Evidence for the Influence of Syntax on Prosodic Parsing

Andrés Buxó-Lugo & Duane G. Watson
University of Illinois at Urbana-Champaign

Corresponding Author:
Andrés Buxó-Lugo
Department of Psychology, University of Illinois at Urbana-Champaign
603 E. Daniel St.
Champaign, IL 61820
buxo2@illinois.edu

Dated: February 16, 2016

Author's Note

This project was supported by Grant Number R01DC008774 from the National Institute on Deafness and Other Communication Disorders, Grant Number T32-HD055272 from the National Institutes of Health, and a grant from the James S. McDonnell Foundation.

Abstract

We investigate whether expectations based on syntactic position influence the processing of intonational boundaries. In a boundary detection task, we manipulated a) the strength of cues to the presence of a boundary and b) whether or not a location in the sentence was a plausible location for an intonational boundary to occur given the syntactic structure. Listeners consistently reported hearing more boundaries at syntactically licensed locations than at syntactically unlicensed locations, even when the acoustic evidence for an intonational boundary was controlled. This suggests that the processing of an intonational boundary is a product of both acoustic cues and listener expectations.

Keywords: psycholinguistics, language processing, prosody

Evidence for the Influence of Syntax on Prosodic Parsing

In this paper, we investigate the types of information listeners use to parse prosodic structure. An important part of parsing prosodic structure is detecting intonational boundaries, which are used to group utterances into smaller constituents that sometimes reflect the syntactic structure of spoken sentences (Cooper & Paccia-Cooper, 1980; Ferreira, 1993; Watson & Gibson, 2004). These boundaries are signaled by pauses, changes in F0 contours, and pre-boundary lengthening, among other cues (e.g., Klatt, 1975; Pierrehumbert & Hirchberg, 1990; Turk & Shattuck-Hufnagel, 2007; Ladd, 2008). Listeners, in turn, can use intonational boundaries to decipher the linguistic structure of a message, as in the case of syntactically ambiguous sentences (Schafer, Speer, & Warren, 2005; Snedeker & Trueswell, 2003).

However, few studies have explored how listeners build their representation of utterances' prosodic structure. Current models that aim to shed light on the relationship between prosody and other levels of representation tend to be unidirectional, often focusing on how prosody can guide the interpretation of other constructs such as syntax (e.g., Price et al., 1991; Kjeelgard & Speer, 1999; Schafer et al., 2000). For example, Schafer (1997) proposes the following relationship between prosody and syntax: "the prosodic representation that is constructed by the phonological component is passed on to higher-level modules in the same way that lexical information is made available to them" (p. 6) such that prosodic information is "part of the computational vocabulary of the syntactic and semantic/pragmatic processing modules" (p. 6). According to such models of prosodic parsing, listeners build prosodic representations from the acoustic cues, and then use these constructs to guide their interpretation of higher-level structures. However, it is possible that prosodic parsing is more interactive, or bi-directional. In such a model, information from higher-level structures and listener expectations,

along with acoustic cues, guide the parsing of prosodic structure. This study investigates whether the detection of intonational boundaries is wholly driven by acoustic features in the speech signal, or whether input from the syntactic context influences listeners' interpretations.

Intonational boundaries provide an ideal opportunity to investigate listeners' parsing of prosodic structure because of the close link between syntactic boundaries and intonational phrasing (Nespor & Vogel, 1986). Many studies have explored the connection between syntactic and prosodic structures. For example, constraints such as Align-XP (Selkirk, 1986; 1995) and Wrap-XP (Truckenbrodt, 1999) argue that there are grammatical constraints that govern the mapping between syntactic structure and prosodic boundaries, resulting in a preference to produce intonational boundaries at syntactic boundaries. Similarly, algorithmic approaches that predict where boundaries occur make use of syntactic information, such as the length of syntactic constituents and the relationship between syntactic dependents (e.g., Cooper & Paccia-Cooper, 1980; Ferreira, 1988; Watson & Gibson, 2004)[1]. Studies have also found that listeners can accurately locate syntactic boundaries based on prosodic cues alone (Beach 1991; de Pijper & Sanderman, 1994; Streeter, 1978). Lastly, listeners use prosodic boundaries to resolve syntactic ambiguities (e.g., Kjeelgard & Speer, 1999; Kraljic & Brennan, 2005; Lehiste, 1973; Price et al., 1991; Schafer, 1997; Schafer et al., 2005; just to name a few). For example, Snedeker and Trueswell (2003) examined productions of sentences with attachment ambiguities such as: "Tap the frog with the flower," where "flower" could be used as an instrument used for tapping, or the prepositional phrase could be interpreted as a modifier of "the frog." Speakers who were aware of the ambiguity produced intonational boundaries that disambiguated the syntax (after the verb for a modifier interpretation, and after the noun "frog" for an instrument interpretation).

[1] These apparent effects of constituent length have also been conceptualized as effects of the phonological length of consistuents (see Jun & Bishop, 2015 for a discussion).

Critically, listeners used these cues to carry out the correct instruction. This suggests that listeners can accurately parse the syntactic structure of a sentence if intonational boundary cues are provided.

Given that there is a strong correlation between intonational boundaries and syntactic structure, it is possible that listeners not only use prosodic structure to make inferences about syntactic structure, but also use syntactic structure to make inferences about prosodic structure. This type of interaction between processing systems is ubiquitous in language processing. For example, perception studies have found that syntax influences where listeners report hearing bursts of noise (Garrett, Bever, & Fodor, 1966), that morphological context affects the perception of ambiguous phonemes (Ganong, 1980), and that top-down knowledge of the speech signal affects whether degraded speech is perceived as speech at all (Remez et al., 1981). More recent studies (e.g., Kim & Osterhout, 2005; Tabor & Tanenhaus, 1999) have proposed parallel-process models where processing streams for semantic interpretation and syntactic interpretation are independent but still interact through cross-talk or attraction. According to some of these models, each processing system (e.g., syntactic processing system, semantic processing system, etc.) attempts to reach likely interpretations of a stimulus based on their input; however, if a processing system does not have sufficient evidence for converging on an interpretation, it is likely to be influenced by other processing streams.

Given that interaction between levels of processing is ubiquitous in the language comprehension system, it would be surprising if listener expectations did not influence their interpretation of prosody. Some studies have found that prosodic information from earlier in an utterance influences how listeners segment words (e.g., Brown et al., 2011; Dilley et al., 2010) and how they interpret lexical stress (Brown et al., 2012) later in an utterance. Also work by

Bishop (2012) suggests that expectations about discourse structure can influence the perception of acoustic prominence. This is further supported by work by Cole, Mo, and Baek (2010), where untrained listeners prosodically transcribed speech from the Buckeye corpus. In their study, both vowel duration and syntactic context were correlated with boundary reports, each factor independent of the other. In fact, syntactic context was the best predictor of boundary detection, suggesting that listeners' judgments were influenced by their expectations of where boundaries should occur.

However, Cole et al. (2010) did not directly manipulate listener expectations of intonational boundaries. Corpus analyses are a useful tool for detecting correlations, such as the one found between syntactic context and boundary detection in Cole et al. (2010). However, a challenge for these approaches is controlling for other potential variables that might be confounded with the theoretical construct of interest. For example, it is possible that boundary detection was driven by acoustic cues that were not accounted for in the analyses. This makes it difficult to definitively establish that syntactic expectations are driving the detection of intonational boundaries. An advantage of investigating this issue through an experimental design is that these potential confounds can be more precisely controlled with the goal of understanding whether syntactic context alone drives the perception of prosody. That is our goal here. If prosodic parsing is guided by expectations, one would expect a greater tendency to report hearing an intonational boundary in locations in which they typically occur. In the current study, we directly manipulated the acoustic evidence for intonational boundaries and the syntactic context in which these possible boundaries were located. By manipulating word duration, F0 contour, and pause duration of potential boundary sites, we were able to make these locations sound more or less boundary-like. These manipulated words were placed at points at

which boundaries were syntactically licensed and at points at which boundaries were syntactically unlicensed, allowing us to independently manipulate acoustic and syntactic cues to the presence of a boundary. Examining this question in the context of a controlled experiment allows us to see the effects of syntax on prosodic parsing while controlling for acoustic factors, and vice-versa. Furthermore, by individually manipulating acoustic cues and syntactic context, we can observe how these factors interact. For example: how strong do the acoustic cues have to be for listeners to report a boundary in an unexpected location? Thus, our study has two main goals: 1) to replicate the findings in Cole et al. (2010) in an experimental context, and 2) to investigate the extent to which acoustic cues and syntactic context both contribute to the boundary detection process.

Understanding whether different processing systems play a role in intonational boundary detection is important for two reasons. The first is that current prosodic coding systems require coders to use both auditory cues from the speech signal and visual information from a pitch track to make judgments about prosodic phenomena. The underlying assumption in these approaches is that prosody is driven by acoustic cues and expert coders can be trained to detect them. If it is the case that expectations influence listeners' representation of prosody, these coding strategies may need to be re-evaluated. The second reason this question is important is because intonational boundaries in the sentence processing literature have traditionally been studied with the goal of understanding whether intonational boundaries disambiguate syntactic structure, with the assumption that these prosodic boundaries can be detected by listeners using bottom up acoustic cues (see Wagner & Watson, 2010 for a review). Although a number of studies have clearly demonstrated that prosody can bias listeners towards specific syntactic interpretations (e.g., Carlson, Clifton, & Frazier, 2001; Kjelgaard & Speer, 1999; Kraljic & Brennan, 2005;

Price et al, 1991; Snedeker & Trueswell, 2003, and many others), if it is the case that the processing of intonational boundaries is partly driven by syntax, the relationship between syntax and intonational boundaries may be more complex than has been previously assumed.

In Experiment 1, we manipulated the acoustic properties of two critical words in various sentences: one word in a location in which a following intonational boundary would be syntactically licensed, and one in which it was syntactically unlicensed. The acoustic manipulation was done in 9 equal-sized steps ranging from cues that suggested that no boundary was present to cues that strongly suggested that a boundary was present. This was inspired by the VOT continua used in phoneme differentiation tasks (e.g., Ganong, 1980). This continuum allowed us to observe effects of syntax-driven boundary expectations when the acoustic cues were more vs. less indicative of boundary presence. If boundary detection is strictly driven by acoustic factors, listeners should report hearing boundaries whenever the acoustic cues indicate the presence of a boundary. Conversely, if the syntactic processing system influences boundary detection, listeners should be more likely to report boundaries at the syntactically licensed location independent of the acoustic information in the critical words.

To preview the results, we find that listeners are more likely to report hearing an intonational boundary at syntactically licensed locations compared to syntactically unlicensed locations, independent of acoustic cues. Experiment 2 was designed to rule out the possibility that the effect we see in Experiment 1 was a product of the type of instructions participants received. Experiment 3 was designed to rule out the possibility that the syntactic effects in Experiments 1 and 2 were the result of listeners building expectations about boundary locations across the course of the experiment.

Experiment 1

Method

Participants. Twenty English speakers from the United States of America participated in the study. Two participants were excluded due to having learned a language other than English from an early age (before 5). This resulted in 18 monolingual English speakers. They were all users of Amazon's Mechanical Turk service, and they all had at least a 95% approval rating for previous task completions. They were paid $6.00 for participating in the study.

Materials. A native English speaker was recorded while producing variants of 14 critical items. Each item was a unique noun-modifier pair (e.g., "green frog," "big bowl," etc.). For each of these item pairs, 2 different sentence structures were produced. One structure included a direct object with a prenominal modifier. In the other structure, the direct object was modified by a relative clause that included the same adjectival modifier. For example:

a.) Put the big bowl on the tray.

b.) Put the bowl that's big on the tray.

The purpose of the two structures was to balance the part of speech that preceded the preferred locations for boundaries. In a), a boundary is syntactically licensed after "bowl" (a noun), while in b) a boundary is syntactically licensed after "big" (an adjective). These locations were chosen because previous work suggests that major syntactic boundaries, such as the boundary between an object phrase and a prepositional phrase, are likelier places for intonational boundaries than non-major syntactic boundaries (e.g., between a noun and a modifier: Gee & Grosjean, 1983; Watson & Gibson, 2004).

Each of these sentences was produced once with a boundary at a syntactically licensed location, and once with a boundary at a syntactically unlicensed location, as in the following:

c.) Put the big bowl | on the tray.

d.) Put the bowl that's big | on the tray.

e.) Put the big | bowl on the tray.

f.) Put the bowl | that's big on the tray.

Examples (c) and (d) have boundaries at syntactically licensed locations while examples (e) and (f) are produced with boundaries at syntactically unlicensed locations. There were 14 items, 2 sentence structures, and 2 boundary locations, resulting in 56 different recordings.

A boundary continuum was constructed by first transcribing the key nouns and modifiers ("big" and "bowl" in the previous example) in all of the items using Praat's textgrid feature (Boersma & Weenink, 2015). The duration of each word, along with the pause that followed it, were measured. In order to measure the F0 contour, the average F0 was sampled from 10 equally-spaced regions throughout the word. The measurements from the naturally produced boundary words and naturally produced non-boundary words were then used as ends of a boundary spectrum. Seven equally-spaced boundary-steps in between these 2 end points were also derived, resulting in 9 steps of boundary-likeness. The boundary steps for F0 contour were created by first smoothing the contours of the end points into the cubic functions that best fit them. The difference between the boundary and non-boundary words at each of the 10 equally spaced points throughout the word was divided by the number of steps, which resulted in an interval by which we could change the curve at each point for each step (illustrated in Fig. 3).

Two key words in each of the original 14 recordings were resynthesized so that one of the words, what we will call the target word, was the primary point of acoustic manipulation and where we varied the boundary spectrum between 1 (non-boundary) and 9 (boundary). The other word, the non-target word, always had acoustic cues that were consistent with the absence of a boundary. The non-target word was re-synthesized to balance the effects of re-synthesis on boundary detection at the target. The target word and non-target word were counter-balanced so that half the time the target word was at the syntactically licensed location and half the time the non-target word was at the syntactically unlicensed location. In order to make the recording as natural as possible, the F0 contour was resynthesized so that the initial point of the contour was matched to the F0 of the corresponding point in the original non-resynthesized word. This prevented sudden changes in F0 as the word started. The rest of the F0 contour values were derived by fitting the appropriate curve to the starting point (the beginning of the curve corresponding to the onset of the word) and calculating the values at 9 other equally-spaced points. The F0 contour was resynthesized based on these values at 10 equally-spaced points throughout the word using Praat's Manipulate function, which is based on the PSOLA algorithm. Words and pauses were lengthened (or shortened) to match the durations given by the desired boundary step. This was done using Praat's Lengthen function, which also makes use of the PSOLA algorithm. In order to control for the effects of the words surrounding the target words, we resynthesized sentences that originally had the boundary in the syntactically licensed location as well as sentences that originally had a boundary in the syntactically unlicensed location. The four most natural sounding items after resynthesis were selected for the experiment. This resulted in a total of 272 recordings (4 items * 2 boundary locations * 9 boundary steps * 2

sentence structures * 2 source sentences – 16, since at boundary-step 1 there is no difference between boundary position).

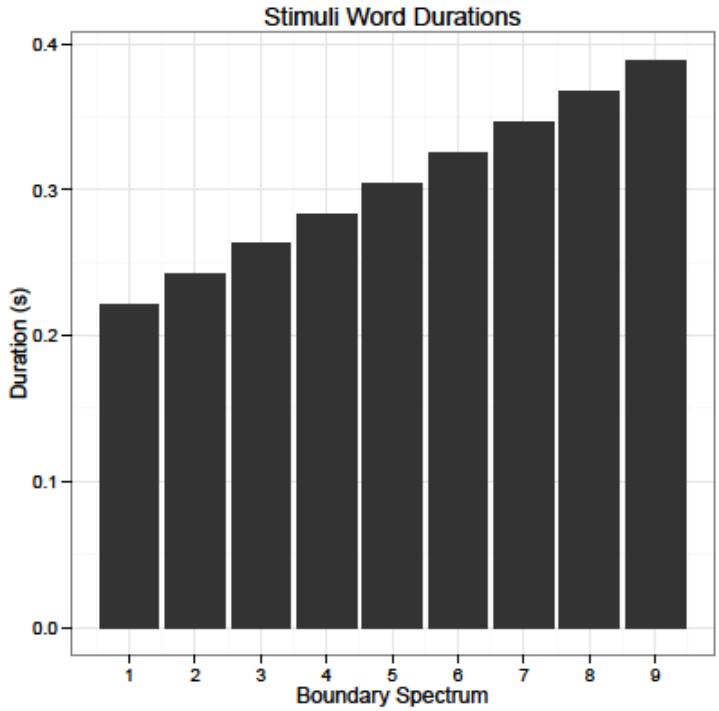

Figure 1: Word durations for critical words at each step of the boundary spectrum.

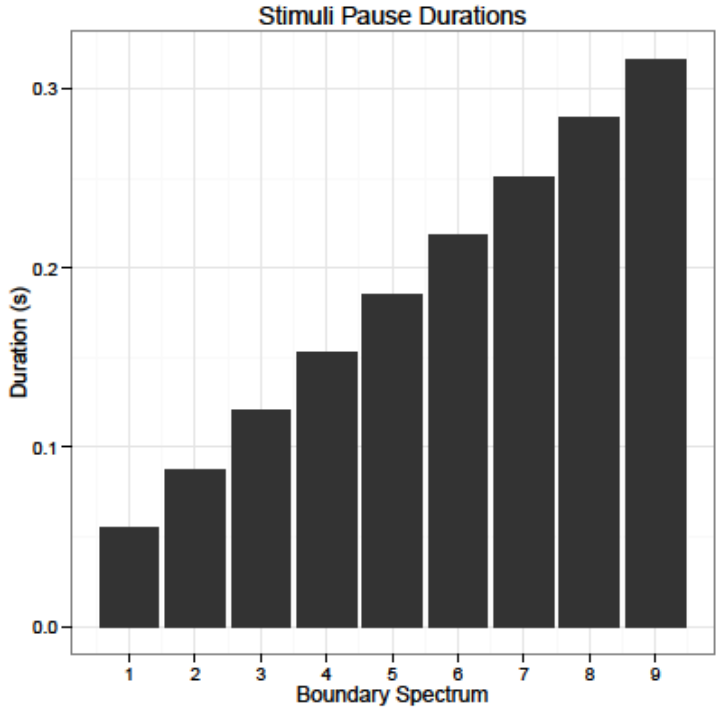

Figure 2: Following pause durations for critical words at each step of the boundary spectrum.

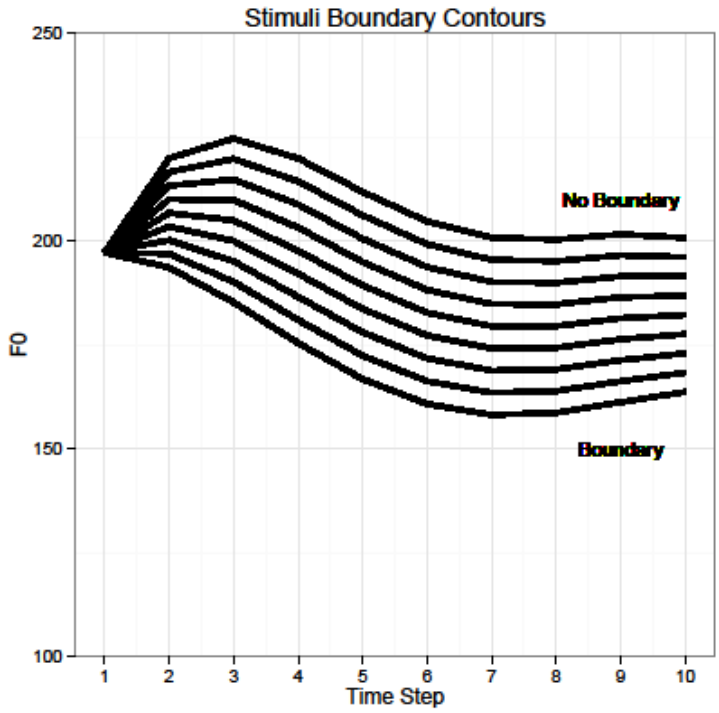

Figure 3: F0 contours for the critical words at each of the steps of the boundary spectrum.

Procedure. All recordings were uploaded to Qualtrics, an online survey service. The survey was posted on Amazon's Mechanical Turk website, where members were able to participate in the survey for pay. The instructions explained that speakers often group utterances into chunks, and that these chunks are often divided by what we call boundaries. They were then told that words that precede boundaries sound "different" than words that do not. Instructions were phrased in this way so that listeners would not explicitly look for cues such as pauses to determine whether there was a boundary or not (see Appendix). There were 2 recordings of sentences with naturally produced boundaries so that listeners could hear them, followed by a sentence indicating where they were likely to have heard a boundary in the examples. The speaker in these recordings was not the same as the speaker who recorded the sentences used in the study.

For each question, participants saw a media player icon of the recording and under it, the sentence in written form. Next to the sentence, the question read: "There is a boundary after:" The participants' task was to check boxes under the word(s) they felt preceded a boundary. Recordings could be played as many times as necessary, and participants could mark as many words as they wanted. The questions were presented in a random order, and all participants heard all 272 recordings. We analyzed the perceived boundary rate after the 2 critical words for each recording.

Data Analysis. We obtained binary boundary ratings for each word of each sentence that was presented. For this study, we limit analyses to the two critical regions. Participants rarely reported hearing boundaries at any of the non-manipulated regions. Only 3.8% of boundary reports were at non-critical regions (compared to 96.2% at critical regions). There were a total of 4,896 sentences, resulting in 9,792 data points.

Results

The data was analyzed using logistic mixed effects models to examine how boundary reports differed as a function of boundary spectrum and critical region (i.e. syntactically licensed vs unlicensed location for a boundary), as well as their interactions. All logistic mixed effect models were built using the lme4 package in R (Bates et al., 2015). Critical regions were effect coded, and random intercepts and slopes were included for subject and item. The models also included fixed effects for source sentence and sentence structure. Because the maximal model did not converge, we used the maximal random effects structure that converged, following conventions proposed in Barr et al. (2013). A summary of the model results is presented in Table 1.

Results are presented in Figure 4. There was a main effect of critical region ($b = 1.196$, $Z = 5.293$, $p < .001$): listeners were more likely to report boundaries at syntactically licensed locations than syntactically unlicensed locations. There was also an effect of boundary spectrum ($b = 0.121$, $Z = 4.647$, $p < .001$), with more boundary-like cues resulting in more boundary reports, as well as an interaction between critical region and boundary spectrum ($b = -0.068$, $Z = -4.580$, $p < .001$): more boundary-like cues increased boundary reports more strongly at the syntactically unlicensed region than at the syntactically licensed region. To explore the interaction, we conducted a post-hoc analysis of effects of the boundary spectrum for syntactically licensed and unlicensed locations. Boundary spectrum had a stronger effect at unlicensed locations ($b = 0.186$, $Z = 10.333$, $p < .001$) than licensed locations ($b = 0.048$, $Z = 2.203$, $p < .05$). This suggests that acoustic cues are more likely to influence the interpretation of boundaries in contexts in which boundaries are not expected. When boundaries are expected, acoustic cues either have a smaller effect or listeners are already at ceiling in perceiving a break. Effects for sentence structure and original source sentence were also investigated in the above

models. These effects are discussed in the Appendix, along with additional post-hoc analyses.

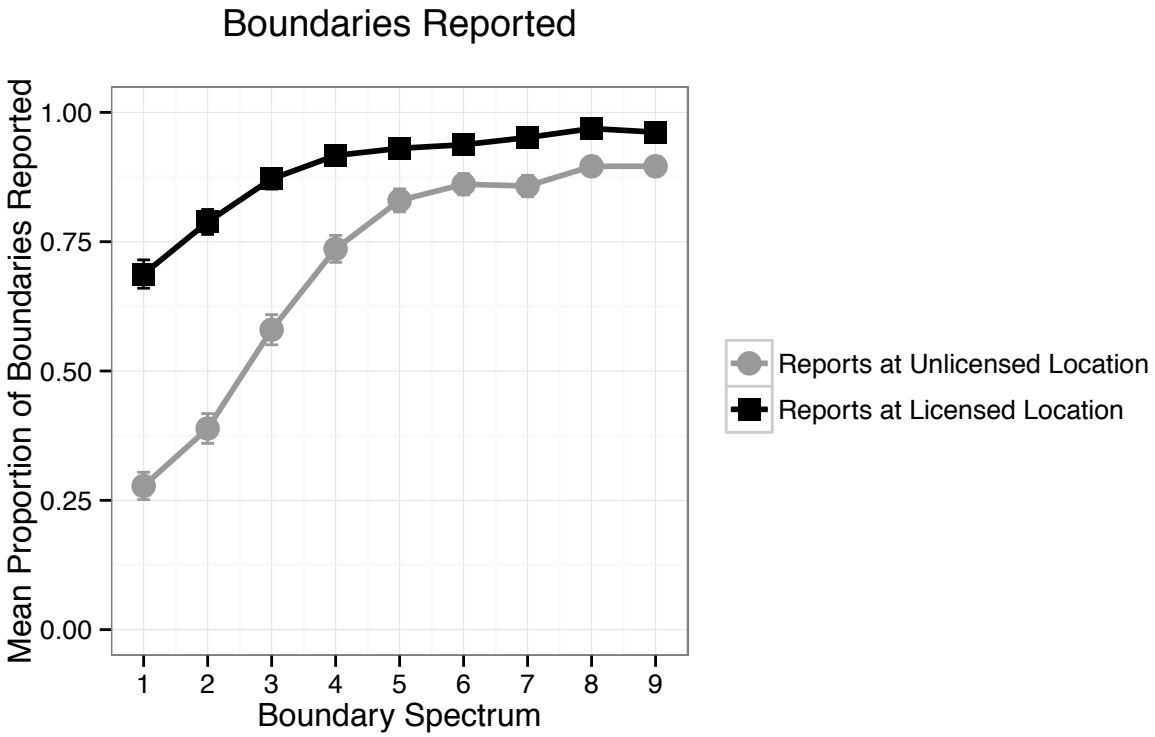

Figure 4: Mean proportion of boundaries reported per condition in Experiment 1. Error bars indicate standard error.

Discussion

These results suggest that listener expectations influence boundary processing. Specifically, whether the position observed was a syntactically licensed location for a boundary or not influenced whether a boundary was reported, even when controlling for acoustic cues. There was also a main effect of boundary spectrum, with listeners reporting more boundaries when words sound more boundary-like. Furthermore, there was an interaction between syntactic structure and boundary spectrum, which suggests that listeners are more likely to use acoustic evidence when it is encountered at syntactically unlicensed locations. Although the results suggest that boundary processing is not strictly a bottom-up process, there is one explanation that needs to be addressed. When participants read the instructions for the study, they were presented

with 2 example recordings, both of which included boundaries produced at syntactically licensed locations. Because of this, it is possible that the instructions created a bias for listeners to report boundaries only at these locations, and dismiss boundaries at unlicensed locations. We address this issue in Experiment 2 by asking participants to complete the very same task preceded by instructions that use examples of intonational boundaries at both syntactically licensed and unlicensed locations in the instructions.

Experiment 2

Methods

Participants. Twenty-one English speakers from the United States of America participated in the study. Three participants were excluded due to having learned a language other than English from an early age (before 5), and two participants were excluded for having participated in Experiment 1 before. This resulted in 16 monolingual English speakers. They were all users of Amazon's Mechanical Turk service, and they all had at least a 95% approval rating for previous task completions. They were paid $6.00 for participating in the study.

Materials. Materials were the same as those in Experiment 1.

Procedure. The procedure was the same as in Experiment 1, except for the instructions. While instructions were phrased in the same way as in the first study, one of the example recordings included a boundary produced in a syntactically unlicensed location. The purpose of this was to remove any possible bias we might have introduced in Experiment 1.

Data Analysis. Data analysis was similar to Experiment 1. Again, participants rarely reported boundaries at non-critical regions, with only 4.8% of boundary reports corresponding to non-critical regions (compared to 95.2% of reports corresponding to critical regions). There were a total of 4,352 sentences, resulting in 8,704 data points.

Results

The results are presented in Figure 5. A visual inspection of the Figure suggests that the patterns are largely the same as those in Experiment 1 although somewhat attenuated. Once again, there was a main effect of critical region ($b = 0.773$, $Z = 2.955$, $p < .01$), suggesting that listeners reported more boundaries at syntactically licensed locations than unlicensed locations, and a main effect of boundary spectrum ($b = 0.032$, $Z = 2.054$, $p < .05$), suggesting that listeners reported more boundaries when there were stronger acoustic cues indicating boundary presence. Furthermore, there was a significant interaction between critical region and boundary spectrum ($b = -0.033$, $Z = -2.203$, $p < .05$). Analyses investigating effects of boundary spectrum in syntactically licensed and unlicensed locations individually revealed an effect of boundary spectrum at unlicensed locations ($b = 0.066$, $Z = 3.372$, $p < .001$), but not at licensed locations ($b = 0.002$, $Z = 0.073$, $p = .942$). This suggests that acoustic factors have a stronger effect on interpretation at locations at which listeners do not expect to hear a boundary. Thus, Experiment 2 replicated the results from Experiment 1.

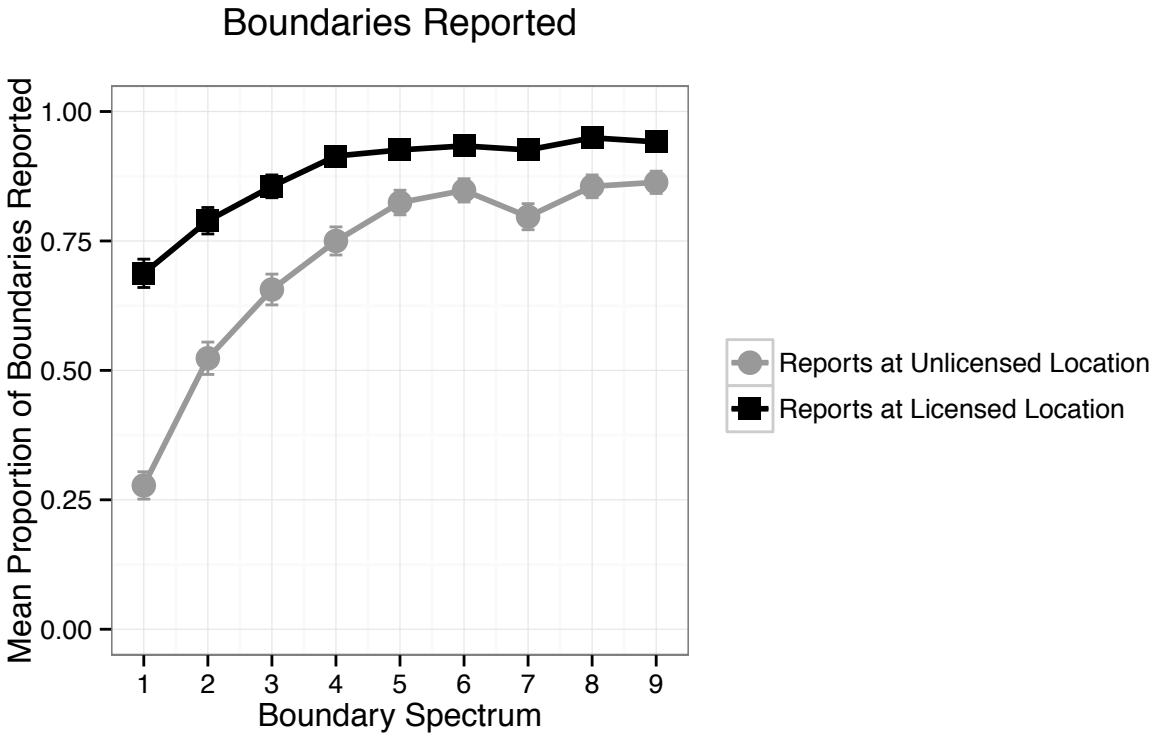

Figure 5: Mean proportion of boundaries reported per condition in Experiment 2. Error bars indicate standard error.

Discussion

The purpose of Experiment 2 was to eliminate the possibility that the results from Experiment 1 were the result of the instructions biasing participants towards reporting only boundaries at natural locations. The results from Experiment 2 were consistent with those of Experiment 1, suggesting that the effects were not driven by instruction bias. This provides further evidence for the presence of syntactic effects on intonational boundary processing.

One alternative explanation for this pattern of results is that these effects are the result of learning across the experiment. Because listeners only ever hear boundaries at two locations in the sentence, participants may be more likely to report boundaries at the two critical locations. There is at least some evidence that learning effects may be a potential confound. In both experiments, participants report hearing boundaries at the syntactically unlicensed location

around 30% of the time, even when there is no acoustic evidence consistent with the presence of a boundary (we discuss this more in the General Discussion). It is possible that because the participant frequently hears boundaries at this location throughout the experiment, they may be more likely to report hearing a boundary there, and may even be more likely to do so in a canonical boundary position.

In post-hoc analyses, we examined whether trial order had any effect on response patterns in Experiments 1 and 2. Treating order as a continuous fixed effect resulted in models that failed to converge, so order was binned into quartiles and included as a fixed effect in the maximal logistic mixed effect model that converged. There was no effect of order in Experiment 1 ($b = -0.036$, $p = 0.162$), and there was only a marginal effect for Experiment 2 ($b = -0.050$, $p = 0.072$). To the extent that there was a numerical trend, in both experiments, participants were less likely to report a boundary the further they progressed through the experiment, which is inconsistent with participants learning to report boundaries at the two critical locations over the course of the experiment. This suggests that participants were not reporting boundaries based on where they had heard them before in the context of the experiment. Nevertheless, it is possible that learning occurred too quickly to be captured by including trial order in the models. Although this learning would be largely consistent with our claim that expectations influence the interpretation of boundaries, the goal of Experiment 3 was to further demonstrate that these expectations were not generated across the course of the experiment. In Experiment 3 we rule this possibility out by exposing each participant to only two trials.

Experiment 3

Methods

Participants. Three-hundred English speakers from the United States of America participated in the study. Thirty-eight participants were excluded due to having learned a language other than English from an early age (before 5). This resulted in 262 monolingual English speakers. They were all users of Amazon's Mechanical Turk service, and they all had at least a 90% approval rating for previous task completions. They were paid $0.75 for participating in the study.

Materials. Materials were a subset of those used in Experiments 1 and 2. Only recordings that had been manipulated so that both critical words were at boundary step 1 were used (16 total). This means that neither of the critical words had acoustic cues that should signal the presence of a boundary. Each participant heard a random subset of 2 from these 16 recordings.

Procedure. The procedure was the same as it was for Experiments 1 and 2. The same instructions and examples from Experiment 2 were used.

Data Analysis. Data analysis was identical to Experiments 1 and 2. Because there was no longer a boundary spectrum manipulation, logistic mixed effect models only tested for the main effect of boundary position. There were a total of 524 sentences, resulting in 1048 data points.

Results

Experiment 3 replicated the main results from Experiments 1 and 2. Participants reported hearing a boundary at syntactically licensed locations 59.4% of the time, as opposed to 42% of the time at syntactically unlicensed locations. This main effect of critical region was significant ($b = 0.506$, $Z = 4.185$, $p < .001$).

Discussion

Experiment 3 was designed to determine whether listeners from Experiments 1 and 2 were developing expectations across the experiment that were driving their reports of hearing an intonational boundary. The results rule this out. In Experiment 3, listeners only heard 2 sentences. This is unlikely to have been enough exposure for them to develop new expectations about likely locations for intonational boundaries. In addition, none of the recordings they heard were supposed to have signaled boundary presence. The two critical words were identical in terms of duration, following pause duration, and F0 contour. Thus, any difference in reports between the two critical regions was due to the syntactic position at which the boundary occurred.

One unexplained puzzle is the relatively high rates of boundaries reported at the syntactically unlicensed location. We think it is likely that these are the result of the acoustic manipulations of the recordings. Although recordings were resynthesized so that they sounded as natural as possible, there were sometimes noticeable changes in speaking rate or F0 from one word to the next due to these manipulations. This could have resulted in the detection of a boundary even for words that were resynthesized to sound like non-boundary words. However, it is important to note that this only explains the overall base rate of hearing boundaries. It does not explain why listeners report hearing a boundary more often in the syntactically expected location than syntactically unexpected location, where the acoustic signal is exactly the same.

General Discussion

The primary goal of this paper was to determine whether syntactic expectations influence the processing of intonational boundaries. We found that listeners use syntactic information to decide where boundaries are likely to occur. They even report hearing boundaries in these locations in the absence of boundary cues. These results suggest that, in addition to acoustic

phonetic information, other linguistic processing streams influence boundary processing.

These results are consistent with those of Cole et al. (2010), who also found that syntactic context was a more reliable predictor of boundary reports than acoustic information. Additionally, our results found critical interactions between syntactic context and acoustic information, suggesting that listeners are likely to report boundaries according to their initial expectations, but strong acoustic evidence for boundaries can diminish the differences caused by these initial biases.

These findings suggest that a complete theory of prosody needs to include a mechanism by which information flows bidirectionally between prosodic structure and "higher-level" structures. Current models of prosodic parsing could benefit from this added information. For example, we agree with Beckman (1996) that prosody itself needs to be parsed, but add that syntactic information plays a role at some point in prosodic parsing. Similarly, a model like the one described in Schafer (1997) could be modified to account for these data. In such a model, prosodic structure would initially be constructed from the acoustic information, and this structure would then help guide syntactic and semantic interpretation. However, these higher-level interpretations then constrain prosodic representations, such that the original prosodic interpretation fits the most plausible interpretation at other levels of linguistic structure. Although our results are not sufficient to specify what type of architecture a complete model of prosody must have, they do suggest that bidirectionality between prosodic processing and other levels of language processing are a necessary feature.

These results also have implications for common practices in the field. For example, one major assumption in the coding approach taken by most prosodic annotation systems is that transcribers accurately mark prosodic events based on the acoustics of words and visual

information from a pitch track. Although coding systems like ToBI make clear that their main goal is to annotate the subjective prosodic perception of the listener, the tools and instructions that are provided are mainly based on acoustic-phonetic information (Beckman & Ayers, 1997). Similarly, the RaP annotation system makes use of heuristics that are based on low-level cues, such as the perception of beats and stress when deciding what words constitute a phrase (Dilley & Brown, 2005). The present study suggests that transcribers might not be guided by acoustics alone. In fact, expert transcribers might have even stronger expectations than non-expert listeners about where intonational boundaries should occur. This may lead to reports of boundaries when little acoustic evidence for a boundary exists. On the other hand, it is also possible that expert coders are less susceptible to syntactic expectations if they are trained to focus only on the acoustic signal. Of course, if acoustic-phonetic cues are not the only factors that determine the percept of a boundary, we need to re-consider how we conceptualize intonational boundaries and their psychological representations. Ultimately, the data from this experiment suggests that coding schemes may need to be reconsidered with expectation biases in mind.

One alternative explanation for these findings is that listeners have preferences for where boundaries should occur in sentences, such as preferring late boundaries over early boundaries, and the results reported above are not actually the result of syntactic expectations. This is possible given that the syntactically expected location always occurred later in the sentence than the syntactically unexpected location. However, we think this explanation is unlikely. First, the critical words were as close to each other as possible. There were no words in between them in one sentence structure (as in “big bowl”) and only one word between them in the other structure (“bowl that’s big”). This ensured that the critical words were roughly in the same positions,

meaning general preferences for positions within a sentence (e.g., "near the middle") should not have made a significant difference in reports between the critical words. We believe that syntactic expectancy is a simpler explanation.

An important open question is understanding why syntactic context would affect prosody processing at all. This question is not surprising if we consider the field of language processing as a whole. A number of studies on language comprehension have proposed parallel-process or constraint-based model accounts in which separate processing systems interact to produce a final interpretation of a sentence (e.g., Kim & Osterhout, 2005; Macdonald et al., 1994; Tabor & Tanenhaus, 1999; Trueswell et al., 1994).

Although these models have primarily been proposed to account for the interaction between syntax and semantics in sentence comprehension, similar architectures might underlie syntactic and prosodic processing. In this case, prosodic structure and syntactic structure are processed separately. When prosodic and syntactic boundaries occur at the same location, the comprehension system has enough evidence to infer that there is a boundary present at that location. However, in cases where acoustic cues are at odds with syntactic expectations, the comprehension system needs to weigh the evidence from both sources in order to make a guess. When prosodic cues are strong enough, the comprehension system will interpret the cues as a prosodic boundary. But, when the acoustic cues are missing or weak, expectations from syntactic cues drive the final interpretation. In either case, the comprehension system's goal is to reach the best global interpretation of the utterance given the evidence. It is possible that in some cases, in order to reach the best possible interpretation, listeners' prosodic representation needs to be revised in light of new syntactic and semantic information.

Although our results are discussed from a parallel-process account, we believe other

frameworks can also account for these findings. For example, noisy-channel models of language comprehension (e.g., Gibson et al., 2013) propose that listeners use Bayesian inference to process linguistic structure. Under this framework, it is assumed that communication is inherently noisy and that a rational listener will try to determine the relative probabilities of a speaker's intended production given the available linguistic cues. These inferences are driven by both the prior probability of the cues and the likelihood of the cues given the intended production. Within this framework, a listener who is trying to determine whether a boundary was produced would calculate the prior probability of syntactic and acoustic cues to the boundary and the likelihood of these cues given the presence, or absence, of an actual boundary. Perhaps the effects we see in these experiments have to do with the calculation of this likelihood. Acoustic cues to prosody may be weighted less heavily than syntactic cues because word durations, pauses, and F0 are affected by so many different factors in English (e.g., Watson, 2010). This might result in listeners weighting syntactic cues more heavily, as syntactic cues may have proven to be more reliable in the past.

Of course, the parallel processing and noisy-channel theories discussed above are general theories of why there might be effects of syntax on detecting prosodic boundaries. They are not theories of the mechanisms by which syntactic structure has its effects. Bishop (2013) has proposed two possible ways in which non-prosodic linguistic knowledge might affect the perception of prosody. One possibility is that the processing system is restorative: it projects grammatical knowledge of the mapping between prosodic structure and syntactic structure onto the perceptual signal, creating the percept of the presence (or absence) of a boundary. Another possibility proposed by Bishop is that listeners' detection of prosodic phenomena are partly an epiphenomenal product of processing a sentence. For the data discussed here, processing the

closure of a syntactic phrase may create the subjective experience of a break, which thereby colors how listeners perceive the sentence's prosody.[2] Cole et al. (2010) propose a similar processing based explanation for effects of word frequency on listener's perception of acoustic prominence. They found that low frequency words were judged to be more prominent than high frequency words even when acoustic information did not predict a difference. They propose that the cognitive effort necessary for processing a low frequency word may have led to the perception of stronger acoustic prominence. Although the current work cannot adjudicate between differing mechanisms driving effects of syntax and prosody perception, these proposals in the literature suggest a means by which these effects occur might. We leave the question of what mechanisms drive the effect to future research.

Perhaps what is most surprising about these data is the strength of listener expectations in driving boundary detections. In syntactically licensed locations, participants report hearing a boundary the majority of the time, even when no acoustic evidence for the boundary exists. If this is representative of how other prosodic phenomena are perceived, this may explain some of the controversy in the prosody literature about the nature of prosodic categories. For example, there is a great deal of debate surrounding what the acoustic correlates are for prosodic phenomena like intonational boundaries and pitch accents (see Wagner & Watson, 2010 for a review). The current study suggests it may not be possible to ask these types of questions without controlling for listener expectations about prosodic structure.

To conclude, syntactic context is an important predictor in whether listeners report a boundary at a given location. These data suggest that information from non-prosodic processing systems influence prosodic processing. Listeners likely form expectations based on their

[2] We would like to thank an anonymous reviewer for pointing out these possible mechanisms for the effects we see in the Experiments.

experiences with language in the past, and these expectations influence what they hear in the present.

References

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *Journal of Memory and Language*, *68*, 255-278.

Bates D., Maechler M., Bolker B., & Walker S. (2015). *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.1-8, http://CRAN.R-project.org/package=lme4.

Beach, C. M. (1991). The interpretation of prosodic patterns at points of syntactic structure ambiguity: Evidence for cue trading relations. *Journal of Memory and Language, 30,* 644-663.

Beckman, M. E. (1996). The parsing of prosody. *Language and Cognitive Processes, 11,* 17-67.

Beckman, M. E. & Ayers Elam, G. (1997). Guidelines for ToBI labeling, version 3: Ohio State University.

Bishop, J. (2013). Prenuclear accentuation: phonetics, phonology, and information structure. Doctoral dissertation, UCLA, Los Angeles, California.

Bishop, J. (2012). Information structural expectations in the perception of prosodic prominence. In G. Elordieta & P . Prieto (Eds.), *Prosody and Meaning* (pp. 239-270). Berlin: Walter de Gruyter.

Boersma, P., & Weenink, D. (2015). Praat: doing phonetics by computer [Computer program]. Version 6.0.05, retrieved 19 November 2015 from http://www.praat.org/

Brown, M., Salverda, A. P., Dilley, L. C., & Tanenhaus, M. K. (2012). Metrical expectations from preceding prosody influence spoken word recognition. *Proceedings of the 34th Annual Conference of the Cognitive Science Society.* Austin, TX: Cognitive Science Society.

Brown, M., Salverda, A. P., Dilley, L. C., & Tanenhaus, M. K. (2011). Expectations from preceding prosody influence segmentation in online sentence processing. *Psychonomic Bulletin and Review 18,* 1189-96.

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

Cole, J., Mo, Y., & Baek, S. (2010). The role of syntactic structure in guiding prosody perception with ordinary listeners and everyday speech. *Language and Cognitive Processes, 25,* 1141-1177.

Cole, J., Mo, Y., & Hasegawa-Johnson, M. (2010). Signal-based and expectation-based factors in the perception of prosodic prominence. *Laboratory Phonology, 1,* 425–452.

Cooper, W. E., & Paccia-Cooper, J. (1980). *Syntax and speech.* Cambridge, MA: Harvard University Press.

de Pijper, J. R., & Sanderman, A. A. (1994). On the perceptual strength of prosodic boundaries and its relation to suprasegmental cues. *The Journal of the Acoustical Society of America, 96,* 2037-2047.

Dilley, L. C. & Brown. M. (2005). The RaP (Rhythm and Pitch) Labeling System, Version 1.0.

Dilley, L. C., Mattys, S. L., & Vinke, L. (2010). Potent prosody: Comparing the effects of distal prosody, proximal prosody, and semantic context on word segmentation. *Journal of Memory and Language*, *63,* 274-294.

Ferreira, F. (1988). *Planning and timing in sentence production: The syntax-to-phonology conversion.* Unpublished dissertation, University of Massachusetts, Amherst, MA.

Ferreira, F. (1993). Creation of prosody during sentence prosody. *Psychological Review, 100,* 233-253.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance, 6,* 110-125.

Garrett, M., Bever, T. G., & Fodor, J. (1966). The active use of grammar in speech perception. *Perception and Psychophysics, 1,* 30-32.

Gee, J., & Grosjean, F. (1983). Performance structures: A psycholinguistic appraisal. *Cognitive Psychology, 15,* 411-458.

Gibson, E., Bergen, L., & Piantadosi, S. T., (2013). The rational integration of noise and prior semantic expectation: Evidence for a noisy-channel model of sentence interpretation, *Proceedings of the National Academy of Sciences*, *11,* 8051-8056.

Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language, 52*, 205-225.

Kjelgaard, M. M., & Speer, S. R. (1999). Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language, 40,* 153-194.

Klatt, D. H. (1975). Vowel lengthening is syntactically determined in a connected discourse. *Journal of Phonetics, 3,* 129-140.

Kraljic, T., & Brennan, S. E. (2005). Using prosody and optional words to disambiguate utterances: For the speaker or for the addressee? *Cognitive Psychology, 50,* 194-231.

Ladd, D. R. (2008). *Intonational phonology* (2nd edn.). Cambride, UK and New York, NY: Cambridge University Press.

Lehiste, I. (1973). Phonetic disambiguation of syntactic ambiguity. *Glossa, 7,* 107-122

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review, 101*, 676-703.

Nespor, M., & Vogel, I. (1986). Prosodic phonology. Dordrecht, the Netherlands: Foris.

Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. Cohen, J. Morgan, & M. Pollack (Eds.), *Intentions in communications*. Cambridge, MA: MIT Press.

Price, P. J., Ostendorf, S., Shattuck-Hufnagel, S., & Fong, C. (1991). The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America, 9,* 2956-2970.

Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981). Speech perception without traditional speech cues. *Science, 22,* 947-949.

Schafer, A. J. (1997). Prosodic Parsing: The Role of Prosody in Sentence Comprehension. Doctoral dissertation, University of Massachusetts, Amherst, MA.

Schafer, A. J., Speer, S. R., & Warren, P. (2005). Prosodic influences on the production and comprehension of syntactic ambiguity in a game-based conversation task. In M. Tanenhaus & J. Trueswell (Eds.) *Approaches to Studying World Situated Language Use: Psycholinguistic, Linguistic and Computational Perspectives on Bridging the Product and Action Tradition* (pp. 209-225). Cambridge: MIT Press.

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

Selkirk, E. O. (1986). On derived domains in sentence phonology. *Phonology, 3,* 371-405.

Selkirk, E. O. (1995). Sentence prosody: intonation, stress and phrasing. In J. Goldsmith (Ed.), *The Handbook of Phonological Theory*. London, UK: Blackwell.

Snedeker, J., & Trueswell, J. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language, 48,* 103-130.

Streeter, L. A. (1978). Acoustic determinants of phrase boundary perception. *The Journal of the Acoustical Society of America, 64,* 1582-1592.

Tabor, W., & Tanenhaus, M. K. (1999). Dynamical Models of Sentence Processing. *Cognitive Science, 23*, 491-515.

Truckenbrodt, H. (1999). On the relation between syntactic phrases and phonological phrases. *Linguistic Inquiry, 30,* 219-255.

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language, 33,* 285-318.

Turk, A., & Shattuck-Hufnagel, S. (2007). Phrase-final lengthening in American English. *Journal of Phonetics 35,* 445-472.

Wagner, M., & Watson, D. G. (2010). Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes, 25,* 905-945.

Watson, D. G. (2010) The many roads to prominence: Understanding emphasis in conversation. In B. Ross (Ed.) *The Psychology of Learning and Motivation,* Vol. 52 (pp. 163-183).

Watson, D. G., & Gibson, E. (2004). The relationship between intonational phrasing and syntactic structure in language production. *Language and Cognitive Processes*, *19,* 713-755.

Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustic Society of America, 91,* 1707-1717.

Table 1: Summary of logistic mixed effect model results.

	Experiment 1	Experiment 2	Experiment 3
Intercept	b = 0.140 SE = 0.201 Z value = 0.698 p = 0.485	b = 0.242 SE = 0.181 Z value = 1.333 p = 0.182	b = 0.047 SE = 0.076 Z value = 0.622 p = 0.534
Critical Region	b = 1.197 SE = 0.226 Z value = 5.293 p < 0.001	b = 0.773 SE = 0.262 Z value = 2.955 p = 0.003	b = 0.506 SE = 0.121 Z value = 4.184 p < 0.001
Spectrum	b = 0.121 SE = 0.026 Z value = 4.647 p < 0.001	b = 0.032 SE = 0.016 Z value = 2.054 p = 0.040	
Critical Region * Spectrum	b = -0.068 SE = 0.015 Z value = -4.850 p < 0.001	b = -0.033 SE = 0.015 Z value = -2.203 p = 0.028	

Appendix A

Items: Boundary locations are indicated by "|"s. Words that were acoustically manipulated are in bold.

1a. Put the **bead** that's **teal** | in the jar.

1b. Put the **bead** | that's **teal** in the jar.

1c. Put the **teal bead** | in the jar.

1d. Put the **teal** | **bead** in the jar.

2a. Put the **bowl** that's **big** | on the tray.

2b. Put the **bowl** | that's **big** on the tray.

2c. Put the **big bowl** | on the tray.

2d. Put the **big** | **bowl** on the tray.

3a. Put the **book** that's **black** | on the chair.

3b. Put the **book** | that's **black** on the chair.

3c. Put the **black book** | on the chair.

3d. Put the **black** | **book** on the chair.

4a. Put the **dog** that's **brown** | on the couch.

4b. Put the **dog** | that's **brown** on the couch.

4c. Put the **brown dog** | on the couch.

4d. Put the **brown** | **dog** on the couch.

Appendix B

Instructions presented for all experiments:

When we speak, we group our utterances into smaller chunks. These chunks are often divided by what we call "boundaries." When words are right before boundaries, they tend to sound different to when they are not. For example, listen to the following sentence:

(Recording of a non-manipulated production: “Put the green frog | in the box.”)

In this sentence, it sounds like there is a boundary after the word "frog." Here is another example sentence:

(Recording of a non-manipulated production: “Put the frog that is green | in the box.” For Experiments 2 and 3 this was changed to “Put the green | frog in the box.”)

In this sentence, it sounds like there is a boundary after the word "green.“

For this task, you will hear some recordings of sentences and will have to specify after what words you hear a boundary. You can listen to the recordings as many times as you like, and can mark more than one word as having a boundary after it.

Appendix C

Example spectrograms of stimuli:

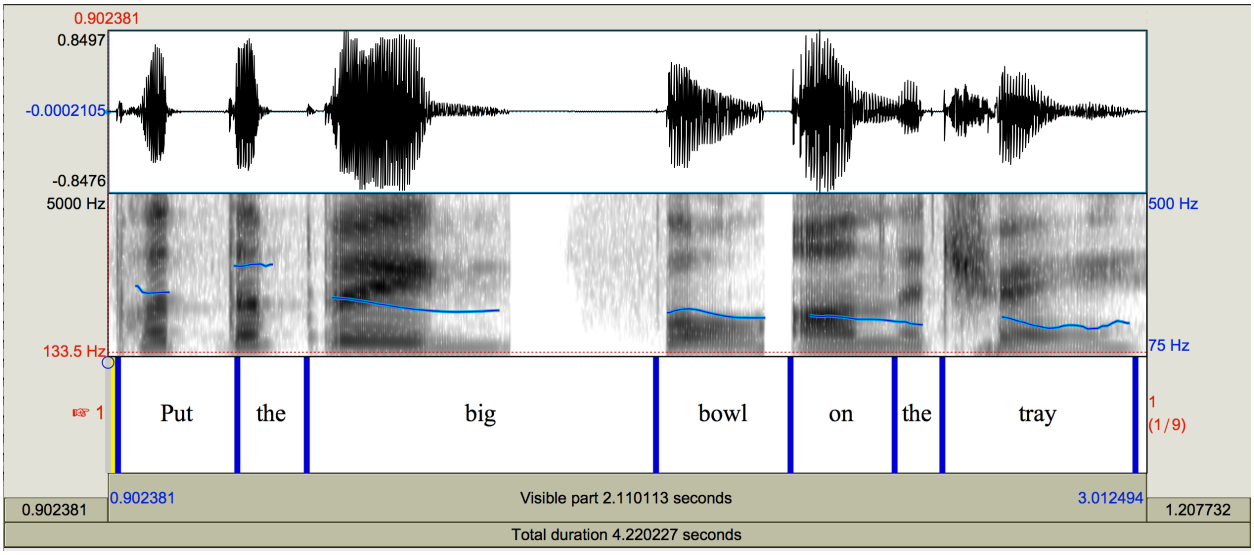

Figure 6. Boundary after "big."

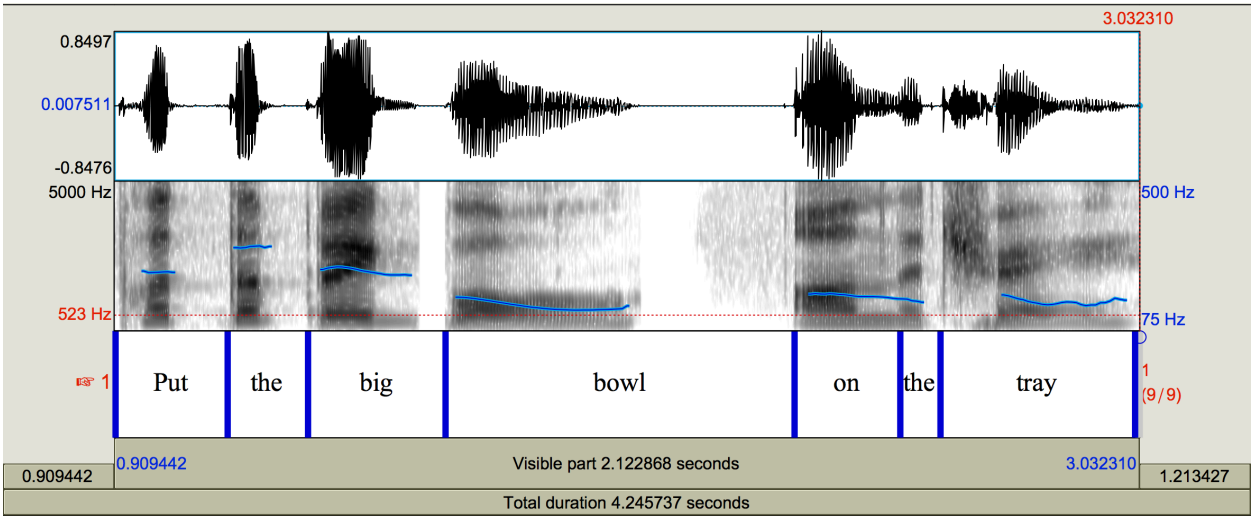

Figure 7. Boundary after "bowl."

Appendix D

In addition to the critical effects discussed in the Results sections, we found a main effect of sentence structure in Experiment 1 ($p < .05$) and Experiment 2 ($p < .001$), such that there were more boundaries reported overall in noun-modifier sentences than modifier-noun sentences. Even so, both structures exhibited the same boundary spectrum by critical region interaction (Experiment 1: modifier-noun structure: $b = -0.067$, $Z = -4.541$, $p < .001$; noun-modifier structure: $b = -0.067$, $Z = -4.540$, $p < .001$; Experiment 2: modifier-noun structure: $b = -0.033$, $Z = -2.202$, $p < .05$; modifier-noun structure: $b = -0.033$, $Z = -2.202$, $p < .05$). There was no effect of sentence structure for Experiment 3.

Additional post-hoc analyses were run to investigate whether listeners reported more boundaries at the unlicensed position for noun-modifier structures than modifier-noun structures. Listeners reported more boundaries at the unlicensed location for the noun-modifier structures than in the modifier-noun structure (Experiment 1: modifier-noun structure mean boundary reports: 0.40, noun-modifier structure mean boundary reports: 0.53; $b = -0.261$, $Z = -9.341$, $p < .001$; Experiment 2: modifier-noun structure mean boundary reports: 0.45, noun-modifier structure mean boundary reports: 0.50; $b = -0.232$, $Z = -7.405$, $p < .001$). This difference could be due to interpreting the relative clause as a non-restrictive. However, the complementizer "that" is not typically used as a complementizer in a non-restrictive relative clause (see Grodner, Gibson, & Watson, 2005), so it is unlikely that this is driving the difference in effect size. Nevertheless, when participants reported hearing a boundary at the unlicensed location of noun-modifier structures in Experiment 1, about 57% of the time they also marked a boundary at the licensed location. In Experiment 2 this occurred about 33% of the time. We leave the source of this difference to future work, but critically, there were effects of syntactic expectation in both

sentence structures.

There was no effect of source sentence, i.e. whether the original structure before re-synthesis was one in which a boundary after the target word was present or not. Additionally, items that resulted in a double-stop closure between the 2 target words (e.g., “big bowl”) were coded because double-stop closures can sometimes create the percept of a boundary. This predictor was added to the models described in the Results sections, but there was no effect of double-stop closure items.