

The relationship between intonational phrasing and syntactic structure in language production

Duane Watson

University of Rochester, Rochester, NY, USA

Edward Gibson

Massachusetts Institute of Technology, Cambridge, MA, USA

In this paper, we evaluate several theories of how syntactic/semantic structure influences the placement of intonational boundaries in language production (Cooper & Paccia-Cooper, 1980; Gee & Grosjean, 1983; Ferreira, 1988). Although the theories that we tested are shown to be quite successful, they are complex, and furthermore, they are incompatible with recent evidence for incrementality in sentence production. In light of these problems, we propose a simpler incremental model called the Left hand side/Right hand side Boundary hypothesis (LRB). According to this hypothesis, two factors that underlie the successful performance of the algorithms from the literature contribute to the likelihood of producing intonational boundaries at word boundaries: (1) the size of the recently completed syntactic constituent at a word boundary; and (2) the size of the upcoming syntactic constituent. These factors are further constrained by syntactic argument relationships. We demonstrate that the LRB performs as well as previous models with respect to the data from Experiment 1. In Experiment 2, we present evidence that the LRB outperforms previous

Correspondence should be addressed to Duane Watson, Meliora 494, Brain and Cognitive Sciences, University of Rochester, Rochester, NY 14627. Email: dwatson@bcs.rochester.edu

This project was supported by NSF grant BCS-0218605 'Intonational boundaries in sentence production and comprehension'. The first author was supported by NSF grant SES-0208484 'Postdoctoral Research Fellowship'. We would like to thank Erin Conwell and Melanie Goetz for help with data coding. We would also like to thank the following people for their comments: Mara Breen, Dan Grodner, Tessa Warren, Doug Rohde, Jennifer Arnold, Stefanie Shattuck-Hufnagel, Steven Pinker, Ken Wexler, Mike Tanenhaus, Charles Clifton, Lyn Frazier, Fernanda Ferreira, an anonymous reviewer, and audiences at the 2003 GLOW Workshop on Intonational Phonology and the 2001 CUNY Human Sentence Processing Conference.

models in certain instances. In Experiment 3, we demonstrate that the discourse status of relative clauses is an additional factor in intonational boundary placement.

INTRODUCTION

A problem for the fields of psychology and linguistics is understanding the role of prosodic information in communication. Understanding this role is a challenge because prosody can communicate many different types of information. For example, pitch accents and prosodic phrasing can communicate information such as discourse structure by signalling which information is given, new, or most salient (Bolinger, 1972; Dahan, Tanenhaus, & Chambers, 2002; Gussenhoven, 2002; Pierrehumbert & Hirschberg, 1990; Schafer, 1997; Selkirk, 1984; Steedman, 2000; Terken & Nootboom, 1987). At the same time, prosody can signal a sentence's syntactic structure, evidenced by listeners' use of prosody to determine the intended interpretation of syntactically ambiguous sentences (e.g. Kjelgaard & Speer, 1999; Marslen-Wilson, Tyler, Warren, Grenier, & Lee, 1992; Price, Ostendorf, Shattuck-Hufnagel, & Fong, 1991; Schafer, 1997; Snedecker & Trueswell, 2003; Speer, Kjelgaard, & Dobroth, 1996).

In this paper we focus on the relationship between intonational phrasing and syntactic structure in sentence production.¹ An intonational phrase is a prosodic unit of speech that contains at least one syllable that receives phrasal stress (pitch accent) and ends with a boundary tone (Pierrehumbert & Hirschberg, 1990). Intonational phrases are often, but not always, separated by pauses, and the final word of the phrase tends to be longer than the same word would be in a phrase-medial position. Below are possible intonational phrasings for the given sentences, where “//” delineate intonational phrase boundaries.

- (1) a. The professor who the students liked // taught the class.
- b. Tony // as you know // can't take care of himself
- c. Take out the trash // Carmela.
- d. It was possible to parse the example sentence // in several different ways
- e. It was possible // to parse the example sentence in several different ways

Intonational phrase boundaries can separate clauses (1a); surround asides (1b); precede vocatives (1c); and occur in more than one position in an utterance, (1d) and (1e) (Nespor & Vogel, 1986; Selkirk, 1978).

¹ Phonologists have classified intonational phrases into full intonational phrases and intermediate intonational phrases (Beckman & Pierrehumbert, 1986) although we will not make this distinction in this paper.

A variety of factors seem to determine where intonational boundaries can occur. Some are semantic and pragmatic in nature. Focused words and new words tend to have phrasal stress, and stressed words tend to mark the end of intonational phrases. Asides, non-restrictive modifiers, sentential adverbs, vocatives, and parentheticals also seem to require their own intonational phrase (Nespor & Vogel, 1987).

The syntactic structure of a sentence also probably plays a role in boundary placement (Selkirk, 1984, 1986; Nespor & Vogel, 1986; Truckenbrodt, 1999). In this paper we focus on the relationship between intonational phrasing and syntactic structure in sentence production.

Although we focus on intonational boundaries in language production, the scope of this work has some important implications for the role of intonational boundaries in comprehension. For instance, this research may shed some light on the hypothesis that some on-line reading effects may be indirectly due to the implicit construction of prosodic structure as a person reads a sentence (Fodor, 1998, 2002; Bader, 1998). The work presented here is also relevant to the large body of work demonstrating that intonational phrase boundaries can serve as cues to differing syntactic structures in language comprehension (see Cutler, Dahan, & Van Donselaar, 1997 for a review). Much of this work suggests that listeners use intonational boundaries as cues to syntactic structure. However, without an understanding of the relationship between intonational phrasing and syntactic structure in production, clear predictions of how boundaries signal syntactic structure to listeners cannot be made.

Below, we review three theories of the syntax-intonational phrasing relationship in production. The first two theories attempt to predict the *size* of a pause between words, rather than the likelihood of producing a pause between words. More recently, Ferreira (1993) observed that the length of a pause between words is driven by lower level properties of the phonology of the words in the local environment as well as whether an intonational boundary occurs at the location or not. Syntactic structure does not determine the size of a pause, although it may affect the probability of a boundary. In addition, as we mentioned above, pauses do not always occur at intonational phrase boundaries. Thus, we will consider variants of the algorithms that apply to predict the *likelihood* of producing intonational boundaries rather than the size of the pauses. To avoid confusion, the models below will be presented as theories of intonational boundary placement. We first summarise the algorithms, and then evaluate them afterwards.

First, Cooper and Paccia-Cooper (1980, CPC) hypothesised that intonational phrase boundaries are roughly correlated with the number of syntactic brackets at a word boundary: the more syntactic constituents that begin or end a phrase, the greater the likelihood of an intonational

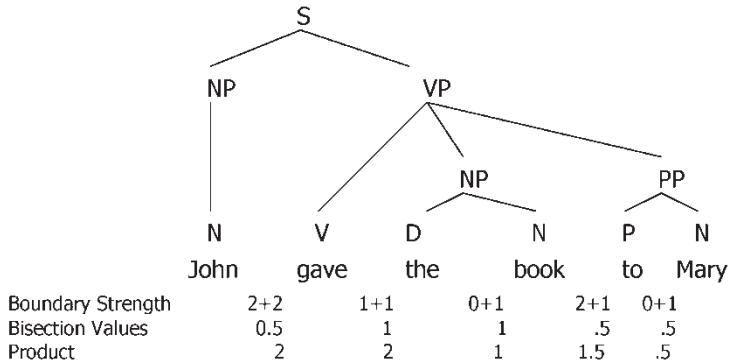


Figure 1. Values for boundary strength and bisection predicted by the CPC algorithm.

boundary. CPC also hypothesised that boundaries occur at the ends of syntactic constituents more than at the beginnings, so right brackets are weighted more heavily than left brackets. To calculate the likelihood of an intonational boundary, the nodes that dominate the words on the left and the right side of the word boundary, but not both, are counted. Nodes that refer to minor categories are not counted, nor are the non-terminal nodes on the left side of the tree that do not branch. The number of nodes on the left side is multiplied by 2. In Figure 1, the boundary between “book” and “to” has a boundary strength value of 3. The dominating nodes on the left are NP and N. Because NP is nonbranching, only N is counted, yielding a value of 2 when the total for the left side is doubled.² On the right side, the dominating nodes are PP and P. P is not counted, so the total value is 1.

The bisection values are calculated by first counting the number of category words in the sentence and dividing them by 2. This number is divided into the number of major category words to the word boundary from either the beginning of the sentence or the end of the sentence, whichever is closest. At the boundary between “book” and “to”, this value is 0.5 ($=1/(4/2)$). The product of the boundary strength and the bisection component is the relative likelihood of the intonational boundary. The algorithm contains further procedures for weighting speech rate and sentence length. Because Gee and Grosjean (1983, GG) found that these components do not improve the performance of the algorithm, we do not present or consider them here.

Second, GG proposed an eight-step algorithm in order to predict intonational boundaries: (1) the sentence is segmented into phonological phrases, (2) for each phonological phrase, a strictly right-branching tree is

² NP does branch into a D and an N, but the D consists of a minor category and is not counted.

formed from the words in the phonological phrase, independent of the syntax of these words; (3) phonological phrase trees that form NPs or Ss are then combined into a right branching tree that extends over the constituents; (4) the phonological phrase tree containing the main verb of the sentence is then adjoined to either the material before or after it depending on which has the fewest phonological phrases; (5) all the subtrees constructed thus far are then adjoined into a left branching tree; (6) the likelihood of an intonational phrase boundary at a word boundary is then calculated by counting the number of nodes dominated by the lowest node in the tree that dominates all of the words on both sides of the boundary; (7) and (8) the word boundary before the final phonological phrase in the sentence and the boundaries before words with two or more feet are weighted by increasing the boundary strength value by 1.

The tree in Figure 2 was constructed by applying the steps of the GG algorithm to the sentence "John gave the book to Mary". The word boundary between "book" and "to" is predicted to be the most likely location for an intonational boundary to occur because it receives a value of 8.

Third, Ferreira (1988) hypothesised that both syntactic structure and semantic structure play a role in intonational phrasing. Ferreira hypothesised that breaks occur in sentences such that the resulting units are as semantically coherent as possible, defining coherence as having a minimal number of dependencies across units. Ferreira observed that this type of semantic/prosodic organisation is advantageous to both the speaker

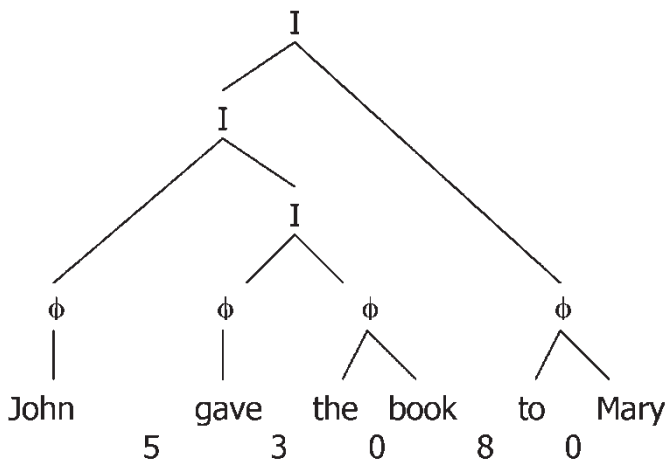


Figure 2. Intonational boundary likelihood quantities predicted by the Gee and Grosjean model. The θ nodes indicate phonological phrases and the "I" nodes indicate larger phonological units.

and the listener because semantically coherent objects are easier to maintain in working memory, and preserving the semantic coherence within an intonational phrase facilitates comprehension for the listener. Ferreira's specific proposal is based on X-bar theory (Jackendoff, 1977). The semantic coherence of adjacent constituents is approximated by making generalisations about the configurations of syntactic trees, which represent argument and modifier relationships explicitly. She hypothesises that the higher two units attach within the tree, the greater their semantic independence, and the more likely they are to be separated by an intonational phrase. The Ferreira algorithm works by scanning the tree to locate the lowest node in the X-bar tree that dominates each word boundary. Once this node is identified, its immediate daughters are matched to a syntactic category template. There are nine syntactic pair templates, each associated with a numerical ranking. Pairs that are relatively low in the tree, such as a pair of heads X^o and Y^o , receive a lower rank than pairs that occurred relatively high in the tree, such as an XP and a YP. Each word boundary receives the ranking associated with its matched template. These rankings are then used to predict the relative likelihood of an intonational boundary.

Figure 3 contains the syntactic structure assumed by Ferreira's algorithm along with the ranking of syntactic templates.³ This model predicts that the word boundary between "John" and "gave" is the most likely location for a boundary. The node that dominates both words at the boundary is IP, and its immediate daughters are NP and I'. These nodes map on to X' XP, which has a ranking of 7. Because this ranking has the highest value in the sentence, this location is the most likely location for a boundary.

Finally, there are two purely linguistic approaches to intonational boundary placement that are worth mentioning. One is an Optimality Theoretic constraint proposed by Selkirk (1986, 2000) called Align XP. She proposes that in English, phonological phrases should align with the right edge of XPs that are headed by lexical elements. The second proposal is Truckenbrodt's (1999) Wrap constraint. Roughly speaking, Wrap favours prosodic phrasings that do not break up syntactic constituents over those that do. This constraint does not apply to adjuncts or clauses. Although these constraints were designed to predict the grammaticality of different prosodic phrasings rather than their relative likelihood, they can also be thought of as performance constraints. As performance constraints, the two hypotheses can be summarised as preferences to place intonational boundaries at syntactic boundaries.

³ In the version of X-bar theory used in the Ferreira algorithm, ternary branching is permitted when there is more than one argument of a head. In this case, "the book" and "to Mary" are both arguments of "gave", so both arguments are sisters of the verb.

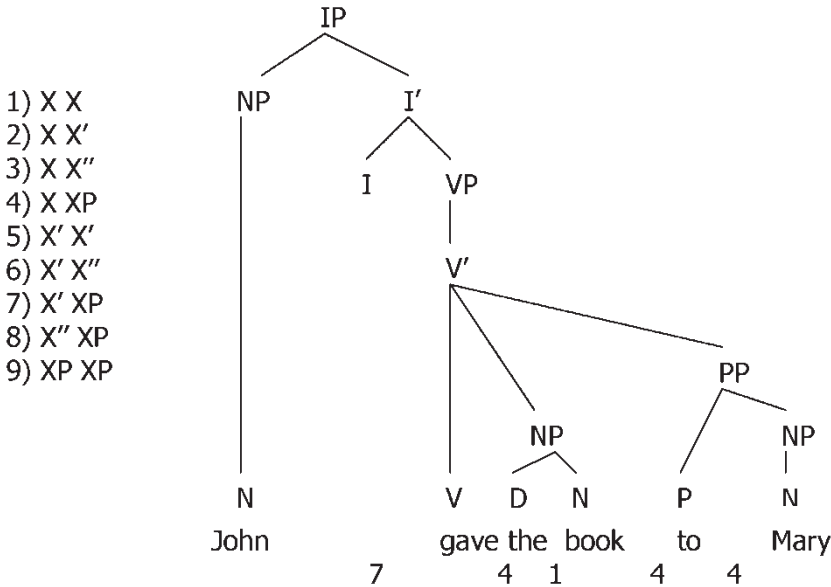


Figure 3. The X-bar representation used in the Ferreira algorithm.

There are strong parallels between the predictions of these two constraints and those of CPC's model. The CPC algorithm hypothesises that the number of constituent boundaries at a word boundary correlates with the likelihood of an intonational phrase. This is a performance-based version of Wrap. In particular, word boundaries with relatively few brackets are more likely to occur in the middle of a syntactic constituent than those with greater numbers of brackets. Placing intonational boundaries at these latter locations are more likely to satisfy the Wrap constraint. In addition, the CPC algorithm captures the Align constraint by weighting right brackets more heavily than left brackets. Furthermore, CPC's requirement that only nodes of constituents headed by lexical items are included in the boundary strength count corresponds to the Align constraint's lexical restriction. Thus, the CPC model embodies Selkirk's and Truckenbrodt's claims that (1) speakers favour placing intonational phrase boundaries at the boundaries of XPs and (2) speakers favour placing intonational boundaries at the right edge of XPs.

GG, CPC, and Ferreira's algorithms are complex, consisting of a number of steps that are potentially difficult to isolate. But before we discuss the relative complexity of the proposals, along with potentially simpler alternatives, we discuss the results of an experiment that was designed to empirically evaluate these theories. Despite the complexity and detail in the theories, they have not previously been evaluated using modern

psycholinguistic rigour, partly because analysing speakers' utterances is both time consuming and difficult. Indeed, CPC never tested their proposal empirically. GG tested 14 sentence types, but with only one token for each type. Furthermore only six participants took part in their experiment. Ferreira tested 10 sentence types, but again only one token per type, and only four participants took part in her experiment. In the latter two studies, only one token of each type was tested, so it is difficult to know whether the findings generalise across multiple tokens. As a result of this gap in the literature, we performed Experiment 1.

EXPERIMENT 1

Eight different structural types were evaluated in Experiment 1, having the format in (2a)–(2h). The intonational boundary predictions of each algorithm are shown in Table 1a and 1b.

- (2) a. Subject-extracted relative clause (RC)
The judge who ignored the reporter fired the secretary.
- b. Object-extracted RC
The judge who the reporter ignored fired the secretary.
- c. Object-extracted RC plus one NP
The judge who the reporter for the newspaper ignored fired the secretary.
- d. Object-extracted RC plus two NPs
The judge who the reporter for the newspaper in the capital ignored fired the secretary.
- e. Subject-extracted RC within an object extracted RC
The judge who the reporter who attacked the senator ignored fired the secretary.
- f. Object-extracted RC within another object extracted RC
The judge who the reporter who the senator attacked ignored fired the secretary.
- g. Two right branching RCs
The reporter ignored the judge who fired the secretary.
- h. Three right branching RCs
The senator attacked the reporter who ignored the judge who fired the secretary.

The sentence types that we tested varied in terms of their length and their syntactic complexity. In particular, we had two- and three-clause sentences in right-branching and nested configurations. The increasingly long and complex sentences were included because these sentences types would likely be produced with boundaries by most people. Thus they would provide many instances to test the predictions of the three theories. The

TABLE 1a

The numeric predictions of Gee and Grosjean's (1983) Phi algorithm (GG), Ferreira's (1988) X-bar algorithm, and Cooper and Paccia-Cooper's (1980) algorithm (CPC) for sentences (2a)–(2d). Each value is the prediction for the word boundary following the word in the corresponding column. Higher numbers indicate greater likelihood of placing an intonational boundary at that point.

(2a)	The	judge	who	ignored	the	reporter	fired	the						
GG	0	5	0	3	0	9	4	0						
Ferreira	3	8	2	4	1	7	4	1						
CPC	0	1.6	2.8	2.4	.8	9.6	1.2	.4						
(2b)	The	judge	who	the	reporter	ignored	fired	the						
GG	0	5	1	0	3	9	4	0						
Ferreira	3	8	2	1	7	7	4	1						
CPC	0	1.6	1.2	.4	3.2	12	1.2	.4						
(2c)	The	judge	who	the	reporter	for	the	newspaper	ignored	fired	the			
GG	0	7	1	0	3	1	0	5	11	4	0			
Ferreira	1	8	2	2	4	4	1	7	7	4	1			
CPC	0	1.98	1.32	.33	3.3	.66	.66	8	11.22	.99	.33			
(2d)	The	judge	who	the	reporter	for	the	newspaper	in	the	capitol	ignored	fired	the
GG	0	8	1	0	5	1	0	4	1	0	7	13	4	0
Ferreira	1	8	2	2	4	4	2	4	4	1	7	7	4	1
CPC	0	1.74	1.16	.29	2.85	1.14	.57	4.3	.86	.86	8.6	8.55	.87	.29

TABLE 1b

The numeric predictions of Gee and Grosjean's (1983) Phi algorithm (GG), Ferreira's (1988) X-bar algorithm, and Cooper and Paccia-Cooper's (1980) algorithm (CPC) for sentences (2e)–(2h).

(2e)	The	judge	who	the	reporter	who	attacked	the	senator	ignored	fired	the
GG	0	3	1	0	7	1	0	3	9	13	4	0
Ferreira	3	8	2	3	8	2	4	1	7	7	4	1
CPC	0	1.74	1.16	.29	3.42	3.99	2.58	.86	13.76	10.83	.87	.29
(2f)	The	judge	who	the	reporter	who	the	sentator	attacked	ignored	fired	the
GG	0	3	1	0	7	1	0	3	9	13	4	0
Ferreira	3	8	2	3	8	2	1	7	7	7	4	1
CPC	0	1.74	1.16	.25	2.28	1.71	.57	3.44	16.34	10.83	.87	.29
(2g)	The	reporter	ignored	the	judge	who	fired	the				
GG	0	5	3	0	9	0	4	0				
Ferreira	1	7	4	3	8	3	4	1				
CPC	0	1.6	3.2	.8	3.2	2.4	1.2	.4				
(2h)	The	senator	attacked	the	reporter	who	ignored	the	judge	who	fired	the
GG	0	5	3	0	9	0	3	0	13	0	4	0
Ferreira	1	7	4	3	8	2	4	3	8	2	4	1
CPC	0	1.16	2.28	.57	3.44	2.58	3.44	1.72	2.28	1.71	.87	.29

structures that we tested were not designed to differentiate among the theories, although the theories did make differential predictions on the structures to some degree. One of the challenges of experimental work in sentence production is eliciting controlled syntactic constructions, while at the same time getting the speaker to use production processes as naturally as possible. Like many researchers in sentence production (Cooper & Paccia-Cooper, 1980; Ferreira, 1988; Gee & Grosjean, 1983), we use a reading task, despite some of the obvious drawbacks, namely, that intonational phrasing in reading and spontaneous speech might differ. Still, some work has shown that the prosodic patterns in read and spontaneous speech are prosodically phrased in similar ways (Blaauw, 1994; Ferreira, 1991), and it is likely that the processes used in natural speech production and in reading aloud highly overlap. Therefore, although a more natural speech task would have been desirable, a reading task is used as a first approximation in order to allow for experimental control.

With these concerns in mind, a novel paradigm was used to elicit the test sentences from the speaker. Each trial was conducted with two participants: a listener and a reader. The reader was given a written list of test sentences and was instructed to read each sentence silently to herself so that she fully understood the meaning of the sentence before producing it out loud. This was to ensure that the production was made with knowledge of the information in the entire sentence, including material at the end. After the reader produced the sentence, the listener was presented with a question on the computer that she had to answer. This task encouraged the reader to communicate the content of the sentence in as natural a manner as possible.

Method

Participants. Sixteen pairs of native English speakers from the MIT community participated in the study for \$5.00 each. One participant was recorded while reading sentences and the other answered visually presented questions about the sentences which appeared on a computer after the speaker had finished the utterance.

Materials. Thirty-two tokens of each of the eight syntactic constructions in (2a)–(2h) were tested. The items were constructed so that their syntactic structures were as in (2a)–(2h) up to the final verb of the sentence (e.g., ‘fired’ in the examples above). For some items, there were small variations between items beginning at this point. The final verb sometimes varied in tense and aspect, but always constituted exactly one phonological phrase (as defined by GG and Nespor & Vogel, 1986). There was also some variation in the material following the verb. In 25 of the items, one

phonological phrase followed the final verb, whereas in the other 7, the verb was followed by two phonological phrases. The syntactic structures of these continuations varied from item to item. To ensure that any correlations between the models' predictions were not due to this variation, a regression was done on a subset of seven items that had identical syntactic structures. The results for all the items grouped together were the same as those for the items with identical syntactic structures.

The materials were divided into eight counterbalanced lists in a Latin Square design, so that each subject saw only one condition for each item and saw four tokens of each of the eight syntactic types. Each list contained a random ordering of the 32 items. There were no filler trials in this initial experiment, partly because of the variety of different structural types. To avoid effects of item position, a second set of eight lists was created by switching the position of the first 16 and final 16 items in each list, resulting in a total of 16 lists.

Procedure and analysis. Two participants were included in each trial. One participant, the reader, was presented with a list of sentences to read to the second participant, the listener. The reader was instructed to read the sentence silently to herself until she was sure that she could produce the sentence correctly. She then read the sentence out loud exactly once. The reader's speech was recorded, and the recordings were digitised at a 16Khz rate.

The listener was engaged in a yes/no question-answering task on a computer. While listening to the sentence produced by the reader, the listener was presented with a blank white screen on a computer screen. After hearing the sentence, the listener pressed the space bar, and a question about the sentence was presented. The listener pressed keys labelled 'yes' and 'no' to indicate their response. The reader did not see the question that was presented, and no feedback was given to the listener or reader about whether the question was answered correctly.⁴

Each production was transcribed by three coders (one of whom was the first author) for intonational breaks using a subset of the ToBI coding system (Silverman et al., 1992), similar to the transcription system used by Price et al. (1991). The strength of a boundary was marked by each of the coders using the following break indices: 4 – intonational phrase boundary, 3 – intermediate phrase boundary, 2 – a pause with no tonal movement

⁴ Because the goal of this experiment was understanding where speakers place intonational boundaries rather than how boundaries affect listener comprehension, we did not systematically vary the contents of the questions. However, question accuracy data were recorded and are presented below.

(often with a P to indicate a hesitation pause), 1 – a normal word boundary, and D – disfluency, such as a word repetition. Intonational and intermediate phrase boundaries were collapsed in the analysis below. Reliability between coders was calculated over the collapsed intermediate and intonational phrases, and was measured by calculating the proportion of the instances two particular transcribers agreed on the label of a word boundary using the method described in Pitrelli, Beckman, & Hirschberg (1994). The reliability for the entire data set was 94%. Since the level of reliability was high, the entire data set coded by the first author is used.

We conducted analyses similar to Gee & Grosjean (1983) and Ferreira (1988). Regression analysis were carried out for each of the three algorithms in order to see how well they performed in predicting the likelihood of intonational boundaries at word boundaries.

Results. Overall, listeners were accurate in the question answering task, answering 81% of the questions correctly overall. In an analysis of all 512 trials in the experiment (32 items \times 16 participants), all three algorithms accounted for a highly significant quantity of variance in predicting the likelihood of an intonational boundary ($N = 85$ for each; GG: $r^2 = .76, p < .0001$; Ferreira: $r^2 = .72, p < .0001$; and CPC: $r^2 = .39, p < .0001$). In an analysis of the differences between the correlations (Olkin, 1967), the GG and Ferreira algorithms performed significantly better than the CPC algorithm $t(1, 82) = 6.08, p < .001$ and $t(1, 82) = -4.15, p < .001$ respectively. GG and Ferreira were not significantly different from each other, $t(1, 82) = .84, p = .20$. A second analysis excluded trials with disfluencies (37 trials) or hesitation pauses (47 trials), or whose comprehension questions were answered incorrectly (95 trials). This exclusion yielded a total of 340 out of 512 trials to be analysed. (A trial may have met one or more of these criteria to be excluded.) The results were very similar with respect to this data set ($N = 85$ for each; GG: $r^2 = .75, p < .0001$; Ferreira: $r^2 = .71, p < .0001$; and CPC: $r^2 = .39, p < .0001$). The variance accounted for by algorithm for each sentence type under this analysis is listed in Table 2. The results were also similar when a regression was performed on the seven items that had identical syntactic structures in all critical regions ($N = 85$ for each; all trials included: GG: $r^2 = .70, p < .0001$; Ferreira: $r^2 = .70, p < .0001$; and CPC: $r^2 = .38, p < .0001$; only correctly answered and non-disfluent trials: GG: $r^2 = .68, p < .0001$; Ferreira: $r^2 = .70, p < .0001$; and CPC: $r^2 = .38, p < .0001$).

Discussion

The results from Experiment 1 demonstrate that the three algorithms are successful at predicting intonational phrase boundaries across the range of

TABLE 2

The variance accounted for by Ferreira's (1988) X-bar model, Gee and Grosjean's (1983) Phi algorithm (GG), and Cooper and Paccia-Cooper's (1980) algorithm (CPC), for each sentence type in Experiment 1, under an analysis of correctly answered and non-disfluent trials.

<i>Sentence in Experiment 1</i>	<i>Algorithm R²</i>			<i>Question accuracy</i>
	<i>Ferreira</i>	<i>GG</i>	<i>CPC</i>	
(2a)	.74	.79	.63	87%
(2b)	.58	.77	.56	92%
(2c)	.57	.81	.45	82%
(2d)	.70	.86	.49	79%
(2e)	.84	.64	.37	71%
(2f)	.72	.67	.39	60%
(2g)	.78	.84	.34 ^a	98%
(2h)	.79	.82	.19 ^a	83%
Overall	.71	.75	.39	81%

^a indicates non-significance.

structures in the target stimuli. The GG and Ferreira algorithms work very well overall, accounting for over 70% of the variance in predicting where people place boundaries. The performance of the CPC model, and by extension, the linguistic models that make predictions based on constituent boundaries alone, was less impressive, accounting for less than 40% of the variance in predicting where speakers place boundaries.

Despite the success of the GG and Ferreira algorithms, it is still unclear why they are as successful as they are. There are two general difficulties with evaluating these algorithms. First, the algorithms are very complex, especially the GG and CPC algorithms. Thus it is hard to isolate the effects of any of the steps in the algorithms. Second, the algorithms are not incremental, which is inconsistent with what we currently know about how human sentence production operates (Ferreira, 1996, Levelt, 1989; Roelofs, 1998). Although the degree to which the production system is incremental is controversial, it is generally agreed that speakers do not plan complete sentences before they start speaking (e.g., Bock & Levelt, 1994). Both the Ferreira and CPC algorithms require that the speaker have complete knowledge of the sentence that she is going to produce, in order for the algorithms to apply.⁵ It is not clear whether these algorithms can

⁵ The task in Experiment 1—spoken production of a sentence that is known ahead of time—makes it possible for the human sentence production mechanism to perform non-incrementally in this task. It is therefore logically possible that the non-incremental nature of the task might induce a task-specific production strategy which is fundamentally different from normal sentence production. But by Occam's razor, we assume that this is unlikely.

operate with limited look ahead. GG claim that their model can be applied incrementally with only limited look ahead but provide no explicit implementation. Finally, the three algorithms assume specific syntactic theories, with assumptions that are not common to all syntactic theories. Thus, it is not clear whether the success of the models depends upon general properties of syntactic structure or idiosyncratic properties of the syntactic theory employed by the algorithms.

In the following sections we seek to motivate an initial set of hypotheses that (1) is simpler than the previous models, so that the components can eventually be testable individually; (2) is incremental; and (3) is based on aspects of syntactic structure which are common to different syntactic frameworks. Furthermore, we would like the proposal to be as empirically successful as the current hypotheses. In order to accomplish these goals, we look for underlying generalities in the current models that give rise to their similar performance.

The size of the most recent constituent

According to CPC's algorithm, the likelihood of an intonational boundary increases with the number of right brackets at a word boundary. But Ferreira (1993) demonstrated that this factor is often confounded with constituent size: Word boundaries with a large number of brackets tend to mark the end of longer constituents than those with fewer brackets because they contain more constituents. Ferreira demonstrated that when constituent size is controlled, the number of right brackets is a poor predictor of the presence of intonational phrase boundaries. Thus the success of CPC's algorithm may be in part because the size of a recently produced constituent, not the number of right brackets at a word boundary, predicts the likelihood of an intonational boundary. We present this generalisation in (3):

- (3) Left-hand side (LHS) constituent size: As the size of a syntactic constituent increases, the likelihood of a boundary following that constituent increases (cf. Cooper & Paccia-Cooper, 1980; Ferreira, 1993).

A possible conceptual motivation for this hypothesis is that the sentence production mechanism may need a refractory period – an intonational phrase boundary – after the production of a particularly long constituent.

In addition to correlating with the predictions of CPC's model, the predictions made by this hypothesis correlate with the predictions of GG and Ferreira's (1988) proposals. Under GG's model, the likelihood of an intonational phrase is predicted by the number of nodes dominated by the boundary node. Two aspects of this model derive results that correlate

with (3). The first is the syntactic constituent rule. This rule groups phonological phrases that form either a clause or a noun phrase together under a single right branching node. An important consequence of this adjunction is that predictive values before and after the resulting constituent are higher than the values between phonological phrases within a resulting constituent. In addition, because longer constituents will have more phonological phrases and, thus, more nodes, it follows that the longer the phonological constituent is, the greater the likelihood of an intonational boundary before and after the constituent.

The predictions of Ferreira's (1988) algorithm correlate with those of (3) in a similar manner. Of the eight template types, the three most likely to coincide with an intonational boundary contain an XP as one of the pairs. Thus, the model predicts that the ends of constituent XPs are likely locations for intonational boundaries.

We show below that (3) is empirically supported by the results of Experiment 1. Furthermore, work by a number of researchers suggests that in ambiguous sentences, speakers produce intonational boundaries at locations that coincide with the right edge of substantial syntactic constituents (Carlson, Clifton, & Frazier, 2001; Cooper & Paccia-Cooper, 1980; Price et al., 1991; Schafer, Speer, Warren, & White, 2001; Snedeker & Trueswell, 2003).

The size of the upcoming constituent

A second generalisation whose predictions correlate with the predictions of the three models is presented in (4):

- (4) Right-hand side (RHS) constituent size: As the size of an upcoming syntactic constituent increases, the likelihood of a boundary before that constituent increases (cf. Ferreira, 1991).

The conceptual motivation behind this hypothesis is that speakers may need extra processing time to plan a longer syntactic constituent. Empirical support for this hypothesis stems from work by Sternberg, Monsell, Knoll, & Wright (1978) who found that speakers' initiation times for a list of words was a function of the number of stressed syllables they were about to produce. Sternberg and colleagues argued that these stress groups serve as a unit of planning in production. Ferreira (1991) followed up this work by studying initiation times of full sentences. She found that the length and the syntactic complexity of a sentence's subject were correlated with the sentence initiation times. In addition, Ferreira found that the probability of pausing before the verb phrase was positively correlated with the complexity of the sentence's direct object. She argued that these pause and initiation time data reflect processes in planning an upcoming

constituent. Wheeldon & Lahiri (1997) produced similar results showing that sentence production latency is a function of the number of phonological words in the upcoming sentence. All of these results together suggest that the phonological and syntactic properties of upcoming structure influence production time. We propose that intonational phrase boundaries provide this time.

Further evidence comes from the three algorithms that have been discussed. The predictions of CPC's algorithm correlate with the predictions of the RHS hypothesis in (4) because, according to CPC, the likelihood of an intonational boundary increases with the number of left brackets at a word boundary, which is correlated with the size of the upcoming constituent. The properties of GG's and Ferreira's algorithms that were discussed in relation to the LHS hypothesis in (3) also make predictions that correlate with the RHS hypothesis in (4). For example, the predictions of GG's algorithm are partly a function of the length of an upcoming noun phrase or clause, and Ferreira's algorithm predicts that intonational boundaries are more likely before XPs. Each of these properties makes predictions that correlate with the predictions of the RHS hypothesis in (4).

The LRB hypothesis

Putting together the LHS and RHS hypotheses, we arrive at an initial statement of our general claim:

- (5) The LHS/RHS boundary (LRB) hypothesis: The likelihood of a boundary between the LHS and RHS constituents increases with the size of each constituent.

In order to formalise the LRB hypothesis, we need to define LHS and RHS constituents. We define these terms in (6) and (7):

- (6) The LHS constituent at a word w_i is the largest completed syntactic constituent that ends with w_i . A syntactic constituent is defined as *completed* if the head of the constituent has no rightward dependents.
- (7) The RHS constituent at a word w_i is the largest syntactic constituent which starts at w_i .

There are two features of note in the LRB hypothesis. First, these definitions do not make reference to a specific syntactic phrase structure theory. This feature of our proposal is an advantage over previous models, which required specific syntactic systems to generate predictions (Cooper & Paccia-Cooper, 1980; Ferreira, 1988; Gee & Grosjean, 1983). The present proposal is consistent with all current phrase structure systems,

including X-bar theory. Second, the LRB hypothesis is initially assumed to quantify the size of the LHS and RHS constituents in terms of the *largest* syntactic constituent on the left and right side of a potential boundary respectively. This is an initial approximation in both cases. A more accurate future version of the size of the LHS constituent would need to take into account where the last intonational boundary had been produced. If a boundary had been produced within the largest constituent to the left, then only the size of the material up to that boundary should be counted in the LHS component. Thus in our initial statement of the LRB hypothesis, the size of the LHS component may sometimes be overestimated. The hypothesis about the size of the RHS constituent given in (7) is also an approximation. A more accurate RHS constituent size would be determined by the size of the constituent on the right that had been planned thus far. Because the language production mechanism is incremental, a speaker does not always plan a sentence to its completion before he or she starts speaking (V. Ferreira, 1996; Levelt, 1989; Roelofs, 1998). Thus the RHS size component may also sometimes be overestimated by assuming that the speaker knows the largest upcoming constituent.

In order for the LRB hypothesis to make specific predictions, a size metric is required for each of the LHS and RHS components. In addition, other constraints on intonational boundary placement may limit the application of the LRB hypothesis. One such constraint, semantic relatedness, is discussed below.

Measuring size

There are many possible units that could be used to measure size including phonemes, syllables, words, and larger syntactic or phonological phrases. In most instances, these units of measurement will be correlated with each other. For example, consider the increasing size of the subject NP in (8):

- (8) a. The reporter wrote a story.
 b. The reporter at the press conference wrote a story.
 c. The reporter at the press conference in Washington wrote a story.

Whether one counts phonemes, syllables, words or phrases, the size of the subject NP increases through (8a), (8b), and (8c). Therefore, any one of these units would probably serve as a good approximation of size in most cases.

As a first approximation, we use phonological phrases as a measure of size as defined by Gee & Grosjean (1983). In English, phonological phrases are defined as all the words within the maximal projection of a lexical head

on the lexical head's left side (Gee & Grosjean, 1983; Nespor & Vogel, 1986). This particular definition of the term phonological phrase was proposed by Nespor & Vogel (1986). However, it should be noted that there is a great deal of controversy surrounding how exactly phonological phrases should be syntactically defined. For example, Selkirk (1986), Ferreira (1988), Truckenbrodt (1999), and others define the term differently. For the purposes of measuring size, how phonological phrases are defined is not crucial to our argument since, as discussed above, most units of size would most likely do just as well. Only a controlled experiment that manipulates potential units of length (e.g., syllables, words, phonological phrases, etc.) can determine what the appropriate metric is.

One ramification of measuring size in terms of phonological phrases is that boundaries will not be hypothesised to occur within phonological phrases. This is a reasonable first approximation, although there can be exceptions (e.g., see Shattuck-Hufnagel & Turk, 1996). In addition, the definition of what counts as a phonological phrase clearly has consequences for the locations to be considered for intonational boundaries. For example, Ferreira (1988) observes that unstressed words do not always adjoin to the stressed word on the right:

- (9) a. John picked-up movie-tickets.
 b. The professor-has forgotten

In (9a), where the dashes signal words that constitute a phonological phrase, the particle "up" cliticises to the main verb "picked" on its left while in (9b), the auxiliary verb "has" adjoins to the subject NP "professor", and is often contracted in this context. As a first approximation, we will use GG's definition of a phonological phrase, with the knowledge that this definition must be extended or revised to account for the effects in (9) and the counterexamples discussed in Shattuck-Hufnagel & Turk (1996).

Semantic relatedness. An additional factor that seems to play a role in determining intonational phrasing is the semantic relationship between dependencies. Selkirk (1984) argued that intonational phrase boundaries in an utterance are subject to a semantic well-formedness constraint called the Sense Unit Condition (SUC). The SUC states that heads that do not have a dependency relationship cannot occur together in an intonational phrase. Formally, Selkirk defines the SUC as follows:

- (10) The Sense Unit Condition of Intonational Phrasing
 The immediate constituents of an intonational phrase must together form a sense unit.

- (11) Two constituents C_i , C_j form a sense unit if (a) or (b) is true of the semantic interpretation of the sentence:
- a. C_i modifies C_j (a head)
 - b. C_i is an argument of C_j (a head)

Thus, the SUC predicts that sentence (12b) should sound distinctly worse than sentence (12a) because “the book” and “to Mary” do not engage in a head-argument/modifier relationship:

- (12) a. [John gave the book] [to Mary]
 b. [John gave] [the book to Mary]

One important implication of the SUC is that semantically related words tend to be grouped together in the same intonational phrase while semantically unrelated words tend not to be. In light of the SUC, we hypothesise that intonational boundaries tend to occur before large syntactic constituents if and only if they are not semantically related to the most recently processed head. As a first approximation at semantic relatedness, we propose that syntactic constituents are considered semantically related to the immediately preceding head if they are one of the head’s arguments (see Schütze and Gibson, 1999 for a discussion of the criteria for argumenthood).⁶

The LRB hypothesis revisited. LHS/RHS boundary (LRB) weight is defined in (13):

- (13) LHS/RHS boundary (LRB) weight: The LRB weight at a word boundary between w_1 and w_2 is defined to be the sum of
- a. the size of the LHS constituent terminating at w_1 , in terms of phonological phrases;
 - b. the projected size of the RHS constituent in phonological phrases starting at w_2 , if this is not an argument of w_1 ;
 - c. 1, if w_1 marks the end of a phonological phrase.

We point out that constituent length rather than constituent type is relevant to calculating the LRB weight. The size of the LHS constituent is computed by determining the size of the largest syntactic constituent that is terminated at w_1 . The size in phonological phrases of the largest RHS constituent is computed if it is not an argument of the most recently produced word.

Because speakers may produce sentences incrementally without a full plan for the material that is coming next (as discussed above), the RHS

⁶ It is possible that modifiers may also play a role suggested by the SUC. We leave this question for future research.

constituent might not be fully known at the point of producing a word w_i . As a result, we test an alternative, incremental version of the LRB in addition to the version presented above. In this version, the size of the RHS constituent can have three values, corresponding to an upcoming argument, a small upcoming non-argument, or a large upcoming non-argument.⁷ Constituents that are arguments receive a RHS value of 0. Constituents consisting of one or two phonological phrases are classified as small and receive a RHS value of 2. Constituents consisting of three or more phonological phrases are classified as large and receive a RHS value of 4.

The third part of the definition of LRB weight reflects the fact that intonational boundaries are more likely at phonological phrase boundaries than at other word boundaries. As a result, an additional value of 1 is added at each phonological phrase boundary location.

The LRB weight hypothesis is then given in (14):

- (14) The LRB weight hypothesis: The LHS/RHS boundary weight is proposed to be correlated with the probability of producing a boundary at a given location.

An example is provided in Figure 4 of how the LRB hypothesis applies to a sentence.

The phonological phrases – all the words up to and including a noun or verb—are shown separated by large spaces in Figure 4. The first number at each phonological phrase boundary is the size of the LHS constituent. Constituents are shown graphically as lines with double arrows. Phonological phrase boundaries that coincide with a right-facing arrow mark the end of a complete constituent, as defined in (6). The second number represents the size of the largest RHS constituent being integrated at that position (as long as it is not an argument of the head to which it is integrating). Phonological phrase boundaries that coincide with a left-facing arrow mark the beginning of a constituent. The number below the first two represents additional weighting for the phonological phrase boundary.

For example, the LRB predicts that the most likely place for a boundary to occur in this sentence is between “[ignored]” and “[fired]”. At this point, the largest completed constituent is the matrix subject of the sentence “[the judge who the reporter for the newspaper ignored]”, which extends over four phonological phrases, creating an LHS value of 4. The size of the RHS constituent to be integrated at this point is 2, reflecting the two phonological phrases in the VP “[fired the secretary]”. Finally, 1 is

⁷ We thank Mike Tanenhaus for this idea.

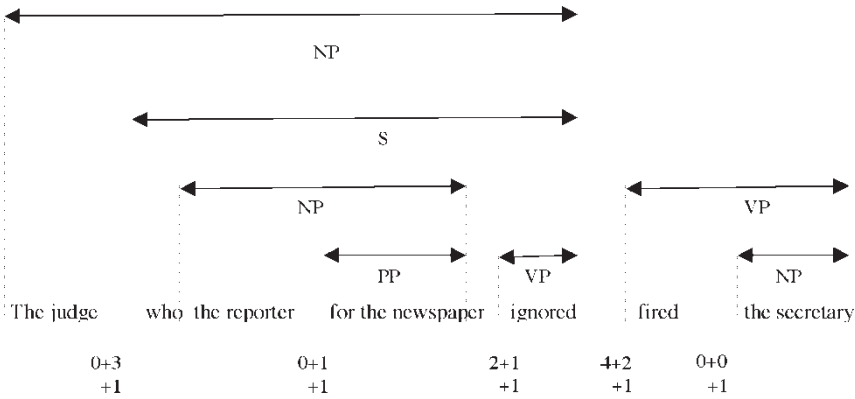


Figure 4. The predictions of the LRB on an example sentence. Phonological phrase boundaries are indicated by large spaces between words. Constituent boundaries that are relevant to the LRB—those that coincide with phonological phrase boundaries—are indicated by dotted vertical lines. Constituent lengths are indicated by the accompanying horizontal arrows. The LRB values for the boundaries are presented below the sentence. The top left value indicates the size of the LHS constituent in phonological phrases and the top right value indicates the size of the RHS constituent. The bottom value (+1) is the additional weighting for phonological phrase boundaries.

added to the LHS and RHS components, reflecting the fact that this is a phonological phrase boundary. Thus the total value at this boundary is 7.

A relatively low likelihood is assigned to a word boundary if the left-hand constituent is not completed or the right-hand constituent is an argument of the most recent head. For example, at the boundary between "[fired]" and "[the secretary]", no syntactic constituents have been completed, so the LHS is 0. Furthermore, the RHS constituent is an argument of the preceding word, the verb "[fired]", so it does not count toward the RHS count. Only the third component of LRB weight makes a contribution (consisting of the value 1), because this location is a phonological phrase boundary.

Evaluating the LRB with respect to the data from Experiment 1

The LRB predictions for the eight syntactic constructions in (2) are provided in Table 3. In an analysis of all 512 trials in the experiment, the LRB accounted for a significant quantity of the variance, $r^2 = .74$ ($p < .001$), similar to the quantity of variance explained by GG's and Ferreira's algorithms. When only those trials without disfluencies, hesitation pauses or incorrect comprehension questions were analysed, the LRB accounted

TABLE 3

The numeric predictions of the LRB. Each value is the prediction for the word boundary following the word in the corresponding column. Higher numbers indicate greater likelihood of placing an intonational boundary at that point.

(2a)	The	judge	who	ignored	the	reporter	fired	the						
LRB	0	3	0	1	0	6	1	0						
(2b)	The	judge	who	the	reporter	ignored	fired	the						
LRB	0	3	0	0	3	6	1	0						
(2c)	The	judge	who	the	reporter	for	the	newspaper	ignored	fired	the			
LRB	0	4	0	0	2	0	0	4	7	1	0			
(2d)	The	judge	who	the	reporter	for	the	newspaper	in	the	capitol	ignored	fired	the
LRB	0	5	0	0	3	0	0	2	0	0	5	8	1	0
(2e)	The	judge	who	the	reporter	who	attacked	the	senator	ignored	fired	the		
LRB	0	5	0	0	3	0	1	0	5	8	1	0		
(2f)	The	judge	who	the	reporter	who	the	senator	attacked	ignored	fired	the		
LRB	0	5	0	0	3	0	0	3	5	8	1	0		
(2g)	The	reporter	ignored	the	judge	who	fired	The						
LRB	0	6	1	0	3	0	1	0						
(2h)	The	senator	attacked	the	reporter	who	ignored	the	judge	who	fired	the		
LRB	0	8	1	0	5	0	1	0	3	0	1	0		

for a similar quantity of variance, $r^2 = .74$ ($p < .001$). The data presented in the tables and charts below use this analysis.

Furthermore, the results were similar when only the seven items that had identical syntactic structures in all critical regions were analysed (all trials: $r^2 = .71$; $p < .001$; only correctly answered and non-disfluent trials: $r^2 = .71$; $p < .001$). Finally, an analysis was conducted on only those word boundaries that coincided with phonological phrase boundaries. Even without these points, the LHS and RHS components of the LRB account for a highly significant quantity of variance ($r^2 = .55$, $N = 41$, $p < .001$).

Figure 5 presents a scatterplot of the relationship between LRB predictions and intonational boundary likelihood, across all 85 word boundary positions in the eight syntactic structure types in (2), using the data from only correctly answered and non-disfluent trials. Figures 6 and 7 present the LRB predictions versus the percentage of the time intonational boundaries were placed at a word boundary for each of the conditions. The variance accounted for by each theory for each sentence type is listed in Table 4.

When the variance accounted for by the LRB is partialled out, the variance accounted for by Ferreira, GG, and CPC drops to $r^2 = .05$ ($N = 85$, $p < .05$); $r^2 = .05$ ($N = 85$, $p < .05$); and $r^2 = .00$ ($N = 85$, $p < .95$), respectively. This analysis demonstrates that the predictions of the LRB are highly correlated with the predictions of the best two existing

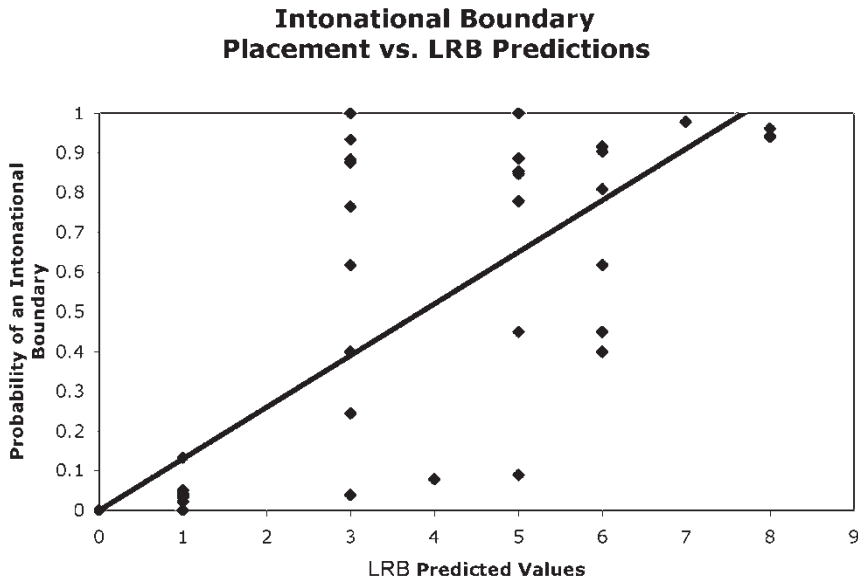


Figure 5. The LRB predictions plotted against the percentage of intonational phrase boundaries that were produced in Experiment 1.

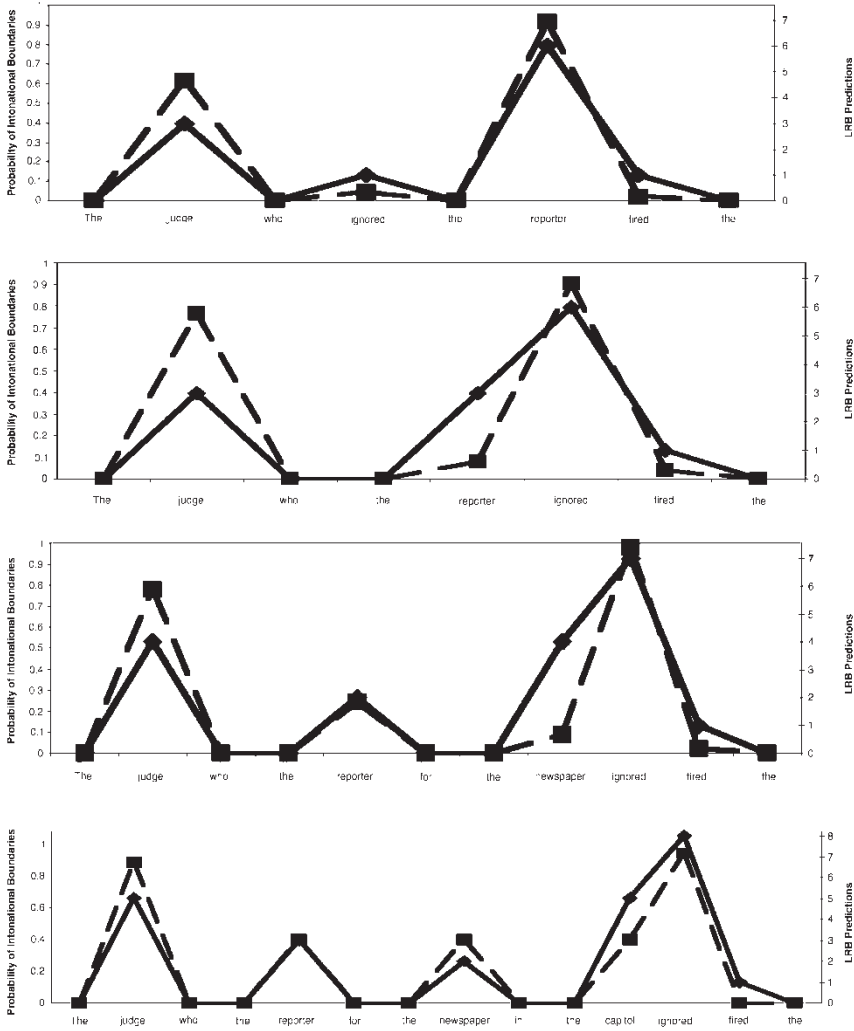


Figure 6. The percentage of intonational boundary placement graphed against the predictions of the LRB for sentences (2a)–(2d) in Experiment 1. —, LRB predictions; - - -, Intonational boundaries.

algorithms, GG’s and Ferreira’s. The variance accounted for by LHS and RHS constituent components of the LRB hypothesis is listed for each sentence type in Table 5. The RHS parameter accounts for more variance than the LHS parameter.

The incremental version of the LRB discussed above that had limited RHS look ahead performed similarly to the original LRB, $r^2 = .74$ ($N =$

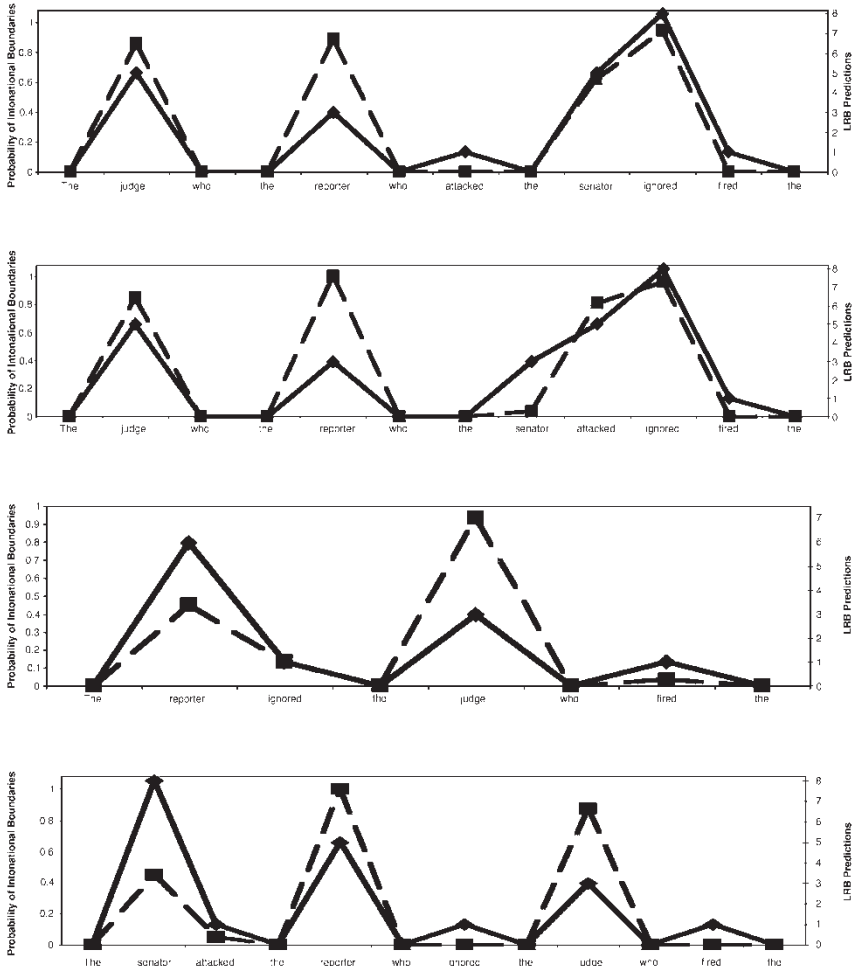


Figure 7. The percentage of intonational boundary placement graphed against the predictions of the LRB for sentences (2e)–(2h) in Experiment 1. —, LRB predictions; ---, Intonational boundaries.

85, $p < .001$), suggesting that an incremental version of the LRB can provide as good an account of the data as the less incremental version.

Discussion

The results from Experiment 1 provide initial support for a model like the LRB. Points that occur before and after large syntactic constituents are highly correlated with intonational phrase boundaries, and intonational

TABLE 4
 The variance accounted for by Ferreira's (1988) X-bar model, Gee and Grosjean's (1983) Phi algorithm (GG), Cooper and Paccia-Cooper's (1980) algorithm (CPC), and the LRB, for each sentence type in Experiment 1.

<i>Sentence in Experiment 1</i>	<i>Model</i>			
	<i>Ferreira</i>	<i>GG</i>	<i>CPC</i>	<i>LRB</i>
(2a)	.74	.79	.63	.94
(2b)	.58	.77	.56	.76
(2c)	.57	.81	.45	.78
(2d)	.70	.86	.49	.89
(2e)	.84	.64	.37	.83
(2f)	.72	.67	.39	.74
(2g)	.78	.84	.34 ^a	.50 ^a
(2h)	.79	.82	.19 ^a	.53
Overall variance	.71	.76	.39	.74

^a indicates non-significance.

boundaries tended not to occur between heads and their arguments. Furthermore, the regression analyses demonstrate that the predictions of the LRB are highly correlated with the predictions of the best two existing algorithms, GG's and Ferreira's. In particular, after the variance accounted for by the LRB was partialled out, the variance accounted for by both the GG's and Ferreira's algorithms dropped to only .05. The variance accounted for by CPC's algorithm dropped to 0. These analyses support the hypothesis that much of what the earlier theories predict is driven by

TABLE 5
 The variance accounted for by the LHS and the RHS constituent components of the LRB for the sentence types in Experiment 1.

<i>Sentence in Experiment 1</i>	<i>LRB Component</i>	
	<i>LHS constituent r²</i>	<i>RHS constituent r²</i>
(2a)	.65	.95
(2b)	.44 ^a	.89
(2c)	.42	.83
(2d)	.40	.84
(2e)	.35	.79
(2f)	.33 ^a	.68
(2g)	.09 ^a	.51
(2h)	.05	.56
Overall variance	.32	.65

^a indicates non-significance.

syntactic constituent size, syntactic boundaries, and semantic cohesion. Although the overall performance of the best three models—the LRB hypothesis, GG’s algorithm, and Ferreira’s algorithm—was very similar, the models performed differentially well on different conditions. In particular, the LRB performed numerically better than the other two algorithms in (2a), (2d), and (2f). The GG algorithm performed best on (2b), (2c), (2g), and (2h). The Ferreira algorithm performed best on (2e), and CPC performed relatively poorly on all the sentences. However, it should be noted that with the exception of (2g) and (2h), in the cases where LRB does not perform better, the difference is minimal.

One potential concern is that the success of the LRB is driven by extreme values. A visual inspection of Figure 4 suggests that there is more variability in mid-range values than in low values (e.g., 0, 1) or high values (e.g., 8). In addition, high values such as 8 occurred mostly in doubly nested structures. To address this concern, we conducted an analysis where the doubly nested structures (2e) and (2f) were excluded. In addition, structures (2g) and (2h) were excluded because results from Experiment 3 that will be presented below suggest that discourse structure may have played a strong role in the intonational phrasing of these sentences. Thus, the analysis was conducted on sentences (2a)–(2d) on word boundaries that were neither 0 or 1, to see how well the LRB performed. The LRB still accounted for a significant amount of the variance, $r^2 = .48$ ($N = 14$; $p < .01$), suggesting that it can successfully predict non-extreme values.

In Experiment 2, we investigated materials in which the LRB makes different predictions from GG’s and Ferreira’s algorithms. Then in Experiment 3, we investigate a potential reason why the LRB may have been less successful in Experiment 1 with respect to the right-branching structure in (2g) and (2h).

EXPERIMENT 2

The performance of the most successful theories—the LRB, GG, and Ferreira algorithms—was very similar in Experiment 1, partly because the three algorithms make very similar predictions for the materials that were tested. However, they make similar predictions for different reasons. Consider (2a), repeated below as (15).

(15) The judge who ignored the reporter fired the secretary.

The LRB predicts that an intonational boundary in this sentence is most likely to occur before the verb “fired” because this point follows a relatively long subject NP in addition to preceding a medium-sized syntactic constituent, the verb phrase. The Ferreira algorithm also predicts a boundary at this point, but does so because it is the boundary between

the subject NP and the VP. The boundary between two XP level categories (i.e., NP and VP) is a likely place for an intonational boundary under this theory. The GG algorithm predicts a boundary before the verb “fired” because it marks the end of a relative clause. In the GG algorithm, the end of a clause is a likely place for an intonational boundary to occur. Experiment 2 was designed to separate some of the predictions of the GG, Ferreira, and LRB theories.

Structures were tested that contained word boundaries that followed large syntactic constituents but were not confounded with factors predicted by the Ferreira and GG algorithm to increase the likelihood of intonational boundaries, as in (16):

- (16) An artist arranged a donation of the paintings of the landscape to the museum.

In (16), the head of the direct object is “donation”, which takes “of the paintings of the landscape” and “to the museum” as arguments. Immediately before the PP “to the museum”, the long PP “of the paintings of the landscape” is completed. Because this word boundary corresponds with the end of a large syntactic constituent, the LRB predicts a high probability of a boundary occurring before the word “to”. The GG algorithm, on the other hand, predicts that this point is an unlikely place for an intonational boundary because the noun phrase “a donation of the paintings of the landscape to the museum”, is adjoined in a right branching tree by the algorithm’s Syntactic Constituent Rule. Thus, intonational boundaries are more likely to occur between phonological phrases at the beginning of the noun phrase than between phonological phrases towards the end of the noun phrase because nodes at word boundaries early in the sentence are located higher in the tree. Furthermore, the Ferreira algorithm also predicts that this location is an unlikely position for a boundary to occur because “to the museum” is an argument of “donation”, and the boundary between a head and its argument is an unlikely location for a boundary to occur.⁸

Method

The same listener-reader paradigm that was used in Experiment 1 was used here.

⁸ Although in this case, the relevant boundary for Ferreira’s (1988) analysis would appear to be the boundary between *landscape* and *to*, Ferreira treats non-adjacent arguments of the verb as if they were adjacent. Thus, the relevant boundary for analysis is between *donation* and *to*.

TABLE 6

Predictions of the GG algorithm, Ferreira algorithm, and LRB for the items in Experiment 2

	An artist arranged a donation of the paintings of the landscape to the												
GG	0	3	11	0	7	1	0	5	1	0	4	1	0
Ferreira	1	7	4	3	4	4	2	4	4	1	4	4	1
LRB	0	7	1	0	1	0	0	1	0	0	4	0	0

Participants. Ten pairs of participants from the MIT community (ten readers and ten listeners) participated for \$5 each.

Materials. Nine sentences with the structure in (16) were tested. Because only one syntactic construction was being tested, thirty-nine unrelated fillers were included. Four separate randomised lists were constructed from the items and fillers.

The predictions of the three models are presented in Table 6. Note that although the theories' numeric values for the word boundary after "landscape" are similar, this value should be interpreted relative to the algorithm's values at other word boundaries. In the LRB framework, this number is relatively large compared with the values at other word boundaries in the sentence. This is not true for the GG and Ferreira algorithms.

Procedure. The same method that was used in Experiment 1 was used to code speakers' productions. The total reliability between coders was 92%.

Results and discussion

In an analysis of all 99 trials, the predictions of the LRB were significantly correlated with participants' intonational boundary placement, $r^2 = .48$ ($N = 13, p < .01$), as can be seen in Figure 8. In an analysis of items that excluded items that contained disfluencies and incorrect responses (12 disfluencies, 4 hesitation pauses, 11 incorrect answers, yielding a total of 25 trials to be discarded), the results did not change, $r^2 = .47$. The LHS component captured most of the variance, $r^2 = .68$ ($N = 13, p < .01$). The RHS component accounted for $r^2 = .15$ ($N = 13, p = .19$).

The GG algorithm's predictions were marginally significant, and the Ferreira algorithm's predictions were not quite significant ($N = 13$ for each; All trials: GG: $r^2 = .25, p = .08$, Ferreira: $r^2 = .23, p = .11$). The overall difference among the models was not significant.

A visual inspection of Figure 8 suggests two things. The first is that the LRB predicts more boundaries after the subject NP than actually occur.

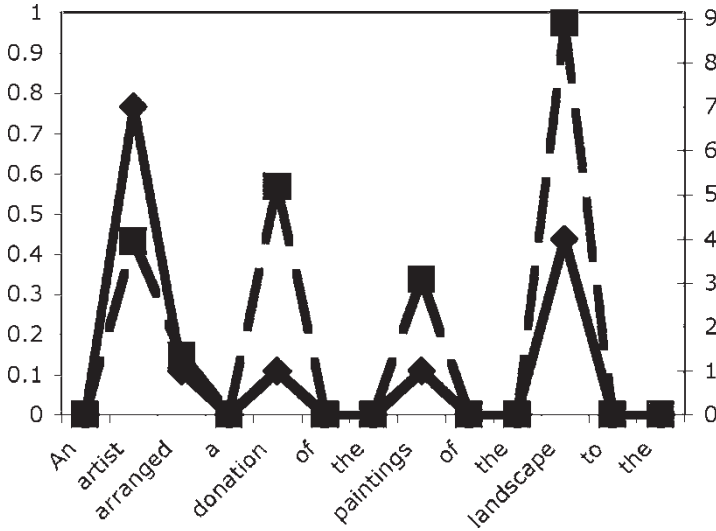


Figure 8. The LRB predictions vs. the percentage of intonational boundaries occurring at each word boundary for items in Experiment 3. —◆—, LRB predictions; —■—, Intonational boundaries.

The LRB predicts that the long VP (five phonological phrases long) makes this a likely location for a boundary. This along with the finding that the LHS contributed more than the RHS in accounting for the variance in this structure suggests that the RHS component of the LRB should be weighted less strongly. Another possibility is that the incrementality of the production system is greater than we originally posited and that the amount of look-ahead is very limited.

There were also more boundaries at points between nouns and their arguments (i.e., between “donation” and “of”, and “painting” and “of”) than predicted by the LRB. This contrasts with the findings in this experiment and in Experiment 1 that boundaries between verbs and a following object were infrequent. There may be a dispreference for producing a boundary between verbs and objects only, which does not generalise to other lexical heads. Such a finding would be consistent with claims in the literature that intonational boundaries are more likely to occur after nouns than other lexical categories (Nespor & Vogel, 1986). In addition, the RHS component predicted a relatively low likelihood of a boundary following the verb in Experiment 1 independently of the semantic constraint. With the exception of (2g) and (2h) in Experiment 1 (which we discuss in the next section), the direct objects in these sentences were relatively short. To explore this question we tested a version of the LRB that lacked the head-argument constraints.

Boundaries between heads and arguments were calculated as they would be in any other position. In addition, the size of the RHS component was capped at two to reflect limited look ahead. Upcoming constituents that were one phonological phrase long received a value of 1 and those two or greater received a value of 2. This algorithm performed quite well on (16), $r^2 = .75$ ($N = 13$, $p < .001$). When we tested this version of the LRB on the structures in Experiment 1, it still accounted for a significant amount of the variance, $r^2 = .64$ ($N = 85$, $p < .0001$). Thus, these data strongly suggest that a simpler version of the LRB that eliminates the semantic component and caps the RHS value may be more successful over a larger range of structures. We leave this question open for further research.

Figure 8 also suggests that there may be a cyclical component to intonational phrase placement. Researchers have argued that speakers may attempt to balance their intonational phrases into units of roughly equal size (Cooper & Paccia-Cooper, 1980; Fodor, 1998; Gee & Grosjean, 1983; Selkirk, 2000). However, the poor performance of the GG model suggest that this cyclicity cannot simply be attributed to a preference for placing intonational boundaries roughly in the middle of the sentence. More work is necessary in understanding how balancing might affect boundary placement.

The CPC algorithm was not discussed above, because it did much worse than the other three algorithms on Experiment 1. But it performs relatively well on the structure in the current experiment ($r^2 = .40$, $p < .05$) although the LRB does numerically better. The CPC algorithm predicts a high probability of an intonational boundary between “landscape” and “to” because this word boundary coincides with a large number of syntactic right brackets. However, as discussed above, Ferreira (1993) demonstrated that when constituent length is controlled, right brackets perform poorly in predicting the likelihood of a boundary. This evidence along with the relatively poor performance of CPC in Experiment 1 suggests that the success of the CPC is due to the high correlation between the number of right brackets and recent constituent length in this structure.

Overall, the results of this experiment suggest that speakers place intonational boundaries after large syntactic constituents, as predicted by the LHS component of the LRB. The Ferreira and GG algorithm do not predict this pattern of results.

EXPERIMENT 3

In Experiment 1, the LRB performed poorly on the right-branching structure in (2g) compared to its performance in other conditions. The LRB predictions for each word boundary in (2g) are shown in Table 7

TABLE 7

The model predictions and the percentage of intonational boundary placement for sentence (2g)

	The	reporter	ignored	the	judge	who	fired	the
GG	0	5	3	0	9	0	3	1
Ferreira	1	7	4	3	8	3	4	1
CPC	0	1.6	3.2	.8	3.2	2.4	1.2	.4
LRB	0	6	1	0	3	0	1	0
Boundary likelihood	0	.44	.13	0	.94	0	.03	0

along with the percentages of intonational boundaries that were produced in Experiment 1 at the relevant position.

The LRB predicts that the largest boundary will occur after “the reporter” because this position is located immediately before the verb phrase, a relatively large syntactic constituent. The second largest boundary is predicted to occur between “judge” and “who”, because the relative clause is two phonological phrases long. The actual results show the opposite trend with respect to these two positions in (2g). That is, speakers were more likely to place a boundary between “judge” and “who”, rather than between “reporter” and “ignored”.

One possible reason for the observed pattern of results in these conditions may have to do with discourse factors rather than syntactic factors. In its written form, an RC is potentially ambiguous between a non-restrictive reading and a restrictive reading. Non-restrictive modifiers typically provide new or aside information about the head that they modify. In the non-restrictive interpretation of the RC in (2g), the listener is informed that there is a judge, and that judge fired the secretary. A restrictive modifier, on the other hand, does not typically provide new information. Rather, it picks out a referent from a contrast set. If one imagines more than one judge (e.g., a judge who fired a secretary, and a judge who did something else), a restrictive reading of the RC informs the listener that the judge who fired the secretary was the judge in question. The two interpretations of the RC can be disambiguated using intonational phrasing. A non-restrictive modifier is generally preceded by an intonational phrase boundary, whereas a restrictive modifier is not preceded by such a boundary. One possible interpretation of the results from Experiment 1 is that speakers are assigning non-restrictive readings to the RCs in these items, by placing boundaries before the RCs.

Suggestive corroborating evidence that is consistent with this hypothesis comes from reading studies in sentence comprehension, where it has been shown that interpretations that presuppose/implicate simpler discourse sets are preferred (Altmann & Steedman, 1988; Grodner, Gibson, & Watson,

in press). In the absence of context, a non-restrictive reading presupposes a simpler discourse set than a restrictive reading. In particular, a restrictive relative clause interpretation requires the construction of a contrast set (i.e., more than one judge in (16)) for interpretation. A non-restrictive reading, on the other hand, implicates no such set and may have been the default interpretation for the speaker.

In Experiment 3 these intuitions about intonational phrasing and relative clauses are empirically tested to see if the restrictiveness of a relative clause can affect intonational phrasing in our two-person reading production task. If so, the presence of intonational phrase boundaries before relative clauses in this experiment may be a result of discourse factors, and the poor performance of the LRB on certain structures in Experiment 1 may be a coincidental result of the discourse structure.

Intuitions suggest that speakers place intonational boundaries before non-restrictive relative clauses but not before restrictive relative clauses (Selkirk, 1978; Nespor & Vogel, 1986; Shattuck-Hufnagel & Turk, 1996). Hirschberg and Avesani (1997) have examined the degree to which English and Italian speakers use prosody to disambiguate a variety of constructions, including restrictive and non-restrictive relative clauses. In their study, participants read sentences containing relative clauses. The sentences were embedded in paragraphs that disambiguated the relative clause towards either a restrictive or non-restrictive reading of the relative clause. Hirschberg and Avesani (1997) found that only 60% of the relative clauses were disambiguated using prosody. This is a surprisingly low number given that intuitions strongly suggest that non-restrictive relatives obligatorily require a pre-clause intonational boundary. The relatively low levels of prosodic disambiguation may have been an artifact of the task. In the experiment, speakers read target sentences to an experimenter who was recording them. It is possible that speakers did not disambiguate the sentences because the task did not involve communicating information to a listener, but rather involved speaking to a researcher in the experiment. In the current study, we investigate the same types of materials as Hirschberg and Avesani (1997) did but use the speaker-listener paradigm that was used in Experiments 1 and 2.

The second goal of the current experiment was to determine whether sentence position influences how a speaker interprets a relative clause. A number of researchers have pointed out that there is a tendency for given information to occur at the beginning of a sentence and for new information to appear towards the end of a sentence (Chafe, 1976, 1987; Diesing, 1992; Du Bois, 1987; Givón, 1979, 1983, 1984; Kratzer, 1995; Prince, 1981). Because restrictive relatives are typically associated with given information (which tends to come earlier in sentences) and non-restrictive relatives are typically associated with new information (which

tends to come later in sentences), the position in which the relative clause appears may influence how the speakers in the experiment interpret it. Corpus work by Fox and Thompson (1990) suggest that the sentential location of a relative clause may be tied to its function in the sentence. Although there was little intonational evidence for non-restrictives in their corpus, they found that relative clauses that modified object NPs were more likely to provide new information about the NP while relative clauses that modified subject NPs were more likely to link the head to entities in the discourse. Thus, in situations where there is no explicit discourse, speakers may be biased towards interpreting RCs that occur at the beginning of a sentence as restrictive and RCs that occur towards the end of a sentence as non-restrictive.

To investigate these questions, the restrictiveness and sentence position of relative clauses were manipulated to determine whether these factors affected intonational boundary placement before the relative clause. Restrictiveness was manipulated by placing sentences in different contexts. In the restrictive context, two referents with the same identifying head noun were introduced to establish a potential contrast set for the referent in the target sentence. Each of the referents was restricted by a prepositional phrase modifier (e.g., *a director at a banquet* and *a director at a film premiere*). Thus, the relative clause in the target sentence would probably be interpreted as restrictive since it was restricting over a set of already introduced discourse referents, providing old information as a pointer. In the non-restrictive contexts, two referents from different categories were introduced without any modification (e.g., *a director* and *a producer*). Thus, the relative clause in the target sentence was not restricting over a set of referents but rather provided new information about an unambiguous referent.

The referents mentioned in the context sentence were either the subject or the object of the target sentence. This enabled the manipulation of the presence of the relative clause at the beginning and the end of the sentence. An example item is provided in (17) below.

- (17) a. Restrictive Interpretation/Subject Modifying Relative Clause
 A group of film critics praised a director at a banquet and another director at a film premiere. *The director who the critics praised at a banquet insulted an actor from an action movie during an interview.*
- b. Restrictive Interpretation/Object Modifying Relative Clause
 A group of film critics praised a director at a banquet and another director at a film premiere. *An actor from an action movie insulted the director who the critics praised at a banquet during an interview.*

c. Non-Restrictive Interpretation/Subject Modifying Relative Clause

A group of film critics praised a director and a producer. *The director who the critics praised at a banquet insulted an actor from an action movie during an interview.*

d. Non-Restrictive Interpretation/Object Modifying Relative Clause

A group of film critics praised a director and a producer. *An actor from an action movie insulted the director who the critics praised at a banquet during an interview.*

Speakers produced these short paragraphs for listeners who then answered questions about the events in the sentences. The sentences that were crucial for analysis are italicised. The crucial word boundary for comparison occurs between *director* and *who*. If restrictiveness plays a role in boundary placement, there will be more boundaries at this position than in the non-restrictive conditions. If position in the sentence plays a role, there will be more intonational boundaries at this point in the object modifying conditions.

Method

The same listener-reader paradigm that was used in Experiments 1 and 2 was used in Experiment 3.

Participants. A total of 11 pairs of participants from the MIT community (11 readers and 11 listeners) participated for \$5 each.

Materials. All items had the same structure as (17) above. Sixty fillers were randomly mixed with the experimental items. The items were presented in four counterbalanced lists in a Latin Square design.

Procedure. The same ToBI coding procedure that was used in Experiments 1 and 2 was used here. Two coders listened for prosodic breaks. For this analysis, intermediate boundaries and intonational boundaries were collapsed. The reliability between coders was 96%. The data coded by the first author was used for analysis.

Results

Analysis revealed two main effects at the word boundary immediately before the relative clause in each condition. Table 8 shows the percentage of pre-relative clause intonational boundaries. First, there was a main effect of restrictiveness with more intonational boundaries occurring

TABLE 8

The percentage of times speakers placed an intonational boundary before the relative clause in Experiment 3. The standard errors are presented in parentheses

	<i>Subject modification</i>	<i>Object modification</i>
Restrictive context	10.0% (6.7)	38.6% (9.4)
Non-restrictive context	27.7% (8.1)	49.1% (10.2)

before non-restrictive relative clauses than restrictive relative clauses, 38.4% vs. 24.3% respectively. This effect was marginal by subjects but significant by items, $F_1(1, 10) = 3.549, p = .09, F_2(1, 19) = 6.84, p < .05$. Second, there was a main effect of the proposition of the relative clause. Intonational boundaries were more likely to occur before relative clauses that modified direct objects than relative clauses that modified subjects, 43.8% vs. 18.9% respectively, $F_1(1, 10) = 8.32, p = .05, F_2(1, 19) = 16.93, p < .001$.

Discussion

As expected, discourse structure appears to play a role in determining intonational phrase placement. In particular, relative clauses with non-restrictive readings and relative clauses that modified a direct object were more likely to be preceded by an intonational phrase boundary than relative clauses with restrictive interpretations and relative clauses that modified a subject. These data suggest that speakers use intonational boundaries to disambiguate relative clauses. They also suggest that speakers are more likely to place boundaries before relative clauses that occur at the end of a sentence.

These results support the hypothesis that non-syntactic factors, namely discourse status and the signalling of information structure via sentence position, probably influenced the likelihood of an intonational boundary before the relative clauses in Experiment 1 and would partly explain the poor performance of the LRB on structure (2g). These factors may have also contributed to the LRB's success on the other structures.

A finding from this experiment that is somewhat puzzling is that subjects did not disambiguate non-restrictive relatives more frequently. Most of the time, speakers preferred not to place a boundary between a relative clause and its head noun, regardless of the restrictiveness and the sentence position of the relative clause. However, this result may not be so surprising in light of the fact that, in reading, commas often signal an intonational break. The absence of commas in the stimuli may have biased the speakers against producing more intonational breaks. In spontaneous speech, speakers' performance may more closely match our intuitions

about intonational phrasing. It is also possible that speakers simply do not consistently disambiguate relative clauses. This result is consistent with Fox and Thompson's (1990) finding that intonational disambiguation of non-restrictives was not very frequent. If so, this could be due to a number of reasons. For instance, the fact that the context of the sentence disambiguates the status of the relative clause may make disambiguation via prosody unnecessary. This issue will be discussed in more depth in the General Discussion.

GENERAL DISCUSSION

We have presented evidence that the intonational phrasing of a sentence is partly a function of the size of upcoming and recently processed syntactic constituents, modulated by the semantic relationships among the constituents' syntactic heads. We have suggested that algorithms by Gee and Grosjean (1983), Cooper and Paccia-Cooper (1980), and Ferreira (1988) are successful (to varying extents) because they incorporate these factors. A theory such as the LRB that is based solely on these factors can do as well as previous models (Experiment 1) and can even outperform these same models in certain instances (Experiment 2).

The LRB theory of boundary placement in production has some important ramifications for comprehension. One of the underlying assumptions of the hypothesis is that intonational boundary placement is the result of planning processes in production, suggesting that boundary placement is a function of the speaker's needs rather than the listener's needs in parsing. But even though the speaker may not be aware of the listener's needs in sentence comprehension, the production of an intonational boundary before and/or after a long constituent can provide the listener with a cue to the appropriate syntactic structure for the sentence that was produced. In particular, if constituents are longer, boundaries are more likely to be produced that will be informative to the listener.

Suggestive evidence that this may be the case is provided by Snedeker and Trueswell (2003) and Kraljic and Brennan (2003). In Snedeker and Trueswell's experiment, speakers did not reliably disambiguate the globally ambiguous sentence structures in (18) with intonational phrasing when the referential context was unambiguous. Consider the LRB boundary placement likelihoods for each of the two structures:

- (18) Tap the frog with the flower
 a. Attachment of "with the flower" to "the frog":
 [Tap [the frog with the flower]].
 1 0 2 0 0

b. Attachment of “with the flower” to “the frog”:

[Tap [the frog] [with the flower]]
 1 0 3 0 0

The only point where a difference between the two structures is predicted by the LRB is at the word boundary of “frog” and “with”. In (18a) the direct object “the frog” is not complete at this word boundary so the LHS value is 0. The upcoming syntactic constituent is the prepositional phrase “with the flower” and it consists of one phonological phrase, so the RHS value is 1. A value of 1 is added because this word boundary coincides with a phonological phrase boundary, yielding a total value of 2. In contrast, the direct object in (18b) is complete, so the LHS value is 1. The RHS value of the upcoming PP is 1 and 1 is added for the phonological phrase boundary, yielding a total of 3. Thus there is a small difference in the LRB likelihoods for a boundary at this location: 3 for (18b) and 2 for (18a), yielding a difference of 1.

Prosodic disambiguation may be more likely when the LRB values of two interpretations of an utterance differ by a larger amount. This can occur in the same kinds of structures as in (18) when the length of the PP modifying the NP is larger, as in (19) below:

(19) a. Put [the dog]_{NP} [in the basket on the star]_{PP}.
 1 0 4 0 0 2 0 0

b. Put [the dog in the basket]_{NP} [on the star]_{PP}.
 1 0 2 0 0 4 0 0

Kraljic and Brennan (2003) report results from a production study investigating materials like (19). In both (19a) and (19b), the LRB predicts a value of 4 at the word boundary between the post-verbal noun phrase and prepositional phrase. In both sentences, these constituents consist of 1 and 2 phonological phrases but vary in which constituent is longer and where the boundary between them occurs. An additional value of one is added for the phonological phrase boundary, yielding an LRB value of 4. Thus, the LRB predicts that an intonational boundary is most likely at the boundary of the post-verbal noun phrase and prepositional phrase. In accordance with this prediction, Kraljic and Brennan (2003) found that speakers placed intonational boundaries between “dog” and “in” in (19a) and between “basket” and “on” in (19b) when giving instructions to another participant. These boundaries were produced regardless of whether the context was ambiguous or not.

Note that the difference in LRB values between (19a) and (19b) at the critical points is greater than in the structures used by Snedeker and Trueswell (2003). The LRB value between “dog” and “in” is 4 in (19a) and only 2 in (19b). The value between “basket” and “on” is 4 in (19b) but

only 2 in (19a), yielding differences of 2 in each case. The fact that people reliably disambiguated the examples in (19) but not the examples in (18) in Snedeker & Trueswell's experiment therefore provides preliminary support for the hypothesis that prosodic disambiguation may be more likely when the LRB values of two interpretations of an utterance differ by a larger amount.

Interestingly, in an additional experiment conducted by Snedeker and Trueswell (2003), speakers did use intonational boundaries to disambiguate sentences like (18) when the referential context supported both readings of the ambiguity. Speakers placed boundaries after "tap" in the NP attachment condition and after "frog" in the VP attachment condition. Other researchers have found similar results, suggesting that the production of prosodically disambiguating cues may depend on factors such as speaker awareness of the ambiguity, the presence of other disambiguating cues, and the type of experimental task (Albritton, McKoon, & Ratcliff, 1996; Fox-Tree & Meijer, 2000; Price et al., 1991). Thus, although speakers place boundaries at points predicted by the LRB, speakers may place intonational boundaries in dispreferred locations in order to signal constituency groupings to listeners. Thus a speaker may place a boundary after the verb in (18a) (i.e., "Tap // the frog with the flower") to signal that the upcoming prepositional phrases are grouped into an NP even though the production system may not require a boundary because of the shortness of the sentence and the semantic relationship between "frog" and "tap".

The complexity of these data suggest that a wide array of factors play a role in determining where an intonational boundary will occur in any given utterance. Not only does syntactic structure influence intonational phrasing, speakers' awareness of listeners' needs may also play a role, as well as factors such as discourse structure, speaker style, and information structure. The LRB provides a first step in characterising some of the structural factors that influence intonational phrasing.

Manuscript received May 2003

Revised manuscript received March 2004

REFERENCES

- Abrittton, D., McKoon, G., & Ratcliff, R. (1996). Reliability of prosodic cues for resolving syntactic ambiguity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 714-735.
- Altmann, G. T. M., & Steedman, M. J. (1988). Interaction with context during human sentence processing. *Cognition*, 30, 191-238.
- Bader, M. (1998) Prosodic influences on reading syntactically ambiguous sentences. In J. D. Fodor & F. Ferreira (Eds.), *Reanalysis in sentence processing*. Dordrecht: Kluwer.

- Beach, C. M. (1991). The interpretation of prosodic patterns at points of syntactic structure ambiguity: Evidence for cue trading relations. *Journal of Memory and Language*, 30, 644–663.
- Beckman, M. E., & Pierrehumbert, J. (1986). Intonational structure in Japanese and English. *Phonology Yearbook*, 3, 255–309.
- Blaauw, E. (1994). The contribution of prosodic boundary markers to the perceptual difference between read and spontaneous speech. *Speech Communication*, 14, 359–375.
- Bock, K., & Levelt, W. (1994). Language production: Grammatical encoding. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 945–984). New York: Academic Press.
- Bolinger, D. (1972). Accent is predictable (if you're a mind reader). *Language*, 48, 633–644.
- Carlson, K., Clifton, C., Jr., & Frazier, L. (2001). Prosodic boundaries in adjunct attachment. *Journal of Memory and Language*, 45, 58–81.
- Chafe, W. (1976). Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In C. Li (Ed.), *Subject and topic* (pp. 25–55). New York: Academic Press.
- Chafe, W. (1987). Cognitive constraints on information flow. In R. Tomlin (Eds.), *Coherence and grounding in discourse* (pp. 21–51). Amsterdam: John Benjamins.
- Cooper, W. E., & Paccia-Cooper, J. (1980). *Syntax and speech*. Cambridge, MA: Harvard University Press.
- Cutler, A., Dahan, D., & Van Donselaar, W. A. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech*, 40 (2), 141–202.
- Dahan, D., Tanenhaus, M. K., & Chambers, C. G. (2002). Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language*, 47, 292–314.
- Diesing, M. (1992). *Indefinites*. Cambridge, MA: MIT Press.
- Du Bois, J. (1987). The discourse basis of ergativity. *Language*, 63, 805–855.
- Ferreira, F. (1988). *Planning and timing in sentence production: The syntax-to-phonology conversion*. Unpublished dissertation, University of Massachusetts, Amherst, MA.
- Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. *Journal of Memory and Language*, 30, 210–233.
- Ferreira, F. (1993). Creation of prosody during sentence prosody. *Psychological Review*, 100, 233–253.
- Ferreira, V. (1996). Is it better to give than to donate? Syntactic flexibility in language production. *Journal of Memory and Language*, 35, 724–755.
- Fodor, J. (1998). Learning to parse? *Journal of Psycholinguistic Research*, 27, 285–319.
- Fodor, J. (2002). Prosodic disambiguation in silent readings. *North Eastern Linguistic Society*, 32, 113–132.
- Fox-Tree, J. E., & Meijer, P. J. A. (2000). Untrained speakers' use of prosody in syntactic disambiguation and listeners' interpretations. *Psychological Research*, 63, 1–13.
- Fox, B. A., & Thompson, S. A. (1990). Relative clauses in English conversation. *Language*, 66, 297–316.
- Gee, J. P., & Grosjean, F. (1983). Performance structures. A psycholinguistic and linguistic appraisal. *Cognitive Psychology*, 15, 411–458.
- Givón, T. (1979). *On understanding grammar*. New York: Academic Press.
- Givón, T. (Ed.). (1983). *Topic continuity in discourse*. Amsterdam: John Benjamins.
- Givón, T. (1984). *Syntax: A functional typological introduction* (Vol. 1). Amsterdam: John Benjamins.
- Grodner, D. J., Gibson, E., & Watson, D. (in press). The influence of contextual contrast on syntactic processing: Evidence for strong-interaction on sentence comprehension. *Cognition*.
- Gussenhoven, C. (2002). *Intonation and interpretation: Phonetics and phonology*. In Proceedings of Speech Prosody 2002, Aix-en-Provence.

- Hirschberg, J., & Avesani, C. (1997). *The role of prosody in disambiguating potentially ambiguous utterances in English and Italian*. Paper presented at the ESCA Tutorial and Research Workshop on Intonation, Athens, 1997.
- Jackendoff, R. (1977). *X-bar syntax*. Cambridge, MA: MIT Press.
- Kjelgaard, M., & Speer, S. (1999). Prosodic facilitation and interference in the resolution of temporary syntactic ambiguity. *Journal of Memory and Language*, 40, 153–194.
- Kraljic, T., & Brenna, S. E. (2003). Prosodic disambiguation of syntactic structure: For the speaker or for the addressee? Paper presented at the 16th annual CUNY conference on Human Sentence Processing in Cambridge, MA.
- Kratzer, A. (1995). Stage level and individual level predicates. In G. Carlson & F. J. Pelletier (Eds.), *The generic book* (pp. 125–175). Chicago: University of Chicago Press.
- Lehiste, I., Olive, J., & Streeter, L. (1976). Role of duration in disambiguating syntactically ambiguous sentences. *Journal of the Acoustical Society of America*, 60, 1199–1202.
- Levelt, W. J. M. (1989). *Speaking. From intention to articulation*. Cambridge, MA: MIT Press.
- Marslen-Wilson, W. S., Tyler, L. K., Warren, P., Grenier, P., & Lee, C. S. (1992). Prosodic effects in minimal attachment. *Quarterly Journal of Experimental Psychology*, 45A, 73–87.
- Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. Dordrecht: Foris Publications.
- Olkin, I. (1967). Correlation Revisited. In J. C. Stanley (Ed.), *Improving experimental design and statistical analysis*. Chicago: Rand McNally.
- Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. Cohen, J. Morgan, & M. Pollack (Eds.), *Intentions in communications*. Cambridge, MA: MIT Press.
- Pitrelli, J., Beckman, M., & Hirschberg, J. (1994). Evaluation of prosodic transcription labelling reliability in the ToBI framework. In *Proceedings of the 1994 international conference on spoken language processing*, Vol. 1, pp. 123–126.
- Price, P. J., Ostendorf, M., Shattuck-Hufnagel, S., & Fong, C. (1991). The use of prosody in syntactic disambiguation. *Journal of Acoustical Society of America*, 90, 2956–2970.
- Prince, E. F. (1981). Toward a taxonomy of given-new information. In P. Cole (Ed.), *Radical pragmatics* (pp. 223–255). New York: Academic Press.
- Roelofs, A. (1998). Rightward incrementality in encoding simple phrasal forms in speech production: Verb-particle combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 904–921.
- Schafer, A. J. (1997). *Prosodic parsing: The role of prosody in sentence comprehension*. University of Massachusetts Doctoral Dissertation, forthcoming. Amherst: GLSA.
- Schafer, A. J., Speer, S. R., Warren, P., & White, S. D. (2001). Prosodic influences on the production and comprehension of syntactic ambiguity in a game-based conversation task. *Fourteenth Annual CUNY Conference on Human Sentence processing*. Philadelphia, PA.
- Schütze, C., & Gibson, E. (1999). Argumenthood and English prepositional phrase attachment. *Journal of Memory and Language*, 40, 409–431.
- Selkirk, E. O. (1978). On prosodic structure and its relation to syntactic structure. In T. Freithem (Ed.), *Nordic Prosody II*, Trondheim: TAPIR.
- Selkirk, E. O. (1984). *The relation between sound and structure*. Cambridge: MIT Press.
- Selkirk, E. O. (1986). On derived domains in sentence phonology. *Phonology Yearbook*, 3, 371–405.
- Selkirk, E. O. (2000). The interaction of constraints on prosodic phrasing. In M. Horne (Ed.), *Prosody: Theory and experiment*. Amsterdam: Kluwer.
- Shattuck-Hufnagel, S., & Turk, A. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research*, 25, 193–247.
- Silverman, K., Beckman, M., Pierrehumbert, J., Ostendorf, M., Wightman, C., Price, P., & Hirschberg, J. (1992). ToBI: A standard scheme for labelling prosody. In *Proceedings of*

- the Second International Conference on Spoken Language processing (ICSLP)*, Bariff, pp. 867–879.
- Snedeker, J., & Trueswell, J. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential contest. *Journal of Memory and Language*, 48, 103–130.
- Speer, S. R., Kjelgaard, M. M., & Dobroth, K. M. (1996). The influence of prosodic structure on the resolution of temporary syntactic closure ambiguities. *Journal of Psycholinguistic Research*, 25, 247–268.
- Steedman, M. (2000). Information structure and the syntax-phonology interface. *Linguistic Inquiry*, 31, 649–689.
- Sternberg, S., Monsell, S., Knoll, R. L., & Wright, C. E. (1978). The latency and duration of rapid movement sequences: Comparisons of speech and typewriting. In G. E. Stelmach (Ed.), *Information processing in motor control and learning*. New York: Academic Press.
- Terken, J., & Nootboom, S. G. (1987). Opposite effects of accentuation and deaccentuation on verification latencies for given and new information. *Language and Cognitive Processes*, 2, 145–163.
- Truckenbrodt, H. (1999). On the relation between syntactic phrases and phonological phrases. *Linguistic Inquiry*, 30, 219–255.
- Wheeldon, L. & Lahiri, A. (1997). Prosodic units in speech production. *Journal of Memory and Language*, 37, 356–381.