leave when Martha returned.
10. Lucy expected that Max would leave when Tony returned.
11. Melinda maintained that the Chairman lied when the questioning started.
12. Martin maintained that the CEO lied when the investigation started.
13. Patricia found out that Lawrence disappeared when the war broke out.
14. Alicia found out that Terrence disappeared when the war broke out.
15. Rodney complained that the bookkeeper was sick whenever an outside audit was scheduled.
16. Linda complained that the secretary was sick whenever an outside review was scheduled.

#### Appendix 4: Sentences used in Experiment 4

- 1a. Susie learned that Bill telephoned after John visited.
2. Sally learned that Pat telephoned after Tim visited.
3. Emmon reported that Sue arrived before Amy returned.
- 4 Brian reported that Sam arrived before Tina returned.
5. Tom discovered that Patty got sick when Lucy went to the Bahamas.
6. Ann discovered that Ellen got sick when Sarah went to the Bahamas.
7. Frieda recorded that the major made a commitment when the President was around.
8. Felix recorded that the mayor made a commitment when the Governor was around.

9. Sally expected that Sam would leave when Martha returned.
10. Lucy expected that Max would leave when Tony returned.
11. Melinda maintained that the Chairman lied when the questioning started.
12. Martin maintained that the boss lied when the investigation started.
13. Patricia found out that Lawrence disappeared when the war broke out.
14. Alicia found out that Terrence disappeared when the war broke out.
15. Rodney complained that the bookkeeper was sick whenever an outside audit
- 16a. Linda complained that the secretary was sick whenever an outside review was scheduled.

## Appendix 5: Sentences used in Experiment 5

- 1a, d. Susie learned that Bill telephoned last night.
- 1b, e. Susie learned that Bill telephoned extremely late last night.
- 1c, f. Susie learned that Bill telephoned after John visited.
- 2a, d. Sally discovered Pat telephoned on Friday.
- 2b, e. Sally discovered Pat telephoned early on Friday.
- 2c, f. Sally discovered Pat telephoned when Susie arrived.
- 3a, d. Emmon reported Sam arrived today.
- 3b, e. Emmon reported Sam arrived very early today.
- 3c, f. Emmon reported Sam arrived before Amy returned.
- 4a, d. Brian concluded Tim had arrived this morning.
- 4b, e. Brian concluded Tim had arrived around seven this morning.
- 4c, f. Brian concluded Tim had arrived after Tina got angry.
- 5a, d. Tom discovered that Bill got sick today.
- 5b, e. Tom discovered that Bill got sick very early today.
- 5c, f. Tom discovered that Bill got sick after Sarah called up.
- 6a, d. Frieda recorded that the major made a commitment this week.
- 6b, e. Frieda recorded that the major made a commitment the week of Ivanov's visit.
- 6c, f. Frieda recorded that the major made a commitment when he met with the President.
- 7a, d. Felix found out the mayor received a bribe yesterday.
- 7b, e. Felix found out the mayor received a bribe the day of the Governor's call.
- 7c, f. Felix found out the mayor received a bribe right after the Governor left.
- 8a, d. Ann learned Patty was ill today.
- 8b, e. Ann learned Patty was ill in the middle of the day.
- 8c, f. Ann learned Patty was ill when Timothy was away.
- 9a, d. Sally began to expect Sam would leave in June.
- 9b, e. Sally began to expect Sam would leave midway through July.
- 9c, f. Sally began to expect Sam would leave when Marie showed up.
- 10a, d. Lucy expected Mark would give up last fall.
- 10b, e. Lucy expected Mark would give up starting late last autumn.
- 10c, f. Lucy expected Mark would give up before Tony returned.
- 11a, d. Melinda maintained the Chairman lied today.



- 11b, e. Melinda maintained the Chairman lied in the press conference today.
- 11c, f. Melinda maintained the Chairman lied when the press conference started.
- 12a, d. Martin maintained the CEO evaded the issue yesterday.
- 12b, e. Martin maintained the CEO evaded the issue the day of the investigation.
- 12c, f. Martin maintained the CEO evaded the issue when the investigation began.
- 13a, d. Patricia found out Laurence disappeared today.
- 13b, e. Patricia found out Laurence disappeared the day of the strike.
- 13c, f. Patricia found out Laurence disappeared when the war broke out.
- 14a, d. Alicia found out Terrence left Sunday.
- 14b, e. Alicia found out Terrence left very early Sunday.
- 14c, f. Alicia found out Terrence left when Carmelina called.
- 15a, d. Rodney complained that the bookkeeper cheated last year.
- 15b, e. Rodney complained that the bookkeeper cheated the year of the audit.
- 15c, f. Rodney complained that the bookkeeper cheated when the audit took place.
- 16a, d. Linda complained that the boss was mean today.
- 16b, e. Linda complained that the boss was mean from morning until night.
- 16c, f. Linda complained that the boss was mean when the auditors came.
- 17a, d. Carolyn complained that the babysitter was cranky last night.
- 17b, e. Carolyn complained that the babysitter was cranky the night of the party.
- 17c, f. Carolyn complained that the babysitter was cranky when the kids were tired.
- 18a, d. Jason concluded Tom went skiing last week.
- 18b, e. Jason concluded Tom went skiing the week of finals.
- 18c, f. Jason concluded Tom went skiing when class was over.

## FOOTNOTES

1. We are following Nespor & Vogel (1983), who state that displaced syntactic constituents, parentheticals, and nonrestrictive relative clauses also form intonational phrases. In Nespor & Vogel (1986), introduction of intonational phrase boundaries was even less constrained than in the 1983 statement.
2. We note that Price et al. (1991) mentioned that the relative size of prosodic break indices could be viewed as guiding the interpretation of their ambiguous sentences, but they did not develop this suggestion to contrast with the possibility that the location of the boundary between intonational phrases is what matters.
3. Our ToBI transcriptions do not present explicit break indices. Beckman and Hirschberg (1994) and Beckman and Elam (1997) describe the use of break indices 3 and 4 as redundant with the marking of phrase accents (H-, L-) and boundary tones (H%, L%) to indicate ip and IPh boundaries, respectively. Break index 2 is used to indicate a clear mismatch between tonal and durational indications primarily at the level of ip vs. no boundary. To avoid redundancy, only the tones will be shown in transcriptions, but durational information may be found in the tables of acoustic measurements.
4. Schafer (1997) showed that an IPh boundary, but not an ip boundary, encourages the processor to finish any incomplete interpretive work in a domain, rendering reanalysis expensive after an IPh boundary has been crossed. For example, if a lexically ambiguous word must be reanalyzed to its less frequent meaning after an IPh boundary, processing is more difficult than if reanalysis occurs after only an ip boundary has been crossed.

Table 1: Durations in ms of selected regions of the sentences used in Experiments 2-5. Standard deviations indicated

in parentheses.

Experiment	Condition	V1 ( <u>learned</u> )	Break 1	V2 ( <u>telephoned</u> )	Break 2
Expt 2	a/c: [ip long ip]	537 (105)	61 (54)	493 (157)	286 (83)
	b/d: [ip ip]	516 (95)	55 (48)	411 (153)	9 (20)
Expt 3	a: [IPh IPh]	619 (94)	267 (42)	510 (143)	370 (76)
	b: [IPh ip]	601 (90)	272 (39)	496 (142)	269 (80)
	c: [IPh 0/ip]	608 (93)	315 (82)	394 (131)	7 (21)
	d: [ip ip]	469 (88)	9 (23)	503 (151)	315 (57)
Expt 4	a: [ip IPh]	597 (90)	353 (58)	539 (103)	382 (58)
	b: [ip ip]	600 (78)	327 (53)	549 (124)	362 (55)
	c: [0 IPh]	424 (73)	8 (16)	530 (104)	402 (75)
	d: [0 ip]	414 (71)	6 (14)	512 (112)	373 (73)
Expt 5	a: [IPh 0], short	597 (65)	421 (83)	345 (96)	0
	b: [IPh 0], long	602 (84)	443 (113)	336 (95)	0
	c: [IPh 0], clausal	614 (94)	490 (99)	334 (91)	0
	d: [0 IPh], short	386 (72)	0	508 (99)	439 (69)
	e: [0 IPh], long	396 (72)	0	521 (93)	437 (75)
	f: [0 IPh], clausal	391 (68)	0	521 (90)	505 (104)

Table 2: Mean F<sub>0</sub> maxima and minima (and standard deviations), in Hz, in selected regions of the sentences used in Experiments 2-5. Separate high and low points indicated where appropriate, with phonological category (or categories, where sentences varied).

Experiment	Condition	V1 F0 targets		V2 F0 targets		N3 H*	Final L-L%
		low	high	low	high		
		L-		L-			
Expt 2	a, b: [ip long ip]	190 (11)		181 (6)		269 (21)	169 (5)
	c, d: [ip ip]	190 (14)		185 (6)		230 (13)	170 (5)
		L-, !H-	H%	L-, !H*	H%		
Expt 3	a: [IPh IPh]	171 (3)	201 (7)	155 (8)	194 (10)	261 (11)	150 (5)
	b: [IPh ip]	169 (5)	199 (7)	185 (6)		228 (7)	152 (4)
	c: [IPh 0/ip]	172 (6)	198 (6)	214 (11)		230 (6)	153 (3)
	d: [ip ip]	236 (7)		175 (6)		218 (10)	151 (7)
		L-	H*	L-	H%		
Expt 4	a: [ip IPh]	193 (9)		172 (5)	201 (9)	256 (11)	159 (5)
	b: [ip ip]	194 (7)		187 (4)		237 (11)	161 (5)
	c: [0 IPh]		229 (9)	176 (5)	205 (7)	243 (9)	161 (5)
	d: [0 ip]		225 (10)	196 (9)		233 (13)	159 (5)
		L-	H%, H*	L-	H*, H%		
Expt 5	a: [IPh 0], short	186 (8)	242 (17)		233 (9)	299 (18)	173 (6)
	b: [IPh 0], long	187 (8)	247 (17)		222 (9)	298 (16)	170 (6)
	c: [IPh 0], clausal	188 (6)	241 (13)		220 (9)	294 (13)	174 (11)
	d: [0 IPh], short		260 (11)	183 (5)	235 (21)	285 (16)	171 (6)
	e: [0 IPh], long		248 (10)	184 (7)	228 (16)	273 (14)	173 (7)
	f: [0 IPh], clausal		250 (11)	184 (5)	226 (15)	274 (12)	173 (10)

Table 3: Proportion of Matrix Modification Interpretations, Experiment 3

Condition	Proportion Matrix
A: [IPh IPh]	.21
B: [IPh ip]	.14
C: [IPh 0]	.15
D: [(ip) ip]	.25

Table 4: Proportion of Matrix Modification Interpretations, Experiment 4

Condition	Proportion Matrix
A: [ip IPh]	.38
B: [ip ip]	.30
C: [0 IPh]	.39
D: [0 ip]	.37

Table 5: Proportion of Matrix Modification Interpretations, Experiment 5

Form of Adjunct Phrase	Prosody	
	[IPh 0]	[0 IPh]
Short	.24	.38
Long	.22	.39
Clausal	.34	.49