Relative Prosodic Boundary Strength and Prior Bias in Disambiguation

Michael Wagner[1], *Serena Crivellaro*[2]

[1]McGill University
[2]Cornell University '09
chael@mcgill.ca

Abstract

Previous research found that the relative rather than absolute size of prosodic boundaries is crucial in disambiguating attachment ambiguities [1, 2]. Furthermore, relative *categorical* differences matter whereas merely *quantitative* ones do not [1]. This paper presents further evidence that relative boundary strength is indeed crucial, but, contrary to earlier findings, gradient quantitative differences affect parsing decisions in gradient ways. Furthermore, varying the plausibility of a given reading in a given context shifts the perceptual boundaries between different phrasings such that quantitatively stronger prosodic cues are necessary to counter-act a prior bias against it.
Index Terms: prosodic boundaries, ambiguity, relative boundary strength, gradience, rational listener

1. Introduction

Past research has shown that the prosodic phrasing can disambiguate structural ambiguities (e.g., [3, 4, 5]). [4] found that listeners can reliably tell the intended reading in productions of a variety of ambiguous linguistic constructions. Some of the prosodic cues used to distinguish the readings were of an absolute nature—speakers used prosodic boundaries of a different categorical *type*, as defined by the ToBI annotation system of American English [6]. But they also found evidence *relative* strength matters: relatively larger prosodic boundaries were associated with relatively larger syntactic junctures. In experiments that manipulated the strength of a boundary preceding a later prosodic boundary, [1, 2, 7] found that what is crucial in parsing is the strength of a boundary *relative to earlier boundaries in the same utterance*, rather than their absolute size. [8] found evidence both for effects of both relative and absolute boundary strengths. Relative Boundary strength in these studies is purely defined in reference to *categorical* differences between prosodic boundaries (intonational phrase, intermediate phrase, no boundary), rather than gradient differences in the acoustic cues encoding prosodic juncture.

In fact, [1] report evidence that gradient differences between boundaries, e.g., durational ones, do not affect parsing decisions—what matters are only the categorical ones. The gradient manipulation was achieved by asking a trained speaker to produce stimuli that varied in final lengthening in gradient ways but did not vary in the category of the prosodic boundaries. However, since it is very difficult to change one prosodic cue and not affect others, maybe inadvertently other acoustic cues were changed as well. It is well known that categorical differences correlate with gradient differences (e.g., [4]), and some studies found that listeners perceive relative boundary strength much more reliably than categorical differences between prosodic boundaries [9], so it seems quite plausible that gradient phonetic differences between boundaries might be important in parsing after all. This paper looks at this question by using synthesized stimuli which allow a fine-grained control of individual acoustic cues. The two questions the present study addresses are: (i) Do gradient differences in boundary rank affect parsing decisions in gradient ways? (ii) How does gradient information about attachment from prosody interact with probabilistic information about which reading is more likely/plausible in a given context?

2. Exp. 1: Choosing a Prosodic Bracketing

In a first study, we looked at the role of relative prosodic boundary strength in disambiguating arithmetic formulae. There are at least three different possible prosodic bracketings, a left-branching and a right-branching one, and one in which the arithmetic operations are presented as a list, with each calculation instruction separated by boundaries of equal strength:

(1) a. B, + C, * D 'flat'
b. (B + C) * D 'left-branching'
c. B + (C * D) 'right-branching'

As a baseline for looking at other ambiguities, we tested (i) whether listeners can distinguish between these different bracketings; and (ii) whether quantitative durational cues are sufficient to manipulate which bracketing is perceived.

2.1. Methodology and Stimuli

We started out with an utterance synthesized with IBM viavoice, a formant-based speech synthesizer. This stimulus was then manipulated in PRAAT [10] using PSOLA resynthesis.[1] In total, 36 stimuli differing in the relative boundary strength were created. Each boundary was strengthened in 5 equal steps (boundary rank 0-5). The total increase in duration at each boundary was 400 ms. Each step in the continuum added 40 ms, 60% of which were added in pause duration and 40% in final lengthening. Since the boundaries were equally spaced we will simply refer to the relative strength by subtracting the rank in the continuum of the second boundary from the first one, resulting in strengths that vary from -5 to +5.

It is possible that in our baseline utterance the synthesizer inadvertently created prosodic cues that point more toward one of the two bracketings. Any such bias, however, would not affect our question whether quantitative manipulation is sufficient

[1]In fact, we ran this study twice, with a constant pitch or using the default pitch contour generated by the synthesizer. This did not seem to affect the outcome. While pitch itself can be an imported cue for phrasing, the lack of a clear pitch scaling cue in our experiment might just increase the reliance on duration as a cue. We plan to test the interaction of pitch and durational cues in future experiments.

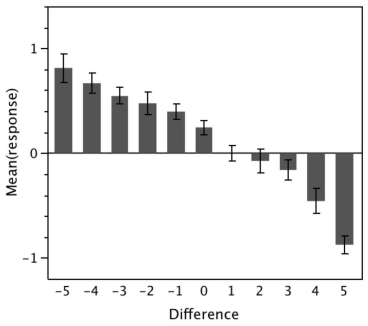

Figure 1: *Average Choice of Bracketing by Relative Boundary Strength. Response = -1: Left-Branching (High Attachment); Response = 0: 'flat' structure; Response = 1: Right-Branching (Low Attachment)*

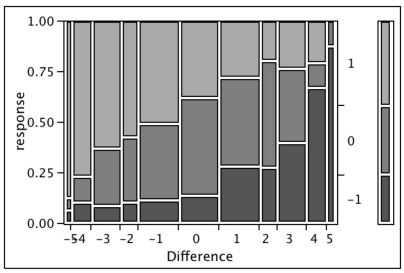

Figure 2: *Choice of Bracketing by Relative Boundary Strength*

to shift the percept, since we are interested in the *relative* size of the boundaries—at worst, our relative boundary strength '0' is not in fact a prosodically neutral starting point. Since there are other sources of bias (e.g., maybe there is a bias toward a certain bracketing) it is in fact impossible to determine which stimulus in the continuum has a 'neutral' prosodic phrasing.

The perception experiment was run using PRAAT. Subjects listened to the acoustic stimuli and had to do a forced choice between three bracketings, and then report their confidence in their choice. The three options were explained to them beforehand in words but without giving an acoustic illustration. The choice was done by clicking one of three buttons labeled with the bracketing. For the left-branching and right-branching choices the formulae had parentheses, and the label of the 'list' choice involved comma-separated calculation instructions. None of the subjects reported a problem in understanding the task. Eight Cornell undergraduates, all native speakers of American English that did not report any hearing disorder, participated in the study. They were payed for their participation.

2.2. Results

The results show a straightforward correlation between relative prosodic boundary strength and choice of bracketing, as illustrated in Fig. 1, which plots the average rating against the difference in boundary strength. The choice in bracketing was significantly influenced by the difference in boundary strength $Pearson's\chi^2(20, N = 630) = 189.5, p < 0.0001$, with an R^2 of 0.13. Looking at the choices in more detail, we see that left-branching was predominantly chosen when the second boundary stronger, right-branching when the first boundary was stronger, and the list-structure when the boundaries were of equal strength (Fig. 2). For the stimuli with boundaries of equal strength (difference=0), 48% chose a list structure, and a big proportion (38%) of participants chose the 'left-branching' structure, while only 14% chose the right-branching one. That list-structure and left-branching structure are more similar to each other and hence more confusable is not surprising, since a list of arithmetic instructions is likely to be interpreted incrementally left-to-right, which leads to the same arithmetic result as the left-branching structure. While the relative boundary difference did not overall affect the confidence rating, there was a small but significant increase in confidence with decreasing boundary difference ($R^2 = 0.04, p < 0.001$) for left-branching structures, and a trend in the opposite direction for right-branching ones—the direction is as expected, given that the prosodic cue is strongest at the end-point of the continuum.

2.3. Discussion

The results how initial evidence that gradient differences in relative prosodic boundary strength influence parsing decisions when listening to ambiguous structures. What we can infer from this is limited for two reasons: The present task directly asks participants which prosodic bracketing they perceive. Maybe gradient cues have an effect in this meta-linguistic task, but are irrelevant in a task that taps their interpretation indirectly. Also, arithmetic formulas may not be representative of linguistic expressions. The following study avoids these problems.

3. Exp. 2: Testing Parsing Decisions

The structural ambiguity between a transitive particle verb and a verb taking a prepositional argument is one of the ambiguities that [4] found was reliably disambiguated by prosodic cues in production, and these cues enabled listeners to disambiguate with high accuracy in perception. We looked at the following 5 ambiguous sentences:

(2)
a. The tourist checked in the bags.
b. The student dropped off the table.
c. The vikings won over their enemies.
d. The tires may wear down the road.
e. The engineers looked up the elevator shaft.

In this ambiguity, there are two different relevant bracketings, not three as above. Note that which reading is intuitively more plausible varies between the sentences depending on pragmatic plausibility, and maybe also on the frequency with which these verbs are used with a particle and a prepositional argument respectively. For evidence for prosodic effects of syntactic frequencies see [11]. We return to this point later.

3.1. Methods

We synthesized the sentences with a formant synthesizer, using the same method as described above. This time, we manipulated boundary strength at each of the two boundaries in *six* steps leading to 49 different stimuli. Participants performed a forced choice between two possible continuations that disambiguated the sentence in either direction. For each choice we had participants rate their confidence. Every subject was tested

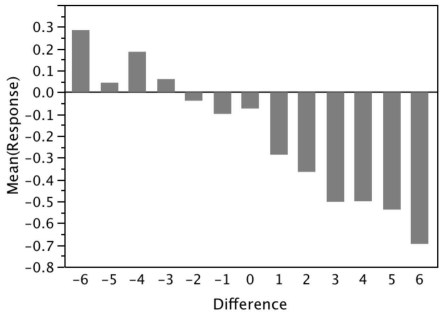

Figure 3: *Response by Relative Boundary Strength. Response -1: Prepositional Phrase; Response 1: Particle Verb.*

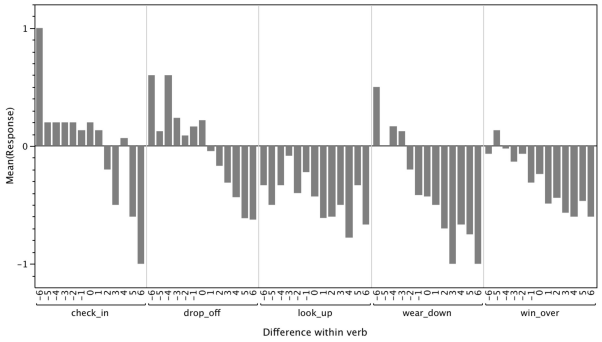

Figure 4: *Response Distribution for the 5 different items.*

on 25 stimuli drawn from the 49. Fillers involving a prominence judgment were interspersed. We aimed for 20 subjects on each item, yielding about 10 responses for each manipulation. For two items, we were only able to recruit 8 subjects.

3.2. Results

Just like in the baseline-case, the responses were significantly influenced by the relative difference between the two boundaries (Fig. 3), a highly significant effect (Pearson's $\chi^2(12, 2228) = 128.6, p < 0001$). There was a trend that confidence increased with the quantitative size of the difference in the expected directions. Since the outcome in this case was a binary decision, we could run a mixed-model logistic model, which is a more appropriate test, and it allows for the control for subject and item effects. Using the lmer function of the statistics package R, we found a highly significant main effect of the boundary difference ($t = -12.1$; [12] recommends a level of an absolute value of $t > 2$ for significance), even after controlling for these random effects, confirming our hypothesis that quantitative difference in relative boundary strength matters.

The prosodic manipulation accounted for less of the variance compared to experiment 1. One reason may be the change in task: In this experiment, subjects did not choose a prosodic bracketing, but were asked indirectly in which of two ways they understood a sentence. More importantly, the by-item plots (Fig. 3) suggest that items differed widely in how biased listeners were toward one reading or another. This is reflected in the graphs as a shift in the response function toward the particle response or toward the preposition response. A model comparison between the original mixed model and a mixed model excluding the random item effect was highly significant (the comparison was done using the 'anova' function in R: $\chi^2 = 46.4, p < 0.001$), showing that the response function differed significantly between items.

3.3. Discussion

The results show that relative prosodic boundary strength matters, and parsing decisions are affected in gradient ways by gradient cues. Whether similar gradient effects have an effect outside of an experimental situation when multiple manipulations of the same ambiguity are used remains to be seen. The difference in response patterns between items suggest that prosodic information interacts with other information about which reading is more likely. The next experiment tests the hypothesis that prior bias can account for some of the item effects.

A complication should be mentioned. The experimental literature (e.g., [3], [4]) assumes a difference in *prosodic phrasing* between the two readings. Another factor, however, that may differentiate between them is the relative *prominence* of the particle/preposition: The particle is expected to be more prominent than the head of a prepositional phrase [13, and references therein]. Since prominence also correlates with duration, this just reinforces the effect of phrasing. However, the prominence (and hence duration) of the particle may be decreased as a result of a rhythmic adjustment due to the adjacency of the accented object. In order to test this more, we have conducted production experiments (not reported here for reasons of space) which suggest that in normal speech the durational effects on the verb are just as expected based on the assumptions made here, while the situation is more complicated—but not incompatible—with respect to the durational effects on the particle/preposition.

4. Experiment 3: Prior Bias

Exp. 1 and 2 show that prosodic cues gradiently influence the likelihood with which participants report perceiving a given reading, and that the interpretation of prosodic cues varies depending on the particular item. One hypothesis why we found by-item differences is that prosodic cues compete with other sources of information such as the frequency of use of the two structures given the verb or the plausibility of the readings. To see this idea, a norming study was conducted in order to quantify the inherent bias in each of our 5 items.

4.1. Methods and Stimuli

40 participants, mostly undergraduates at McGill, were asked to paraphrase the sentences in (2).[2] 3 participants were excluded because based a language questionaire they did not qualify as native speakers of North American English. 18 participants filled out the paraphrases on paper in the lab, 19 submitted them by email. All participants were compensated for participation (the email-responders were also run on other experiments in the lab and received compensation then). An RA coded which reading they paraphrased. If the paraphrase did not disambiguate between the two readings, the trial was ignored.

4.2. Results & Discussion

The average responses (averaged over all bracketings) reveal an apparent inherent bias for each of the items. The estimate

[2]They were asked to paraphrase more sentences, in fact, as part of a bigger norming study for a related production experiment, not reported here.

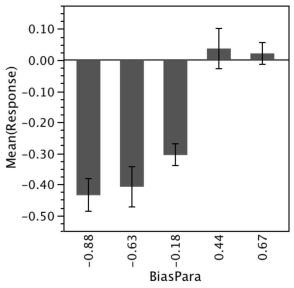

Figure 5: *Proportion of particle response in completion study plotted against average response in perception study.*

of the prior bias based on the norming study yielded a good correlation with the perception results, as Fig. 5 illustrates. We tested the significance of prior bias in predicting the responses by adding our norming estimate as a factor to the mixed model. A model comparison between this model and the original model (controlling for item and subject effects) was highly significant ($\chi^2(4, 6) = 58.5; p < 0.001$). The results show that prior bias affects the interpretation of prosodic cues in perception. Prior bias (as measured by the paraphrase-study) correlates with the average response in experiment 2: Stronger prosodic cues are necessary to counter-act a prior bias against a particular reading.

5. Conclusions

The results of this study lend more support to the view that the *relative strength* of a boundary matters in parsing, as proposed in [1, 2, 8]. They differ from previous studies in that purely quantitative differences were shown to be also relevant, rather than just categorical differences in boundary type, in contrast to [1]. This effect will need to be further confirmed in experiments in which participants are not exposed to multiple manipulations of the same ambiguity. The findings fit well with production evidence that syntax influences the relative ranks of boundaries rather than their absolute category [13], and boundaries later in the utterance can be scaled gradiently relative to earlier ones.

Experiment 3 tested whether fine-grained prosodic cues compete with pragmatic information: Does the likelihood of a parsing decision depends on world knowledge and plausibility? Both prosodic and pragmatic sources of information contributed to predicting the likelihood of perceiving a particular reading in listening to ambiguous stimuli, as would be expected under the assumption that listeners generally make rational use of acoustic cues in perception [14], and prosodic cues compete with semantic and pragmatic ones in guiding parsing decisions.

Future experiments could try to test finer grained hypotheses about the nature of various biases and how they are factored in with prosodic cues: How much of the bias stems from plausibility, and how much from syntactic probabilities? Also, it would be interesting to investigate how boundary strength is used in on-line interpretation. A strong or weak first boundary might affect early stages in the parsing of ambiguous stimuli even if its effect is later 'neutralized' by an even stronger or an even weaker boundary later in the utterance. Furthermore, it would be interesting to let prosody interact with estimated expectedness of upcoming material at particular points in the utterance, which current information-theoretic models of processing found to affect grammatical choices and phonetic reduction ([15] and references therein). In our examples, either the last word of an utterance could be used to manipulate one's overall bias (suppose we replaced 'elevator shaft' with 'phone number'), or an early word could (suppose we replace 'tourist' with 'policeman'). In other words, we can play with the time course at which certain pragmatic biases are introduced, and independently vary the strength of prosodic cues at various points, and thus gain a more sophisticated understanding of how prosodic and other sources of information interact in parsing in real time.

6. Acknowledgements

This research was supported by a seed grant of the Institute for Social Science at Cornell University, by FQRSC grant NP-132516, and a SSHRC Canada Research Chair (Project 212482). Thanks for comments to Mara Breen, Ted Gibson, the audiences at a prosody workshop at the University of Potsdam in January '09 and at UCLA in February '09. Thanks to Steffani Scheer for help with running the norming study.

7. References

[1] K. Carlson, J. Charles Clifton, and L. Frazier, "Prosodic boundaries in adjunct attachment," *Journal of Memory and Language*, vol. 45, pp. 58–81, 2001.

[2] C. J. Clifton, K. Carlson, and L. Frazier, "Informative prosodic boundaries," *Language and Speech*, vol. 45, pp. 87–114, 2002.

[3] I. Lehiste, "Phonetic disambigation of syntactic ambiguity," *Glossa*, vol. 7, pp. 107–122, 1973.

[4] P. J. Price, M. Ostendorf, S. Shattuck-Hufnagel, and C. Fong, "The use of prosody in syntactic disambiguation," *Journal of the Acoustical Society of America*, vol. 9, pp. 2956–2970, 1991.

[5] C. W. Wightman, S. Shattuck-Hufnagel, M. Ostendorf, and P. J. Price, "Segmental durations in the vicinity of prosodic phrase boundaries," *Journal of the Acoustical Society of America*, vol. 92, pp. 1707–1717, 1992.

[6] K. Silverman, M. Beckman, M. Ostendorf, C. Wightman, P. Price, J. Pierrehumbert, and J. Hirschberg, "ToBI: A standard for labeling English prosody," in *Proceedings of the 1992 international conference conference of spoken language processing*, vol. 2, 1992, pp. 867–870.

[7] L. Frazier, K. Carlson, and C. Clifton, "Prosodic phrasing is central to language comprehension," *Trends in Cognitive Sciences*, vol. 10, no. 6, pp. 244–249, 2006.

[8] J. Snedeker and E. Casserly, "Is it all relative? effects of prosodic boundaries on the comprehension and production of attachment ambiguities," 2010, to appear. Language and Cognitive Processes.

[9] J. R. d. Pijper and A. A. Sanderman, "On the perceptual strength of prosodic boundaries and its relation to suprasegmental cues," *The Journal of the Acoustical Society of America*, vol. 96, p. 2037, 1994.

[10] P. Boersma and D. Weenink, "PRAAT, a system for doing phonetics by computer. report 132." 1996, institute of Phonetic Sciences of the University of Amsterdam.

[11] S. Gahl and S. M. Garnsey, "Knowledge of grammar, knowledge of usage: Syntactic probabilities affect pronunciation variation," *Language*, vol. 80, no. 4, pp. 748–775, 2004.

[12] R. Baayen, D. Davidson, and D. Bates, "Mixed-effects modeling with crossed random effects for subjects and items," *Journal of Memory and Language*, vol. 59, no. 4, pp. 390–412, 2008.

[13] M. Wagner, "Prosody and recursion," Ph.D. dissertation, MIT, 2005.

[14] M. Clayards, M. Tanenhaus, R. Aslin, and R. Jacobs, "Perception of speech reflects optimal use of probabilistic speech cues," *Cognition*, vol. 108, no. 3, pp. 804–809, 2008.

[15] R. Levy and T. F. Jaeger, "Speakers optimize information density through syntactic reduction." in *Advances in neural information processing systems (NIPS)*, J. P. B. Schlökopf and T. Hoffman, Eds., vol. 19, 2006, pp. 849–856.