Order Effects in Production and Comprehension of Prosodic Boundaries

Foltz, A.; Maday, K.; Ito, K.; Frota, S.; Elordieta, G.; Prieto, P.

Prosodic Categories: Production, Perception and Comprehension: Production, Perception and Comprehension (Studies in Natural Language and Linguistic Theory)

Published: 27/02/2011

Order Effects in Production and Comprehension of Prosodic Boundaries

Anouschka Foltz, Kristine Maday and Kiwako Ito

Ohio State University

Correspondence:

Anouschka Foltz
Department of Linguistics
Ohio State University
222 Oxley Hall, 1712 Neil Ave.
Columbus, OH 43210-1298
anouschka@ling.ohio-state.edu
Phone: (614) 292-1841

Abstract

Two experiments investigate the effect of sentence familiarity and constituent length on the production of prosodic boundaries in syntactically ambiguous sentences such as *The brother*$_{N1}$ *of the bridegroom*$_{N2}$ *who swims*$_{RC}$ *was last seen on Friday night*, where the person swimming can be either the brother (high attachment) or the bridegroom (low attachment). Participants read aloud sentences with short and long N1s and RCs either on the fly or after forced disambiguation. Productions were coded for prosodic boundary strength at N1 and N2 using the ToBI annotation system. The results suggest that when reading aloud unfamiliar sentences, local syntactic cues drive prosodic structure and this structure does not guide sentence interpretation after the sentence is fully parsed. When reading familiar sentences, readers make rhythmic adjustments and often produce prosodic phrasing that informs their interpretation of the sentence. These findings suggest that prosodic phrasing of read speech only informs the message if the sentence had previously been fully parsed.

X.1 Introduction

A vast literature in psycholinguistics deals with the processing of ambiguity in language (see Altmann 1998 for an overview). Studies have focused mainly on lexical and structural ambiguity. One extensively studied structural ambiguity involves a complex noun phrase modified by a relative clause (RC), as in *Someone shot the servant of the actress who was on the balcony* (see Miyamoto 2008 for a list of studies). Here, the RC *who was on the balcony* can either modify the first noun (N1: *servant*) or second noun (N2: *actress*) of the complex NP, giving the respective interpretations that the servant was on the balcony or that the actress was on the balcony. The RC attaches higher in the syntactic tree (*high attachment*) when modifying the first noun than when modifying the second noun (*low attachment*). We will call this construction the RC attachment ambiguity.

This construction has generated interest for two reasons. First, the preferred interpretation varies across languages (Cuetos and Mitchell 1988), posing a problem for universal parsing principles such as Late Closure (Frazier 1978; Kimball 1973). Late Closure proposes that lexical items – if possible – are universally attached into the clause or phrase currently being processed. Here, the second noun is being processed as the parser encounters the RC and thus the RC should be preferably attached to it. Second, the length of the RC affects interpretation preferences, with more high attachment interpretations for long RCs (e.g. Fernández and Bradley 1999 for English). Prosody is claimed to play a crucial role in both of these phenomena (e.g. Fodor 1998, 2002a, 2002b). The cross-linguistic differences in preferred interpretation of ambiguous sentences may emerge as readers adopt language-specific prosodic phrasing during parsing. Also, the length of the RC may correlate with the likelihood of a prosodic juncture occurring before the RC, which in turn affects the likelihood of attachment choices (cf. Fodor 1998; 2002a; 2002b; Carlson, Clifton and Frazier 2001; Clifton, Carlson and Frazier 2002). Since many previous studies on attachment preferences are based on silent reading, where the implicitly produced prosodic patterns are not measurable, the relationship between such *implicit* prosody and the interpretation of sentences has been mostly speculative.

This study focuses on the issue of constituent length and *overt* prosodic phrasing in English – which shows an overall preference for low attachment (Carreiras and Clifton 1993; 1999; Clifton and Frazier 1996; Cuetos and Mitchell 1988; Fernández 2000; Fodor 1998; Frazier 1978). In particular, the study examines (a) whether constituent length affects prosodic phrasing when reading aloud, (b) whether readers' comprehension judgments are consistent with the prosodic phrasings they produce, and (c) whether the order of the two tasks (reading aloud and providing comprehension judgments) modulates these effects.

X.1.1 Production of ambiguous sentences.

Previous studies suggest that speakers consistently produce prosodic cues to convey the intended meaning of structurally ambiguous sentences in both conscious disambiguations and task-oriented natural speech. Clifton et al. (2002) have termed this the *rational speaker hypothesis*.

In an early study, Lehiste (1973) instructed speakers to produce sentences containing various structural ambiguities in a way that disambiguated them. Phonetic analyses revealed that speakers used prosody – in particular, lengthening – to disambiguate sentences with more than one possible surface bracketing (e.g. *[Steve or Sam] and Bob will come* vs. *Steve or [Sam and Bob] will come* or *[The old men] and women stayed at home* vs. *The old [men and women] stayed at home*). Listeners also successfully identified the intended interpretations.

Schafer, Speer, Warren and White (2000) used a cooperative game task with a predetermined set of utterances to elicit naturalistic speech. Speakers produced temporarily ambiguous sentences such as *When that moves the square will encounter a cookie* (Early Closure construction, where *moves* is intransitive) vs. *When that moves the square it should land in a good spot* (Late Closure construction, where *moves* in transitive). Speakers consistently disambiguated these temporarily ambiguous utterances, even when the context fully disambiguated them: Over 80% of utterances with Early Closure syntax (intransitive verb) had the stronger prosodic boundary after *moves*, and over 70% of utterances with Late Closure syntax (transitive verb) had the stronger boundary after *square*. A following comprehension experiment revealed that listeners successfully determined the intended syntactic structure.

Bradley, Fernández and Taylor (2003) tested how constituent length affects word duration in productions of the RC attachment ambiguity. Speakers produced the target utterances by combining two sentences (1 a and b) into a more complex one (1c). Results showed longer N2 (*prince*) durations for sentences with long RCs compared to those with short RCs. The longer word durations suggested that speakers more often placed a prosodic boundary at N2 when the RC was long (cf. Watson and Gibson 2004; 2005, who found more frequent prosodic boundaries after longer than after shorter constituents in unambiguous sentences).

(1a) The (unusual) plot concerns the guardian of the prince.

(1b) The prince was exiled (from the country for decades)

(1c) The (unusual) plot concerns the guardian of the prince who was exiled (from the country for decades).

Jun (2003) investigated the prosody of the sentence *Someone shot the servant of the actress who was on the balcony* in five production conditions across seven languages. In the *default reading* condition, informants read the sentence silently, answered the comprehension question *Who was on the balcony?*, and then produced the sentence twice. In all languages, the prosodic phrasings of the sentence reflected the previous RC attachment decisions. Informants who gave high attachment interpretations produced a prosodic boundary between N2 and the RC, whereas those who gave low attachment interpretations did not produce a prosodic boundary after N2. Additional informants produced a *default reading* of the sentence before giving an interpretation, and their phrasing followed the patterns described above. However, Jun does not report if they subsequently interpreted the sentence. Instead, she claims that readers' default phrasings – before or after interpretation – reflect their attachment preferences.

These production studies demonstrate that structural ambiguity and constituent length both affect lengthening of phrase-final words and insertion of boundaries. At first glance, they all seem to confirm a rather straightforward mapping between the sentential semantics and prosodic phrasing. However, the speakers in these studies always had a concrete interpretation of the sentences in mind before producing them (whether or not they were aware or made aware of the ambiguity). Thus, it remains unclear whether the constituent length has a clear effect on the prosodic pattern even when speakers must read aloud the sentence while parsing its structure.

X.1.2 Prosody of Reading Aloud

When reading a text aloud, people generally pronounce each word as soon as it is recognized within their visual span. While reports on the prosody during reading aloud are sparse, they do suggest that readers' prosody is generated before the global syntactic structure of a sentence is determined.

Kondo and Mazuka (1996) measured eye-voice span (the distance in letters/characters between where the reader is looking and what the reader is producing) as participants were reading aloud Japanese sentences. They found that eye-voice span was only about 2.5 characters[1] regardless of the sentence's syntactic complexity. They concluded that the readers' prosody was based on limited, local syntactic information rather than a global syntactic analysis. Levin (1979) found that the eye-voice span for English is about 18 letters for good adult readers – which on average corresponds to about three or four words. Since the speaker would utter the words before grasping the overall sentential structure and meaning, the prosodic phrasing of globally ambiguous sentences may not reflect the ultimate message structure the parser achieves. Rather, it may reflect the sensible grouping of words at hand according to the local syntactic relations generated through the incremental process.

Koriat, Greenberg and Kreiner (2002) also suggest that reading prosody represents local structural analysis, and precedes more complete semantic analysis. Their reading study used normal (e.g. *The fat cat with the gray stripes ran quickly…*) and nonsensical (e.g. *The sad gate with the small electricity went carefully…*) Hebrew sentences. Speakers produced prosodic patterns consistent with the local syntactic structure when reading unfamiliar sentences. In addition, semantic modifications (normal vs. nonsense sentences) did not significantly change pause patterns, suggesting that the structure of the sentence rather than its content or semantic coherence modified reading prosody. Their study thus shows that when complete semantic analysis is impeded, local structural analysis alone guides reading prosody. However, their study does not rule out that semantic analysis may modify reading prosody, e.g. in ambiguous sentences.

In sum, the above studies suggest that the prosodic patterns produced after the sentence is fully interpreted may not mirror the prosodic patterns produced while the sentence is parsed. Rather, when first reading *unfamiliar* ambiguous sentences, the local syntactic structure alone may determine readers' prosodic phrasings. Readers' subsequent comprehension judgments (after sentences are fully parsed) may not be consistent with the reading prosody produced during parsing.

X.1.3 Effect of prosody on interpreting ambiguous sentences.

A series of research by Clifton and colleagues demonstrated the effects of boundary strengths and locations on ambiguity resolution (Carlson et al. 2001; Clifton et al. 2002; 2006). They found that the interpretation of the ambiguous structures depended on the *relative* strengths of boundaries at all relevant locations and not on one major boundary's absolute strength. Clifton et al. (2002) manipulated the strength of the prosodic boundary after the head noun of the complex NP, as shown in (2).

(2) I met the daughter$_{[0/ip/IP]}$ of the colonel$_{ip}$ who was on the balcony.

(*0* = no boundary, *ip* = intermediate phrase boundary, IP = intonation phrase boundary, where *ip* is perceptually weaker than *IP*)

If only the local boundary at *colonel* affected the RC attachment preferences, the number of high attachment interpretations would be the same across the three conditions. However, the [0 ip] sequence elicited more high attachment choices than [ip ip], which in turn elicit more high attachment interpretations than [IP ip]. The results suggest that the interpretation of ambiguous structures is guided not by the absolute strength of a critical local boundary (as suggested by Price, Ostendorf and Shattuck-Hufnagel 1991 or Marcus and Hindle 1990), but by the relative strength of the boundary in comparison to other relevant boundaries (cf. Schafer et al. 2000 above).

Clifton et al. (2006) manipulated constituent length and found that prosodic boundaries affected listeners' interpretations more when the ambiguous part of the sentence contained shorter constituents. They proposed that listeners considered prosodic boundaries to be less informative when they were produced with long constituents than with short constituents. In particular, speakers tend to produce long constituents with a preceding and following prosodic boundary. Watson and Gibson proposed that this phenomenon is related to performance: long constituents require more time for planning and recovery – time that can be gained by producing a prosodic boundary before and after the constituent. Listeners may thus

[1] Japanese orthography consists of logographic (Chinese Kanji) and syllabic (Kana/Katakana) characters.

take boundaries preceding or following long constituents to reflect constituent length and the performance requirements associated with it rather than syntactic structure that disambiguates the sentential meaning.

X.1.4 Silent Reading Studies.

Silent reading studies on sentence ambiguity have focused on how RC length affects the interpretation of the RC attachment ambiguity (for English, see Bradley et al. 2003; Fernández and Bradley 1999; Hemforth, Fernández, Clifton, Frazier and Konieczny 2005). In these studies, participants silently read sentences and answered a comprehension question that gauged how the ambiguous constituent was attached. Across languages, longer RCs elicited more high attachment interpretations than shorter RCs. To explain these results, Fodor (2002a; 2002b) proposed the Implicit Prosody Hypothesis (IPH), given in (3).

(3) "In silent reading, a default prosodic contour is projected onto the stimulus, and it may influence syntactic ambiguity resolution. Other things being equal, the parser favors the syntactic analysis associated with the most natural (default) prosodic contour for the construction. (2002b, p. 113)" "Even in reading, prosody is present. Even in silent reading, and even if prosody-marking punctuation is absent. Prosody is mentally projected by readers onto the written or printed word string. And – the crucial point – it is then treated as if it were part of the input, so it can affect syntactic ambiguity resolution in the same way as overt prosody in speech does (2002a, p. 83)"

The implicit prosody account assumes that long RCs are more likely to be set off in their own prosodic phrase than short RCs (cf. Clifton et al. 2006). This leads to a prediction that readers are more likely to project an implicit prosodic boundary after N2 (i.e. immediately before the RC) if the following RC is long. The presence of a prosodic boundary in this location may prompt more high attachment interpretations[2].

[2] Many of the silent reading studies focused on local boundary strength (presence or absence of a boundary at N2) or assumed that a prosodic boundary would be projected *either* after N1 *or* after N2. Nevertheless, the results are consistent with an approach focusing on relative boundary strength since the presence of a boundary after N2 heightens the likelihood of it being the stronger of the relevant boundaries.

Since implicit prosody in silent reading cannot be measured directly, production data has often been used to support silent reading results. The implicit prosody during silent reading is assumed to match the measurable overt prosody from a production task (e.g. Fodor 1998; 2002a; 2002b; Jun 2003). However, most production studies provided interpretations of the ambiguous sentences before participants produced them (e.g. Bradley et al. 2003; Jun 2003).

X.1.5 The present study.

The present study comprises two experiments on the production and comprehension of sentences with an RC attachment ambiguity, for example, *The brother*$_{N1}$ *of the bridegroom*$_{N2}$ *who swims*$_{RC}$ *was last seen on Friday night*. The production experiments investigate whether constituent length and the presence or absence of firm interpretation affects prosodic boundary placement while reading aloud such sentences. Across the two experiments, participants read aloud ambiguous sentences either before or after they chose the interpretations. In the pre-interpretation reading condition novel sentences are read aloud on the fly and are afterwards interpreted. It is assumed that the eye-voice span during this task mimics that reported in Levin (1979), and that the prosodic pattern produced during this task better simulates the prosody during silent reading than the prosody elicited by a post-interpretation production task, where novel sentences are read silently, interpreted, and then read aloud.

The length of N1 and RC was manipulated within each experiment. If constituent length modulated prosody during reading on the fly as well as during the articulation after comprehending the sentence, it would suggest that RC length shapes implicit prosody during silent reading, as the IPH proposes. If the length effect was confirmed only in the prosodic pattern produced after comprehension, it would suggest that constituent length modulates prosody only when the global sentential structure and its meaning are fully established.

The comprehension experiments tested how the prosodic patterns elicited during the two production experiments affect the interpretation of the sentences. If the relative strength of prosodic boundaries at N1 and N2 consistently reflect the message structures, a tight correspondence between the interpretation preferences and the prosodic patterns should be confirmed in the comprehension-first experiment. It remains a question of great interest whether the prosodic patterns produced before knowing the entire structure affects the following interpretation of the sentence. If the interpretation of a sentence is guided by the prosodic phrasing of the articulation on the fly, prosody is confirmed to give robust cues to sentence comprehension, as suggested by Fodor (2002b). In contrast, if the interpretation of the sentence is independent of the preceding prosody, such a result would indicate that prosodic patterns generated during the initial parsing are not processed as an informative cue to the comprehension of the entire sentential structure and meaning. In other words, prosody may be selectively used for sentence comprehension depending on how it is generated.

X.2 Experiment 1:

Experiment 1 examined the prosody of ambiguous sentences unfamiliar to the reader by asking participants to read aloud sentences on a computer screen before answering a comprehension question gauging interpretation. In this task, readers were expected to build an interpretation of the sentence as they read it aloud. The experiment tested whether the length of N1 and the RC affect prosodic phrasing and whether differences in prosodic phrasing affect readers' own subsequent interpretation of the sentences.

Based on previous work (Carlson et al. 2001; Clifton et al. 2002; 2006; Fodor 1998; 2002a, 2002b; Watson and Gibson 2004; 2005), we hypothesized that long N1s and RCs should more frequently be set-off in their own prosodic phrase than short N1s and RCs. Thus, sentences with a long N1 and short RC were predicted to elicit relatively more frequent insertion of a strong boundary after N1 than after N2. Similarly, sentences with a short N1 and a long RC were predicted to elicit relatively more frequent strong boundaries after N2 than after N1. When N1 and RC are either both short or both long, frequent insertions of equal boundaries were predicted.

If the prosodic phrasing of participants' own speech serves as input for their subsequent sentence interpretation, a stronger prosodic boundary after N1 than after N2 should result in more frequent low attachment judgments than other reading prosodies. Also, a stronger prosodic boundary after N2 than after N1 should lead to fewer low attachment interpretations than other prosodic phrasing patterns. When the sentence is produced with equally strong boundaries at these locations, the number of low attachment interpretations should fall in the intermediate range between the other two phrasing patterns. These predictions are summarized in Table 1:

[insert Table 1 about here]

X.2.1 Participants

Sixteen undergraduate students at the Ohio State University participated in the study for partial course credit. They were all native speakers of, mostly Midwestern, American English. None of them reported any speech or hearing disabilities.

X.2.2 Materials

Twenty-four ambiguous target sentences and comprehension questions were constructed. Each sentence contained an RC modifying a complex NP, as in *The brother of the bridegroom who swims was last seen on Friday night*. Here, either the brother (high attachment) or the bridegroom (low attachment) can be the one who swims. Length of N1 (*brother*) and RC (*who swims*) were manipulated to construct four versions of each sentence, as in (4). A short N1 always consisted of two syllables, while a long N1 had four syllables. A long N1 either contained one pre-nominal modifier before the same head noun as that in the short N1 (e.g. lawyer → defense lawyer) or was a four-syllable (compound) noun semantically related to the short N1 (e.g. nanny → babysitter). A long RC contained an adjunct phrase that modifies the same verb as that in the short RC. All short RCs consisted of two or three syllables; all long RCs had five or six syllables. The target comprehension question always gauged participants' interpretation of the ambiguity, e.g. *Who swims (like a fish)?*

(4a) Short N1 and Short RC:

The *brother* of the bridegroom *who swims* was last seen on Friday night.

(4b) Short N1 and Long RC:

The *brother* of the bridegroom *who swims like a fish* was last seen on Friday night.

(4c) Long N1 and Short RC:

The *second cousin* of the bridegroom *who swims* was last seen on Friday night.

(4d) Long NP1 and Long RC:

The *second cousin* of the bridegroom *who swims like a fish* was last seen on Friday night.

In addition, 76 filler sentences with various syntactic structures were constructed. Filler sentences were either unambiguous, contained a temporary ambiguity, or contained a global ambiguity different from the RC attachment ambiguity. Many globally ambiguous filler sentences were heavily semantically biased. None of the filler comprehension questions referred to any ambiguous part of the sentence. The most common interrogative pronoun in the filler questions was *what* (47 questions), followed by *who* (24 questions) to ensure that participants could not anticipate the interrogative pronoun and adjust their prosody in anticipation of the comprehension question. Four lists of 100 sentences were created such that each list contained only one version of the 24 target sentences.

X.2.3 Procedure

Participants were seated in a soundproof booth in front of a computer screen and a response box and wore a head-mounted microphone. Their productions were recorded while reading aloud 100 sentences presented one by one on the computer screen. The sentences and the comprehension questions were presented using E-Prime v.1.0 (Psychology Software Tools 2003). Participants initiated each trial by pushing a button on the response box in front of them. Upon pushing the button, a sentence appeared on the computer screen. Even though they were not given any specific instruction as to when they should begin reading, the timing of the button push and sentence production onset reveal that all participants began reading aloud each sentence immediately after it appeared on the screen, without practicing.

After reading each sentence, participants pressed a button on the response box to move on to a comprehension question about the sentence. Three response options were presented underneath the question. Participants responded by pushing the button that corresponded to their answer choice. The question always inquired about who performed the action in the RC, e.g. *Who swims (like a fish)?* The response options always corresponded to NP1 (e.g. *the brother / the second cousin*), NP2 (e.g. *the bridegroom*), and "I don't know". Participants' response to the comprehension question prompted the beginning of the next trial. Before the experiment, participants had the opportunity to practice the task and ask any questions about the instructions given to them.

X.2.4 ToBI Coding

The productions were coded using the ToBI (Tones and Break Indices; see Beckman and Hirschberg 1993; Beckman and Ayers 1997; Silverman et al. 1992) intonation transcription system for Standard American English (SAE). The ToBI system for SAE is based on the autosegmental metrical theory of intonation originally proposed by Pierrehumbert (1980) (see Beckman, Hirschberg and Shattuck-Hufnagel for a history of the ToBI framework). The theory assumes two levels of phrasal boundaries for English: perceptually weaker intermediate phrase (ip) boundaries and perceptually stronger intonational phrase (IP) boundaries.

Only the ToBI break indices were used to code the productions[3]. Every word was given a break index that indicated the strength of prosodic juncture at its right edge. The following indices were used:

(8) ToBI Break Indices:

1 = word boundary

2 = hesitation

3 = intermediate phrase (ip) boundary (perceptually weaker boundary)

4 = Intonational Phrase (IP) boundary (perceptually stronger boundary)

The first author of the paper, who is trained in the ToBI system, transcribed all productions. A second coder unaffiliated with the project transcribed a small subset of productions. Intertranscriber reliability of the coders is 89% for all words and 79% for the critical words (N1 and N2) (cf. intertranscriber reliabilities in Syrdal and McGory 2000 and Dilley et al. 2006). For data-internal consistency, the data reported here is based on the first coder's transcriptions.

X.2.5 Phonetics

Even though ToBI labeling yielded relatively stable results across labelers, it is a subjective measure. One of the most consistent phonetic cues to a prosodic boundary is pre-boundary lengthening (e.g. Lehiste 1973). We therefore measured duration of N1 and N2 to support the ToBI coding of the data (Figure 1). The words that received higher break indices tended to have longer durations: The mean and median value of the distribution rises for each higher break index. Pairwise comparisons (Tukey contrasts) revealed that at N1 words with no boundary were reliably shorter than words with an ip boundary ($z = 2.9$, $p < 0.05$) or

[3] In a production study that used the same task as Experiment 1, Bergmann, Armstrong, and Maday (2008) compared the production of sentences like *Someone shot the servant of the actress who was standing on the balcony* in English and Spanish. They coded prosodic boundary strength and prosodic boundary type at the verb, N1 and N2. They found that boundary strength in English different at the two sentence locations N1 and N2 with more IP boundaries at N2 than N1. Boundary type at N1 and N2, on the other hand, were comparable: Almost all ip boundaries at N1 and N2 had a H- phrase accent and over 80% of IP boundaries had the patterns H-L% (floor-holding pattern) or L-H% (continuation rise). These boundary types all convey that the speaker is not done speaking (cf. Pierrehumbert & Hirschberg, 1990). Prosodic boundaries at the end of the sentence were overwhelmingly L-L%, indicating finality. The boundaries at N1

IP boundary ($z = 7.0$, $p < 0.001$). Words with an ip boundary were marginally shorter than words with an IP boundary ($z = 2.2$, $p = 0.07$). At N2 words with an IP boundary were reliably longer than words with either an ip boundary ($z = 7.8$, $p < 0.001$) or no boundary ($z = 7.4$, $p < 0.001$).

[insert Figure 1 about here]

The results of a correlation analysis are found in Table 2. The analysis showed only a weak positive correlation of prosodic boundary strength and word duration. Only between 12% and 14% of the variability in the data can be accounted for by the strength of prosodic boundary. However, word duration correlates rather strongly with the words' numbers of segments, which accounts for 73% of the variability in duration of N1 and 51% of the variability of N2. Finally, word frequency of N2 (using Kučera-Francis written frequency counts, Kučera and Francis 1967) showed a weak negative correlation with word duration, accounting for 23% of the variability of N2 durations. Together, strength of prosodic boundary, number of segments, and word frequency account for at least 85% of the variability in word duration. We suggest that much of the remaining variability can be accounted for by individual differences in reading speed.

[insert Table 2 about here]

X.2.6 Results

X.2.6.1 Prosodic phrasing

The experiment elicited a possible total of 384 utterances. However, participants failed to correctly produce a total of six utterances. Six further utterances were excluded because they contained a disfluency at a critical location. The prosodic patterns of the remaining 372 utterances were categorized into three groups according to the location and relative strength of the boundaries. Sentences with a stronger prosodic boundary at N1 than at N2 (*IP*/*ip* or *ip*/*0*) were categorized as Stronger Break Follows (SBF) N1, and those with a stronger boundary at N2 than at N1 (*ip*/*IP* or *0*/*ip*) were labeled SBF N2. Sentences with equally strong prosodic boundaries at both sentence locations (*IP*/*IP*, *ip*/*ip* or *0*/*0*) were categorized as Equally Strong Breaks (ESB). These global prosodic patterns, which capture the relative boundary strengths at the relevant

---

and N2 thus seem to reflect the fact that the sentence is not over. We therefore focus on boundary strength in our analysis.

sentence locations, should be related to how the sentences are interpreted (cf. Carlson et al. 2001, Clifton et al. 2002). Figure 2 compares the occurrences of these prosodic patterns across the four sentence types. Table 3 shows the results from a multinomial logistic regression predicting the location of the stronger boundary from N1 and RC length. The results reveal that sentences with Short N1s (as opposed to Long N1s) were produced reliably more often with ESB ($p < .01$) and SBF N2 ($p < .001$) than with SBF N1. No such differences were found for the RC length manipulation. The results suggest that the length of N1, but not the length of the RC, affects the relative strength of prosodic boundaries at N1 and N2 when people are reading the sentences on the fly.

[insert Figure 2 about here]

[insert Table 3 about here]

Next, we coded the local boundaries at N1 and N2 according to their strength (no boundary, *ip* boundary or *IP* boundary) across the different length conditions to see if longer constituents are more likely to be set off in their own prosodic phrase (as suggested by Fodor 1998; 2002a; 2002b; Watson and Gibson 2004; 2005). Figure 3a compares the distribution of boundary strengths at N1 for the Short N1 and Long N1 sentences, whereas Figure 3b compares the distribution of boundary strengths at N2 for the Short RC and Long RC sentences. Tables 4a and 4b show the results from multinomial logistic regressions predicting local boundary strength at N1 from N1 length and local boundary strength at N2 from RC length. Sentences with a long N1 were more likely to be produced with a prosodic boundary (either ip ($p<.05$) or IP ($p<.001$) as opposed to no boundary) after N1 than sentences with a short N1. This suggests that long N1s were more frequently produced as their own prosodic phrase than short N1s, as suggested by Fodor (1998; 2002a; 2002b) and Watson and Gibson (2004; 2005). However, the same relation between constituent length and prosodic phrasing did not hold for RCs: The insertion of a boundary after N2 was equally frequent across the sentences with a long RC and those with a short RC, i.e., a long RC did not increase the likelihood of a *preceding* prosodic juncture. Rather, participants produced reliably more IP boundaries than ip ($p<.001$) and more ip boundaries than no boundary ($p<.01$) at N2, regardless of RC length. In addition, the presence or absence of a prosodic boundary after N1 did not affect boundary placement after N2. Rather, RCs were mostly preceded by a prosodic boundary regardless of their length and regardless of whether readers had just produced a boundary at N1.

[insert Figures 3a and 3b about here]

[insert Tables 4a and 4b about here]

The present results do not support the hypothesis that long constituents are unconditionally more likely to be produced as their own prosodic phrase. While the length of N1 affected the location and relative strength of prosodic boundaries for the entire sentence, the length of the RC did not. In addition, only the length of N1 predicted the likelihood of a prosodic boundary after N1, but the length of the RC did not predict the likelihood of a prosodic boundary after N2. This suggests that prosodic phrasing of a sentence may be modulated by the length of only certain types of constituents.

One potential factor behind this discrepancy is the semantic and syntactic independence of the morphemes and words added to N1 and RC. While long N1s were nouns or compound nouns with merely more syllables than the nouns of the short N1s, long RCs always had additional words that formed a constituent on their own. As a result, there may have been more opportunities to produce a prosodic boundary within the RC (e.g. who swims // like a fish) than within N1 (e.g. second // cousin). This additional boundary between the verb and the adjunct phrase may have weakened the effect of the overall length of the RC. However, an additional multinomial logistic regression revealed that this was not the case. The Long RC sentences that were produced with a boundary after the verb were not less likely to be produced with a prosodic boundary after N2 than those without a boundary after the verb (all p-values > 0.1).

A more plausible reason for the asymmetric effect of the constituent length is the differences in the availability of length information during production. Note that a prosodic boundary after N1 was inserted *after* participants uttered the entire noun phrase, while a boundary after N2 was inserted *before* the RC for which the length was manipulated. Thus, participants may have had different degrees of certainty about the length of the constituents at those two locations. The null effect of the RC length on the likelihood of a boundary after N2 may reflect that participants had not established the length of the upcoming RC as certainly as they had established the length of N1. As readers produced N2, the entire RC was likely not always within their eye-voice span. It is plausible that participants could anticipate an RC following N2 due to the relative pronoun 'who', but they could not produce a prosodic boundary that would reflect the size of the following RC at this juncture.

X.2.6.2 Sentence Comprehension

A total of 372 responses to the comprehension questions were coded according to the participants' interpretation judgments: high-attachment ('the brother / second cousin swims'), low-attachment ('the bridegroom swims'), and *don't know*. Only six sentences received *don't know* responses: They were excluded from the analysis. The distributions of interpretation judgments were compared across the three prosodic patterns to examine the relation between the prosodic phrasing and the following interpretation of the sentences (Figure 4). Results from a multinomial logistic regression are found in Table 5. The results indicate that interpretation choices were not guided by the prosodic patterns participants produced, i.e., a stronger boundary after N1 than after N2 did not elicit more low attachment responses than a stronger boundary after N2 than after N1. Instead, all prosodic patterns elicited comparable proportions of low attachment and high attachment responses, and low attachment was the dominant interpretation after any of the three prosodic renditions. These results were somewhat surprising since they conflict with Jun's (2003) production data and since relative boundary strength exhibited much clearer effects on attachment preferences in Carlson et al. (2001) and Clifton et al. (2002).

[insert Figure 4 about here]

[insert Table 5 about here]

X.3 Experiment 2:

Experiment 2 examined the prosody produced after readers became familiar with the sentences. In this experiment, readers silently read the sentence and answered the comprehension question before they read it aloud. The experiment tests whether the length of N1 and the RC affect the interpretations of the sentences and whether the interpretations in turn affect readers' subsequent prosodic phrasing.

Experiment 1 showed that only the length of N1, but not of the RC, affected prosody when unfamiliar sentences were read aloud. If we assume that the implicit prosody produced during silent reading (i.e. in this experiment) is similar to the overt prosody in reading aloud on the fly (i.e. in Experiment 1), only the length of N1 should affect readers' implicit prosody. We would then expect more implicit prosodic boundaries after long N1s than after short N1s and frequent implicit prosodic boundaries after N2 regardless of RC length. According to the result from Experiment 1, RC length should not affect interpretation decisions.

Alternatively, implicit reading prosody may differ from overt reading prosody. In particular, silent reading may require fewer processing resources than reading aloud and may allow for more preview of the upcoming structure. If so, readers may have a better idea of the length of the RC when N2 is parsed in silent reading. As a result, even RC length may matter for the implicit prosodic phrasing. Thus, in silent reading, both long N1s and RCs may be set-off in their own prosodic phrase more often than short N1s and RCs. If this implicit prosody guides the interpretation, which in turn feeds the overt prosody, the phrasing patterns produced after silent reading and the comprehension response may better mirror the implicit prosody than the explicit prosody during reading on the fly.

If RC length modulates the implicit reading prosody that guides the interpretation, sentences with a short N1 and a long RC are expected to have a stronger implicit boundary after N2 than after N1 and elicit fewer low attachment interpretations than other sentence types. Sentences with a long N1 and a short RC are expected to have a stronger implicit boundary after N1 than after N2, and thus should elicit more low attachment interpretations than other sentence types. Sentences with a short N1 and a short RC or with a long N1 and a long RC should elicit an intermediate number of low attachment judgments. If sentence interpretation affects subsequent reading prosody low attachment interpretations should induce more productions of a stronger prosodic boundary after N1 than after N2 than high attachment interpretations. High attachment interpretations are expected to lead to more productions of a stronger boundary after N2 than after N1 than low attachment interpretations. These predictions are summarized in Table 6.

[insert Table 6 about here]

X.3.1 Participants

Sixteen undergraduate students from Ohio State University participated in the study for partial course credit. They were all native speakers of, mostly Midwestern, American English. None of them reported any speech or hearing disabilities. No participant in Experiment 2 had participated in Experiment 1.

X.3.2 Materials

The same 24 target items and 76 filler sentences as in Experiment 1 were used for Experiment 2. The comprehension questions were also identical to those used in Experiment 1.

X.3.3 Procedure

The procedure for Experiment 2 differed from Experiment 1 mainly in the order of events. During each trial, participants first read each sentence silently. They were instructed to make sure they understood the meaning of the sentence, and there was no time limit for reading the sentence. After silently reading the sentence, participants pushed a button on the response box to move on to the comprehension question with the response options. After responding to the question, the sentence reappeared on the screen and participants were recorded reading the sentence aloud. They were instructed to concentrate on getting the meaning of the sentence across.

X.3.4 ToBI Coding

The productions were coded as in Experiment 1.

X.3.5 Phonetics

The duration of N1 and N2 were measured for each relevant break index (Figure 5). The words that received higher break indices tended to have longer durations: The mean and median value of the distribution rises for each higher break index. Pairwise comparisons (Tukey contrasts) revealed that at N1words with an IP boundary were reliably longer than words with either an ip boundary ($z = 5.3$, $p < 0.001$) or no boundary ($z = 7.8$, $p < 0.001$) and that at N2 words with an IP boundary were reliably longer than words with either an ip boundary ($z = 5.1$, $p < 0.001$) or no boundary ($z = 8.5$, $p < 0.001$).

[insert Figure 5 about here]

Table 7 shows the results of a correlation analysis. Again we find only a weak positive correlation of prosodic boundary strength and word duration. Between 12% and 15% of the variability in the data can be accounted for by the strength of prosodic boundary. Again number of segments correlates fairly strongly with word duration, such that the strength of prosodic boundary, number of segments, and word frequency (for N2) together account for 83% or more of the variability in word duration.

[insert Table 7 about here]

X.3.6 Results

X.3.6.1 Sentence Comprehension

Only eight of the 384 interpretation judgments elicited *don't know* responses. These were therefore excluded from the analysis, leaving a total of 376 interpretation judgments. No sentence was excluded for disfluencies at critical locations. Figure 6 compares the distribution of interpretation judgments across the four types of sentences. While the overall dominance of low attachment interpretations was confirmed, the four sentence types elicited the predicted differences in interpretation patterns only numerically. Sentences with short N1s and long RCs elicited the fewest number of low attachment interpretations (76.8%). Sentences with long N1s and short RCs elicited the highest number of low attachment interpretations (87.2%). Finally, sentences for which both N1 and RC were either short or long elicited an intermediate number of low attachment interpretations (83% and 82.8%, respectively). However, these differences were not statistically reliable (see Table 8). Therefore, the effect of constituent length on the attachment preference does not seem to be as robust as previously reported (e.g. Bradley et al. 2003; Fernández and Bradley 1999; Hemforth et al. 2005).

[insert Figure 5 about here]

[insert Table 8 about here]

X.3.6.2 Sentence Interpretation and Prosodic Patterns

Recall that Experiment 1 showed no link between the interpretation choices and the global prosodic patterns that readers previously produced. In Experiment 2, readers' interpretation choices affected the global prosodic patterns they subsequently produced. Figure 7 shows the distribution of prosodic patterns produced for the sentences of each interpretation choice. The results of a multinomial logistic regression show that low attachment interpretations (compared to high attachment interpretations) more often elicited productions with the stronger boundary following N1 than with the stronger boundary following N2 (Table 9). They also elicited more productions with equally strong boundaries than with the stronger boundary following N2 ($p < 0.001$). That is, participants were reliably less likely to produce sentences with the stronger boundary following N2 when they had given the sentence a low attachment interpretation than when they had given it a high attachment interpretation.

[insert Figure 7 about here]

[insert Table 9 about here]

Contrary to our prediction, low attachment interpretations elicited a number of productions with equally strong boundaries after N1 and N2 (28.1%) and even with the stronger boundary following N2 (27.7%). In other words, low attachment interpretations did not elicit productions with the stronger boundary after N1 (only 44.2%) as often as expected. Recall that Figure 4 shows that without having been forced to provide an interpretation prior to production, readers strongly disprefer this prosodic pattern (15.8% of total productions). This general dispreference may explain the lower than expected number of productions with the stronger boundary following N1.

X.3.6.3 Constituent Length and Prosodic Patterns

The results further reveal that the global prosodic patterns depended on the lengths of the constituents. Figure 8 shows the distribution of the prosodic patterns elicited by each one of the four sentence types. As in Experiment 1, sentences with short N1s elicited fewer productions with the stronger prosodic boundary after N1 than with the stronger boundary after N2 than sentences with Long N1s. Unlike Experiment 1, RC length affected the global prosodic pattern. Sentences with Long RCs reliably elicited more productions with the stronger prosodic boundary after N2 than equally strong boundaries or the stronger boundary following N1 than sentences with Short RCs (Table 10). This suggests that when the lengths of the constituents are known before production and when the readers try to convey their interpretations, constituent length overall affects the prosodic patterns of the sentences.

Contrary to predictions, sentences with Short N1 and RC and Long N1 and RC did not elicit an 'intermediate' production pattern. Rather, sentences with Short N1 and RC elicited a pattern similar to sentences with Long N1 and Short RC, and sentences with Long N1 and RC elicited a pattern similar to sentences with Short N1 and Long RC.

[insert Figure 8 about here]

[insert Table 10 about here]

Next, the distributions of the boundary types were grouped according to the length of N1 (Figure 9a) and RC (Figure 9b). Statistical analysis confirms the effect of N1 length found Experiment 1, i.e., sentences with a long N1 were more likely to be produced with an IP boundary after N1 (compared to no boundary ($p < 0.001$) or an ip boundary ($p < 0.001$)) than sentences with a short N1(Table 11a). Long N1s are thus more frequently produced as their own prosodic phrases than short N1s.

Figure 9b demonstrates a pattern unlike that found in Experiment 1. Contrary to any predictions, sentences with Long RCs were less likely to be produced with a prosodic boundary after N2. Instead, sentences with Short RCs were more often produced with either an ip boundary ($p < 0.001$) or an IP boundary than no boundary ($p < 0.001$) than sentences with Long RCs (Table 11b).

Notice also that regardless of RC length, productions in Experiment 2 had much fewer prosodic boundaries after N2 (Figure 9b) than in the productions from Experiment 1 (Figure 3b). This suggests that readers' overall low attachment preference for these sentences is reflected in the absence of a boundary at

N2 when the sentences were previously interpreted (Experiment 2), but not when the sentences were not previously interpreted (Experiment 1).

[insert Figures 9a and 9b about here]

[insert Tables 11a and 11b about here]

Finally, unlike Experiment 1, boundary placement at N2 was modulated by the presence or absence of a prosodic boundary at N1 (Figure 10). Readers were less likely to produce an ip or IP boundary at N2 (as opposed to no boundary) if they had produced an ip or IP boundary at N1 (as opposed to no boundary) (Table 12).

[insert Figure 10 about here]

[insert Table 12 about here]

The results of Experiment 2 do not support the hypothesis that long constituents in general are more likely to be produced as their own prosodic phrase: while the length of N1 affected the local placement of a prosodic boundary after N1, it was the presence or absence of a boundary after N1, rather than the length of the RC, that affected the presence or absence of a boundary after N2. However, the results show that both the length of N1 and RC affect the relative strength of boundaries within a sentence.

X.4 Discussion

Combining the simple tasks to read aloud sentences and answer comprehension questions, the present study investigated how constituent length and the familiarity with the target sentences modify explicit reading prosody. Experiment 1 revealed that in the case of reading aloud unfamiliar sentences, constituent length affected the likelihood of boundary insertion only when information about the size of the constituent was fully established. In Experiment 2, the knowledge of constituent length affected the relative strength of prosodic boundaries more than the presence or absence of local boundaries. In addition, the location of a preceding prosodic boundary and readers' sentence interpretations both modulated reading prosody.

These results suggest different mechanisms for reading familiar and unfamiliar sentences. For unfamiliar sentences, local syntactic cues that become available within the span of preview seem to primarily guide the prosodic phrasing, as suggested by Kondo and Mazuka (1996) and Koriat et al. (2002). In Experiment 1, only the length of N1, but not of the RC, predicted the presence or the absence of boundaries and the relative strength of boundaries across the two locations (i.e., after N1/N2). Thus, a longer constituent was produced with a boundary only afterward, but not beforehand. We suspect that this asymmetry emerged because the certainty of locally available syntactic structure differed across the two locations. When readers fixated the first content word of the sentence, the size of N1 was probably fully noticed, as the following preposition 'of' must have been reviewed within the para-foveal window – which is known to span 14-15 character spaces (for a review, see Rayner 1975; 1992; 1998). When the reader detected that the subject noun phrase contained a following PP, the longer letter string or another content word that required the articulation of additional two syllables might have prompted the insertion of a boundary after N1, especially because the size of following modifier PP was probably not yet available. Assuming that the eye-voice span of our participants was up to 18 character spaces (Levin 1979) and that the frequent functional words such as articles and copulas were often not fixated (see the review on past text-reading studies in Rayner and Liversedge 2004), readers' eyes must have already been viewing N2 for sentences with short N1s, whereas they were fixating 'of' or the region beyond it for sentences with long N1s when their voicing started. Therefore, we suspect that the insertion of a post-N1 boundary was possibly planned even before the first word was uttered.

On the contrary, the full length of the RC may not have been available to the readers when they were about to finish articulating N2. Readers were most likely previewing the relative pronoun *who* and the following verb as they started producing N2. Readers therefore knew that N2 was followed by an RC, but the word following the verb that determined the length of the RC may not have been fully processed to establish the constituent size while planning the prosodic phrasing upon articulating N2. Even if the word following the verb was within the preview window, the RC pronoun *who* may have warned readers about the potential complexity of the upcoming constituent: It is plausible that readers overwhelmingly placed a boundary after N2 in preparation for a potentially complex and semantically dense constituent.

When the sentence had been comprehended and had become familiar, both the lengths of constituents and the global syntactic structure that was based on the interpretation seemed to shape the prosody (cf. Jun 2003, where participants' interpretation also affected the subsequent production of the prosodic boundaries). In Experiment 2, the length of N1 affected both the local boundary placement and the relative strength of boundaries across the two locations. Readers again separated a long N1 from the following PP, and did so more often when the following RC had been interpreted as low-attached and thus was not separated from N2. Overall, the preference for low attachment interpretations led to much fewer boundaries after N2, supporting Carlson and Frazier's (2002) rational speaker hypothesis. In addition, the presence and the strength of the prosodic boundary after N1 affected the presence and the strength of a boundary after N2 more than the length of RC. These results together suggest that the knowledge of the lengths of all constituents and the interpretation-driven syntactic structure prompted readers to use prosodic boundaries more informatively and to avoid close boundary placement within a sentence. Thus, higher familiarity with the sentence seems to allow not only better control over the informative phrasing of the word string, but also a better rhythmic adjustment of the utterance.

While the ShortN1/LongRC and LongN1/ShortRC sentences respectively induced the expected productions of stronger breaks after N2 and after N1 in Experiment 2, the two sentence types with either both short or both long N1 and RC were not produced with similar prosodic patterns. Instead, sentences with Short N1 and Short RC were likely to be produced with the stronger boundary after N1, while sentences with Long N1 and Long RC were often produced with the stronger boundary after N2. Thus, the sentences with short constituents were produced as the low-attached structure, which in Jun's (2003) study was the default reading for English. When constituents were longer, however, the largest prosodic boundary was more frequently inserted before the beginning of the complex RC. Although it may not be directly motivated by the attachment preference, a prosodic juncture following and preceding a large constituent may have been required for rhythmically felicitous speech planning.

In addition to the effect of familiarity on the overt prosodic patterns, the present study demonstrated the limits of prosodic effects on the interpretation of ambiguous sentences. In contrast to past findings of listeners' sensitivity to relative prosodic boundary strength (Clifton et al., 2002; 2006), Experiment 1 showed that the prosodic patterns produced while simultaneously parsing the sentences did not guide the subsequent interpretation of the sentence. As discussed above, the prosodic phrasings while reading aloud unfamiliar sentences were guided by local syntactic information rather than global semantic information. When readers were asked to provide interpretations of the sentences after reading them on the fly, they could access to the complete syntactic structure and reconstruct the semantic relations among the constituents. During this reconstruction of the message, readers seemed to ignore the prosody they previously produced. Thus, prosodic phrasing may serve as an input to the sentence interpretation only when the listener believes that the prosody cues the syntactic/semantic structure of the sentence.

Finally, an account must be provided for the lack of an effect of constituent length on sentence interpretation. Across the two experiments, we found a robust low-attachment preference that was not modulated by the length of the RC. Such results do not conform to those of previous studies (Bradley et al. 2003; Fernández and Bradley 1999; Hemforth et al. 2005). One possible reason for the absence of a length effect in the present study is insufficient difference in length across the sentence types. Long RCs in previous studies were often over five syllables longer than their short counterparts. In our study, long RCs were only between three and five syllables longer than the corresponding short RCs. Hemforth et al. (2006) propose that longer RCs are usually also more informative than shorter RCs (e.g. *who swims* vs. *who swims like a fish*), and require more costly semantic processing. They further argue that the presence of a semantically more demanding long RC can be justified if it modifies the central element of the proposition, such as the *head* of the subject noun phrase (i.e., N1). According to this "information load account", more high-attachment responses may have been observed if our Long RCs were semantically heavier.

X.5 Conclusions

The present study exhibited how familiarity with the sentence and semantic structure affects the prosodic patterns of structurally ambiguous sentences. Although the analyses focused on the strengths of boundaries at the two RC attachment sites and their relations to sentence interpretations, the prosodic cues to the sentence structure are not limited to the size of boundaries. For example, Schafer et al. (2000) found that listeners successfully determined the intended syntax of the ambiguous sentences in spontaneous speech even when the relative strengths of the boundaries at relevant locations did not support this interpretation. Thus, future studies need to examine how listeners make use of other prosodic cues such as phrasal accents, boundary tones, the types and distribution of pitch accents, rhythm, and microprosody to achieve the comprehension of sentence structure intended by the speaker.

Using the utterances produced in the present study, we are currently investigating whether listeners are sensitive to the speaker's familiarity with the sentences they produced. If naïve listeners relied solely on the boundary strength to achieve the interpretation of the sentence, utterances of the same prosodic patterns would lead to identical interpretations regardless of the speaker's familiarity. Preliminary data suggests that listeners find the unfamiliar speakers' prosody less reliable than that of the familiar speakers. We plan to further investigate whether the perceived degree of speaker's certainty about the sentential message relates to perceived fluency, as well as what prosodic cues predict the perception of speaker's certainty or fluency.

References

Altmann, Gerry T. M. 1998. Ambiguity in sentence processing. *Trends in Cognitive Sciences* 2: 146-152.

Beckman, Mary E. and Julia Hirschberg. 1993. *The ToBI Annotation Conventions*. Unpublished Manuscript.

Beckman, Mary E. and Gayle M. Ayers. 1997. *Guidelines for ToBI labeling*. Unpublished Manuscript.

Beckman, Mary E., Julia Hirschberg and Stefanie Shattuck-Hufnagel. 2007. The Original ToBI System and the Evolution of the ToBI Framework. In *Prosodic Typology: The Phonology of Intonation and Phrasing*, ed. Sun-Ah Jun, 9-54. Oxford: Oxford University Press.

Bergmann, Anouschka, Meghan Armstrong and Kristine Maday. 2008. Relative Clause Attachment in English and Spanish: A Production Study. *Proceedings of the Conference on Speech Prosody 2008*, Campinas, Brazil, 505-508.

Bradley, Dianne, Eva M. Fernández and Dianne Taylor. 2003. *Prosodic weight versus information load in the relative clause attachment ambiguity.* Paper presented at the 16th CUNY Conference on Human Sentence Processing. Cambridge, MA.

Carlson, Katy, Charles Clifton, Jr. and Lyn Frazier. 2001. Prosodic Boundaries in Adjunct Attachment. *Journal of Memory and Language* 45: 58-81.

Carreiras, Manuel and Charles Clifton, Jr. 1993. Relative clause interpretation preferences in Spanish and English. *Language and Speech* 36: 353-372.

Carreiras, Manuel and Charles Clifton, Jr. 1999. Another word on parsing relative clauses: Eye-tracking evidence from Spanish and English. *Memory and Cognition* 27: 826-833.

Clifton, Charles Jr. and Lyn Frazier. 1996. *Construal.* Cambridge, MA: MIT Press.

Clifton, Charles Jr.; Katy Carlson and Lyn Frazier. 2002. Informative Prosodic Boundaries. *Language and Speech* 45: 87-114.

Clifton, Charles Jr.; Katy Carlson and Lyn Frazier. 2006. Tracking the what and why of speakers' choices: Prosodic boundaries and the length of constituents. *Psychonomic Bulletin and Review* 13: 854-861.

Cuetos, Fernando and Don C. Mitchell. 1988. Cross-linguistic differences in parsing: Restrictions on the use of the Late Closure strategy in Spanish. *Cognition* 30: 73-105.

Dilley, Laura, Mara Breen, Edward Gibson, Marti Bolivar and John Kraemer. 2006. A comparison of inter-coder reliability for two systems of prosodic transcriptions: RaP (Rhythm and Pitch) and ToBI (Tones and Break Indices). *INTERSPEECH-2006*, paper 1619-Mon2A3O.6.

Fernández, Eva M. 2000. *Bilingual sentence processing: Relative clause attachment in English and Spanish*, CUNY Graduate Center: Unpublished Doctoral Dissertation.

Fernández, Eva M. and Dianne Bradley. 1999. *Length Effects in the Attachment of Relative Clauses in English.* Poster presented at the 12th Annual CUNY Conference on Human Sentence Processing. New York, NY

Fodor, Janet Dean. 1998. Learning to Parse? *Journal of Psycholinguistic Research* 27: 285-319.

Fodor, Janet Dean. 2002a. Prosodic Disambiguation in Silent Reading. *Proceedings of the North East Linguistic Society (NELS)* 32: 113-132. Amherst, MA.

Fodor, Janet Dean. 2002b. Psycholinguistics Cannot Escape Prosody. *Proceedings of Speech Prosody*, 83-88. Aix-en-Provence, France.

Frazier, Lyn (1978). *On comprehending sentences: Syntactic parsing strategies.* University of Connecticut: Unpublished doctoral dissertation.

Hemforth, Barbara, Susana Fernández, Charles Clifton, Jr., Lyn Frazier and Lars Konieczny. 2005. *Relative Clause Attachment in German, English, Spanish and French: Effects of Position and Length.* Manuscript.

Hemforth, Barbara, Caterina Petrone, Mariapaola D'Imperio, Joël Pynte, Saveria Colonna and Lars Konieczny. 2006. Length effects in PP-attachment: Prosodic balancing or balancing information load. *Proceedings of the Annual Conference of the Cognitive Science Society*. Vancouver, Canada.

Jun, Sun-Ah. 2003. Prosodic Phrasing and Attachment Preferences. *Journal of Psycholinguistic Research* 32: 219-249.

Kimball, John. 1973. Seven Principles of Surface Structure Parsing in Natural Language. *Cognition* 2: 15-47.

Kondo, Tadahisa and Reiko Mazuka. 1996. Prosodic Planning While Reading Aloud: On-Line Examination of Japanese Sentences. *Journal of Psycholinguistic Research* 25: 357-381.

Koriat, Asher, Seth N. Greenberg and Hamutal Kreiner. 2002. The extraction of structure during reading: Evidence from reading prosody. *Memory and Cognition* 30: 270-280.

Kučera, Henry and W. Nelson Francis. 1967. *Computational Analysis of Present-Day American English.* Providence: Brown University Press.

Lehiste, Ilse. 1973. Rhythmic and syntactic units in production and perception. *Journal of the Acoustical Society of America* 54: 1228-1234.

Levin, Harry. 1979. *The Eye-Voice Span.* Cambridge, MA: MIT Press.

Marcus, Mitchell and Donald Hindle. 1990. Description theory and intonation boundaries. In *Cognitive models of speech processing: Psycholinguistic and computational perspectives*, ed. Gerry T. M. Altmann, 483-512. Cambridge, MA: MIT Press.

Miyamoto, Edson T. 2008. *Relative clause attachment*. http://etm4rc.googlepages.com/table.html. Accessed 21 April 2010.

Pierrehumbert, Janet B. 1980. *The phonology and phonetics of English intonation*. Ph.D. Dissertation. MIT, Cambridge, MA.

Pierrehumbert, Janet B. and Julia Hirschberg. 1990. The meaning of intonational contours in the interpretation of discourse. In *Intentions in communication*, eds. Philip R. Cohen, Jerry Morgan and Martha E. Pollack, 271-311. Cambridge, MA: MIT Press.

Price, Patti J.; Mari Ostendorf and Stefanie Shattuck-Hufnagel. 1991. The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America* 90: 2956-2970.

Psychology Software Tools, Inc.. 2003. *E-Prime*. [computer software].

Rayner, Keith. 1975. The Perceptual Span and Peripheral Cues in Reading. *Cognitive Psychology* 7: 65-81.

Rayner, Keith. 1992. *Eye Movements and Visual Cognition: Scene Perception and Reading.* Springer Series in Neuropsychology. New York, NY: Springer-Verlag.

Rayner, Keith. 1998. Eye Movements in Reading and Information Processing: 20 Years of Research. *Psychological Bulletin* 124: 372-422.

Rayner, Keith and Simon P. Liversedge. 2004. Visual and Linguistic Processing during Eye Fixations in Reading. In *The Interface of Language, Vision and Action: Eye Movements and the Visual World*, eds. Fernanda Ferreira and John M. Henderson: 59-104. New York: Psychology Press.

Schafer, Amy J., Shari R. Speer, Paul Warren and S. David White. 2000. Intonational Disambiguation in Sentence Production and Comprehension. *Journal of Psycholinguistic Research* 29: 169-182.

Silverman, Kim; Mary E. Beckman, John Pitrelli, Mari Ostendorf, Colin Wightman, Patti J. Price, Janet Pierrehumbert and Julia Hirschberg. 1992. ToBI: A standard for labeling English prosody. *Proceedings of the 1992 International Conference of Spoken Language Processing* 2: 867-870.

Syrdal Ann K., Julie McGory. 2000. Inter-transcriber reliability of ToBI prosodic labeling. *Proceedings of the International Conference on Spoken Language Processing*. Beijing, China, 235–238.

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

Watson, Duane and Edward Gibson. 2005. Intonational Phrasing and Constituency in Language Production and Comprehension. *Studia Linguistica* 59: 279-300.

Figures and Captions:

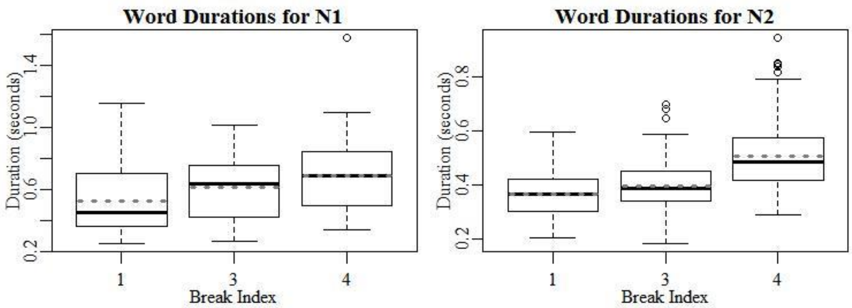


Figure 1: Boxplots showing the distribution of duration measurements across boundary types at N1 and N2. For each break index, the solid black line shows the median and the gray dotted line shows the mean of the distribution. The box below the median shows the second quartile, the one above the median shows the third quartile. The bottom whisker shows the first quartile, the top whisker shows the fourth quartile. The dots represent outliers. The position of the median within the box indicates the skew of the distribution (skewed right if the median is near the bottom of the box and vice versa).

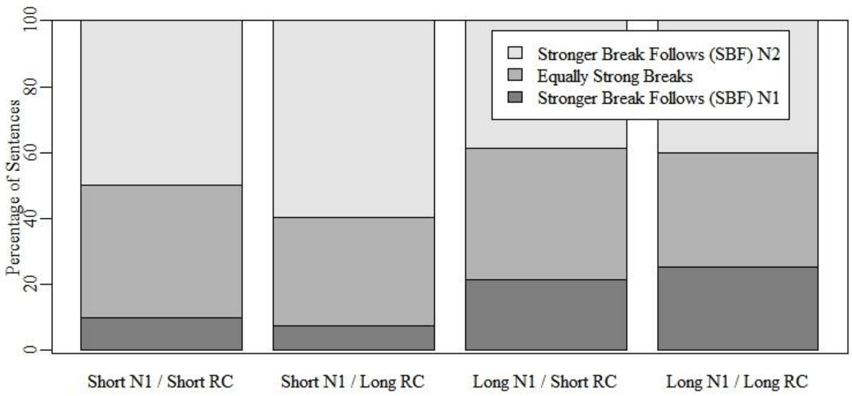


Figure 2: Global prosodies of interest elicited by each sentence type. The x-axis shows the sentence types, and the y-axis shows the number of sentences produced with each prosodic pattern of interest.

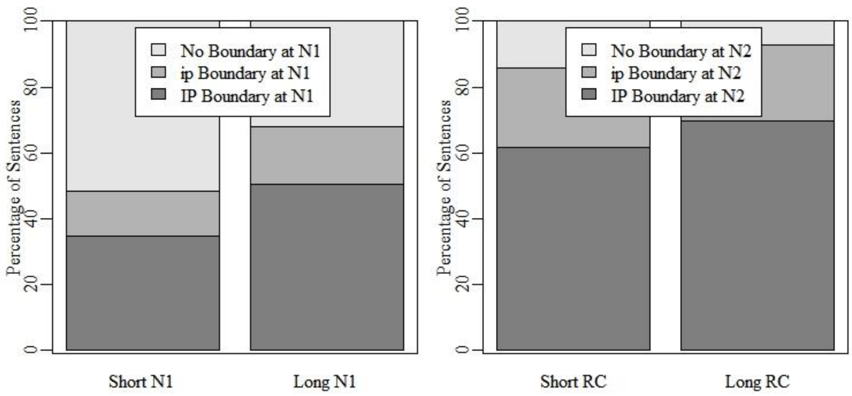


Figure 3a: Number and Type of Prosodic Boundaries at N1 from Sentences with Short and Long N1s.

Figure 3b: Number and Type of Prosodic Boundaries at N2 from Sentences with Short and Long RCs.

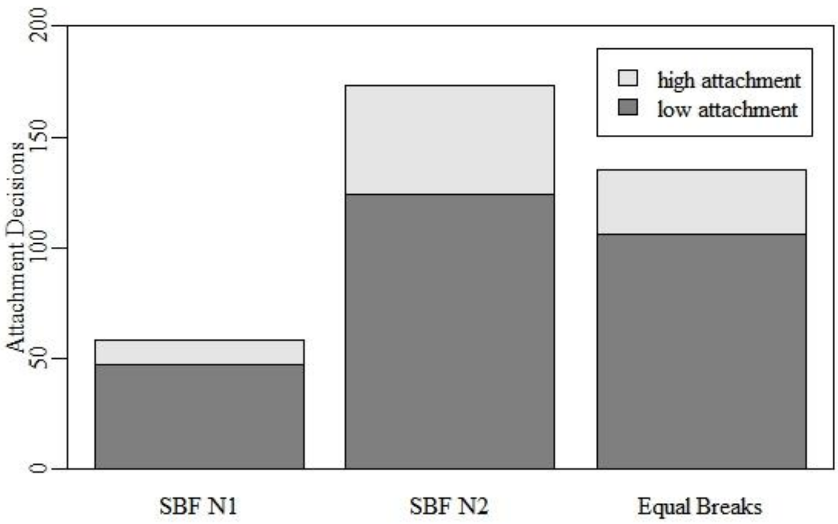


Figure 4: Sentence interpretations by global prosodic patterns. The x-axis shows the prosodic patterns produced; the y-axis shows the sentence interpretations.

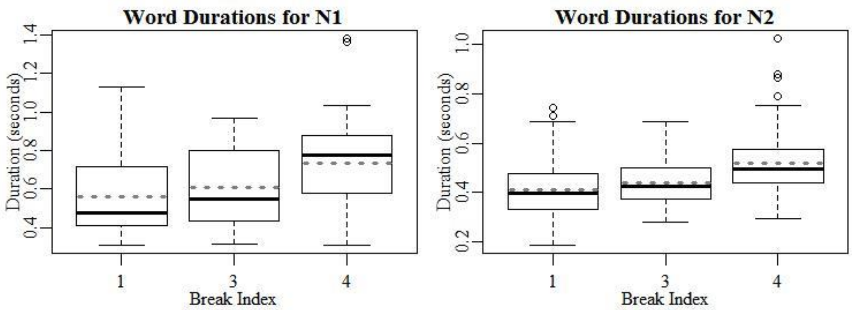


Figure 5: Boxplots showing the distribution of duration measurements across boundary types at N1 and N2. For each break index, the solid black line shows the median and the gray dotted line shows the mean of the distribution. The box below the median shows the second quartile, the one above the median shows the third quartile. The bottom whisker shows the first quartile, the top whisker shows the fourth quartile. The dots represent outliers. The position of the median within the box indicates the skew of the distribution (skewed right if the median is near the bottom of the box and vice versa).

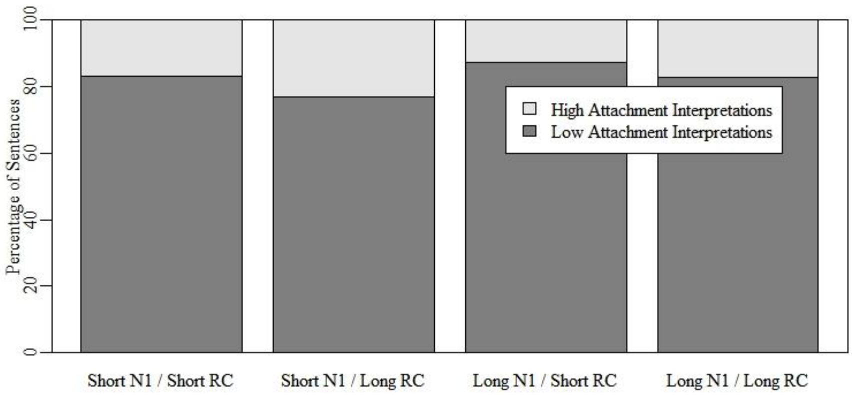


Figure 6: Interpretations of each sentence type. The x-axis shows the sentence types, and the y-axis shows the percentage of sentences given a high attachment vs. low attachment interpretation.

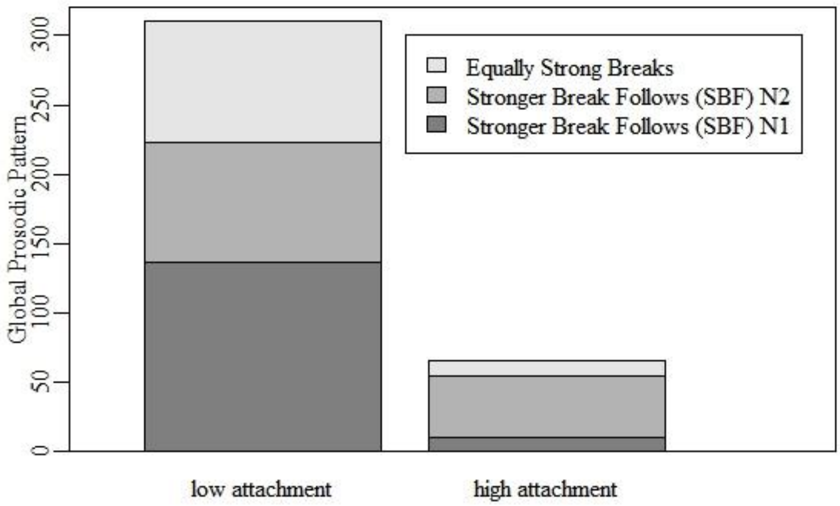


Figure 7: Sentence interpretations by global prosodic patterns. The x-axis shows the prosodic patterns produced; the y-axis shows the sentence interpretations.

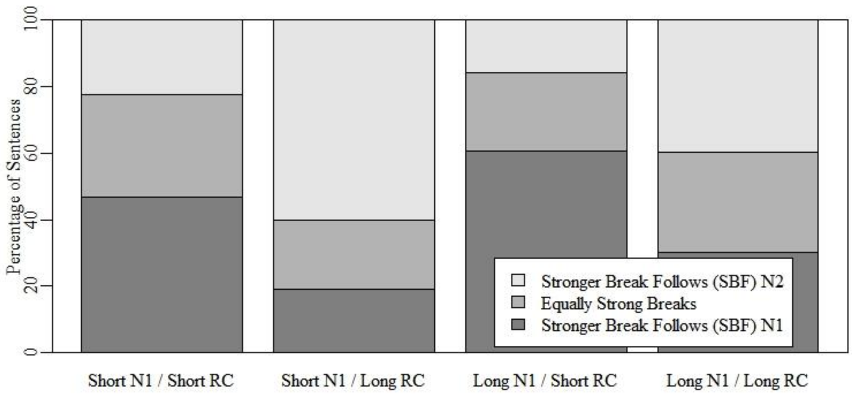
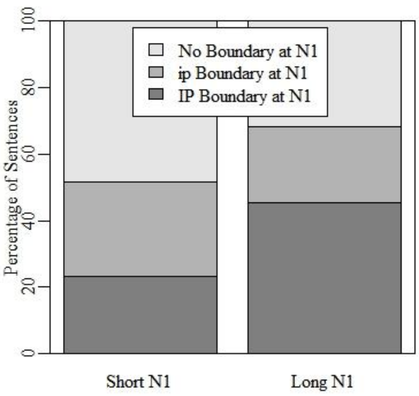


Figure 8: Global prosodies of interest elicited by each sentence type. The x-axis shows the sentence types, and the y-axis shows the number of sentences produced with each prosodic pattern of interest.



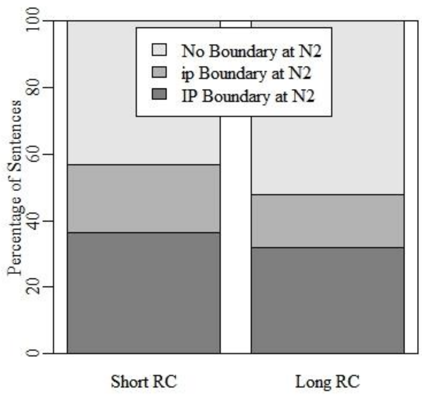


Figure 9a: Number and Kind of Prosodic Boundaries at N1 from Sentences with Short and Long N1s.

Figure 9b: Number and Kind of Prosodic Boundaries at N2 from Sentences with Short and Long RCs.

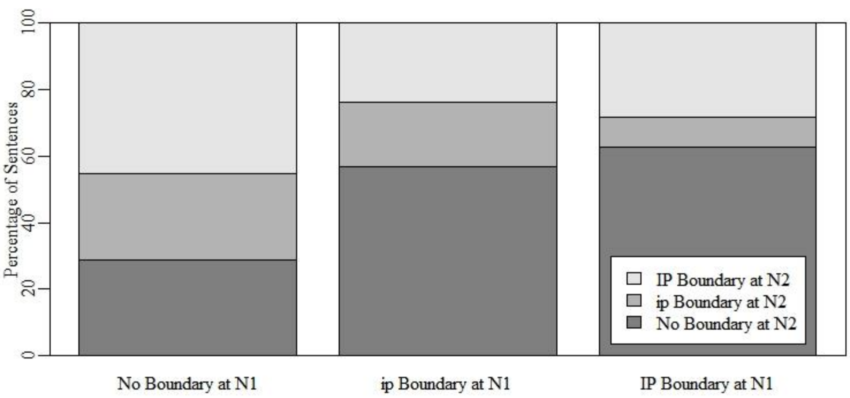


Figure 10: Boundary Placement at N2 as a Function of Boundary Placement at N1. The x-axis shows boundary types at N1; the y-axis shows boundary types at N2.

Tables:

Table 1: Predictions for Experiment 1

Sentence Type	Short N1 / Long RC	Long N1 / Short RC	Long N1/ Long RC Short N1/ Short RC
⇩	⇩	⇩	⇩
Reading Prosody	Boundary Strength N1 < N2	Boundary Strength N1 > N2	Boundary Strength N1 = N2
⇩	⇩	⇩	⇩
Sentence Comprehension	fewer low attachment decisions	more low attachment decisions	intermediate low attachment decisions

Table 2: Correlation Analysis of Word Durations from Experiment 1

	Prosodic Boundary Strength	**Number of Segments**	**Word Frequency**
Word Duration N1	$r = 0.34$; $r^2 = 0.12$	$r = 0.86$; $r^2 = 0.73$	not available[4]
Word Duration N2	$r = 0.38$; $r^2 = 0.14$	$r = 0.71$; $r^2 = 0.51$	$r = -0.48$; $r^2 = 0.23$

[4] Due to English spelling conventions, where many compound nouns are spelled as two separate words (e.g. *defense lawyer*), it was not possible to obtain frequency counts for many N1s.

Table 3: Multinomial Logistic Regression for Figure 2: Coefficients of Predicted Factors. (For all *X vs. Y*, *X* represents the baseline and *Y* represents the alternative. N = 372; Log-Likelihood = –368.57; McFadden R^2 = 0.025694; Likelihood ratio test: χ^2 = 19.44, p < 0.001)

Factors	**SBF N1 vs. ESB**			**SBF N1 vs. SBF N2**			**ESB vs. SBF N2**		
	Est.	t	p <	Est.	t	p <	Est.	t	p <
Long N1 vs. Short N1	0.965929	2.8537	**.01**	1.323249	4.0142	**.001**	0.35732	1.5540	n.s.
Long RC vs. Short RC	0.220067	0.7017	n.s.	-0.069763	-0.2275	n.s.	-0.28983	-1.2625	n.s.

Table 4a: Multinomial Logistic Regression for Figure 3a: Coefficients of Predicted Factors. (For all *X vs. Y*, *X* represents the base-line and *Y* represents the alternative. N = 372; Log-Likelihood = –371.08; McFadden R^2 = 0.019965; Likelihood ratio test: χ^2 = 15.119, p < 0.001)

Factors	**no boundary vs. *ip* boundary at N1**			**no boundary vs. *IP* bounda-ry at N1**			***ip* boundary vs. *IP* boundary at N1**		
	Est.	t	p <	Est.	t	p <	Est.	t	p <
Long N1 vs. Short N1	-0.73716	-2.3609	**.05**	-0.85453	-3.6998	**.001**	-0.11736	-0.3779	n.s.

Table 4b: Multinomial Logistic Regression for Figure 3b: Coefficients of Predicted Factors. (For all *X vs. Y*, *X* represents the base-line and *Y* represents the alternative. N = 372; Log-Likelihood = –317.67; McFadden R^2 = 0.0040533; Likelihood ratio test: χ^2 = 2.5857, p=0.27449)

Factors	**no boundary vs. *ip* boundary at N2**			**no boundary vs. *IP* bounda-ry at N2**			***ip* boundary vs. *IP* boundary at N2**		
	Est.	t	p <	Est.	t	p <	Est.	t	p <
Long RC vs. Short RC	-0.31449	-0.8128	n.s.	-0.52035	-1.4983	n.s.	-0.20585	-0.8267	n.s.

Table 5: Multinomial Logistic Regression for Figure 4: Coefficients of Predicted Factors. (For all *X vs. Y*, *X* represents the baseline and *Y* represents the alternative. N = 372; Log-Likelihood = –201.51; McFadden R^2 = 0.0074332; Likelihood ratio test: χ^2 = 3.0182, p = 0.22111)

Factors	**low attachment vs. high attachment**		
	Est.	t	p <
ESB vs. SBF N1	0.15611	0.3951	n.s.
ESB vs. SBF N2	-0.36768	-1.3666	n.s.
SBF N1 vs. SBF N2	-0.52379	-1.3966	n.s.

Table 6: Predictions for Experiment 2

Sentence Type	Short N1 / Long RC	Long N1 / Short RC	Long N1/ Long RC Short N1/ Short RC
⇩	⇩	⇩	⇩
Silent Reading	Implicit Boundary Strength N1 < N2	Implicit Boundary Strength N1 > N2	Implicit Boundary Strength N1 = N2
⇩	⇩	⇩	⇩
Sentence Comprehension	fewer low attachment decisions	more low attachment decisions	intermediate low attachment decisions
⇩	⇩	⇩	⇩
Reading Prosody	More: Boundary Strength N1 < N2	More: Boundary Strength N1 > N2	All Patterns: N1 < N2; N1 > N2; N1 = N2

Table 7: Correlation Analysis of Word Durations from Experiment 2

	Prosodic Boundary Strength	**Number of Segments**	**Word Frequency**
Word Duration N1	$r = 0.34$; $r^2 = 0.12$	$r = 0.87$; $r^2 = 0.75$	not available[5]
Word Duration N2	$r = 0.39$; $r^2 = 0.15$	$r = 0.68$; $r^2 = 046$	$r = –0.47$; $r^2 = 0.22$

Table 8: Multinomial Logistic Regression for Figure 6: Coefficients of Predicted Factors. (For all *X vs. Y*, *X* represents the baseline and *Y* represents the alternative. N = 372; Log-Likelihood = –172.9; McFadden $R^2 = 0.010162$; Likelihood ratio test: $\chi^2 = 3.55$, p = 0.16949)

Factors	**low attachment vs. high attachment**		
	Est.	t	p <
Long N1 vs. Short N1	0.35678	1.3002	n.s.
Long RC vs. Short RC	-0.36970	-1.3474	n.s.

Table 9: Multinomial Logistic Regression for Figure 7: Coefficients of Predicted Factors. (For all *X vs. Y*, *X* represents the baseline and *Y* represents the alternative. N = 372; Log-Likelihood = –389.86; McFadden $R^2 = 0.045011$; Likelihood ratio test: $\chi^2 = 36.751$, p < 0.001)

Factors	**SBF N1 vs. ESB**			**SBF N1 vs. SBF N2**			**ESB vs. SBF N2**		
	Est.	t	p <	Est.	t	p <	Est.	t	p <
high att. vs. low att.	-0.63639	-1.4155	n.s.	-1.94724	-5.1737	**.001**	-1.310844	-3.6471	**.001**

[5] Due to English spelling conventions, where many compound nouns are spelled as two separate words (e.g. *defense lawyer*), it was again not possible to obtain frequency counts for many N1s.

Table 10: Multinomial Logistic Regression for Figure 8: Coefficients of Predicted Factors. (For all *X vs. Y*, *X* represents the baseline and *Y* represents the alternative. N = 372; Log-Likelihood = –379.3; McFadden R^2 = 0.070872; Likelihood ratio test: χ^2 = 57.865, p < 0.001)

Factors	**SBF N1 vs. ESB**			**SBF N1 vs. SBF N2**			**ESB vs. SBF N2**		
	Est.	t	p <	Est.	t	p <	Est.	t	p <
Long N1 vs. Short N1	0.335394	1.2664	n.s.	0.816579	3.1001	**.01**	0.48118	1.7478	n.s.
Long RC vs. Short RC	-0.744738	-2.7615	**.01**	-1.792098	-6.6434	**.001**	-1.04736	-3.7040	**.001**

Table 11a: Multinomial Logistic Regression for Figure 9a: Coefficients of Predicted Factors. (For all *X vs. Y*, *X* represents the baseline and *Y* represents the alternative. N = 372; Log-Likelihood = –395.3; McFadden R^2 = 0.025662; Likelihood ratio test: χ^2 = 20.823, p < 0.001)

Factors	**no boundary vs. *ip* boundary at N1**			**no boundary vs. *IP* boundary at N1**			***ip* boundary vs. *IP* boundary at N1**		
	Est.	t	p <	Est.	t	p <	Est.	t	p <
Long N1 vs. Short N1	-0.11857	-0.4442	n.s.	-1.04767	-4.2338	**.001**	-0.92910	-3.3327	**.001**

Table 11b: Multinomial Logistic Regression for Figure 9b: Coefficients of Predicted Factors. (For all *X vs. Y*, *X* represents the baseline and *Y* represents the alternative. N = 372; Log-Likelihood = –360.16; McFadden R^2 = 0.071194; Likelihood ratio test: χ^2 = 55.214, p < 0.001)

Factors	**no boundary vs. *ip* boundary at N2**			**no boundary vs. *IP* boundary at N2**			***ip* boundary vs. *IP* boundary at N2**		
	Est.	t	p <	Est.	t	p <	Est.	t	p <
Long RC vs. Short RC	-1.53248	-5.0597	**.001**	-1.62779	-6.4913	**.001**	-0.09531	-0.2990	n.s.

Table 12: Multinomial Logistic Regression for Figure 10: Coefficients of Predicted Factors. (For all *X vs. Y*, *X* represents the baseline and *Y* represents the alternative. N = 372; Log-Likelihood = –367.05; McFadden $R^2 = 0.053442$; Likelihood ratio test: $\chi^2 = 41.447$, $p < 0.001$)

Factors	**no boundary vs. *ip* boundary at N2**			**no boundary vs. *IP* boundary at N2**			***ip* boundary vs. *IP* boundary at N2**		
	Est.	t	p <	Est.	t	p <	Est.	t	p <
no vs. *ip* at N1	-0.95929	-2.7376	**.01**	-1.32917	-4.1711	**.001**	-0.36987	-0.9845	n.s.
no vs. *IP* at N2	-1.80118	-4.7482	**.001**	-1.24572	-4.5056	**.001**	0.55547	1.4319	n.s.
***ip* vs. *IP* at N2**	-0.841892	-2.0422	**.05**	0.083448	0.2593	n.s.	0.92534	2.0127	**.05**