The Influence of Prosodic Boundaries in Syntactic Interpretation

By

Caoimhe M. Harrington Stack

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Psychology

May 8th, 2020

Nashville, Tennessee

Approved:

Duane G. Watson, Ph.D.

Sarah Brown-Schmidt, Ph.D.

Megan Saylor, Ph.D.

Emily Phillips Galloway, Ed.D.

Table of Contents

List of Figures

List of Tables

I. Prosodic Boundaries and Syntax

Introduction

Upon hearing a story that ends with an unlikely outcome, in this case that a man bit a dog, a speaker might ask incredulously, “So did the man really bite the dog?”. To produce an utterance even this short requires the coordination of many aspects of language. First, the speaker accesses the words that cover the content of the situation -dog, bite, human- and then uses English syntax to order the content to appropriately communicate information about the relation between the words and to indicate that they are asking a question. To produce the message out loud, the speaker articulates the needed English phonemes to form each word of the sentence intelligibly. After doing all this, the speaker can say “So did the man really bite the dog?” But as an additional layer to help convey meaning the speaker will use prosody, suprasegmental sounds which go beyond word choice, to indicate that they are questioning whether the man *really* bit the dog. To indicate that a question is being asked, this speaker likely produced their phrase with a rising intonation, with a noticeable rise in pitch at the end of ‘dog’. This intonation is often used in American English to indicate that a question is being asked (Pike, 1945) and serves as an additional cue on top of the lexical content that the listener should provide an answer.

This use of intonation to indicate a question or a statement is just one example of how prosody aids communication in spoken language. Beyond intonation, prosody includes speech phenomena such as rhythm, word grouping, pauses, pitch of speech, and prosodic boundaries (Price et al., 1991; Shattuck-Hufnagel & Turk, 1996; Wagner & Watson, 2010). Often, prosody serves as a ‘layer’ on top of language that provides additional information about the speaker and the context (Banse & Scherer, 1996; Lima & Castro, 2011; Simpson, 2009), and may help to facilitate the comprehension of speech (Sanderman & Collier, 1997; Veenendaal, Groen, & Verhoeven, 2014; Wingfield, Lombardi, & Sokol, 1984) . Certain prosodic features are associated with speaker identity. For example male speakers often have lower pitches than female speakers (Simpson, 2009). Prosodic features can also help to convey emotions the speaker is feeling. For example, fear is associated with an increased F0 mean and range, and an increased speech rate (Banse & Scherer, 1996). And prosody that matches the speaker's message may improve comprehension while mismatched prosody may hinder it (Braun, Dainora, & Ernestus, 2011). One can imagine in the above dog-man conflict that, after hearing the story, a friend might ask "So did the man bite the dog?" but with a flat intonation that falls slightly. In this case, the listener will still likely realize that a question is being asked, but will have to rely on the syntactic information alone to determine whether a question or statement is being produced. Prosody may not be explicitly necessary to understand the content of a conversation but it facilitates communication by providing information about the speaker and helping to structure the message they are conveying.

The focus of this dissertation is investigating how the prosodic feature of prosodic boundaries influence syntactic parsing. Prosodic boundaries occur where utterances are broken down into smaller sub-groupings of words by a speaker (Lehiste, 1973). For example, a speaker may produce the sentence, "I got the information about the fundraiser when I went to the store//, ran into Mrs. Smith who is on the school's committee//, and she gave me the information about the fundraiser." where // indicates the presence of a prosodic boundary. These boundaries can be marked by a variety of cues that occur together or alone. These cues include pauses, increased word duration before the boundary, boundary tones, and pitch accents (Lehiste, Olive, &

Streeter, 1976; Pijper and Sanderman, 1994; Price et al., 1991; Swerts, 1996; Wightman et al., 1992).

One feature of prosodic boundaries is that their placement within phrases is often associated with certain syntactic structures (Cutler et al., 1997; Frazier, Carlton, & Clifton, 2006; Kraljic & Brennan, 2005b; Price et al., 1991; Schafer et al., 2000). This can be most clearly demonstrated in globally ambiguous sentences. For example, consider the sentence:

1. I met the daughter of the colonel who was on the balcony.

Sentence (1) has two grammatically acceptable interpretations which a speaker could convey– that the daughter is on the balcony, or that the colonel is on the balcony. Which interpretation a listener ultimately follows is dependent on what they believe the speaker intended to attach the phrase ‘on the balcony’ to – the daughter or the colonel. These two attachment options are referred to as high-attachment and low-attachment, based on how the phrase in question would syntactically attach to other elements of the sentence. The low-attachment interpretation occurs when the comprehender attaches ‘the balcony’ to the closer noun phrase, ‘the colonel’. The high-attachment interpretation occurs when ‘the balcony’ is instead attached farther away and ‘the daughter’ is understood to be the attachment site. While native English speakers have an overall preference for the low-attachment interpretation (Carreiras & Clifton, 1999), the strength of this preference can be influenced by where a speaker places a prosodic boundary within the phrase.

The influence of boundaries on syntactic interpretation of sentences is often tested by asking listeners to provide interpretations of ambiguous sentences with the location and size of boundaries within the phrase manipulated. Carlson, Clifton, and Frazier (2001) designed an experiment that tested listeners' comprehension of sentences like the one shown below, where (1) and (2) indicate potential boundary locations.

2. "I met the daughter (1) of the colonel (2) who was on the balcony."

Reliably, a larger boundary cue at location (1) biased listeners towards an early-attachment interpretation of the sentence, where the colonel was on the balcony. When the boundary instead occurred at location (2), listeners increased their bias towards the high-attachment interpretation. This shift in attachment-bias suggested that prosodic boundaries can provide information to listeners about the syntactic structure of utterances. The existence of this relationship between prosodic boundaries and syntactic interpretation suggests a systematic relationship between the two, but this relationship is not deterministic.

Prosodic boundaries tend to coincide with syntactic boundaries (Price et al., 1991), but this is not a one-to-one relationship. Even in Carlson, Clifton, and Frazier (2001), an early boundary biased listeners to a low-attachment interpretation only about 70% of the time. So while the effect of boundary placement on sentence interpretation is robust, it does not guarantee that a listener will interpret a sentence in a specific way. Further complicating the link between prosodic boundaries and syntax is the fact that boundary use can be optional. For example the ambiguous sentence, "The bus driver stopped the rider with a mean glare" can have a Verb Phrase (VP) interpretation, where the driver used a mean glare to stop the rider, or a Noun Phrase (NP) interpretation, where the rider had a mean glare. This sentence can be produced with boundaries in many or no locations, as shown below.

3(a) The bus driver stopped // the rider with a mean glare.

3(b) The bus driver // stopped the rider with a mean glare.

3(c) The bus driver // stopped //the rider // with a mean glare.

3(d) The bus driver stopped the rider with a mean glare.

Schafer (1997) asked subjects to decide whether sentences such as 3(a) – 3(d) had a VP or an NP interpretation. The rate of VP interpretation was influenced by where a boundary occurred, but importantly all forms of the sentence sounded acceptable to listeners regardless of boundary placement.

There also exist cases where the placement of a boundary does not help a listener discern the underlying syntax of a sentence. In the ambiguous sentence ‘The shooting of the hunters was terrible’, placing a prosodic boundary at different locations has no effect on how listeners interpret the sentence (Lehiste, Olive, & Streeter, 1976; Shattuck-Hufnagel & Turk, 1996). This suggests that prosodic boundaries are not necessary for interpreting syntactic structure, but are more helpful in some contexts than in others.

The combination of these data that hint at a systematic, but optional, relationship between prosody and syntax suggests a probabilistic relationship. Prosodic boundaries *tend* to be produced in predictable locations in certain syntactic structures, but there is a great deal of variability in their production. This may be to help the speaker structure their language for themselves, or it may be to serve as an indicator of the syntactic structure for the benefit of the listener. While the question of why speakers produce prosodic boundaries is interesting, it was not the focus of this dissertation. Instead, the experiments here focus on how *listeners* respond to prosodic boundaries to make syntactic interpretations.

Understanding how prosodic boundaries are used for language comprehension is a challenging process. Since listeners' syntactic preferences are altered by where a prosodic boundary occurs, it is clear that boundary location is used to aid language comprehension to some extent. However, the variability of this relationship, and the flexibility speakers have in deciding where to place boundaries, indicates that listeners are often not relying on these boundaries to understand what is happening in a conversation. This variability makes it difficult to understand how the parser uses prosodic boundaries in online processing. In the following experiments two language processing models that provide possible explanations of how listeners use boundaries, the Visibility Hypothesis and the Ideal Adapter Framework, were explored. The goal was to investigate the influence that boundaries have on syntactic interpretation in order to understand how the language parser integrates boundary presence into language comprehension.

The Visibility Hypothesis, explained in more detail below, broadly states that where prosodic boundaries are placed influences how listeners process the connection between informational units within an utterance (Frazier & Clifton, 1997). This constrains how likely a listener is able to make one interpretation over another when faced with multiple potential interpretations. Under this hypothesis, the language parser uses processing and/or attentional resources to track the information contained in a sentence. Within a sentence, prosodic boundaries parcel pieces of linguistic information together, making language within the same parcel more accessible to the parser. Thus, the parser is biased to interpret sentences in a way that reduces the demand on processing resources, and so it is sensitive to where prosodic boundaries occur and how they divide up utterances.

On the other hand, the Ideal Adapter Framework (Kleinschmidt & Jaeger, 2015) proposes that language comprehension is the result of the parser being sensitive to the distribution of linguistic cues in the environment. That is, listener expectations about the boundary-syntax relationship may exist due to experiences listeners have with prosodic boundaries being produced in conjunction with certain syntactic structures. Importantly, these statistics can be continuously updated and are sensitive to recent changes in the linguistic distribution. In the case of boundaries, this would mean that listeners tend to have a certain boundary-syntax bias due to speakers producing them in this pattern. However, this relationship should be flexible so that if the parser were to encounter evidence that the boundary-syntax relationship has changed, it can update its expectations. These two theories are discussed in more detail below.

Prosodic Visibility Hypothesis

The Visibility Hypothesis (VH) is a sentence processing theory proposed by Frazier and Clifton (1997) which proposes that both written and auditory sentence processing is highly impacted by the *visibility* of information within a sentence relative to other information. This hypothesis is easy to illustrate through the use of ambiguous sentences, such as sentence 3 which is repeated below.

(3) The bus driver stopped the rider with a mean glare.

As discussed before, this sentence has two interpretations, the VP and the NP interpretation. The interpretation the comprehender has of this sentence is based on whether they believe the prepositional phrase (PP) “with the mean glare” is what “the bus driver” used, or if it is instead an attribute of “the rider”. The VH specifies that it is easier for comprehenders to make attachments that use fewer processing resources. One of the main determinants of processing resources in ambiguous sentences is how *visible* each potential attachment site is to the comprehender. That is, after encountering the PP “with a mean glare”, is it easier for the listener to attribute this to the bus driver or to the rider? In the case of speech, the focus is on the role that boundaries have in making attachment sites more or less visible during auditory sentence processing.

Prosodic boundaries often appear within sentences, breaking them into smaller subgroupings that are referred to as intonational phrases. An intonational phrase is bounded by a prosodic boundary on either side of it, as illustrated in 3(a) which is repeated below.

3(a) The bus driver stopped // the rider with a mean glare.

The prosodic boundary, indicated by //, results in an utterance with two intonational phrases, the second of which contains the PP that the comprehender must make an attachment decision about. The VH proposes that intonational phrases encapsulate information into perceptual packages that make information within the phrase more visible than information outside of it. When information is more visible, the parser has easier access to it when engaging in online comprehension. This means that in 3(a), when the comprehender encounters the PP and must make a decision about where it attaches to, ‘the rider’ is a more visible attachment site than the ‘bus driver stopped’. Because “the rider” is contained within the same intonational phrase as the PP, attaching them requires fewer resources than if the parser tried to cross over the boundary to attach the PP to “the bus driver stopped”. In contrast, listeners who hear a sentence such as 3(d), where both ‘the rider’ and ‘the bus driver’ exist within the same intonational phrase as the PP, should find both attachment sites equally plausible as an interpretation. Both attachment sites remain equally visible to the parser so there is no extra processing cost required to pick one over the other. As predicted by the VH, Shafer (1995) found that listeners increase their rate of VP interpretations in 3(d) as compared to 3(a). The predictions of the VH are not limited solely to the constructions discussed here and successfully predicts how readers and listeners process many sentence types.

The VH describes a relationship between prosody and syntax as the result of natural constraints on the language parser. Attentional or processing resources are required for the comprehender to interpret the meaning of a sentence, and it costs the parser more of these resources to work across intonational phrases. Thus, comprehenders are biased to make interpretations that are contained within a phrase as attachment sites remain more visible than those that are outside the phrase. If this hypothesis is correct, the boundary-syntax relationship is due to limitations imposed by the parser in language processing. As a result, this use of prosodic boundaries to interpret syntax should be inflexible.

Adaptation

The Ideal Adapter Framework (IAF) is a language processing model proposed by Kleinschmidt and Jaeger (2015). This error-driven model focuses on the role that predictions, and the errors that occur when predictions are wrong, play in language comprehension. For instance, when hearing a speaker say "I hate ?ugs" where the ? represents an uncertainty of whether a /p/ or /b/ was produced, the listener may believe that the speaker was stating their general disgust towards insects. However, if context quickly makes it clear that the speaker was referring to pugs, the listener will have experienced a *prediction error*. The parser notes that the ambiguous p/b was interpreted incorrectly, and tells the language system that in the future /p/ is more likely to occur. Now, the parser will update its understanding of how this specific speaker may produce /p/ in the future based on the recent information it just received from the speaker themselves. Upon encountering another p/b production from this speaker, the parser is now more likely to predict that it was a /p/ that was produced. Importantly, IAF claims that incorrect predictions result in rapid updating of cue distributions so that language comprehension with a speaker can improve within the course of a short conversation.

This process of the parser changing its predictions of the likelihood of certain productions is called *adaptation* and is one hypothesis as to how listeners successfully navigate the wide range of variability present in speech. Listeners who are able to adapt to a speaker's speech patterns can use this knowledge to interpret phonetic categories, facilitate speech comprehension, and make predictions about upcoming content that is more specific to the current speaker and context. The ability to adapt to speech in such a way could allow listeners to overcome the challenges inherent in processing speech from a variety of speakers.

The need for the language system to have some form of adaptation is clear. While language learning is often discussed in terms of childhood development, it is obvious that adults must have a way to handle changing language input. If adults had a static language processing system, they would struggle to comprehend a nonnative speaker with an accent they never encountered while a child and would be flummoxed at the use of new terms such as "google it" (technological misunderstandings aside). But we know intuitively that this is not the case. Adult speakers update how they use and comprehend language, interacting with new people and understanding new terms all the time. And in the last 20 years, a wave of experiments has found robust evidence that supports the hypothesis that at least some of this updating is as rapid as Kleinschmidt and Jaeger (2015) argue for in the IAF (Clarke & Garrett, 2004; Baese-Berk et al., 2013; Eisner & McQueen, 2005; Kraljic, Brennan, & Samuel, 2008; Kraljic & Samuel, 2005a; Sidaris et al., 2009).

Classic perceptual adaptation experiments expose listeners to speakers who produce one or more phonemes in unusual ways (Eisner & McQueen, 2005; Kraljic & McQueen, 2005a; Maye, Aslin, & Tanenhaus, 2008 for instance). For instance, a listener might hear a speaker who instead of producing an /s/ in words produces an ambiguous sound between /s/ and /ʃ/. When initially encountering such a speaker, a listener would not know if the speaker had just said 'sip' or 'ship'. However, after a short exposure to this speaking pattern, listeners will show that they have adapted to understand that s/ʃ is actually /s/ for this speaker . After exposure, listeners will successfully categorize the ambiguous sound as /s/, and show facilitated processing of words containing the ambiguous phoneme. This effect is well-demonstrated and known to occur to both consonants and vowels, and to more than one altered phoneme at a time (Dahan, Drucker, & Scarborough, 2009; Eisner & McQueen, 2005; Kraljic & Samuel, 2005a; Kraljic & Samuel, 2006; Kraljic & Samuel, 2007; Maye, Aslin, & Tanenhaus, 2008;). Similar perceptual adaptation has also been observed to accented, nonnative speakers (Clarke and Garrett, 2004; Floccia, et al., 2006). Initially, listeners are slower to process nonnative speech in comparison to native speech, but within a couple minutes this difference can be overcome simply through exposure to the speaker. Processing costs that were present at the beginning of the experiment can disappear altogether.

This evidence for perceptual adaptation to both native and nonnative speech suggests that the language parser can rapidly updates its expectations about incoming speech after encountering unexpected linguistic cues. This aids the parser when encountering variability both within and between speaker by allowing it to make more accurate predictions about upcoming speech specific to the context it is in. These more accurate predictions reduce the processing cost associated with language processing and allows communication to become easier and more efficient.

Beyond phoneme production, speech may also vary by lexical items, grammatical structures, and prosodic cues. If the parser adapts to phoneme production in language, it is reasonable to think it may adapt to these other aspects of language as well. A language system that could adapt to all these cues would explain how it is that listeners successfully deal with such a wide range of variability on an everyday basis. In recent years, there has been a move in research to investigate whether adaptation is present in other areas of speech (Harrington Stack, James, & Watson, 2018; Fine et al., 2013; Fraundorf & Jaeger, 2016; Kamide, 2012; Kurumada, Brown, & Tanenhaus, 2012). So far, the evidence has suggested that adaptation extends to many areas of language processing. In recent years, researchers have found evidence that adaptation occurs in response to speaker-specific attachment preferences (Kamide, 2012), written syntactic structures (Fine et al., 2013; Wells et al., 2009), dialectical variations in spoken syntactic structures (Fraundorf & Jaeger, 2016), verb structures (Ryskin et al., 2016), reliability of adjective use (Grodner & Sedivy, 2011), and quantifier processing (Yildirim et al., 2013).

The IAF has consolidated this research on adaptation with the goal of providing a comprehensive language processing model. According to IAF, language comprehension is achieved through three means. First, when a listener encounters a linguistic situation, they successfully understand the speech if it is recognized as a previously encountered situation. If instead the situation is new, listeners can either generalize information about their previous experience to the new situation if it is sufficiently similar, or if the situation is novel, they employ adaptation. This adaptation is proposed to be a Bayesian belief-updating model that balances a systematic understanding of prior linguistic experiences with updated expectations that alter expectations incrementally. These updated expectations are a result of prediction errors noted by the parser. When the parser notes that a mistaken prediction was made, an adjustment is made to its expectations to incorporate what the correct prediction would have been for future encounters. In the bug/pug example introduced before, when the parser realizes that the ambiguous sound was a /p/ instead of the /b/ that was assumed, this realization will allow the parser to weight the ambiguous p/b sound as being more likely to be a /p/. The larger a prediction error is, the more the language system will adjust its expectations of what will occur next. Thus, the model of the speaker's speech is updated online as the listener gathers more data from them. Kleinschmidt and Jaeger (2015) propose that adaptation extends to most, if not all, domains of language.

However, one area in which adaptation remains understudied is prosody. To date, adaptation has only been studied in contrastive accenting (Kurumada et al., 2014) and prosodic boundaries (Buxo-Lugo et al., 2020). This makes prosody an ideal place to continue the research of adaptation for two reasons. The first is to better answer the question of whether listeners adapt to prosody when in conversations. This will provide insight into how applicable IAF is to all domains of language. The second is that considering prosody through the lens of adaptation makes very different predictions from the VH about the influence of boundaries in interpreting syntax. If comprehension of prosodic boundaries is adaptable, it would suggest that the parser is not relying on perceptual units for processing as posited by the VH. If listeners are able to alter their interpretation of the boundary-syntax relationship based on boundaries occurring in new locations, it would indicate that the parser was not reliant on the visibility of information, but rather on experience with where boundaries occur. Finding evidence for adaptation in prosodic boundaries would be evidence against the VH. Successful adaptation would suggest that the role of prosodic boundaries in syntactic interpretation is not about how visible attachment sites are, but simply about how often a pattern of attachment has occurred in the past. Thus, listeners should be able to change their understanding of this pattern. If adaptation extends to prosodic boundaries, comprehenders will not be limited by processing constraints while interpreting sentences. Instead, their syntactic interpretation will be dependent on recent experiences with syntax and boundaries, and this will be malleable.

Experiments

The following experiments in this dissertation were designed to explore the validity of the Visibility Hypothesis and the Ideal Adapter Framework as explanations for how boundary-syntax relationships are represented and maintained in the parser. The goal was to gain insight into why the relationship between prosodic boundaries and syntax exists. If the presence of a boundary serves to make information within a phonological phrase more visible than information outside of it, the syntactic interpretation associated with a boundary's placement should be resistant to adaptation. Listeners who are exposed to new boundary-syntax relationships should be resistant to changing how they interpret boundaries and experience with new boundary-syntax distributions should not alter the basic constraints on processing resources that exist in the parser. In addition, if the VH is correct, it should be possible to induce similar effects with non-prosodic information. Providing listeners with new acoustic information that separates sentences into subgroupings should continue to make information that is perceptually packed together more accessible than information outside of this package.

On the other hand, if prosodic boundaries are open to adaptation, as many other aspects of language are, listeners should have flexible boundary-syntax representations. While the parser may have initial preferences for which syntactic interpretations are cued by which boundary locations, this information should be updated when presented with alternative evidence. If the IAF is correct, boundary-syntax representations should change as a result of encountering a new pattern of use.

Six experiments were completed to explore the boundary-syntax relationship as proposed by the VH and IAF. Experiments 1 - 3 tested the predictions of the VH through the use of new acoustic cues replacing prosodic boundaries to create new perceptual units. Experiments 4 -6 tested whether adaptation occurs to boundaries and to the new acoustic cues created to test the VH.

II. Testing the Visibility Hypothesis

Experiment 1

The goal of Experiments 1 – 3 was to investigate whether perceptual units, as predicted by the VH, would influence sentence analysis. Experiment 1 presented listeners with naturally produced sentences with prosodic boundaries that test the predictions of the VH. The goal of this study was to check if listeners interpreted sentences produced with natural boundaries in the way that the Visibility Hypothesis predicted. If so, then these stimuli would be suitable to manipulate to further test the predictions of the VH. For ease of presentation, these experiments are not presented in the order they were completed in. In actuality, Experiment 1 was a control condition of Experiment 3.

The goal of Experiment 1 was to ensure that the critical items that were used in Experiments 2 and 3 were interpreted by listeners as the VH predicts when produced with natural boundaries. By testing these stimuli, it would ensure that any lack of effect in Experiments 2 and 3 was not due to the base critical items.

Methods

Participants. Subjects were recruited from Amazon Mechanical Turk and paid $4.00 for completion of the experiment. Thirty subjects completed the experiment and were retained for analysis.

Stimuli. Twenty-eight critical items were adapted from Carlson, Clifton, and Frazier (2001). Critical items had PPs that were ambiguous as to where they should be attached. An example is given below.

4. Susie learned that Bill telephoned after John visited.

This sentence has two potential interpretations, referred to as low- and high-attachment. In the low-attachment interpretation, the PP is attached to the second VP of “Bill telephoned”. In this case, the interpretation is that *Bill* did something after John visited. Alternatively, in the high-attachment interpretation, the PP is attached farther away to the first VP of “Susie learned”. Here, the sentence is interpreted so that *Susie* learned something after John visited. Prior studies (Carlson, Clifton,& Frazier, 2001) have shown that native English speakers are biased to interpret these sentences as low-attachment with an early boundary as in 4(a), but that this bias is reduced when they hear the sentence with a late boundary as in 4(b).

4(a) Susie learned // that Bill telephoned after John visited.

4(b) Susie learned that Bill telephoned // after John visited.

Frazier and Clifton (1997) argued that this bias pattern emerges due to the constraints proposed by VH. The bias to make a low-attachment interpretation when a boundary occurs early, as in 4(a), happens because a perceptual unit containing both the second VP and the PP is created. This results in the second VP being more accessible than the first VP to the parser when it must make the decision of where to attach the PP. However, in the case of 4(b) where a late boundary occurs, the low-attachment bias is reduced because now both VPs are in the same, separate phonological phrase from the PP. Since both potential attachment sites are in the prior phonological phrase they are both equally valid as attachment sites. The visibility of the first VP is not reduced as much as it was in the case of 4(a). As a result, listeners should reduce the number of low-attachment interpretations and increase the number of high-attachment interpretations they make.

The goal of this experiment was to test that the sentences used followed the predictions made by the VH. To do this, each of the 28 critical items was recorded twice, once with an early boundary and once with a late boundary. In addition, 42 filler sentences that did not contain a PP ambiguity were recorded. All the critical and filler items were recorded by a female native English speaker with a Midwestern U.S. accent.

Forced-choice comprehension questions which gave subjects two answer choices were created for each item. Comprehension questions for critical items always probed whether subjects had interpreted the sentence as low- or high-attachment. For example, after hearing the critical sentence "Susie learned that Bill telephoned after John visited", subjects would be asked the following:

What happened after John visited?

1. Susie learned something after John visited.
2. Bill telephoned after John visited.

In the above example, the first answer represents the high-attachment interpretation and the second choice represents the low-attachment interpretation. Comprehension questions for filler items always had a correct and incorrect option to choose from.

Procedure. The critical and filler items were combined to make four lists of 28 critical items and 42 filler items. Each list was counterbalanced for boundary location and answer presentation. In each list, subjects heard half of the critical items with an early boundary and half with a late boundary. The presentation of answer choices to each comprehension question was counterbalanced so that half of the low-attachment interpretation answers occurred first, and half of high-attachment interpretation answers occurred first. Likewise, filler item answers were counterbalanced so that half of the correct answers appeared first and half of the incorrect answers appeared first. Within each list, the presentation of filler and critical items was randomized for each subject. In total, each subject heard 70 sentences and answered a comprehension question after each one. Subjects did not receive feedback about whether the question was answered correctly or not.

Results

Subject responses for critical items were coded based on whether their answer indicated low- or high-attachment interpretations and whether the boundary occurred in an early or late position in the sentence. The results indicated that sentences with an early boundary were interpreted as low-attachment 78.57% of the time. Sentences with a late boundary were interpreted as low-attachment 59.05% of the time. These results are displayed in Figure 1. To analyze these results, a multi-level model was constructed which analyzed the effect of boundary location on syntactic interpretation. There was a main effect of boundary location ($b=-.07$, $t=-3.33$, $p<.05$) indicating that subjects were influenced by where a boundary occurred when making syntactic interpretation decisions. Subjects made significantly more low-attachment interpretations when boundaries were in an early position.

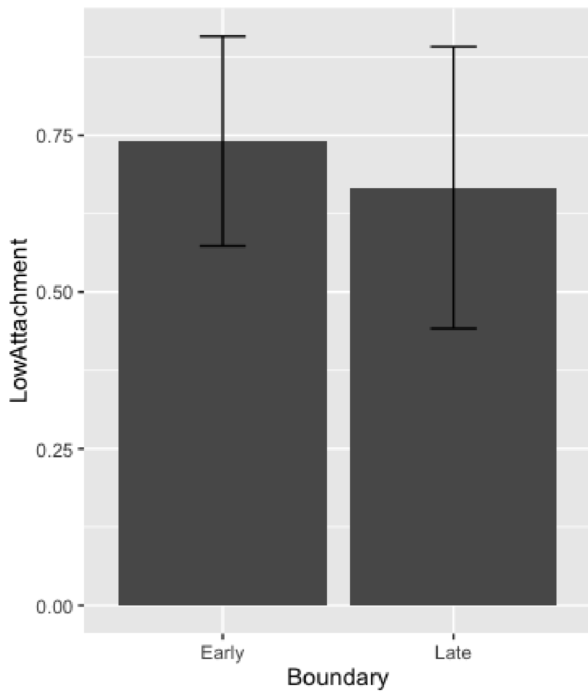

Figure 1. Proportion of low-attachment interpretations by boundary location in Experiment 1.

Discussion

The data showed that the critical items chosen for this experiment were consistent with the predictions of the VH when produced with natural boundaries. Subjects were more likely to make low-attachment interpretations than high-attachment interpretations when there was an early boundary. Given that these stimuli produced the effects predicted by the VH, they were deemed suitable to be used in Experiments 2 and 3 to investigate how listeners interpret sentences with perceptual units created in novel ways.

If the Visibility Hypothesis is correct and sentence analysis is influenced by perceptual units, then altering the sentences from Experiment 1 so that the perceptual units are created in new ways should still result in the same pattern of sentence analysis. That is, when listeners encounter a perceptual break early in the sentence, they should be influenced to make a low-attachment interpretation. When a perceptual break occurs later in the sentence, listeners should increase their high-attachment interpretations, mirroring the pattern seen in Experiment 1. To test this, Experiments 2 and 3 create perceptual units that are marked by buzzers and inserted silence, rather than by naturally produced prosodic boundaries.

Experiment 2

The goal of Experiment 2 was to investigate whether sentence analysis would depend on perceptual units, as predicted by the VH, when the perceptual phrases were created through non-prosodic cues. Sentences were edited so that prosodic boundaries were removed and then a non-linguistic cue, a basketball buzzer, replaced them. This resulted in sentences with a clear auditory cue appearing in the place of the prosodic boundaries. The goal of this was to create perceptual units that were bounded by non-linguistic information. If the boundary-syntax relationship is due to the how information is distributed across perceptual units, the interpretation of ambiguous sentences should be identical no matter what cue is used to create perceptual units. Subjects should use a non-linguistic cue presented in the same manner as prosodic boundaries to interpret syntactic structure. However, if listeners fail to use the non-linguistic cue to make interpretation decisions, it suggests that the boundary-syntax relationship may not be due to the visibility constraints on the parser. In such a case, other models of sentence processing should be considered to explain the boundary-syntax relationship.

Methods

Participants. Subjects were recruited from Amazon Mechanical Turk and paid $4.00 for completion of the experiment. Forty subjects completed the experiment, but one was excluded for indicating that English was not their native language, resulting in 39 subjects being retained for analysis.

Stimuli. The twenty-eight critical items adapted from Carlson, Clifton, and Frazier (2001) and used in Experiment 1 were used again. The goal of this experiment was to create the same perceptual units produced by the prosodic boundaries in Experiment 1 but with an entirely new cue. Taking the naturally produced sentences with early and late boundaries from Experiment 1, they were edited in order to remove the boundary and replace it with a non-linguistic sound. To remove the natural boundaries, the sentences were edited so that the first half of sentence 4(b) was spliced to the second half of sentence 4(a) so as to create a natural sounding sentence with the major prosodic boundaries removed. Then a basketball buzzer noise was inserted where the boundary had originally occurred to create perceptual units of speech separated by a non-prosodic cue. This resulted in sentences that were interrupted partway through with the sound of the buzzer. Two sentences were created for each critical item so that each item had one sentence with an early buzzer and one with a late buzzer. In addition, the 42 filler sentences without the PP ambiguity were edited so that the buzzer occurred at random points. The same forced-choice comprehension questions from Experiment 1 were used for both critical and filler items.

Procedure. The sentences were combined to make four lists as in Experiment 1. In each list, subjects heard half of the critical items with an early buzzer and half with a late buzzer. The answers to each comprehension question was counterbalanced so that half of the low-attachment interpretation answers occurred first, and half of high-attachment interpretation answers occurred first. Filler item answers were counterbalanced so half of the correct answers appeared first and half of the incorrect answers appeared first. Within each list, the presentation of filler and critical items was randomized for each subject. In total, each subject listened to the 70 sentences and answered a comprehension question after each one. Subjects did not receive feedback about whether the question was answered correctly or not.

Results

Subject responses were coded based on whether their answer indicated low- or high-attachment interpretations and whether the buzzer occurred in an early or late position. The results indicated that sentences with an early buzzer were interpreted as low-attachment 68.13% of the time. Sentences with a late buzzer were interpreted as low-attachment 69.96% of the time. To analyze these results, a multi-level model was constructed which analyzed the effect of buzzer location on syntactic interpretation. There was no effect of buzzer location ($b=-.02$, $t=-.731$, $p>.05$), indicating that subjects did not alter their rates of low-attachment interpretations based on where a buzzer occurred within a sentence.

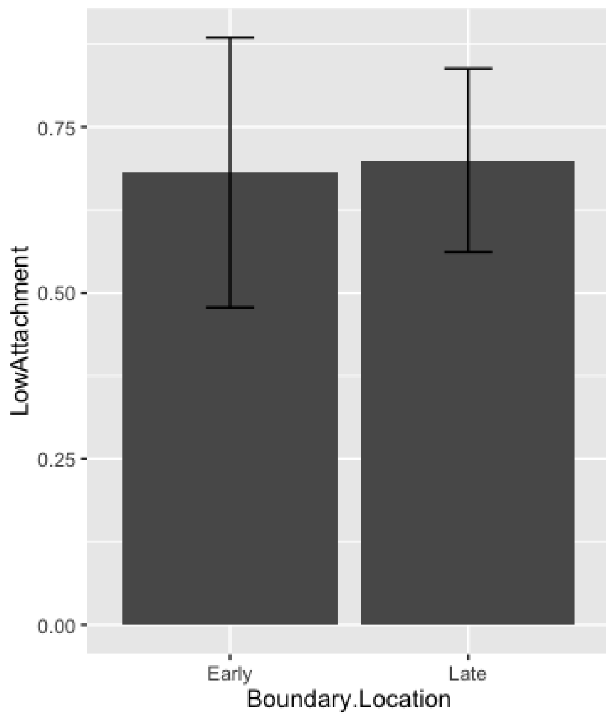

Figure 2. Proportion of low-attachment interpretations by buzzer location in Experiment 2.

Discussion

The data showed that subjects did not use the perceptual units created by the buzzer to make interpretation decisions about the sentence. Subjects who heard the late buzzer were just as likely to make low-attachment interpretations as subjects who heard the early buzzer. Subjects showed a bias for low-attachment interpretations of critical sentences regardless of boundary placement. This suggests that the interpretation of ambiguous sentences is not influenced by processing constraints that focus syntactic analysis within perceptual units.

The results of this experiment provided preliminary evidence that the VH is incorrect. If sentence processing is influenced by constraints on the processing system, specifically by which units of information occur within perceptual units together, then comprehenders should continue to be influenced by these constraints regardless of how these processing units are created. Because subjects were not sensitive to processing units created by the use of the non-linguistic buzzer, it suggests that the parser is not making decisions based on the information contained within a perceptual unit.

However, there are two potential explanations that must be considered for why the buzzers may have failed to provide perceptual units for the subjects. First, it may be the case that the buzzer was a highly distracting cue for subjects to listen to. Perhaps the parser generally *does* focus its processing within perceptual units regardless of how these units are created, but particularly distracting acoustic information diminishes this tendency. If sentence interpretation is influenced by constraints on processing resources, it may be the case that these processing resources get distracted by particularly odd acoustic cues. It would not matter what attachment sites are more or less visible to the parser if the parser does not have the resources to analyze them properly. If this is the case, it would explain why subjects showed a low-attachment bias across all sentences. Native English speakers have a low-attachment bias overall (Carreiras and Clifton, 1999) and listeners may have just reverted to this bias when overwhelmed with a distracting cue.

It may also be the case that the buzzers did not create the perceptual units they were intended to make. The goal of the buzzer was to place a strong marker that ‘chunked’ sentences apart in the same places that prosodic boundaries would occur. In Experiment 2, this was a basketball buzzer that occurred *over* the spoken audio where the boundary would have occurred. It may be that subjects, while hearing the buzzer, were able to filter the noise out as being unimportant to the task at hand. After all, it is common in conversation that a listener must focus on their partner’s speech over the sound of a typing keyboard, music in the background, the barking of a dog, or any other number of unrelated sounds. In these cases, it would not make sense for a listener to perceive perceptual units at each of these interruptions that are clearly not a part of the conversation they are having.

Research on auditory stream analysis, which looks at how listeners organize and perceive multiple sounds, provides some insight that may explain the results of Experiment 2. The auditory system is thought to use both bottom-up cues and top-down knowledge to integrate and stream multiple sounds (Bregman, 2008). Listeners in Experiment 2 may have been able to use both as a way of focusing on the incoming speech as one unit, rather than as two units separated by a buzzer. As a bottom-up strategy, the auditory system has a bias to group similar fundamental frequencies together as a single sound, and treat distinctly different fundamentals as a different sound (Bregman, 2008). In addition, sounds that begin at a later time are less likely to be considered part of an already ongoing sound. In the case of Experiment 2, both of these expectations were violated by having a distinctly different audio cue occur partway through the speech. As a top-down mechanism, listeners have experience that speech is a part of communication, but very little (if any) experience with buzzers occurring in a communicative context. This experience with speech would allow listeners to focus in on the streaming of speech and treat the buzzer as a distraction (Dowling, Lung, & Herbold, 1987). If the auditory system is able to easily categorize a distinctively different sound, such as the basketball buzzer, as separate from the speech stream, the parser may simply ignore the presence of the intrusion in the speech stream. The null results of Experiment 2 may be due to both bottom-up and top-down analyses of the speech and buzzer audio being employed by listeners. When listeners encountered a strongly non-linguistic tone in the middle of a sentence, it was likely quite easy to parse out of the speech stream they were focused on.

Experiment 3 was then designed to test the VH while removing these potential confounding explanations. The buzzer that played over the sentences was replaced instead by an inserted silence. This inserted silence was linguistic-like, in that pauses do occur in conversation, and also was a true break in the speech. If stimuli in Experiment 2 were regarded as the buzzer being a sound occurring *over* the sentence, the pause in Experiment 3 stimuli now presented a clear break *within* the sentence that created two clearly separate units.

Experiment 3

The goal of Experiment 3 was to provide a more natural, though still non-linguistic, cue to break sentences into perceptual units. While Experiment 2 created perceptual units within the critical sentences, the specific stimuli used may have prevented the parser from paying attention to these perceptual units. Experiment 3 was thus designed to be similar to Experiment 2, but replaced the buzzer with a less distracting cue that more certainly indicated a break in perceptual units. The buzzer was replaced with a silence that was inserted where the boundary would have occurred. The use of added silence as a cue had the benefit of overlapping with a property of prosodic boundaries (that is, a salient pauses) while still being a non-linguistic cue like the buzzer.

If the VH is correct, listeners should use the perceptual units created by the silence to parse sentences. Along with the results from Experiment 1 and 2, this would suggest that sentence processing relies on the visibility of attachment sites, but that this visibility may be ignored when the parser is faced with confusing or difficult acoustic loads. However, if listeners still fail to use these perceptual units for sentence processing, it would be hard to argue that the VH is a correct regarding the boundary-syntax relationship. Since the Visibility Hypothesis does not provide an explanation for why only perceptual units created by prosodic boundaries would be considered by the parser, it would suggest that a different language model is needed to explain the boundary-syntax relationship.

Methods

Subjects. 40 subjects were recruited from Amazon Mechanical Turk and paid $4.00 for completing the experiment. 1 subject was excluded from analysis for indicating that they were not native speakers of English. This resulted in 39 subjects being retained for analysis.

Stimuli. The stimuli for this experiment were similar to those from Experiments 1 and 2. The same 28 critical items were used, but edited so that instead of boundaries or buzzers, silences created the perceptual units. As in Experiment 2, the naturally recorded items with an early and late boundary were spliced together to create natural sounding sentences with major prosodic boundary cues removed. Then each sentence had a silence of 1.015 seconds inserted into the early or late boundary location.

The use of inserted silence to create perceptual units gave subjects a cue that was similar to prosodic boundaries which are sometimes indicated by a pause in the speech stream. However, the other markers of a prosodic boundary, such as pitch and durational changes, were absent at these breaks. In addition, the use of a computer-generated silence sounded markedly different from a human-produced pause. The resulting effect was that the silence created a cue that had some similarities to prosodic boundaries but was still clearly non-linguistic, like the buzzer. However, the silence was not a jarring cue within the middle of the speech stimuli, making it less distracting than the buzzer. The 42 filler items had silence of the same length inserted in between two words. The same comprehension questions for critical and filler items from Experiments 1 and 2 were used in Experiment 3.

Procedure. The critical and filler sentences were combined to make four lists similar to Experiment 2. In each list, subjects heard half of the critical items with an early silence and half with a late silence. The answers to each comprehension question were counterbalanced so that half of the low-attachment interpretation answers occurred first, and half of high-attachment interpretation answers occurred first. Filler item answers were be counterbalanced so that half of the correct answers appear first and half of the incorrect answers appeared first. Within each list, the presentation of filler and critical items were randomized for each subject. Subjects listened to each of the 70 sentences and answered a forced choice comprehension question.

Results

Subject answers were coded based on whether they made low- or high-attachment interpretations to critical sentences. The data was analyzed to compare the proportion of high-attachment interpretations made when the silence occurred in the early position as compared to when the silence occurred in the late position. Sentences with an early pause were interpreted as low-attachment 70.13% of the time while sentences with late pauses were interpreted as low-attachment 73.81% of the time. A multi-level model was constructed to analyze the effect of pause location on sentence interpretation.

The results failed to find an effect of pause location on sentence interpretation ($b=.04$, $t=1.327$, $p>.05$). This indicated that subjects' sentence interpretation was not biased by the location of a pause in a sentence. Regardless of where the pause occurred in a sentence, subjects were just as likely to make a low-attachment interpretation.

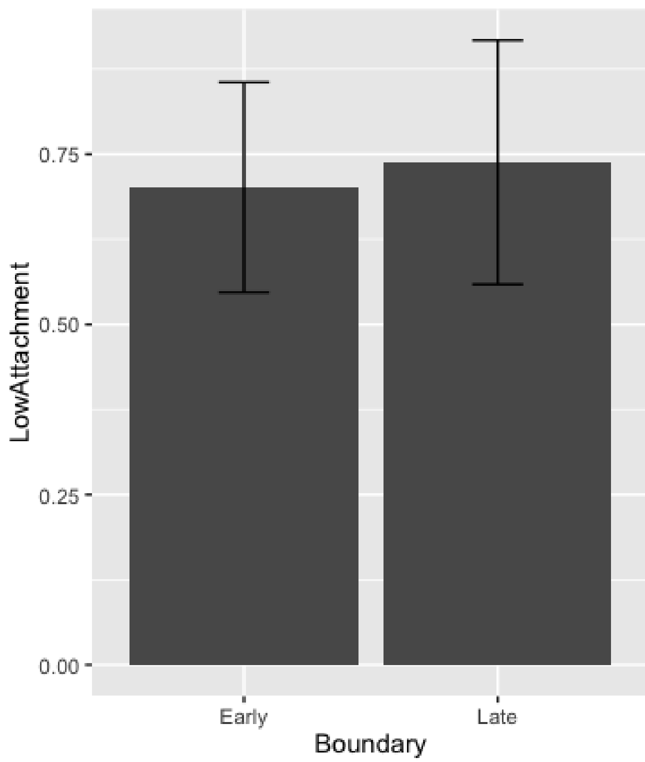

Figure 3. Proportion of low-attachment interpretations by pause location in Experiment 3.

Discussion

Subjects did not use the perceptual units created by computer-generated silences to analyze sentences. Regardless of whether the silence appeared in the early or late position, subjects were equally likely to say that the sentence had a low-attachment interpretation. This indifference to the location of the perceptual units suggests that the Visibility Hypothesis does not explain the boundary-syntax relationship appropriately. If processing is about processing resources and the visibility of linguistic information within perceptual units, then sentence analysis should still be influenced by perceptual units created in novel ways. The failure for this to occur suggests that processing resources and visibility alone cannot explain the boundary-syntax relationship.

Experiment 3 addressed the issues present in Experiment 2's stimuli and still failed to find evidence that listeners were processing sentences based on perceptual units. This suggests that the constraints proposed by the VH on sentence processing do not provide a complete picture of how boundaries and syntax interact. It may be the case that the foundation of what the VH is correct, but some details are inaccurate. I explore some possible explanations below. These are cautious speculations as the experiments presented here were not designed to answer these questions.

The main claim of the VH is that visibility of units is what drives interpretation. The effect of visibility is thought to exist both in written and spoken language such that many variables can influence visibility. For instance, written language interpretation can be influenced by self-paced reading which gives multiple words at a time. Frazier and Clifton (1997) report that altering which groups of words appear together induces visibility effects. Frazier and Clifton (1997) suggest that perhaps working memory or attentional processes may explain these visibility effects. The failure of alternatively-created units to induce these effects in Experiments 2 and 3 suggests that *if* visibility is the driving factor in syntactic interpretation, then boundaries are a necessary component of visibility in spoken language. This is to say that *if* visibility is a real effect, it may require boundaries in spoken language and the VH would have to state this specifically. This would suggest that these effects are specifically linguistically-driven in nature, rather than due to general cognitive constraint.

What it is that is special about boundaries such that induces these effects is unclear. It may be that perceptual units play a key role in interpretation, but that these units require all the cues associated with boundaries (pitch accents, word-lengthening, etc.) to affect visibility. Prosodic boundaries as a whole may send a signal to the parser that the speaker is structuring or marking their language production in a specific way. If this is the case, the null results of Experiment 3 might be explained by the inserted silence missing the other cues associated with boundaries. If visibility is impacted by all the cues associated with boundaries, having only the physical pause in Experiment 3 may not have been enough of a marker for listeners.

It may be that the production of a boundary implies speaker intent. Boundaries, and all their associated cues, are chosen by a speaker and listeners may be very well aware of this. Rather than these boundary effects being induced by the limits of memory or attention, it may be that listeners assume a speaker produced a boundary to structure language in some way. In contrast, when a speaker's speech is edited to be structured in the same way, as occurred in Experiment 3, listeners may assume that the perceptual unit information is less useful as it is not what the speaker intended to produce and thus is not pertinent structural information. The VH is concerned with how listeners engage in sentence analysis, but understandably does not delve into why speakers decide to produce boundaries where they produce them. It may be important that when listeners hear language they are aware that someone is producing it with the intent of communicating. The presence of a boundary may indicate to a listener that there is a reason the speaker produced a boundary and that is what the listener attends to.

The results of Experiments 2 and 3 indicate that the boundary-syntax relationship is not accurately explained by the visibility of attachment sites within an utterance. If the VH is to be accurate, it would need to explain what is unique about the presence of prosodic boundaries that divides phrases into perceptual units that causes processing to be altered. Frazier and Clifton (1997) indicate that attention or working memory may explain the effects of the Visibility Hypothesis. If this is the case, then more research needs to be done to understand how prosodic boundaries act in this manner.

There is of course, also the possibility that the Visibility Hypothesis is not correct at all. At this point an alternative model, such as adaptation, should then be considered to explain the link between boundaries and syntax. Given that Experiments 2 and 3 failed to find evidence for the Visibility Hypothesis, I turned to exploring Adaptation. Experiments 4, 5, and 6 investigate the predictions of adaptation as a potential way of explaining the syntax-boundary relationship.

III. Testing Adaptation

Experiment 4

The results of Experiments 1 - 3 provided preliminary evidence that the Visibility Hypothesis does not adequately explain the boundary-syntax relationship. Experiment 4 was designed to test whether the Ideal Adapter Framework could explain this relationship instead. In this experiment, sentences were produced that had prosodic boundaries either in appropriate and inappropriate locations given their syntactic structure. Subjects in the experimental condition were exposed to sentences with inappropriate boundaries, while the control group listened to filler sentences. At the end of the experiment, all subjects listened to sentences with inappropriate boundaries and were asked to rate how natural the speech sounded. If sensitivity to the boundary-syntax relationship is tied to the distribution of cues, then subjects with exposure to inappropriate boundaries should rate such sentences as more acceptable due to their exposure to these sentences. These results would provide evidence of adaptation in prosody. If subjects do not improve their rating of inappropriate boundaries after exposure, it would suggest that the linguistic system does not update its boundary-syntax understanding after exposure to a new distribution.

Methods

Participants. Subjects (N=101) were recruited from Amazon Mechanical Turk and were paid $1.59 for completing the experiment. All subjects were self-reported native speakers of English who were currently living in the United States.

Stimuli. Twenty-four critical items were adapted from Gibson, Pearlmutter, and Torrens (1999). Critical items ended with a relative clauses (RC) that had two potential attachment sites earlier in the sentence. Critically, the sentence was manipulated so that the attachment site was disambiguated syntactically, resulting in only one grammatically acceptable interpretation. Four example sentences created from one critical item are shown below, with the attachment site and the critical verb bolded.

5(a) The man shot the servant of the actresses who were on the balcony.

5(b) The man shot the servant of the actresses who was on the balcony.

5(c) The man shot the servants of the actress who were on the balcony.

5(d) The man shot the servants of the actress who was on the balcony.

In the critical sentences, the relative clause (RC), such as ‘who was on the balcony, had only one attachment site which was determined by the plurality of the verb was/were in the phrase. Each sentence therefore could therefore only be interpreted as low-attachment, when the RC attached to the later NP, or high-attachment, when the RC attached to the earlier NP. For example, 5(a) and 5(d) have low-attachment interpretations, where the RC attaches to the closer NP. 5(b) and 5(c) have a high-attachment interpretation, where the RC attaches to the NP farther up in the sentence. Each critical item had four sentence variations, as shown by 5(a) –5(d) so that each item had all possible attachment x verb sentences.

Each of those four sentences were recorded twice, once with an early boundary and once with a late prosodic boundary, as shown below.

5(e) The man shot the servants // of the actress who was on the balcony.

5(f) The man shot the servants of the actress // who was on the balcony.

Early boundaries occurred directly after the first NP, and later boundaries occurred directly after the second NP. This resulted in having two sentences that had identical wording, but varied in whether the boundary appeared in an appropriate or inappropriate place. The decisions as to what was considered appropriate placement was based on prior research (i.e. Kraljic & Brennan, 2005b; Snedeker & Trueswell, 2003). In total, each critical item had eight sentence variations that were produced. In addition, 40 fillers that did not utilize the relative clause construction were created. These were recorded with the prosodic boundary produced in a natural place as determined by the experimenter.

All sentences were recorded by a female speaker with a Midwestern U.S. accent. The prosodic cues were produced naturally, but with the goal of being easily identified by a naïve listener. The sentences did not undergo any editing or alteration.

Procedure. Subjects were randomly assigned to the control (n=50) and the experimental (n=51) condition. For both groups the study was divided into two blocks, though to subjects there was no indication of where Block 1 ended and Block 2 began. Block 1 had 35 sentences, and Block 2 had 9 sentences.

In the experimental condition, subjects in Block 1 heard 20 of the critical items. Sixteen of these items had a boundary in the unexpected location, and four of them had a boundary in the expected location. In the control group, subjects listened to 35 fillers in Block 1. In Block 2, both groups heard four critical items with boundaries in unexpected locations and five fillers.

After each sentence, subjects were asked to rate each sentence on how natural it sounded. Subjects were not asked to focus specifically on where prosodic boundaries occurred, but rather on their subjective sense for how the sentence sounded. Subjects were provided with a 7-point scale where 1 indicated least acceptable and 7 indicated most acceptable.

Results

To confirm that participants were sensitive to the appropriate vs inappropriate boundary location manipulation, a linear mixed effect analysis compared ratings of appropriate and inappropriate critical items in Block 1 for the experimental group. There was a main effect of appropriateness on sentence ratings. Subjects rated appropriate boundary items as significantly more natural than inappropriate items (b=-.56, t=-6.03, p<.05), indicating their preference for appropriate boundary placement. This is shown in Figure 4.

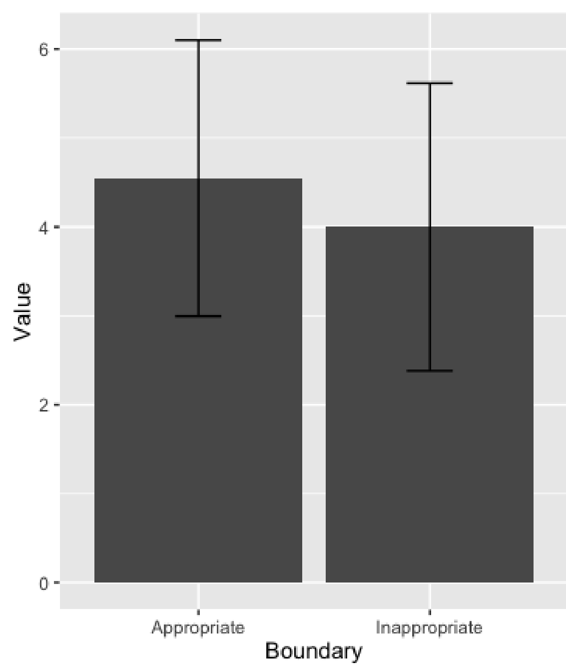

Figure 4. Naturalness ratings of appropriate and inappropriate boundary placements within Block 1.

To test for adaptation, a linear mixed effect analysis compared the ratings for inappropriate items for the experimental group from Block 1 to Block 2. There was no effect of Block. Inappropriate sentences in Block 2 were not rated as significantly different from inappropriate sentences in Block 1 ($b=.04$, $t=.122$, $p>.05$), indicating that adaptation did not occur. A comparison of the two sentence types is shown in Figure 5. In addition, a comparison of Block 2 ratings for inappropriate sentences for the control and experimental group revealed that there was no significant difference in ratings ($t=-.032$, $p>.05$). This suggests that the Experimental group's increased experience with inappropriate boundary placement did not result in an improved rating of these sentences compared to the Control group who had no experience with them.

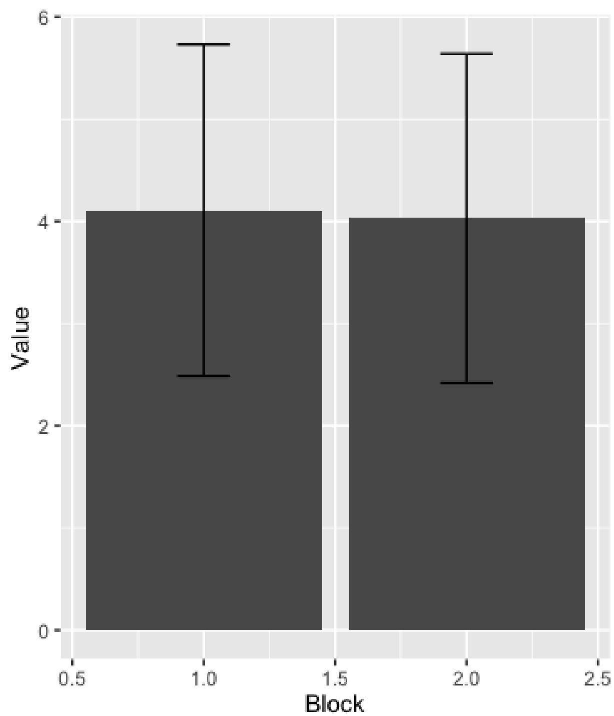

Figure 5. Sentence rating of inappropriate boundary placement across Blocks 1 and 2 for Experimental group.

Discussion

There was no evidence of adaptation to prosodic boundaries in this experiment. The experimental group, which was exposed to prosodic boundaries in unexpected locations, rated these sentences just as unnatural at the beginning of the experiment as at the end of the experiment. Furthermore, their ratings for these sentences were not significantly different from the control group's ratings despite the experimental group having more experience with this sentence type than the control group. These results fail to provide support for the Ideal Adapter Framework. The acceptability of prosodic boundaries did not change as a result of increased exposure to inappropriate boundaries. These results cannot be explained by subjects not being sensitive to the appropriateness of the boundary since they rated inappropriate boundary sentences as significantly worse than appropriate boundary sentences.

However, recent work by Buxo-Lugo et al. (2020) that found adaptation to prosodic boundaries suggests that several experimental design choices may explain why subjects failed to adapt to prosodic boundaries in Experiment 4. Buxo-Lugo et al. found that listeners *do* adapt to prosodic boundaries when they are given feedback on how to interpret these boundaries. In Buxo-Lugo et al., subjects listened to ambiguous sentences that had either an early or late prosodic boundary and were asked comprehension questions that probed their interpretation of the sentence. After answering the comprehension question, subjects were told whether their answer was correct or not. This provided them with feedback as to whether they had interpreted the sentence correctly, which could be tied to what the boundary placement indicated about the syntactic structure of the sentence. Over the course of the experiment, subjects' answer performance improved, indicating that they were successfully adapting their understanding of prosodic boundaries.

There are several key differences between Buxo-Lugo et al. (2020) and Experiment 4 presented here. First, in Experiment 4 subjects were asked to provide a naturalness rating of speech, rather than answer a comprehension question. This may have impacted adaptation in several ways. Having their task be to only to consider how natural language sounded, rather than interpret the sentence, may have resulted in shallower processing of the sentences and the boundaries within them. However, the difference in naturalness ratings between appropriate and inappropriate sentences suggests that subjects were paying attention to the odd placement of boundaries. Boundary location was counterbalanced so that, across items, an equal number of early-inappropriate boundaries and late-inappropriate boundaries occurred. For both early- and late-inappropriate boundaries to be rated as less natural indicates that subjects were noting the relationship between the syntax and the boundary. This suggests that subjects were engaging with the task and the sentences. The engagement subjects had with the task without evidence of adaptation may also be explained if adaptation shows up as improved comprehension, but preference for prosodic patterns is slower to change. That is, subjects in Experiment 4 may have begun to track the way in which the speaker produced boundaries, but still maintained a preference for hearing boundaries in their appropriate place. This would explain why null results were found for Experiment 4 even if adaptation was occurring. This does not necessarily run counter to claims of adaptation, as a successful parser would be one that knows when to engage in adaptation and when to maintain broader distributions (Kleinschmidt, 2018).

Secondly, subjects in Experiment 4 were not given explicit feedback about how they should interpret sentences. While the grammatical constraints of the sentences in Experiment 4 should have told subjects how to interpret sentences, this may not be as powerful as explicit feedback telling them that they are correct or incorrect, as in Buxo-Lugo et al. Explicit feedback may better prompt the parser to consider what features of production might be helpful in coming to the correct decision in future sentences. Strongly related to this is that the stimuli in this experiment were not globally ambiguous, but rather temporarily ambiguous. This meant that any confusion the parser had about the meaning of the sentence should have been resolved by the end of the sentence. In comparison, globally ambiguous sentences would require the parser to decide between two plausible interpretations itself. Without disambiguating syntactic information, the parser is likely forced to weight other cues more heavily than usual. In this case, the location of a prosodic boundary would be a key cue to sentence interpretation. Both the presence of feedback and persisting ambiguity in Buxo-Lugo et al. might have pushed subjects to consider the communicative intent of the prosodic boundaries more strongly.

Lastly, the items in Buxo-Lugo et al. were globally ambiguous, in contrast to the stimuli in Experiment 4 which were disambiguated syntactically by the end of the sentence. When utterances are ambiguous the listener is forced to make a decision about what the speaker intended. This was part of the task as they were asked a comprehension question afterwards that probed which interpretation of the sentence they had heard. Having to make a decision about the intended message may result in stronger adaptation as error signals are theorized to be the mechanism which drives adaptation (Kleinschmidt & Jaeger, 2015). In this case, if a listener strongly believes the speaker intended a low-attachment interpretation, and then receives feedback that a high-attachment interpretation was intended instead, they would have had a large error-signal which indicates to them that adaptation is required. In Experiment 4, listeners did not have to consider two possible interpretations. By disambiguating the sentences syntactically, subjects likely were not in an uncertain state where they were making errors regarding sentence interpretation.

If prosodic adaptation does occur in situations similar to Buxo-Lugo et al.'s design, then Experiment 4 offers up potential limitations and constraints to adaptation. While these results cannot provide conclusive answers as to the role of adaptation in prosody, they do point in a clear direction for future research on adaptation. Future research on prosodic adaptation should focus on the role that the type of stimuli, feedback, and test questions have on how successful adaptation is.

To further explore potential limitations of adaptation, Experiment 5 was designed to test whether listeners could adapt to the use of non-linguistic information presented with clear feedback on how it mapped to syntactic interpretation. The goal was to explore whether adaptation was powerful to any cues in the linguistic environment, or is restricted to likely linguistic cues.

Experiment 5

The results of Experiment 4 indicated that listeners do not adapt their interpretation of prosodic boundaries based on mere exposure to altered prosodic patterns. However, later work by Buxo-Lugo et al. (2020) found that listeners who were given feedback about how they should have interpreted ambiguous sentences adapted to prosodic boundaries, lending support to the hypothesis that boundary-syntax relationships may be modified through adaptation. Experiment 5 was designed to test if this adaptation is unique to linguistic cues or a general adaptation that listeners can apply to any information present in a linguistic context. Using the non-linguistic buzzer cue from Experiment 2, which listeners did not use to interpret sentences, subjects in Experiment 5 were provided feedback on how to interpret the cue within critical sentences. If listeners could learn how to incorporate the buzzer into sentence processing as a result of receiving feedback, it would suggest that adaptation is a powerful mechanism which is not constrained to just linguistic cues. If listeners do not adapt, it would suggest that linguistic adaptation is limited to linguistic cues. Adaptation theories such as the IAF should consider such constraints when modeling language processing.

Methods

Subjects. 120 subjects were recruited from Amazon Mechanical Turk and paid $4.00 for their participation. Seven participants were excluded for indicating that English was not their native language, resulting in 113 subjects being used for data analysis.

Stimuli. The stimuli used in this experiment were largely identical to those in Experiment 1. Twenty-eight critical items with PP ambiguities were created. Each item was recorded twice, once with an early boundary and once with a late boundary. As in Experiment 2, the two versions of the sentence were spliced together to create a sentence with no major prosodic boundary. This sentence was retained for use in a control condition. Then, the same buzzer from Experiment 2 was added to items, once in the early position and once in the late position. This resulted in each critical item having three sentences – one control sentence without a buzzer or prosodic boundary, and two experimental sentences with a buzzer occurring in the late and early position. The same 42 fillers from Experiment 1 were used in the experimental conditions, and fillers without buzzers were used in the control condition. The same speaker from Experiment 1 recorded these stimuli.

Comprehension questions were similar to Experiment 1. Comprehension questions for the critical items probed whether subjects made the low- or high-attachment interpretation. Comprehension questions for the filler items had one correct and one incorrect answer to choose from. All questions had two answers for subjects to choose from.

Procedure. Subjects were randomly assigned to one of two experimental conditions or the control group. Subjects in the experimental groups were in either the Congruent or Incongruent condition. In the Congruent condition, subjects were given feedback that indicated that buzzers followed the same pattern that one would expect from prosodic boundaries. That is, an early boundary was associated with low-attachment interpretations, and a late boundary was associated with high-attachment interpretations. In the Incongruent condition, subjects encountered the opposite pattern. Early boundaries were associated with high-attachment interpretations, and late boundaries were associated with low-attachment interpretations. For each comprehension question, subjects completed a forced-choice question that probed whether they interpreted the sentence as high or low attachment. After answering the question, subjects received feedback telling them either that they were correct or incorrect.

Subjects in the control condition listened to the same sentences, but without the buzzer edited in. This resulted in subjects listening to critical sentences without a prominent boundary. Feedback was given randomly so that half the time subjects were told the sentences should have been interpreted as low-attachment, and the other half of the time that it should have been interpreted as high-attachment.

Within each list, subjects received feedback that sentences were low-attachment half the time and high-attachment half the time. All answers were counterbalanced so that low-attachment interpretations were displayed first half the time, and high-attachment interpretations were displayed first half the time. Filler items had the correct answers displayed first half the time. This resulted in twelve counterbalanced lists. Each experimental condition had four lists which counterbalanced early vs late buzzers, and answer order. The control condition had four lists which counterbalanced low- and high-attachment feedback and answer order. Item order was randomized within each list for each participant.

Results

A mixed effect model was constructed to analyze the effect of buzzer location, congruency, and trial order on sentence interpretation. The results are displayed in Figure 6 showing the rate of low attachment interpretations made in each quarter of the experiment by group and buzzer location. There was no effect of buzzer location ($b=-.04$, $t=-.671$, $p>.05$) or congruency condition ($b=-.12$, $t=-1.72$, $p>.05$) on sentence interpretation. There was a main effect of trial order on sentence interpretation ($b=-.001$, $t=-2.262$, $p>.05$), but no significant two-way or three-way interactions. As subjects experienced more critical items, they moved towards interpreting sentences as low-attachment 50% of the time, regardless of where the buzzer occurred.

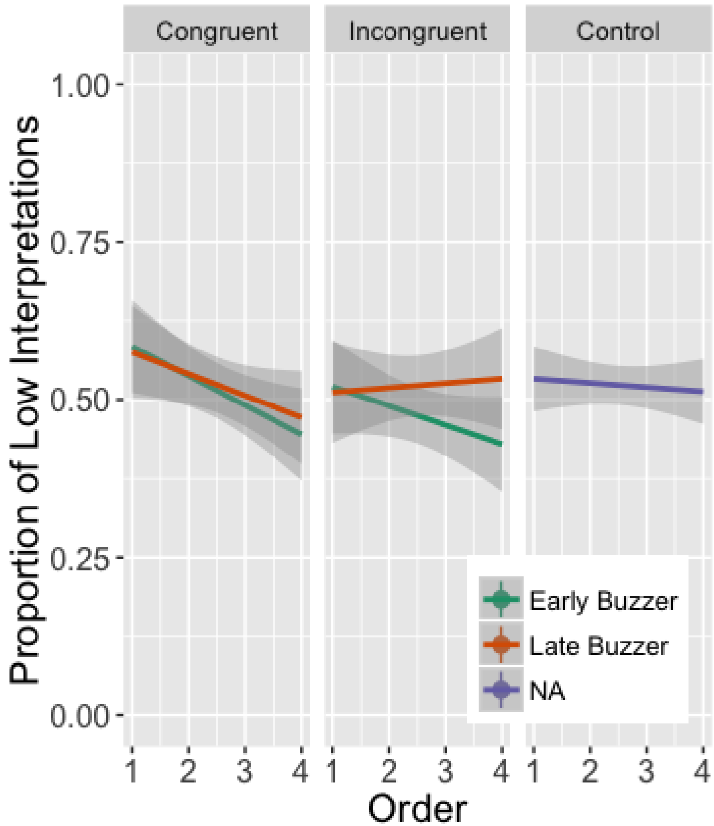

Figure 6. Proportion of correct interpretations to the Congruent and Incongruent buzzer conditions and the control condition.

Discussion

Subjects failed to show evidence of adaptation in any of the conditions. Despite the buzzers systematically indicating whether a listener should interpret a sentence as high- or low-attachment, subjects did not use this information to alter their interpretations as they moved through the experiment. Interestingly, there was a significant effect of item order on low-attachment interpretations. Figure 6 shows that as subjects moved through the experiment, regardless of the group they were in, they moved to making low-attachment interpretations about half of the time. This is evidence of potential adaptation to the statistics of the experiment itself. All subjects received feedback that indicated that low-attachment and high-attachment sentences occurred 50% of the time in this experiment. So, while subjects ignored the information that buzzers could tell them about the interpretation of the sentence, they may have been tracking the overall likelihood of certain syntactic structures occurring. This suggests that the adaptation seen to prosodic boundaries in Buxo-Lugo et al. (2020) was due to something special about the linguistic information contained in prosodic boundaries and may generally indicate that linguistic adaptation is limited to specific, linguistic cues. While not an inherently surprising claim, this is an important limitation suggested by Experiment 5. To more accurately understand adaptation, future research should consider what these limitations may be both in prosody and other domains of language.

Unfortunately, both Experiments 4 and 5 offer only potential limitations on prosodic adaptation. The null results make it difficult to draw firm conclusions about the nature of this adaptation and when it is successful. Thus, Experiment 6 was designed for three purposes. The first was to attempt to replicate the findings of Buxo-Lugo et al. (2020) in a similar study. As Experiments 4 and 5 failed to find any evidence of adaptation, it was important to replicate the basic adaptation effect to be confident in its existence. As Buxo-Lugo et al. is the first to find evidence of adaptation, replicating the effect is necessary in order to conclude that the null results in Experiments 4 and 5 were due to experimental design. The second purpose was, assuming that adaptation occurs in prosody, to begin to investigate how quickly prosodic adaptation occurs. As the discussion of both prior experiments has focused on questioning the limits of adaptation, Experiment 6 was designed to explicitly test one possible constraint. The last goal of Experiment 6 was to design an experiment that could broadly compare the Visibility Hypothesis and Adaptation in boundaries.

Experiment 6

The results of Buxo-Lugo et al. (2020) suggested that adaptation occurs to boundaries when subjects have 24 exposures to the critical boundary-structure pattern and are provided with feedback. There are many potential constraints to adaptation that should be investigated in the future. Experiment 6 focuses on one of them, namely what effect the frequency of cue exposure has on adaptation. Experiment 6 manipulated how often critical structures contained the boundary cue. Subjects were assigned to either the control condition, where all critical sentences had boundaries to help influence syntactic interpretation, or the experimental condition, where only 25% of the critical sentences had boundaries. In the experimental condition, the remaining 75% of critical trials had the boundary removed in the same way that boundaries were removed in Experiments 2 and 3. The distribution of critical items is shown in Table 1 below.

Group	Items with boundaries	Items without boundaries
Control (High Frequency)	32	0
Experimental (Low Frequency)	8	24

Table 1. Critical item distribution in Experiment 6

The goal of this experiment was to compare two groups, one of which should show adaptation, and one which may not depending on how sensitive adaptation is. If prosodic adaptation is rapid, both groups should show evidence of improving their syntactic interpretation by the end of the experiment. For the low-frequency group, this would be after only eight exposures of the critical sentence structure with the critical boundary. In both groups, every encounter with the cue was tied to explicit feedback. What differed was how many times the cue occurred across the critical items. If adaptation is a slower process, a graded response should be seen to the boundaries where subjects who receive consistent disambiguating cues should show the most learning in how to use these prosodic cues, while those who receive a lower frequency of disambiguating cues should show more variability in understanding how to apply boundaries to syntactic interpretation.

If the VH is correct and visibility of attachment sites is a natural constraint in sentence processing, subject should show similar interpretations of early and late boundaries regardless of the experimental group they belonged to. Importantly, these rates of interpretation should not vary over the course of the experiment. Because the VH proposes that visibility is a natural constraint in sentence processing, subjects should still have similar sensitivity to the perceptual units created by prosodic boundaries at the end of the experiment regardless of whether the cue appeared eight times or thirty-two times.

Methods

Subjects. Eighty subjects were recruited from Amazon Mechanical Turk and were paid $4.00 for completing the experiment. All subjects were retained for analysis.

Stimuli. Thirty-two critical items similar to those in Experiment 1 were used. These items were globally ambiguous sentences with RC attachment ambiguities. Each item was produced twice by a native speaker of American English, once with an early and once with a late boundary, as shown below.

6(a) The artist judged the painting // of the statue that was by the window.

6(b) The artist judged the painting of the statue // that was by the window.

Generally, an early boundary, as shown in 6(a), is associated with low-attachment, where the statue is understood to be by the window. A late boundary, as shown in 6(b), is associated with high-attachment, where the painting is understood to be by the window. A native English speaker produced these sentences with boundaries naturally produced. In addition to having two sentences with naturally produced boundaries, a third sentence with the prosodic boundary spliced out, as in Experiments 2 and 3, was created for each item. This resulted in a natural-sounding ambiguous sentence with no prosodic boundary cues.

Each critical item was paired with a free-response comprehension question that probed the interpretation of the sentence. For example, subjects who heard either sentence 6(a) or 6(b) above were asked "What was by the window? "and provided a written response to the question. All filler items were also paired with a free-response comprehension question that probed information contained in the sentence.

Procedure. The experiment was designed as one block where subjects listened to and answered 32 critical items and 29 filler items. At the beginning of the experiment, all subjects completed two practice items to familiarize themselves with the process. For all critical and filler items, subjects were probed on information contained in the sentence that was answered by a simple one-word or two-word answer. Subjects submitted their answer into the online experiment, and then received feedback based on whether or not their answer matched the correct answer. For instance, if subjects heard the sentence "The knight gave the roses to the lady", they would be asked "Who gave roses away?". An upper or lower-case answer with or without an article was accepted, as were popular misspellings (such as 'night'). If a subject correctly entered one of these answers, they received the feedback "Correct!'. If a subject did not enter one of these answers, they received the feedback “Incorrect. The answer was the knight gave the roses away.” For critical items, the questions specifically probed whether a subject believed the sentence had a high- or low-attachment interpretation.

Subjects were assigned to either the high-frequency cue group or the low-frequency cue group. In the high-frequency group, half of all critical items had an early boundary and the other half had a late boundary. The feedback for each critical item was consistent so that subjects were told that early-boundary items should have had a low-attachment interpretation, and late-boundary items should have had a high-attachment interpretation. In the low-frequency group, four critical items had early boundaries, four critical items had late boundaries, and the remaining 24 critical items had no obvious boundary. For the critical items with early-boundaries subjects received low-attachment feedback, while late-boundary items were always paired with high-attachment feedback. The 24 critical items with no boundary were randomly assigned to receive low-attachment or high-attachment feedback so that the subject was told each attachment type was correct 50% of the time. Four lists for each of the groups were created, resulting in eight experimental lists that varied which of critical items had early, late, or no boundary. The presentation of individual items was randomized completely for each subject.

Results

Subjects’ answers were coded as to whether they indicated a low- or high-attachment interpretation for each critical item. A mixed effect model was constructed to analyze the effect of boundary location, experimental condition, and overall trial order on sentence interpretation.

There was a main effect of boundary location on sentence interpretation, such that early-boundaries resulted in more low-attachment interpretations ($b=.31$, $t=-6.233$, $p<.05$). These results are shown in Figure 7 below. There was no effect of experiment condition ($b=-.05$, t=-

.665, $p>.05$) or trial order ($b=.0001$, $t=.125$, $p>.05$) on sentence interpretation, suggesting that sentence interpretation to boundary location was similar across groups and time.

As a post-hoc analysis, critical items were coded to reflect the order in which critical items only were seen. This resulted in a critical item order, referred to as rank, from 1 – 32. A mixed-effect model was constructed to analyze whether this rank order would show evidence in favor of adaptation. This model analyzed the effect of boundary location, experimental condition, and critical item order on sentence interpretation. Again, there was a main effect of boundary location on sentence interpretation ($b=-.30$, $t=-6.052$, $p<.05$), such that early-boundaries resulted in higher rates of low-attachment interpretations. However, there was still no main effect of experiment condition ($b=.04$, $t=-.528$, $p>.05$) or trial order ($b=.0003$, $t=.155$,

To test whether adaptation occurred between the beginning and end of the experiment but was not detectable on an item by item basis, the order in which subjects saw the critical items was broken into blocks of four items each. This allowed for a comparison of how subjects interpreted the first four items they encountered and the last four they encountered. This was analyzed in the high frequency group only as all critical items they listened to had boundaries. A mixed effect model was constructed which analyzed the effect of boundary location and block on proportion of low-attachment interpretations. As before, there was a main effect of boundary location ($b=-.28$, $t=-4.673$, $p<.05$), but no effect of block ($b=-.02$, $t=-.385$, $p>.05$).

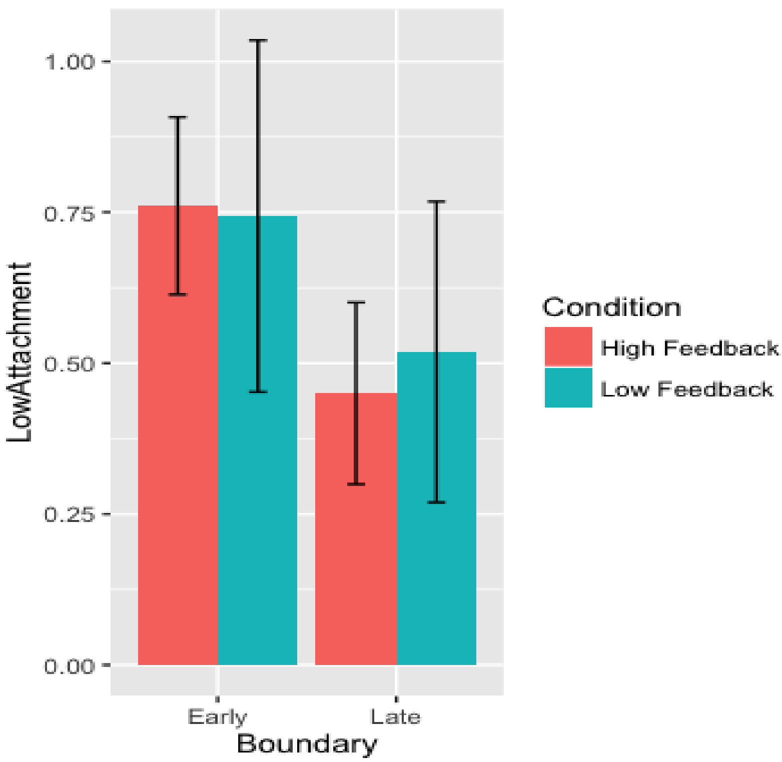

Figure 7. Proportion of low-attachment interpretations by boundary location in both feedback conditions.

Discussion

The data of Experiment 6 showed that the proportion of low-attachment interpretation across early and late boundary sentences was similar in the high- and low-frequency groups. The rate of low-attachment across boundaries did not change as a function of the frequency of which subjects encountered the critical boundary cue. Even when analyzed alone, the high-frequency group failed to show evidence of adaptation. Grouping item presentation by blocks and comparing the earliest and last block to look for adaptation still failed to find evidence of adaptation. These null results make it difficult to draw firm conclusions, but cast questions regarding how powerful adaptation is and how adaptation works for prosodic information. Prior research had found evidence that adaptation to intonational boundaries occurs, but the failure to conceptually replicate this in Experiment 6 raises questions about how and when prosodic adaptation is successful.

Instead, Experiment 6 provides results that would be predicted by the Visibility Hypothesis. Listeners interpreted prosodic boundaries in a similar pattern regardless of whether they had encountered this cue eight times or thirty-two times. This suggests that listeners were not updating or altering their knowledge of what a boundary meant as a result of experience with the cue. Unfortunately, the design of this experiment did not allow for further investigation of why listeners interpreted the boundaries as they did. Because of this, these results were not seen as evidence that the visibility of processing units, as the VH hypothesizes, were what accounted for the results.

IV. General Discussion

The goal of this dissertation was to research the influence that prosodic boundaries have in syntactic interpretation. Two main theories of language processing, Adaptation and The Visibility Hypothesis, were used to frame the research questions of Experiments 1 -6.

Visibility Hypothesis

The Visibility Hypothesis proposes that natural constraints in processing are what drive the syntax-boundary relationship. The data presented in this dissertation lend little support for the VH. Experiment 1 was completed to ensure that the stimuli used to test the VH were valid. In Experiment 1, listeners' syntactic interpretations were influenced by the position of naturally produced boundaries as predicted by the VH, allowing for the use of the same sentences in Experiments 2 and 3. In Experiments 2 and 3, artificially created perceptual units were introduced through the use of buzzers and inserted silences. The goal in both cases was to create the same perceptual units within sentences that boundaries provide, but through alternate means. In both cases, subjects failed to use these cues as a means of interpretation. This suggested that visibility of sentence information due to how perceptual units package information is not what drives the boundary-syntax relationship.

However, it is clear that listeners have a bias to interpret prosodic boundaries in line with the predictions of the VH. Experiment 6 was designed to broadly compare the predictions of the VH and the IAF. There, listeners failed to adapt to boundary location, and instead showed similar syntactic interpretation patterns regardless of how often they experienced the critical cue. These results raise several questions about boundaries. The first, which was explored in this dissertation, was whether adaptation could provide a better explanation for how listeners decide to interpret boundaries. However, alternative explanations for these results should be investigated. It may be that specific aspects of boundaries, which were not explored thoroughly in Experiments 2 and 3, are what create processing units and impact visibility.

Adaptation

Adaptation proposes that language processing in general is driven by the parser's ability to constantly monitor the language input which allows it to update its expectations of what is likely to occur. By updating its expectations, the parser eases comprehension when it correctly anticipates incoming input. Initially, evidence of prosodic adaptation, found by Buxo-Lugo et al. (2020), suggested that this may be a better explanation than the VH for how boundary-syntax mappings are created, altered, and used. The researchers found evidence that listeners exposed to boundary-syntax mappings with feedback on how to interpret the boundaries altered their interpretation of boundaries in the direction that the feedback had given them. If the results of Buxo-Lugo et al. (2020) were correct, it suggests that the boundary-syntax relationship may not be constrained due to processing limitations, but rather by the experience a listener has with other speakers' production patterns. Experiments 4-6 were designed to test prosodic adaptation further to understand the circumstances under which it is successful.

Initially, both Experiments 4 and 5 failed to find evidence of adaptation. Due to the design of these experiments, they could not answer how successful prosodic adaptation was, but rather pointed to reasonable constraints that might exist. Experiment 4 failed to find evidence of adaptation when listeners had syntax to disambiguate the sentences they were hearing and were simply asked to rate how natural language sounded. This suggested that adaptation requires a more fine-grained measure to accurately capture it. Experiment 5 had subjects receive feedback on how to interpret buzzers that were used in a boundary-like fashion. When subjects failed to adapt to these non-linguistic cues, it suggested that linguistic adaptation is limited to reasonable communicative cues. However, following these two failures to find evidence of adaptation, Experiment 6 was designed with a condition that sought to replicate Buxo-Lugo et al.'s finding of prosodic adaptation.

The failure of Experiment 6 to replicate Buxo-Lugo et al. (2020) makes it difficult to draw strong conclusions about the role that adaptation may play in using prosodic boundaries for syntactic interpretation. It suggests that either a Type II error was made in Experiment 6, a Type I error was made by Buxo-Lugo et al., or that adaptation, at least in prosodic boundaries, is a weak effect that is hard to capture. Making it difficult to conclude which of these three explanations is accurate is that both experiments had similar sample sizes, item counts, and designs. Experiment 6 investigated adaptation over 32 items with 40 subjects, while Buxo-Lugo et al. had 50 subjects in each group and 24 items. Both experiments used the same sentence construction and the same native speaker to record the stimuli. Importantly, the similarity in design and subject and item size does suggest that the failure to replicate was not due to substantial experimental changes. Much of the difficulty in drawing conclusions from Experiment 4 was due to using a method and measurement that differed from Buxo-Lugo et al. But the failure to replicate prosodic adaptations in Experiment 6 means it is reasonable to being to question whether adaptation occurs to boundaries.

If prosodic adaptation does occur, it appears that it is a weak effect which requires experiments to have high power to reliably be detected. This is not surprising because even ignoring the failure to replicate adaptation in Experiment 6, the null findings in Experiment 4 indicate that prosodic adaptation may be difficult to induce or measure. When subjects were given exposure to temporarily ambiguous sentences with unusual boundary placements, they rated the sentences as sounding unnatural throughout the experiment, and did not improve their ratings as a function of experience. This suggests that mere exposure to altered prosodic patterns is not enough to induce adaptation, at least on a short time-scale. Future research should increase sample size and, if adaptation is found, begin to explore the constraints of adaptation. The limitations suggested by Experiments 4 and 5, that type of cue, type of feedback, type of exposure, and type of task may play a role, would be an excellent place to begin.

Prosodic Boundaries and Syntax

The six experiments in this dissertation failed to find evidence that either the VH or the IAF, as they currently are hypothesized, accurately describe the influenced that boundaries have on syntactic interpretation. Both Experiments 1 and Experiments 6 showed that the stimuli used throughout this dissertation were appropriate for testing the boundary-syntax relationship. This points to a need for a theory to better describe the relationship. Because neither theory of language processing was supported, I consider what else these experiments may indicate about prosodic boundaries and syntax.

The first is simply that these six experiments provide further evidence of the strong relationship between syntax and boundary placement. Experiments 1 and 6 replicated prior work showing the general relationship listeners assume exists between early boundaries and low attachment in English. In Experiment 4, listeners rated grammatically correct sentences produced by a native speaker as being less natural if boundaries were put in syntactically unexpected locations. While there is variability in how listeners use boundaries to make syntactic interpretations, all three experiments provided evidence of this robust relationship. Importantly, these experiments replicated prior work showing that syntactic interpretation is strongly influenced by boundaries even though it is not a 1:1 relationship.

It seems that there is something special about prosodic boundaries that listeners are attuned to. This suggests that, broadly, listeners may be sensitive to what types of information certain linguistic elements can convey. In the case of boundaries, it could be that they are known for structuring language. That is, a speaker who produces a boundary is often placing it at a point where an idea is coming to an intermediate or final end (Pike, 1945). If this is the experience listeners have with boundaries, then the presence of a boundary may indicate to a listener that a speaker is structuring information together in a helpful way and act as a potential trigger for listeners to interpret the syntax in one way or another.

In some ways, this is similar to what the VH predicts, which is that the presence of prosodic boundaries is packaging semantic information together in a way that impacts listeners' interpretations. However, the process it proposes differs significantly. Rather than boundary effects occurring due to *cognitive constraints* on information processing, boundary effects would be due to metalinguistic knowledge that boundaries are produced to create structure. This would mean that there is something specific about boundaries that induces these effects, rather than something general about language processing.

The question of why adaptation was not seen to prosodic boundaries cannot be answered with the experiments completed here. But if boundaries are seen as markers of structural intent, rather than having natural processing constraints impose their effects, it would follow that listeners *should* be able to adapt their understanding of what these boundaries mean when provided with sufficient evidence. And, since there is ample evidence of adaptation in other areas of language such as phoneme perception, it is pertinent to consider why boundaries would be different. Considering the ways in which boundaries differ from phonemes in general language use may indicate several reasons why a difference in adaptation occurs.

There are three ways in which phonemes and boundaries differ conceptually that may impact adaptation – necessity, priority, and variability. The first, necessity, is how necessary the production of a linguistic aspect is to comprehending speech. One of the interesting aspects of the boundary-syntax relationship is that boundaries often bias listeners towards an interpretation, but they are not always produced or used in this manner (Kraljic & Brennan, 2005b; Snedeker & Trueswell, 2003). This is one of the reasons that understanding the boundary-syntax relationship is so difficult. When a strong bias exists, but is not always true, it suggests that a linguistic component isn't *needed* for comprehension, but is still helpful and used by listeners. And so, a listener may find it useful for a speaker to structure an utterance with boundaries to produce "I'm going to Nashville // and I'm letting you come.", but not vital to understanding the message. However, hearing a speaker say the same phrase but with missing or mispronounced phonemes could lead a listener astray into hearing "I'm going to Asheville and I'm getting you gum." Referring back to Frazier and Clifton (1997), the sentence "The bus driver stopped the rider with a mean look" can be produced with a boundary in many different places or with no boundary at all. While the boundary placement may influence listeners to make one interpretation over the other, all forms of the sentence, including the one absent a boundary, are acceptable. But when listeners encounter a speaker with altered pronunciation, they tend to note the mispronunciation and initially have slowed processing (see Clarke & Garrett, 2004 for instance). This may be an important factor in the parser's decision of what to adapt to first and foremost. If a speaker produces a *necessary* feature in an altered manner, the parser may be able to note the difference faster and update its distribution for the feature more quickly than it would for an unnecessary, though still helpful, feature. In addition, an altered phoneme may have the ability to change the meaning of a message far more than an altered boundary location. Minimal pairs, words which differ by just one phoneme, can cause confusion if mistaken for each other. If a listener encounters a speaker who is unclear of their production of /s/ or a /sh/, it may be a high priority for the parser to decode which is meant in order for communication to continue successfully. On the contrary, if a listener encounters a speaker who sometimes moves their boundaries to an unlikely location, this may not stop successful decoding of the message.

It could be argued though that most communication holds enough disambiguating information and context that it is unlikely that one feature is every truly necessary. That is, a listener who is unsure if a speaker produced a /d/ or /t/ in a word, can deduce from the context of the conversation on camping and the syntactic structure of the sentence that the speaker was saying “I think I will pack the big *tent*”, and not “I think I will pack the big *dent*.” In this way, the phoneme production may be no more necessary than a boundary in a locally ambiguous phrase that is disambiguated by the context that occurred before hand. If this is the case, the parser still may assign priority to understanding the distribution of features that it attends to first for communication. This may be determined by how often ambiguous and potentially ambiguous utterances contain the feature in question. In the case of phonemes, minimal pairs are not uncommon in English which may cause phonemes to be a high priority for the parser. For boundaries, there may be few instances where the placement of a boundary is crucial to understanding a message. In this case, the parser may assign lower priority to tracking the distribution of prosodic boundaries.

How quickly adaptation takes place may also be dependent on how much variability a particular feature tends to display. The more variable a feature is, the harder it may be for the parser to update a distribution. This would be comparable to simple hypothesis testing where the null hypothesis is rejected when observed data is far enough away from the expected mean. As the standard error of the mean increases due to more variability of the samples in question, an observed collection of data has to occur farther out from the mean in order for the null to be rejected. In the case of boundaries, their placement and even existence within phrases may vary enough that the parser requires quite a bit of evidence that a speaker is producing them in a unique way before it 'rejects the null' that a speaker is following a normal pattern and begins learning a new distribution for a speaker. Kleinschmidt (2018) proposes similar constraints to explain differing levels of adaptation to phonemes. He argues that realistically a listener cannot track and use all available knowledge of linguistic cue distributions, so the *utility* and *informativity* of adapting to a phoneme within a context are considered by the parser. The utility of phoneme is high when it is useful in allowing a listener to successfully cope with a speaker's variability. The informativity of a phoneme is high when a listener can make accurate predictions about cue-mappings related to the phoneme for a group of speakers. Kleinschmidt discusses in-depth the impact that these two categories have on phoneme adaptation specifically, but suggests that their use is not limited to just phonemes.

An initial first step to exploring these factors would be to analyze spoken corpora for instances of potential ambiguity in phoneme production and boundary location. To look for the potential influence of necessity and priority, finding how often words with minimal-pairs are produced and how often ambiguous sentences that boundaries could disambiguate are produced would be a first step. If there are more cases of minimal pair words occurring, it would suggest that, in real world interactions, listeners have a reason to prioritize phoneme perception. To look for the potential influence of variability, a more detailed look would have to be taken at phoneme production and boundary placement within certain syntactic structures. This would be a harder comparison to make as the variability that exists in phoneme and boundary production do not seem comparable. Phoneme production would most often vary in pitch production and Voice Onset Time, whereas boundaries would vary in the strength and placement of boundary. To take a broader look at all of prosody, it would be possible (though tedious) to note the times in which altered phoneme production hindered communication and when prosodic features seemed to do the same. For instance, a speaker who says "I'm so excited" in a low-volume with a flat intonation might be asked "Really? You don't seem excited." by a listener who notes the mismatched prosody and message. An exploratory analysis of natural conversational corpora would give insight into how often these conversational hiccups arise.

The creation of an experimental corpus may also be very useful here. The collection and transcription of speech that is directed towards minimal pairs and boundary production would allow for collecting natural language but with greater chance of capturing the features in question. Giving subjects case studies to talk about that have the potential for minimal pair subjects and ambiguous phrases would be ideal. This would allow for measuring how often ambiguous words and phrases occur and how well listeners interpret these situations, without constricting speech choices too much. Along with the natural corpora, this would allow for an exploratory analysis of how often listeners encounter prosodic boundary variability and how often listeners *need* to use boundaries rather than them supplementing speech.

Lastly, a simple experiment that measures how subjects process phonemically and prosodically difficult audio may provide a useful measure of how important each aspect is to speech comprehension. Subjects could be asked to transcribe audio or provide a value judgment about audio that has been phonemically altered or has inappropriate prosody. In the phonemic case, subjects might hear "?eaches are my favorite!" where it is unclear if the ? is a /p/ or /b/. In the prosodic case, subjects might hear "Peaches are my favorite." with prosody associated with anger or fear. The reaction time of subjects to transcribe the audio or make a value judgment may reveal what effect unexpected prosody has on processing speech and how it compares to unexpected phoneme production.

Conclusion

The goal of this dissertation was to research the relationship that exists between syntax and prosodic boundaries, specifically focusing on understanding what mechanisms play a role in their relationship. The link between boundaries and syntax is replicated several times, confirming the robust relationship between them. The experiments here fail to find evidence in favor of the Visibility Hypothesis. There is also a failure to find evidence for adaptation in prosodic boundaries, but the findings point to clear limitations which may play a part in limiting the power of prosodic adaptation. These avenues of research should be explored in the future.

References

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of personality and social psychology*, *70*(3), 614.

Braun, B., Dainora, A., & Ernestus, M. (2011). An unfamiliar intonation contour slows down online speech comprehension. *Language and Cognitive Processes*, *26*(3), 350-375.

Bregman, A. S. (2008). Auditory scene analysis. *The senses: A comprehensive reference*, *3*, 861-870.

Buxo-Lugo, A., Harrington Stack, C.M., Kurumada, C., & Watson, D.G. (in prep). Variability and adaptation in prosody-based syntactic comprehension.

Carreiras, M., & Clifton, C. (1999). Another word on parsing relative clauses: Eyetracking evidence from Spanish and English. *Memory & cognition*, *27*(5), 826-833.

Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *The Journal of the Acoustical Society of America*, *116*(6), 3647-3658.

Cutler, A., Dahan, D., & Van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and speech*, *40*(2), 141-201. https://doi.org/10.1177/002383099704000203

Dahan, D., Drucker, S. J., & Scarborough, R. A. (2008). Talker adaptation in speech perception: Adjusting the signal or the representations?. *Cognition*, *108*(3), 710-718. https://doi.org/10.1016/j.cognition.2008.06.003

Eisner, F., & McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. *Attention, Perception, & Psychophysics*, *67*(2), 224-238. https://doi.org/10.3758/BF03206487

Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PloS one*, *8*(10).

Floccia, C., Goslin, J., Girard, F., & Konopczynski, G. (2006). Does a regional accent perturb speech processing?. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(5), 1276-1293. 10.1037/0096-1523.32.5.1276

Fraundorf, S. H., & Jaeger, T. F. (2016). Readers generalize adaptation to newly-encountered dialectal structures to other unfamiliar structures. *Journal of Memory and Language*, *91*, 28-58.

Clifton, C., Carlson, K., & Frazier, L. (2006). Tracking the what and why of speakers' choices: Prosodic boundaries and the length of constituents. *Psychonomic Bulletin & Review*, *13*(5), 854-861.

Frazier, L. & Clifton, C. (1997). Sentence reanalysis, and visibility. In J. Fodor and F. Ferreira (EDS.) *Reanalysis in Sentence* Processing (pp. 143-176).

Gibson, E., Pearlmutter, N. J., & Torrens, V. (1999). Recency and lexical preferences in Spanish. *Memory & Cognition*, *27*(4), 603-611.

Grodner, D., & Sedivy, J. C. (2011). 10 The Effect of Speaker-Specific Information on Pragmatic Inferences. In *The processing and acquisition of reference* (Vol. 2327, pp. 239-272). MIT Press.

Harrington Stack, C. M., James, A. N., & Watson, D. G. (2018). A failure to replicate rapid syntactic adaptation in comprehension. *Memory & cognition*, *46*(6), 864-877.

Kamide, Y. (2012). Learning individual talkers' structural preferences. *Cognition*, *124*(1), 66-71.

Kleinschmidt, D. F. (2019). Structure in talker variability: How much is there and how much can it help?. *Language, cognition and neuroscience*, *34*(1), 43-68.

Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological review*, *122*(2), 148.

Kraljic, T., & Samuel, A. G. (2007). Perceptual adjustments to multiple speakers. *Journal of Memory and Language*, *56*(1), 1-15. https://doi.org/10.1016/j.jml.2006.07.010

Kraljic, T., & Samuel, A. G. (2006). Generalization in perceptual learning for speech. *Psychonomic bulletin & review*, *13*(2), 262-268. https://doi.org/10.3758/BF03193841

Kraljic, T., & Samuel, A. G. (2005a). Perceptual learning for speech: Is there a return to normal?. *Cognitive psychology*, *51*(2), 141-178. https://doi.org/10.1016/j.cogpsych.2005.05.001

Kraljic, T. & Brennan, S.E. (2005b). Prosodic disambiguation of syntactic structure: For the speaker or for the addressee? *Cognitive Psychology, 50*(2), 194-231. https://doi.org/10.1016/j.cogpsych.2004.08.002

Kurumada, C., Brown, M., Bibyk, S., Pontillo, D., & Tanenhaus, M. (2014, January). Rapid adaptation in online pragmatic interpretation of contrastive prosody. In *Proceedings of the Cognitive Science Society* (36).

Kurumada, C., Brown, M., & Tanenhaus, M. (2012). Pragmatic interpretation of contrastive prosody: It looks like speech adaptation. In *Proceedings of the Cognitive Science Society* (34).

Lehiste, I. (1973). Phonetic disambiguation of syntactic ambiguity. *The Journal of the Acoustical Society of America*, *53*(1), 380-380.

Lehiste, I., Olive, J. P., & Streeter, L. A. (1976). Role of duration in disambiguating syntactically ambiguous sentences. *The Journal of the Acoustical Society of America*, *60*(5), 1199-1202.

Lima, C. F., & Castro, S. L. (2011). Speaking to the trained ear: musical expertise enhances the recognition of emotions in speech prosody. *Emotion*, *11*(5), 1021.

Maye, J., Aslin, R. N., & Tanenhaus, M. K. (2008). The weckud wetch of the wast: Lexical adaptation to a novel accent. *Cognitive Science*, *32*(3), 543-562. 10.1080/03640210802035357

Pike, K. L. (1945). The Intonation of American English. University of Michigan Publications.

Price, P. J., Ostendorf, M., Shattuck-Hufnagel, S., & Fong, C. (1991). The use of prosody in syntactic disambiguation. *the Journal of the Acoustical Society of America*, *90*(6), 2956-2970.

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*(4), 2037-2047.

Ryskin, R. A., Qi, Z., Duff, M., & Brown-Schmidt, S. (2016). Constraints on adaptation to syntactic variability between and within speakers. In *Poster presented at the 29th Annual CUNY Conference on Human Sentence Processing, Gainesville, FL*.

Sanderman, A. A., & Collier, R. (1997). Prosodic phrasing and comprehension. *Language and Speech*, *40*(4), 391-409.

Schafer, A. J., Speer, S. R., Warren, P., & White, S. D. (2000). Intonational disambiguation in sentence production and comprehension. *Journal of psycholinguistic research*, *29*(2), 169-182.

Shattuck-Hufnagel, S., & Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of psycholinguistic research*, *25*(2), 193-247.

Simpson, A. P. (2009). Phonetic differences between male and female speech. *Language and linguistics compass*, *3*(2), 621-640.

Snedeker, J., & Trueswell, J. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and language*, *48*(1), 103-130.

Swerts, M. (1997). Prosodic features at discourse boundaries of different strength. *The Journal of the Acoustical Society of America*, *101*(1), 514-521.

Veenendaal, N. J., Groen, M. A., & Verhoeven, L. (2014). The role of speech prosody and text reading prosody in children's reading comprehension. *British Journal of Educational Psychology*, *84*(4), 521-536.

Wagner, M., & Watson, D. G. (2010). Experimental and theoretical advances in prosody: A review. *Language and cognitive processes*, *25*(7-9), 905-945.

Wells, J. B., Christiansen, M. H., Race, D. S., Acheson, D. J., & MacDonald, M. C. (2009). Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive psychology*, *58*(2), 250-271.

Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *The Journal of the Acoustical Society of America*, *91*(3), 1707-1717.

Wingfield, A., Lombardi, L., & Sokol, S. (1984). Prosodic features and the intelligibility of accelerated speech: Syntactic versus periodic segmentation. *Journal of Speech, Language, and Hearing Research*, *27*(1), 128-134.

Yildirim, I., Degen, J., Tanenhaus, M., & Jaeger, F. (2013). Linguistic variability and adaptation in quantifier meanings. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 35, No. 35).