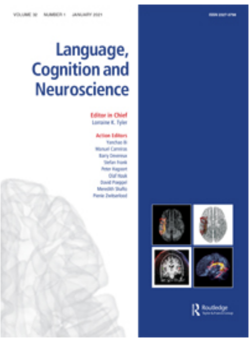


Routledge
Taylor & Francis Group

Language, Cognition and Neuroscience

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/plcp21

Integrating prosody in anticipatory language processing: how listeners adapt to unconventional prosodic cues

Chie Nakamura, Jesse A. Harris & Sun-Ah Jun

LANGUAGE, COGNITION AND NEUROSCIENCE
https://doi.org/10.1080/23273798.2021.2010778

Routledge
Taylor & Francis Group

REGULAR ARTICLE

OPEN ACCESS Check for updates

Integrating prosody in anticipatory language processing: how listeners adapt to unconventional prosodic cues

Chie Nakamura [a], Jesse A. Harris[b] and Sun-Ah Jun[b]

[a]Global Center for Science and Engineering, Waseda University, Tokyo, Japan; [b]Department of Linguistics, University of California Los Angeles, Los Angeles, CA, USA

ABSTRACT

A growing body of research suggests that language users integrate diverse sources of information in processing and adapt to the variability of language at multiple levels. In two visual-world paradigm studies, we explored whether listeners use prosody to predict a resolution to structures with a PP that is structurally ambiguous between a modifier and an instrument interpretation. The first study revealed that listeners predict a referent that is most compatible with the location of a prosodic boundary, casting anticipatory looks to the appropriate object even before the onset of a disambiguating word. The second study indicated that listeners failed to anticipate instrument resolutions when the prosody of non-experimental filler items was unconventional, even though experimental items remained identical to the first study. The results suggest that listeners adjust their predictive processing to the utility of prosodic information according to whether a speaker reliably conforms to the conventional use of prosody.

ARTICLE HISTORY
Received 4 September 2020
Accepted 10 November 2021

KEYWORDS
Prosody; reliability; prediction; anticipatory eye-movements; attachment ambiguity

1. Introduction

When comprehending spoken language, listeners must integrate diverse sources of linguistic and non-linguistic information to form a coherent representation. In this study, we concentrate on the integration of prosodic information, broadly defined acoustically as changes in pitch, duration, and amplitude accompanying an utterance, which is phonologically categorized as stress, pitch accent, intonation, and prosodic structure, etc. (Ladd, 1996/2008; Cutler, Dahan, & Donselaar, 1997; Wagner & Watson, 2010, for review). Prosodic information can be used to signal syntactic constituency, prominence relations among words, and sentence types as well as the information structure status of a constituent in discourse (Beckman & Pierrehumbert, 1986; Pierrehumbert & Hirschberg, 1990; Price, Ostendorf, Shattuck-Hufnagel, & Fong, 1991; Shattuck-Hufnagel & Turk, 1996; Ladd, 1996/2008; Jun, 2005, 2014). For instance, an utterance of the form *You're a doctor* has a very different interpretation when rendered with a rising question intonation as opposed to a falling statement intonation.

Structural ambiguity can also be resolved by prosodic information. Although the relation between prosodic structure and syntactic structure is not strictly isomorphic, prosodic groupings, especially those larger than words, are often aligned with syntactic groupings (Nespor & Vogel, 1986/2007; Ferreira, 1988; Selkirk, 1986, 2000, 2007, 2011; Hayes, 1989; Jun, 1998; Truckenbrodt, 1999). Typically, the boundary of a prosodic unit or prosodic phrase is cued by lengthening of the phrase-final syllable and the tonal changes (e.g. rising, falling) on the final syllable, called a boundary tone. The last syllable of the prosodic unit can also be followed by a pause. For example, sentence (1) is structurally ambiguous between two meanings depending on how the prepositional phrase (PP) is construed. If the PP is attached to the modifier noun (*the musicians*), the PP conveys that the musician has lousy equipment (henceforth *modifier interpretation*). If the PP is attached to the verb (*recorded*), the PP is understood as conveying that the engineer used lousy equipment to record the musician (henceforth *instrument interpretation*). In this example, the location of a prosodic boundary (indicated by the '%' symbol) can contribute to disambiguating the structure, and hence its meaning. This prosodic boundary is typically cued by intonational and durational patterns. In example (1a), the last syllable *-ed* in *recorded* would be much longer in duration than that in example (1b) and carry a boundary tone, while in

CONTACT Chie Nakamura cnakamura@aoni.waseda.jp Waseda University Build. 61-502B, 3-4-1 Okubo, Shinjuku, Tokyo 169-8555, Japan

example (1b), the last syllable –*cians* in *musicians* would be much longer than that in example (1a) and carry a boundary tone. In both examples, the prosodic boundary could be made stronger by adding a pause.

(1) The engineer recorded the musicians with lousy equipment.
a. The engineers recorded % the musicians with lousy equipment. (modifier interpretation compatible)
b. The engineers recorded the musicians % with lousy equipment. (instrument interpretation compatible)

In an auditory interpretation study, Schafer (1997) found that the location of a prosodic boundary influenced how the PP attachment ambiguity was resolved.[1] The study showed that there was a slight overall preference for the instrument interpretation in PP attachment ambiguity resolution in processing sentences like example (1), but a significant decrease in instrument interpretations when a prosodic boundary was located before the modifier noun as in (1a). The results, as well as those of others using measures such as reading times, questionnaires, and judgment tasks (Beach, 1991; Schafer et al., 1996; Schafer, 1997), indicate that prosodic information provides good diagnostic cues to syntactic constituency (Price et al., 1991; Shattuck-Hufnagel & Turk, 1996; Beckman, 1996; Schafer 1997; Watson & Gibson, 2004; Watson et al., 2006; Breen et al., 2011 and references therein, see also Kjelgaard & Speer, 1999; Schafer, Speer, Warren, & White, 2000; Snedeker & Casserly, 2010; Speer, Kjelgaard, & Dobroth, 1996, among others).

However, while these studies provide convincing evidence that comprehenders integrate prosodic information with syntactic information when processing spoken language, they were not specifically designed to address the point at which listeners use prosodic cues to identify syntactic structure. Thus, it was not clear from these studies whether prosodic information has an immediate effect on initial parsing decisions or if prosodic considerations are attended to relatively late in comprehension. To more directly address the question of timing, studies that followed have often used sensitive online techniques such as eye-tracking (Ito & Speer, 2008) or electroencephalography (EEG) (see Bögels, Schriefers, Vonk, and Chwilla, 2011 for review).

An early example is Snedeker and Trueswell (2003), who examined globally ambiguous sentences, as in example (2). Similar to the previous example, the PP (*with the flower*) may attach to the NP, yielding a modifier interpretation (2a), or it may attach to the VP, yielding an instrument interpretation (2b). In a series of experiments, Snedeker and Trueswell (2003) investigated how speakers prosodically disambiguate such sentences in production, and the extent to which listeners were sensitive to these prosodic cues in adopting the modifier interpretation or the instrument interpretation in a visual world eye-tracking study.

(2) Tap the frog with the flower.
a. Tap the frog that is holding the flower. (modifier interpretation)
b. Using the flower, tap the frog with it. (instrument interpretation)

The production study showed that speakers tended to insert a prosodic boundary between the noun and ambiguous PP by lengthening the direct object noun (*the frog*) to convey an instrument interpretation (2b). To convey a modifier interpretation (2a), speakers inserted a prosodic boundary after the verb by lengthening the verb (*tap*). In the visual world eye-tracking study, Snedeker and Trueswell (2003) also observed that these prosodic markers influenced how listeners processed visual scenes shortly after the onset of the direct object and prior to the onset of the ambiguous PP. Their results demonstrated that listeners rapidly integrate prosodic information to determine syntactic structure, and that cues to prosodic boundaries also lead listeners to make predictions about upcoming speech.

Snedeker and Trueswell (2003) findings suggest that prosodic information is immediately accessed at the earliest stages of online processing. A growing number of studies indicate that listeners use prosodic cues even before they encounter lexical material, *anticipating* information that is likely to come next. The influence of prosody in anticipatory processing has been observed in different domains of language comprehension, such as structural prediction (Nakamura et al., 2012; Weber et al., 2006), interpretation of discourse (Kurumada, Brown, Bibyk, Pontillo, & Tanenhaus, 2014), and prediction in information structure (Dahan, Tanenhaus, & Chambers, 2002; Ito & Speer, 2008). These studies indicate that language users generate real-time expectations about upcoming linguistic input by rapidly integrating prosody, as well as other constraints such as discourse context and world knowledge.

Interestingly, the view that language processing draws on anticipation has been argued to ease communicative interactions and to facilitate efficient processing (Pickering & Garrod, 2007; Chang, Dell, & Bock, 2006). Recent studies suggest that anticipatory processing plays a critical role in explaining how language users cope with variability in the linguistic environment, finding that participants change how they process the experiment stimuli in response to the exposure they receive within or prior to the experiment, an effect sometimes referred to as *linguistic adaptation*. Linguistic adaptation has been observed on multiple levels of linguistic representations, such as phonetic and phonological processing (Clayards, Tanenhaus, Aslin, & Jacobs, 2008; Kleinschmidt & Jaeger, 2015; Kraljic, Samuel, & Brennan, 2008; Maye, Aslin, & Tanenhaus, 2008; Norris, McQueen, & Cutler, 2016), sentence level syntactic processing (Fine et al., 2013; Jaeger & Snider, 2013; Kaschak & Borreggine, 2008; Kaschak & Glenberg, 2004), and discourse context (Ryskin et al., 2020; Schuster & Degen, 2020; Van Berkum et al., 2005). Flexibility in adapting to speaker-specific information has been also observed in the perception of differences in pronunciation such as foreign accents and dialects (Bradlow & Bent, 2008; Maye et al., 2008), lexical choice (Brennan & Hanna, 2009), and the role of prosody in making pragmatic inferences (Kurumada, et al., 2014).

Collectively, these studies provide compelling evidence that comprehenders are acutely sensitive to how speakers use prosodic cues to signal a resolution to linguistic or perceptual ambiguity, presumably with the assumption that speakers reliably use such cues to be consistent and informative. However, one issue that has not been addressed by previous studies is whether comprehenders attend to the consistency with which a speaker uses prosody to signal syntactic structure, which we refer to here as the overall *reliability* or *informativity* of prosodic information. For example, while there is already substantial evidence that listeners are adept at predicting likely continuations of input they have received so far, it is not clear whether the overall reliability of prosody also feeds into the structural prediction process. Multiple studies have already shown that listeners engage in less anticipatory processing when contrastive accent (L+H*) is used unreliably or infelicitously (Tanenhaus, Kurumada, & Brown, 2015; Roettger & Franke, 2019; Roettger & Rimland, 2020; Foltz, 2020; Nakamura et al., 2019, 2021), suggesting that listeners are sensitive to the reliability of how contrastive accent is used by a speaker when making pragmatic inferences or in resolving referential ambiguity. In the present study, we address the extent to which language users consider how informative prosodic cues are when prosodic boundary information does not consistently provide useful cues for processing. Understanding whether listeners are sensitive to the overall reliability of prosody would address several current issues in sentence processing literature, including: the conditions and the extent to which the language comprehension system is adaptive, how comprehenders recover intended messages from noisy input, and how listeners might modify their language processing strategies to meet variation in the speech signal. As language users rapidly integrate various sources of information to decode speech and text in real-time, it is reasonable to think that the language processing system may not only use cues that are informative, but also reduce reliance on cues that are uninformative or idiosyncratic.

In support of this general view, Tanenhaus, et al. (2015) investigated the effect of rising contrastive pitch accent on pragmatic meaning. Prior to testing, participants were exposed to either a speaker who used contrastive accents in a consistent and reliable way (the *reliable speaker* group), or a speaker who used contrastive accents in an inconsistent and unreliable way (the *unreliable speaker* group). During the test phase, the reliable speaker group used intonational cues to choose a picture that was likely to reflect the speaker-intended referent. In contrast, the unreliable speaker group appeared to place less weight on the intonational cues in their judgments. Crucially, their results indicate that listeners assess how informative prosodic cues are in comprehending the speaker's intended meaning. However, since the study used an offline judgment task, the results could not address how listeners use the reliability of information accumulated over time in generating expectation for the upcoming input in real-time online language comprehension. More recently, Foltz (2020) used the visual world paradigm to address the questions regarding how listeners would respond to inconsistent prosodic cues. Her results with native German speakers showed that exposure to infelicitous use of contrastive accent (e.g. *Click on the red duck. Click on the GREEN*$_{\mathrm{L+H}}$ *banana*.) decreased participants' engagement in predictive processing over the course of the experiment.

Roettger & Franke (2019) also addressed a closely related issue using a forced-choice mouse-tracking task. They manipulated prosodic realization of sentences like example (3), which was an answer to a topic-question such as *Hat der Wuggy dann die Geige aufgesammelt?* ("Did the Wuggy then pick up the violin?"). Two types of prosody in the responses to the question, verb accent prosody and object accent prosody, were compared. The verb accent prosody was characterized by a rising-falling pitch accent on the auxiliary verb *hat* ("has"), which provides the discourse-given referent interpretation of the object. On the other hand, the object accent prosody had a rising-falling pitch accent on the object, which provides the discourse-new referent interpretation of the object. Therefore, as an answer to the question *Hat der Wuggy dann die Geige aufgesammelt?* ("Did the Wuggy then pick up the violin?"), the discourse-given prosody was more natural for (3a) because "violin" has been already introduced in the discourse by the question. In contrast, the discourse-new prosody was more natural for (3b) because the answer "pear" corrects the topic "violin".

(3) a. Der Wuggy hat dann die Geige aufgesammelt.
(The Wuggy then picked up the violin.)
b. Der Wuggy hat dann die Birne aufgesammelt.
(The Wuggy then picked up the pear.)

Overall mouse trajectories indicated that participants used prosodic information to make predictive inferences (violin or pear) during sentence comprehension in Experiment 1. Crucially for our purposes, Roettger & Franke (2019) also tested how unreliable prosody impacted responses in Experiment 2. When listeners were exposed to items with unconventional prosody (e.g. discourse-given prosody placed on discourse-new material in (3b)), anticipatory mouse responses toward the target referent were disrupted. These results provide good evidence that listeners evaluate the reliability of the mapping between prosodic patterns and their function, and that they adjust the extent to which they rely on prosodic information in making pragmatic inferences about the intended message of the utterance (see also Roettger & Rimland, 2020).

One possible criticism for the previous studies that tested adaptation to unreliable use of prosody is that it is not clear whether the listeners in fact evaluated the overall informativity of prosody, or they simply learned the speaker-specific form-function mapping of the contrastive accent. Roettger and Franke (2019), for example, showed that when exposed to unconventional use of pitch accent placement, participants initially discarded the informational value of intonation. Crucially, however, the more listeners were exposed to unconventional use of pitch accent, the more they associated pitch accents with unconventional meanings (see also Tanenhaus et al., 2015; Roettger & Rimland, 2020; Foltz, 2020, 2021; Nakamura et al., 2019, 2021). Therefore, it is not yet clear how listeners respond to unconventional use of prosody when it cannot be mapped onto any alternative meaning.

1.1. Research questions for present study

Prosodic boundaries provide strong cues that can disambiguate structures and their corresponding meanings, and are rapidly used during online comprehension (e.g. Snedeker & Trueswell, 2003). As reviewed above, converging evidence indicates that language users use prosodic information to predict upcoming input. In addition, listeners adapt to variation in how prosodic cues are used by a particular speaker. However, less is known about whether predictive processes are subject to adaptation, especially with respect to the effect of prosodic boundaries on disambiguation. Our studies were designed to address three central questions that begin to fill this gap. First, do listeners predict different syntactic structures depending on the location of a prosodic boundary? Second, do listeners reduce the degree to which they use prosodic information in anticipatory processing when a speaker places prosodic boundaries in positions that do not consistently align with sentence structure? Third, as listeners acquire evidence that prosodic boundaries do not reliably align with syntactic constituency, is the cost for violating prosodically-based structural predictions also diminished in turn?

We conducted two eye-tracking experiments using the visual world paradigm, designed to test how listeners use prosodic boundary placement to parse globally ambiguous PP attachment sites. Experimental materials were identical in each experiment but different filler materials were used between the two experiments. For the experimental items, a structurally ambiguous PP was prosodically disambiguated with either modifier interpretation compatible or instrument interpretation compatible prosody, as shown in example (1) above. In addition, the noun within the PP was pragmatically biased either towards an instrument interpretation (*The boy will see the tiger with the binoculars*) or against it (**The boy will see the tiger with the popcorn*). Experiment 1 addresses the first question of whether listeners predictively resolve PP ambiguities from the location of a prosodic boundary. As a linking hypothesis, we assume that anticipatory eye-movements to different objects in the visual scene reflect anticipatory structural analysis. We also assume that the cost for violating prosodic alignment would be observed when the noun is pragmatically incompatible with the structure, as in the condition that paired instrument interpretation compatible prosody with an implausible instrument noun (*popcorn*). Thus, in the analyses, we concentrate on looks made to a modifier object reflecting predictive modifier analysis, and looks made to an instrument object reflecting predictive instrument analysis, in two time windows, the anticipatory time window (*with the*) and the window containing the PP noun (*binoculars / popcorn*). Results in these time windows are further supported by an additional permutation analysis, which does not assume windows of *a priori* interest.

In the anticipatory time window, we predicted that early boundaries signaling a modifier interpretation would generate anticipatory looks to a plausible modifier object, whereas late boundaries signaling an instrument interpretation would generate anticipatory looks to a plausible instrument object (Experiment 1). In the sentence-final object noun window, we predicted that, if listeners anticipated an instrument interpretation from the location of the prosodic break, but later encountered an implausible instrument noun (*The boy will see the tiger % with the popcorn*), they would exhibit signs of processing difficulty, defined as more looks to a plausible but unmentioned instrument object (e.g. a pair of binoculars), compared to the other condition that carries modifier-compatible prosody and ends with the same NP (*The boy will see % the tiger with the popcorn*).

Experiment 2 addresses the second and third questions: whether listeners adjust their parsing predictions after repeated exposure to unreliable prosodic information, and, if so, whether the processing cost incurred by a mismatch between prosodic expectations and plausibility reduces over the course of the experiment. The same experimental recordings from Experiment 1 were used in this experiment, but filler sentences unrelated to the manipulation were produced with prosodic boundaries that failed to align with syntactic structure, discussed in detail in section 2 below. Specifically, we predicted that the atypical use of prosodic boundaries in the filler items would reduce the predictive value of prosodic information overall, and that participants would make fewer anticipatory eye movements as the experiment progressed. Provided that listeners evaluate and track how informative prosodic cues are, adjusting the degree to which they use prosody during parsing, we predicted that listeners in Experiment 2 would also show reduced penalties in response to prosodic mismatch – i.e. when the object noun was incompatible with the parse expected from the location of the prosodic boundary.

As mentioned above, previous studies have reported anticipatory effect of prosody in online structural ambiguity resolution. However, to the best of our knowledge, no study has yet investigated the influence of prosodic boundary information in anticipatory processing in relation to the processing difficulty listeners would experience when semantic information of the input does not match the expectations generated from the location of a prosodic boundary. Evidence that listeners use prosodic information to anticipate the structure of upcoming speech, and consequently exhibit processing difficulty when they receive input that does not match their expectations, would support models in which prosodic information strongly guides predictive processing. Further, finding that listeners rely on prosodic information less when speakers do not consistently align prosodic boundaries with syntax constituency would suggest that listeners are highly attuned to how individual speakers use prosodic cues, and adjust, perhaps strategically, the extent to which they utilize prosodic information in anticipatory processing.

2. Methods and design of the experiments

2.1. Participants

We recruited 64 native speakers of English with normal visual acuity and hearing for the study.

Participants either received course credit or monetary compensation ($20/hour) for participating. Half of the participants (32 participants) participated in Experiment 1, and the other half of the participants (32 participants) participated in Experiment 2.

2.2. Materials

The current study examines the impact of the location of a prosodic boundary in resolving a structurally ambiguous PP to a modifier interpretation or an instrument interpretation such as example (2). The first factor that was manipulated was the placement of the prosodic boundary, characterized by a continuation rise boundary tone (L-H%), located either after the verb (modifier prosody) or between the noun and the PP (instrument prosody), as shown in Figure 1. To make the large boundary in the middle of the sentence sounded as balanced and natural as possible, the same type of boundary with a continuation rise was added after the subject noun in both conditions. The second factor that was manipulated was the final word of the sentence, which was either plausible (4; *binoculars*) or implausible (5; *popcorn*) as an instrument to the verb (*see*). If listeners generate expectations for upcoming syntactic structure based on the location of prosodic boundary, they should exhibit processing difficulty upon encountering an implausible instrument object paired with Instrument prosody.

The original recordings were spliced up to the onset of the sentence-final prosodic phrase for each type of prosody (e.g. (4a) and (5a) for the modifier prosody, (4b) and (5b) for the instrument prosody). As a result, the modifier prosody sentences were identical until the onset of the second NP (*the tiger with the binoculars/popcorn*), and the instrument prosody sentences were identical until the onset of the ambiguous PP (*with the binoculars/popcorn*).

(4) *Plausible instrument condition*
a. *Modifier prosody:*
The boy $_{L\text{-}H}$% will see $_{L\text{-}H}$ % the tiger with the binoculars.
b. *Instrument prosody:*
The boy $_{L\text{-}H}$% will see the tiger $_{L\text{-}H}$% with the binoculars.

(5) *Implausible instrument condition*
a. *Modifier prosody:*
The boy $_{L\text{-}H}$% will see $_{L\text{-}H}$% the tiger with the popcorn.
b. *Instrument prosody:*
The boy $_{L\text{-}H}$% will see the tiger $_{L\text{-}H}$% with the popcorn.

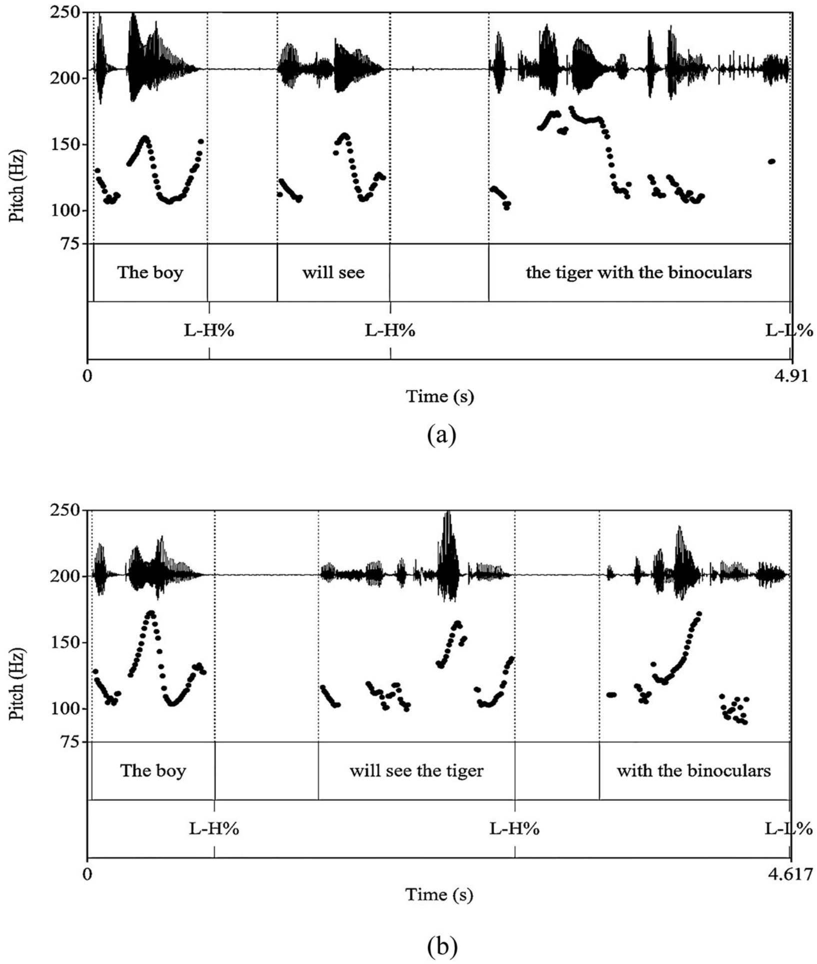


Figure 1. Waveform and pitch track of the sentence in (4), labeled with boundary tones for modifier prosody (4a, top) and Instrument prosody (4b, bottom)

Since the recordings were not spliced across the two types of prosody, the duration of the prosodic boundary after the subject NP was not identical between the modifier prosody (M = 434ms, SD = 164ms) and the instrument prosody (M = 659ms, SD = 256ms).[2] This difference in the duration of the prosodic boundary between the two conditions following the subject NP will be discussed in the results section.

In addition, the filler items were manipulated and presented with the target items in a between-participants design. In Experiment 1, filler items were produced with normal, conventional prosody, such as (6a). In Experiment 2, filler items were produced with various kinds of unconventional prosody, as in example (6b), where a L-L% boundary, usually used sentence or phrase finally, was produced between the determiner and the sentence-final NP. Crucially the target items remained identical between Experiment 1 and 2. Figure 2

(6) a. *Conventional prosody (Experiment 1)*
The boy $_{\text{L-H}}$% will show the balloon $_{\text{L-H}}$% to the dog.
b. *Unconventional prosody (Experiment 2)*
The boy $_{\text{L-H}}$% will show the balloon to the $_{\text{L-L}}$% dog.

Twenty-four experimental items were created, and the same 24 items were used in Experiment 1 and 2. Each item consisted of audio files of the sentence and corresponding visual scenes (Figure 3). The auditory stimuli were recorded by a native speaker of English trained in intonational phonology and English ToBI (Beckman et al., 2005). The average speech rate of our items (3.97 syllables/second) was estimated as the rate of articulation, calculated by the number of syllables divided by the total time minus pausing time (De Jong & Wempe, 2009), and approximates normal speaking tempo (typically around 4 syllables/second, with considerable variation for articulatory content, speech genre, and individuals; see Trouvain, 2004, for review).

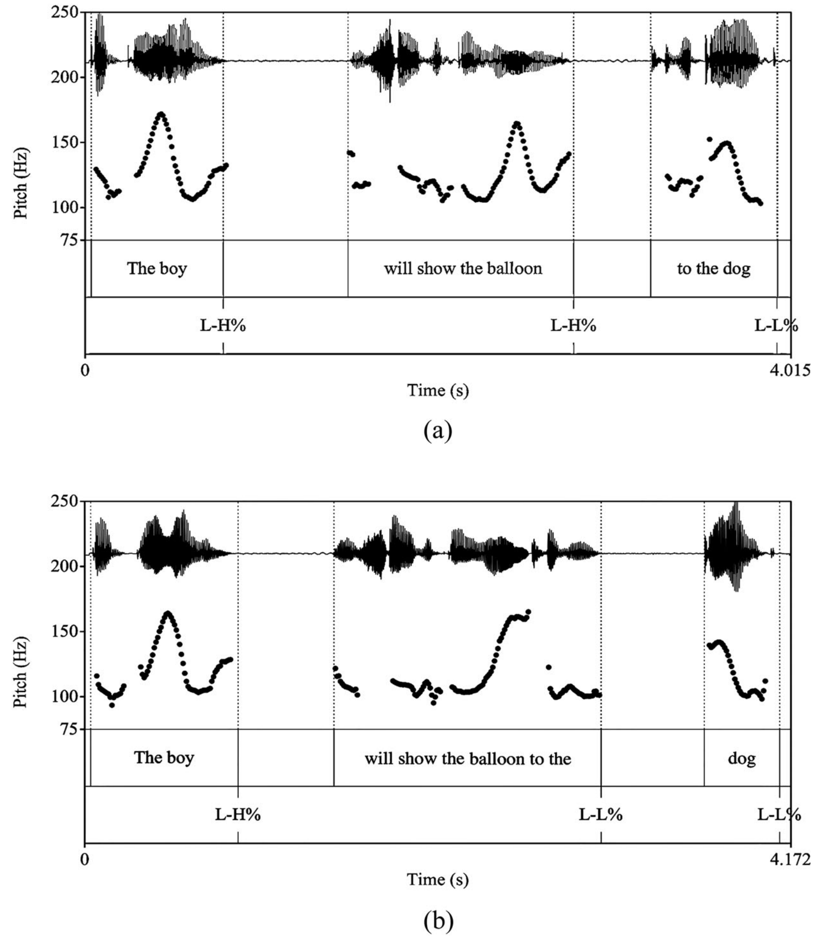


Figure 2. Waveform and pitch tracks of sample filler item illustrating between-participant manipulation in Experiments 1 (conventional prosody, 6a, top) and Experiment 2 (unconventional prosody, 6b, bottom).

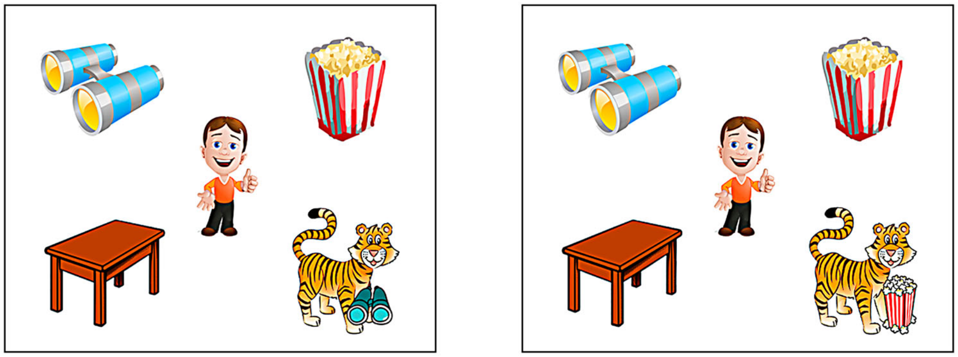
The visual scenes were prepared using clipart images. The position of objects was counter-balanced across the pictures. Four experimental lists were created following a Latin square design including 48 filler sentences unrelated to the structural ambiguity of the target items. The 72 items (24 target and 48 filler sentences) in each list were presented in pseudo-random order. To encourage participants to engage in careful listening, a comprehension question appeared after each trial and participants were asked to click on the correct answer (e.g. "Which of the words was NOT mentioned in the sentence?" followed by a list of possible answers "binoculars", "spider", and "tiger" for (4)).

2.3. Procedure

Participants were told to listen to the sentences carefully while paying attention to the picture on the computer monitor. Auditory stimuli were presented via loud speakers. In each trial, a sentence was presented 2500ms after the picture onset. Eye-movements were recorded via an SR Research EyeLink 1000 Plus tower mount affixed to the table, approximately 55cm away from a 21″ LCD screen. Viewing was binocular, but only the right eye was recorded. Sampling rate was set to 500 Hz. Participants were calibrated to the tracker using 9-point calibration at the beginning of the experiment, and as needed throughout the experiment. A drift correction was performed between each trial. Participants were encouraged to rest whenever they needed a break. Each experimental session took approximately 30 minutes.

2.4. Analysis to be presented

Statistical analyses of eye-movement data from the two experiments were conducted in two different time windows of interest using linear mixed effects regression models (e.g. Baayen et al., 2008). We conducted separate analyses for the looks made to a plausible modifier object (e.g. a tiger with binoculars/popcorn in Figure 3, modifier interpretation compatible) and those made to a plausible instrument object (e.g. a pair of binoculars in Figure 3, instrument interpretation compatible) to



(a) (b)

Figure 3. Visual scenes presented in (a) plausible instrument condition and (b) implausible instrument condition.

examine whether participants predicted different structural resolutions depending on the location of prosodic boundary. The logits of looks to the two target objects were calculated and used as dependent variables in the analyses.

The analyses for the two target objects were first conducted in the "anticipatory time window", defined as the onset of the preposition (*with*) until the onset of the final, critical word (*binoculars/popcorn*), to determine whether participants adopt a structural analysis of the PP based on the location of prosodic boundaries before encountering the critical word. If listeners use the boundary location to anticipate the structural resolution of the PP attachment ambiguity, they should look at the picture most compatible with the expected structure within this window.

Next, we analyzed the looks made to the instrument object for the duration of the sentence final word, which we call the "sentence-final word time window". In the critical case, listeners first hear instrument prosody followed by a noun that is incompatible with an instrument interpretation (*popcorn*), as in (5b). If listeners have anticipated an instrument interpretation of the PP in this case, encountering an incompatible instrument noun within the PP should generate a mismatch between the input and their expectations, delaying looks to the instrument incompatible noun, presumably because mention of the instrument compatible object (e.g. a pair of binoculars) had been expected. In other words, looks made to the instrument object in the sentence-final word time window will be also analyzed to examine whether participants exhibited processing difficulty when the expectation for the upcoming structure does not match the noun within the PP.

Also within the sentence-final word time window, we explored the possibility that participants engage in different strategies over the course of the experiment by analyzing the changes in the patterns of eye-movements to each object in the first and second halves of the experiment. The same analyses for the results of Experiment 2 were conducted to examine whether the degree to which participants in Experiment 2 use prosodic cues in structural analysis is modulated by the reliability of prosody.

In addition to the main analyses using LME models, we further explored the difference in the time course of the effects by employing a non-parametric permutation analysis (Dautriche, Swingley & Christophe, 2015; Von Holzen & Mani, 2012). The permutation analysis allows us to assess when looks to critical objects in the visual scene began to diverge between conditions without any a priori assumption regarding a time window of interest (Groppe, Urbach & Kutas, 2011; Maris, 2012; Maris & Oostenveld, 2007). Thus, we used this analysis to directly compare the results of Experiment 1 and 2 in how much participants rely on prosodic cues in anticipatory structural processing.

3. Data analysis and results

3.1. Experiment 1 (conventional filler prosody)

The fixation coordinates from the eye-tracker were converted into gaze locations corresponding to the five entities and the background in the visual scene. The onset of the ambiguous PP (*with the binoculars/popcorn*) and the onset of the noun within the PP (*binoculars/popcorn*) in each target sentence were determined manually. For the analysis, we summed the looks to each object in the scene and calculated the logit of looks to each entity out of looks to all the objects in the scene, including looks to the background (Barr, 2008).[3]

Statistical analyses with linear mixed effect regression models were conducted for the time window that was time-locked to the onset of the ambiguous PP (*with*) in each item, up to the mean onset of the final noun

(341ms, *anticipatory time window*). In addition to the anticipatory time window analysis, we also analyzed the time window time-locked to the onset of the final noun (*binoculars / popcorn*) in each item until the minimum offset of the sentence (618ms, *sentence-final word time window*) for the looks to the instrument object. This was to examine whether participants exhibited processing difficulty when there is a mismatch between (i) the prediction they might have made from the location of the prosodic boundary and (ii) the actual noun provided in the input, e.g. when the instrument prosody was paired with an implausible-instrument noun.

Linear mixed effect regression models were constructed in decreasing complexity and compared for goodness-of-fit as determined by the Akaike Information Criterion (Akaike, 1973). The initial model consisted of Prosody (modifier or instrument prosody) and Noun Type (plausible or implausible instrument), and their interaction as fixed effects. We also included Trial Block (first or second block of the experiment) in the model to explore whether participants engaged in different processing strategies in the first and second halves of the experiment. There were two steps in the process of selecting the best-fitting model for each analysis; we first determined the fixed effects and then determined the random slope effects. For the selection of the fixed effects, the initial model included the categorical predictors of Prosody, Noun Type, Trial Block, as well as interactions between the factors. We used the AIC value to determine if the model fit improved by eliminating Trial Block. We then determined the random slope effects using a backward selection approach (Bates et al., 2015). Full random effect structures were specified in the initial model, and the AIC value was used in choosing the best-fitting random slope structure. Coefficients, standard errors, *t*-value, and their *p*-values from the best-fitting model are reported. The final model structure is provided in the caption of each table.

3.1.1. Modifier object analysis in the anticipatory time window

For the duration of the anticipatory time window, we predicted that a sentence carrying modifier prosody would entice participants to look at the modifier object more often before encountering the final noun in the PP. Thus, we first analyzed the proportion of looks made to the modifier object (the tiger holding the binoculars/popcorn in Figure 3) in the anticipatory time window (*with the*). Figure 4 shows the proportions of looks to the modifier object from the onset of the ambiguous PP until the sentence offset. The first vertical line marks the onset of the ambiguous PP, the second vertical (dotted) line (341ms) indicates the mean onset of the noun within the PP, and the third vertical (dotted) line (959ms) corresponds to the minimum offset of the sentence. Table 1 shows the results from the best-fitting model. The analysis showed that there was a main effect of Prosody, suggesting that

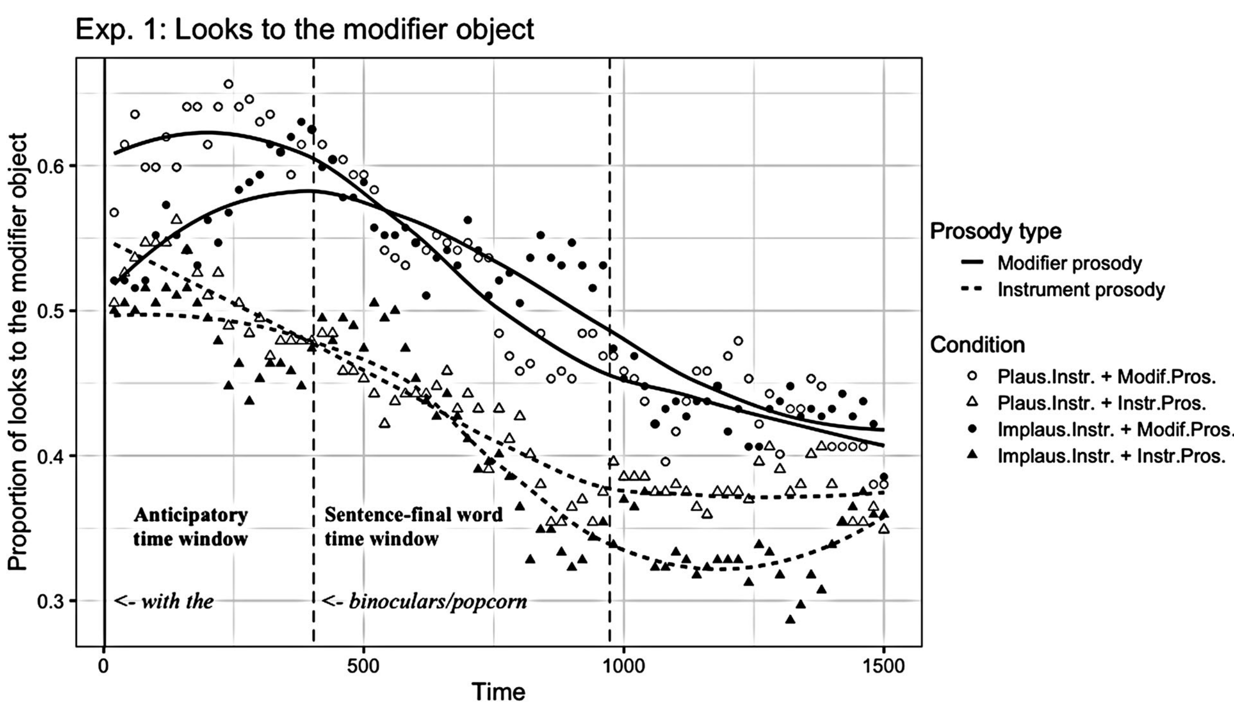


Figure 4. Proportion of looks to the modifier object (Experiment 1), time-locked to the onset of the ambiguous PP.

Table 1. Linear mixed effect model results for the looks to the modifier object in the anticipatory time window in Exp. 1.

	β	*SE*	*t*-value	*p*-estimate
Intercept	2.03	0.36	5.56	<.001
Prosody	−1.34	0.36	−3.68	<.001
Noun Type	−0.29	0.36	−0.79	0.43
Trial Block	−0.07	0.36	−0.20	0.84
Prosody*Noun Type	0.51	0.73	0.70	0.48
Prosody*Trial Block	1.08	0.73	1.48	0.14
Noun Type*Trial Block	0.57	0.73	0.79	0.43
Prosody*Noun Type*Trial Block	0.86	1.45	0.59	0.55

The final model structure: Logit gaze ~ Prosody * Noun Type * Trial Block + (1 | Participant) + (1 | Item).

participants looked at the modifier object significantly more when presented with modifier prosody than with instrument prosody.[4]

3.1.2. Instrument object analysis in the anticipatory time window

We next report the analysis on the looks made to the instrument object (the binoculars in Figure 3) in the anticipatory time window. As with the analysis for the modifier object in the anticipatory window, if participants use prosodic information in anticipatory processing, we expected to find more anticipatory looks to the instrument object when the sentence was produced with instrument prosody. Figure 5 shows the proportion of looks to the instrument object from the onset of the ambiguous PP until the sentence offset. Table 2 summarizes the results from the regression model. The analysis showed there was a main effect of Prosody; participants looked at the instrument object significantly more when it was paired with instrument prosody than with modifier prosody.[5]

Crucially, analyses of looks to (i) the modifier object and (ii) the instrument object in Experiment 1 both support the claim that Prosody influenced processing before the onset of the object noun in the PP. Participants made more anticipatory eye movements to the modifier object (the tiger) when the sentence was presented with modifier prosody, prior to encountering the critical object noun. Similarly, they made more anticipatory eye movements to the instrument object (the binoculars) when the sentence had instrument prosody. The results in the anticipatory time window showed that participants adopted different expectations for upcoming information depending on the location of prosodic boundary.

In addition to the anticipatory time window analyses reported above, we also addressed the possibility that durational differences of the pause following the subject NP between the modifier prosody and the instrument prosody may have influenced the results. If participants used the differences in pause length after the subject NP to anticipate the structure, a difference in looks to each object should appear before the anticipatory time window. To address this possibility, we analyzed the looks to the modifier object and those to the instrument object, time locked to the onset of *will* until the mean offset of the following verb. The results confirmed that there was no effect of Prosody, indicating

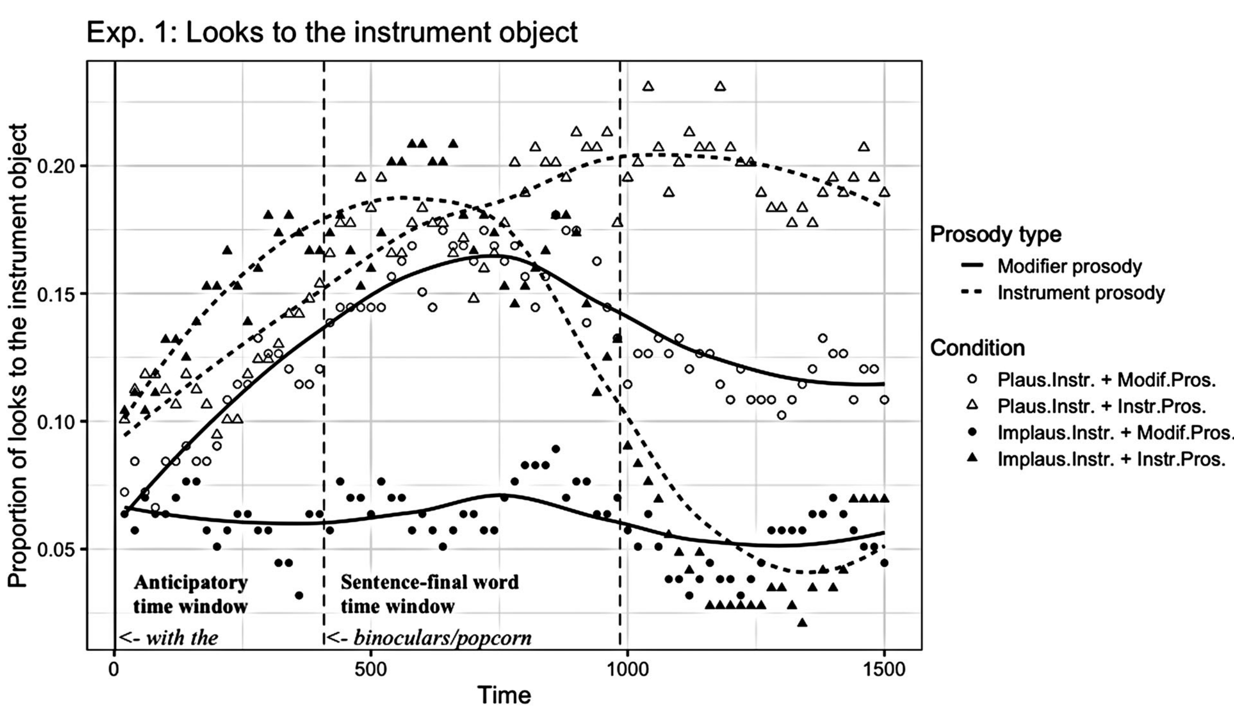


Figure 5. Proportion of looks to the instrument object (Exp. 1), time-locked to the onset of the ambiguous PP.

that differences in the pause duration after the subject NP did not affect participants' interpretation of the sentence.

3.1.3. Instrument object analysis in the sentence-final word time window

For the duration of the sentence-final word time window, we predicted that participants would exhibit processing difficulty upon hearing an implausible instrument noun when the sentence carries instrument prosody, because they had already committed to an instrument construal of the PP from the location of the prosodic boundary. The processing cost due to the mismatch between prosody and semantic input would be reflected in looks to the plausible, but unmentioned, instrument object (i.e. a pair of binoculars in Figure 3) on hearing the sentence-final implausible instrument NP when the sentence carries instrument prosody. Thus, we analyzed the proportion of looks made to the instrument object in the sentence-final word time window. Table 3 summarizes the results from the best-fitting model.

The analysis revealed separate main effects of Prosody and Noun Type. Participants looked at the instrument object more often when they were presented with instrument prosody compared to modifier prosody, regardless of picture type. They also looked at the instrument object more often when two of the same kinds of instrument objects were presented in the visual scene (e.g. two binoculars, Figure 3a), regardless of prosody type. A marginally significant main effect of Trial Block (p=0.06) also suggests that there was a tendency for participants to look at the instrument object more as they progressed through the experiment. Some studies have observed a general preference for instrument interpretations of structurally ambiguous PPs in English (Frazier 1987; Schafer, 1996; Spivey-Knowlton & Sedivy 1995; Rayner, Carlson, & Frazier, 1983; Rayner, Garrod, & Perfetti, 1992; but see also Schütze & Gibson, 1999 and Taraban & McClelland, 1988, for evidence that the relation between the noun and the status of the PP is crucial, and detailed discussion in Schütze, 1995), and our results might reflect that participants were more likely to attach the ambiguous PP to the verb than the NP, at least with these particular materials.

In addition, there was a weak trend towards an interaction between Prosody and Noun Type (p = 0.08). In order to further explore the interaction, we conducted separate post-hoc analyses on the effect of Noun Type in the modifier and instrument prosody conditions, respectively. The results showed that the effect of Prosody was significant only in the implausible instrument condition (p < .001) and not in the plausible instrument condition (p = 0.25). Upon hearing the sentence-final critical noun, participants were slower to look at the implausible instrument object image (the popcorn) when presented with the instrument prosody conditions compared to the modifier prosody conditions. No such difference between the two types of prosody was observed when participants were presented with a plausible instrument noun. The results are highly compatible with the idea that listeners anticipated an instrument object when presented with the instrument prosody, and were later forced to redirect their attention upon hearing an instrument-incompatible noun.

Table 2. Linear mixed effect model results for the looks to the instrument object in the anticipatory time window in Exp. 1.

	β	*SE*	*t*-value	*p*-estimate
Intercept	−4.78	0.21	−22.77	<.001
Prosody	0.67	0.32	2.10	<.05
Noun Type	−0.22	0.28	−0.78	0.44
Prosody*Noun Type	0.62	0.49	1.26	0.21

The final model structure: Logit gaze ~ Prosody * Noun Type + (1 + Prosody + Noun Type | Participant) + (1 | Item).

Table 3. Linear mixed effect model results for the looks to the instrument object for the sentence-final word time window in Exp. 1.

	β	*SE*	*t*-value	*p*-estimate
Intercept	−4.58	0.28	−16.34	<.001
Prosody	0.89	0.34	2.65	<.05
Noun Type	−1.00	0.29	−3.43	<.01
Trial Block	0.87	0.44	2.00	0.06
Prosody*Noun Type	0.88	0.50	1.77	0.08

The final model structure: Logit gaze ~ Prosody * Noun Type + Trial Block + (1 + Prosody + Noun Type | Participant) + (1 | Item).

3.1.4. Discussion

Previous studies have observed that the location of a prosodic boundary strongly influences interpretation, in both offline (Schafer, 1997) and online (Snedeker & Trueswell, 2003) tasks. The results of Experiment 1 support the additional claim that listeners generate online predictions for resolving PP attachment ambiguities on the basis of the location of the prosodic boundary, manifesting as increased looks to an object compatible with the predicted structure. Analyses of the sentence-final time window suggest that participants exhibited processing difficulty when there was a mismatch between the structural prediction and the actual input, reflected in delayed looks to the referent corresponding to the sentence-final noun.

Experiment 1 established that listeners use prosodic information to generate predictions about upcoming material when the alignment between prosodic boundary information and syntactic structure is consistent and conventional. Experiment 2 addresses whether listeners adjust the extent to which they predict the resolution to an ambiguous PP when the alignment is inconsistent and unconventional, achieved by adding a prosodic boundary in an unconventional location only in filler items. We predicted that listeners would adapt to the use of uninformative prosodic boundaries in filler items, reducing how much they use the location of the prosodic boundary to predict a resolution to the PP attachment ambiguity in the experimental items. If confirmed, such results would indicate that not only do listeners attend to speaker variation in prosodic cues, but that they also consider the overall informativity of prosody, potentially reducing the extent to which they engage in anticipatory processing.

3.2. Experiment 2 (unconventional filler prosody)

In Experiment 2, the target items (examples 4-5) were identical to those used in Experiment 1, but Experiment 2 included filler sentences with unusual or unconventional prosody. In the filler items, the boundary location did not align with syntactic constituency structure and the boundary tone imposed an unconventional prosody onto the sentence, rendering prosody an overall less reliable or informative indicator of syntactic phrasing in Experiment 2.

3.2.1. Modifier object analysis in the anticipatory time window

If the unconventional use of prosody for the filler items influences how much listeners rely on prosodic information in parsing, it was expected that participants in Experiment 2 would reduce the extent to which they use prosodic information in anticipating upcoming structure. The same analyses described in Experiment 1 were conducted in Experiment 2. Figure 6 shows the proportion of looks to the modifier object from the onset of the ambiguous PP until the sentence offset, indicating the two time windows used in the analysis. Table 4 summarizes the results from the model used in the analysis of looks to the modifier object in the anticipatory time window. There was a main effect of Prosody, in which participants looked at the modifier object significantly more often when the sentence was produced with the modifier prosody than with the instrument prosody. There was also a main effect of Trial Block, indicating that participants looked more often at the modifier object in the second half of the experiment. This suggests that the more participants encountered filler items whose prosody was not conventional, the more they adopted a modifier interpretation, regardless of prosody type.[6]

3.2.2. Instrument object analysis in the anticipatory time window

We next analyzed the proportion of looks made to the instrument object in the anticipatory time window (*with the*), shown in Figure 7. As with the modifier object analysis in the anticipatory time window, we predicted that participants would rely less on prosodic cues when filler items provided unconventional prosody.

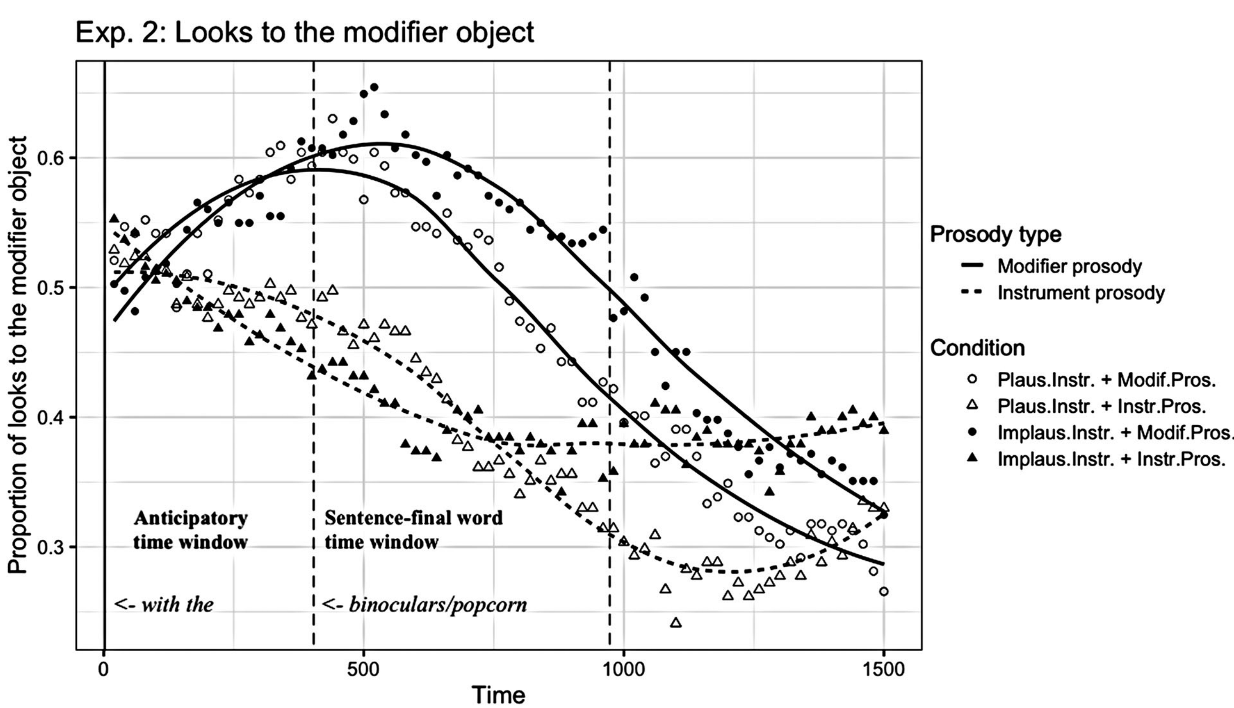


Figure 6. Proportion of looks to the modifier object (Exp. 2), time-locked to the onset of the ambiguous PP.

Table 4. Linear mixed effect model results for the looks to the modifier object in the anticipatory time window in Exp. 2.

	β	*SE*	*t*-value	*p*-estimate
Intercept	1.72	0.39	4.47	<.001
Prosody	−0.95	0.36	−2.64	<.01
Noun Type	−0.21	0.36	−0.60	0.55
Trial Block	1.12	0.450	2.50	<.05
Prosody*Noun Type	0.40	0.72	0.56	0.68
Prosody*Trial Block	0.73	0.72	1.02	0.31
Noun Type*Trial Block	1.42	0.72	1.99	0.47
Prosody*Noun Type*Trial Block	−0.61	1.43	−0.42	0.67

Final model structure: Logit gaze ~ Prosody * Noun Type * Trial Block + (1 | Participant) + (1 | Item).

Table 5 summarizes the results. In contrast to the results observed in Experiment 1, the analysis showed no effect of Prosody, suggesting that participants did not anticipate an instrument noun with Instrument prosody when filler items were presented with unconventional prosody.

The analysis also showed a main effect of Noun Type, in which participants looked more at the instrument object in the plausible-instrument noun condition than in the implausible-instrument noun condition. One possible account of this result is that global viewing strategies were influenced by items in the visual world array. Participants in Experiment 2 looked at the instrument object more when there were two same objects (e.g. two sets of binoculars in the plausible-instrument condition in Figure 3 left) compared to when there was only one instrument object (e.g. one set of binoculars in the implausible instrument condition in Figure 3 right). We interpret this pattern as potentially reflecting a rational strategy adopted over the course of the experiment, in which participants learned to ignore anticipating the sentence completion from prosodic information, relying instead on the objects in the visual array. The effect is not present in Experiment 1; as prosody was always conventional, there would presumably have been no reason to adopt an alternative strategy.[7]

3.2.3. Instrument object analysis in the sentence-final word time window

To address the prediction that the processing cost due to the mismatch between prosody and semantic input would be reduced in Experiment 2, we analyzed looks to the instrument object for the duration of the sentence-final word time window. Table 6 summarizes the results. There was a main effect of Prosody, in which participants looked at the plausible-instrument object more with instrument prosody than with modifier prosody. We also observed a main effect of Noun Type, in which participants looked at the plausible-instrument object more in the plausible-instrument condition than in the implausible-instrument condition. The interaction between Prosody and Trial Block indicates that the effect of Prosody was significant only in the first half of the experiment (β = -0.93, SE = 0.47, t = -1.98, p < .05 in Trial Block 1), but not in the second half of the experiment (β = 0.37, SE = 0.35, t = 1.05, p = 0.30 in Trial Block 2). However, unlike Experiment 1 in which we found a

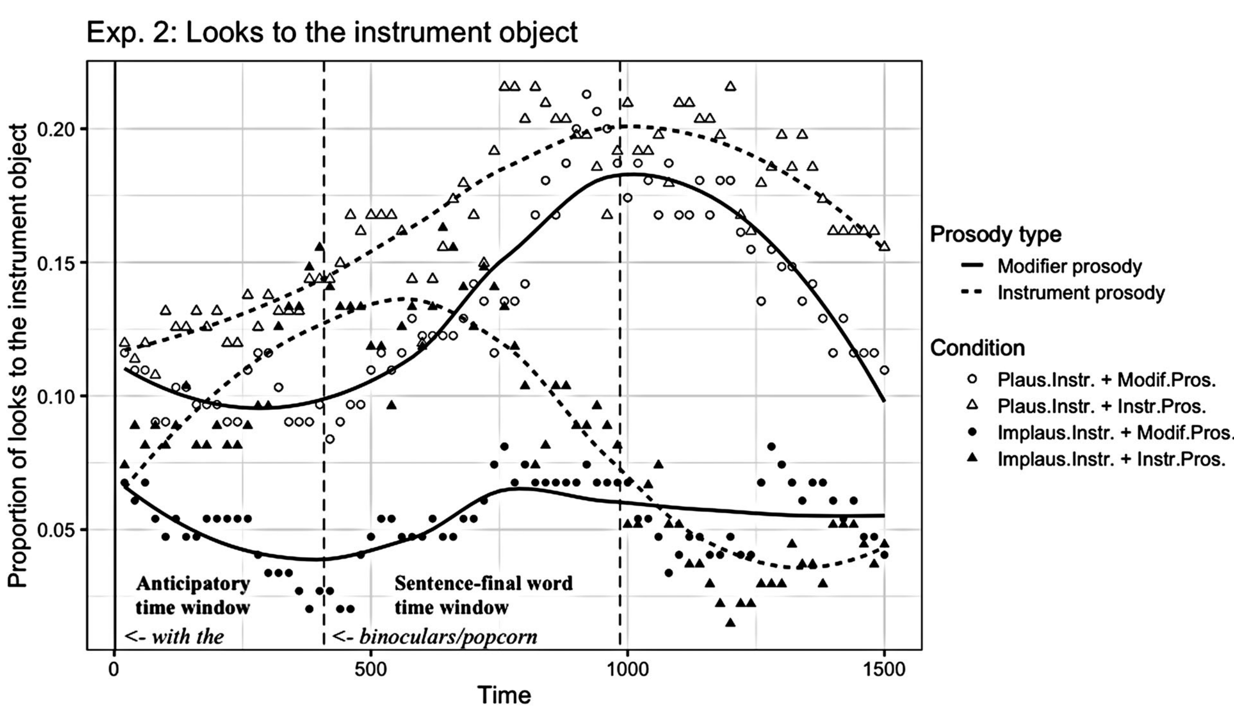


Figure 7. Proportion of looks to Instrument object (Exp. 2), time-locked to the onset of the ambiguous PP.

Table 5. Linear mixed effect model results for the looks to the instrument object in the anticipatory time window in Exp. 2.

	β	*SE*	*t*-value	*p*-estimate
Intercept	-4.95	0.19	-26.11	< .001
Prosody	0.27	0.23	1.16	0.25
Noun Type	-0.74	0.27	-2.70	< .05
Prosody*Noun Type	0.05	0.46	0.10	0.92

Final model structure: Logit gaze ~ Prosody * Noun Type + (1 + Noun Type | Participant) + (1 | Item).

Table 6. Linear mixed effect model results for the looks to the instrument object in the sentence-final word time window in Exp. 2.

	β	*SE*	*t*-value	*p*-estimate
Intercept	-4.93	0.20	-24.71	<.001
Prosody	0.83	0.24	3.58	<.001
Noun Type	-1.42	0.24	-4.91	<.001
Trial Block	0.02	0.31	-0.06	0.95
Prosody*Noun Type	0.02	0.47	0.04	0.97
Prosody*Trial Block	-0.93	0.47	-2.00	<.05
Noun Type*Trial Block	0.05	0.47	0.11	0.91
Prosody*Noun Type*Trial Block	-1.12	0.93	-1.20	0.23

Final model structure: Logit gaze ~ Prosody * Noun Type * Trial Block + (1 + Noun Type | Participant) + (1 | Item).

marginally significant interaction between Prosody and Noun Type, there was no interaction between the two factors in Experiment 2, suggesting that the penalty for mismatching prosody was reduced in Experiment 2. Participants in Experiment 2 may have come to rely less on prosodic information for structural disambiguation when filler items had unconventional prosody. The differences between Experiment 1 and 2 will be explored in greater depth in the combined analysis in Section 3.3, and with the permutation analysis in Section 3.4 below.

3.2.4. Modifier object analysis in the sentence-final word time window

The above analysis showed that there was an interaction between Prosody and Trial Block. If the results reflect the fact that participants in Experiment 2 changed their interpretations of prosodic cues over the course of the experiment, the same pattern of results would be also observed in the looks to the modifier object. To address this interpretation, we analyzed the looks made to the modifier object for the duration of the sentence-final word time window in Experiment 2. The analysis in the sentence-final word time window for the looks to the modifier object revealed that there was an interaction between Prosody and Trial Block ($\beta = 1.87$, $SE = 0.68$, $t = 2.74$, $p < .01$). The planned comparison analyses on the simple effect of Prosody in Trial Block 1 and Trial Block 2 confirmed that the effect of Prosody was greater in Trial Block 1 ($\beta = -3.15$, $SE = 0.48$, $t = -6.62$, $p < .001$) than in Trial Block 2 ($\beta = -1.29$, $SE = 0.47$, $t = -2.71$, $p < .01$). The results suggest that the more participants in Experiment 2 encountered unconventional prosody with filler items, the less they used prosodic information to anticipate the upcoming information in processing target items.

3.2.5. Discussion

The results of the analyses in the anticipatory window of Experiment 2 showed that participants looked more to the modifier object when presented with modifier prosody, but did not appear to anticipate the instrument object in conditions with instrument prosody. The results also indicate that eye-movement patterns changed over the course of the experiment. The results in both time windows showed that the effect of prosody was weaker in the second half of the experiment compared to the first half, suggesting that participants engaged in predictive processing less often as the experiment progressed. The sentence-final word time window analysis also indicated that participants may have committed less often or less strongly to a particular structure, as shown by a reduced penalty for mismatching prosody. Participants in Experiment 2 did not appear to exhibit any strong processing costs upon encountering an implausible instrument object in the instrument prosody condition.

The results of Experiment 2 were different from those of Experiment 1 in that listeners in Experiment 2 adjusted their processing strategies when prosodic boundaries did not consistently align with syntactic constituency. Taken together, the experiments suggest that listeners are highly attuned to the reliability of prosodic information, reducing the extent to which prosodic cues drive anticipatory language processing. In order to more directly address differences between studies, we present a combined analysis of gaze locations upon presentation of the disambiguating noun.

3.3. Combined analysis between Experiment 1 and 2 in the sentence-final word time window

The main analyses reported above revealed that there was an interaction between Prosody and Trial Block in the sentence-final word window only in Experiment 2 but not in Experiment 1. The pattern suggests that only participants in Experiment 2 changed the degree to which they used prosodic cues in structural analysis over the course of the experiment, possibly reflecting that the processing cost due to the mismatch between prosody and the sentence-final noun was reduced in the second half of the experiment. To more directly investigate whether participants showed a greater reduction of anticipatory processing in later trials of

Experiment 2 compared to Experiment 1, we combined the data across the two studies into a between-participant analysis.

Models of looks within the sentence-final time window to the modifier and instrument objects were constructed as before, except that Experiment (Experiment 1 or Experiment 2) was added as a categorical between-participant predictor. As predicted, there was an interaction between Trial Block and Experiment ($\beta = 1.13$, $SE = 0.51$, $t = 2.23$, $p < .05$ for the looks to the modifier object, $\beta = -0.89$, $SE = 0.36$, $t = -2.51$, $p < .05$ for the looks to the instrument object), in which the difference between the two types of prosody for the looks to the target objects was reduced between the first and second halves of Experiment 2 compared to Experiment 1. Figure 8 shows the changes in the proportion of looks to the modifier object and those to the instrument object in the sentence-final word time window throughout Experiment 2. The figure illustrates that the difference between the two types of prosody reduced towards the end of the experiment, and the participants began to favor the modifier object towards the end of the experiment.

3.4. Permutation analysis

The central analysis suggests that the mismatch between the instrument prosody and the implausible-instrument nouns elicited a smaller processing cost in Experiment 2 than in Experiment 1, even though the experimental items were identical. The linear-mixed-effects analysis in separate time-windows, reported above, are standard, well-established approaches to analyzing visual world eye-tracking data (e.g. Altmann & Kamide, 1999; Altmann & Kamide, 2007; Nakamura et al., 2012), especially when there are clear expectations about when the effects are predicted to appear (i.e. in the case of this study, the anticipatory time window and the sentence-final word time window). However, there are other methods for analyzing time course data without any *a priori* assumptions regarding a time window of interest (Groppe, et al., 2011; Maris, 2012; Maris & Oostenveld, 2007). To further explore the differences in the time course of the effects, we employ a non-parametric permutation analysis (Dautriche, et al., 2015; Von Holzen & Mani, 2012). The analysis allows us to better assess when looks to critical objects in the visual scene began to diverge between conditions in the absence of a predefined analysis window.

We conducted the permutation test following the procedure used in Chan et al. (2018). We first created a dataset that included implausible-instrument + modifier prosody and implausible-instrument + instrument prosody conditions in Experiment 1 and 2. Within this dataset, 20ms time bins were created for 1500ms following the onset of the ambiguous PP. We then compared the difference between the two types of prosody (modifier prosody and instrument prosody) for each time bin using the *t*-test. Clusters were created by combining adjacent time bins in which the difference between the two prosody types was significant ($p <$

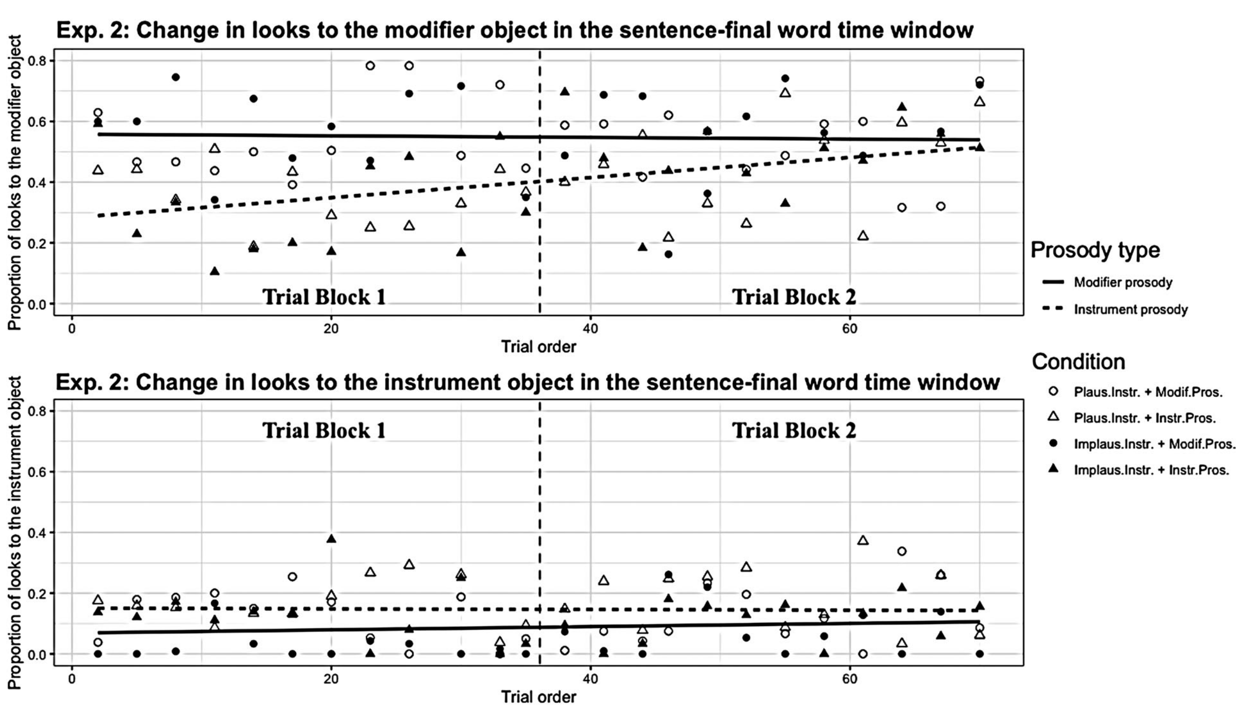


Figure 8. Changes in looks to modifier and instrument objects throughout Experiment 2 in the sentence-final word time window.

.05). For each cluster, we ran a simulation test 1000 times for each experiment by randomly permuting the modifier prosody and the instrument prosody labels. This procedure allowed us to obtain the distribution of t-values without presupposing any particular link between the conditions of the stimuli and the eye-tracking data. We then summed the t-values produced for each time bin within each cluster for each experiment, and constructed a histogram of the simulation data. Finally, we computed p-values by calculating the proportion of the summed t-values in the simulation distribution that were larger than the summed t-values for each of the cluster in the original dataset. If the proportion was less than 0.025, the difference was considered to be significant under a two-tailed t-test.

Figure 9 shows the proportion of looks to the instrument object in the implausible instrument conditions in Experiment 1 and 2, time-locked to the onset of the ambiguous PP of each item (i.e. start of the anticipatory time window). The solid grey line denotes the proportion of looks to the instrument object in the modifier prosody condition, and the dashed black line denotes the proportion of looks to the instrument object in the instrument prosody condition. The time clusters where two types of prosody differ significantly from one another are indicated by grey blocks shown on top of each Experiment panel. The small bars on the bottom of each Experimental panel shows p-values for the difference between the two types of prosody for individual 20ms time bin before the clusters were made. If the bar is extended downward in grey, the p-value is greater than .05. If the bar is extended upward in black, the p-value is significant.

The permutation analysis revealed that the timing of processing difficulty in the implausible instrument + instrument prosody condition differed between the two experiments. In Experiment 1, there was a single time cluster in which the two types of prosody significantly deviated from each other. Participants in Experiment 1 looked significantly more at the instrument object with Instrument prosody than with modifier prosody between 260ms to 740ms (total window time = 480ms, $p < .001$). Considering that the average onset of the sentence-final word falls around 340ms after the onset of the anticipatory time window, it is likely that participants in Experiment 1 started to make anticipatory eye-movements towards the instrument object in Instrument prosody conditions, where they continued to look until hearing the implausible-instrument noun. In Experiment 2, a similar effect was observed in two, shorter time clusters: first from 300ms and 460ms (total window time = 160ms, $p < .001$), and then from 600ms to 660ms (60ms, $p < .05$).

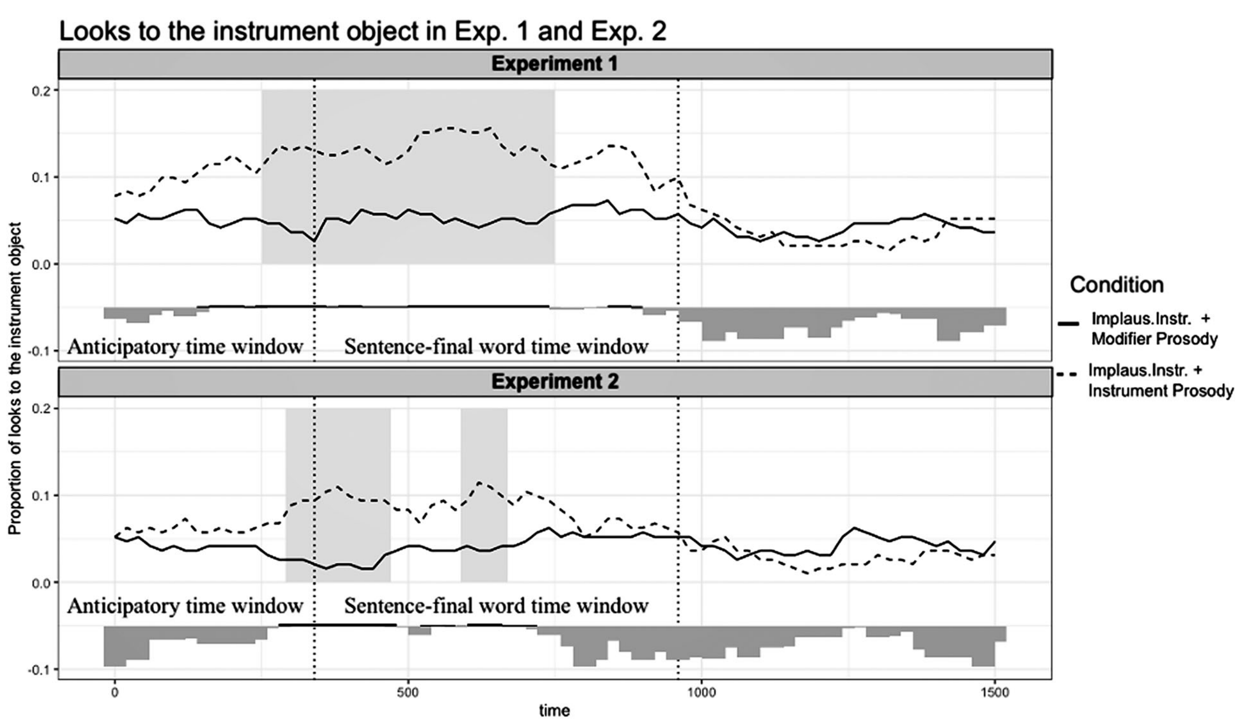


Figure 9. Average proportion of looks to the instrument object in the implausible-instrument conditions in Experiment 1 (top) and Experiment 2 (bottom), time-locked to the onset of the ambiguous PP. The large grey blocks in each panel represent the time clusters when the two different prosody conditions are significantly different by the permutation analysis.

Taken together, the results suggest that participants in Experiment 1 looked at the instrument object earlier and for a longer duration, compared to participants in Experiment 2. The results most likely indicate that Instrument prosody led to stronger expectations for an instrument noun in Experiment 1, and that participants subsequently exhibited greater processing difficulty on an implausible-instrument noun, as a consequence.

4. General discussion

We conducted two visual world paradigm studies manipulating the location of prosodic boundaries to determine how and when listeners use boundary locations in disambiguating PP attachment sites during online auditory sentence comprehension. Experiment 1 showed that gaze location was guided by the location of a prosodic boundary, in support of previous claims that listeners use prosodic boundaries to disambiguate sentence meanings (Carlson et al., 2001; Schafer, 1997; Snedeker & Trueswell, 2003, among many others). In particular, participants looked at the object that was most compatible with the location of the prosodic boundary: modifier objects garnered more looks when the PP was given modifier prosody, and instrument objects received more looks when the PP was presented with instrument prosody. Crucially, these effects were observed *before* participants heard the object noun, further suggesting that participants used cues to prosodic structure to anticipate likely syntactic structure.

Assuming that listeners predicted an attachment site for the PP on the basis of a prosodic boundary location, processing at the sentence final noun was expected to be disrupted when it was pragmatically incompatible with the predicted structure. The results are highly compatible with this prediction. When the sentence ended with an implausible-instrument object, participants were slower to look at the implausible-instrument object on hearing the sentence-final NP in the instrument prosody condition compared to the modifier prosody condition. The evidence suggests that participants expected an instrument interpretation of the PP when presented with instrument prosody; receiving a noun that did not match their expectation (i.e. an implausible instrument noun) elicited a delay in looks to the mentioned target noun. In contrast, nouns in the instrument-plausible condition were also compatible with a modifier construal, and thus did not elicit a processing penalty.

Experiment 2 presented the same experimental materials as Experiment 1, but included filler sentences with unusual or unconventional prosody, rendering prosody a less reliable or informative indicator of syntactic phrasing. If listeners attend to the overall informativity of prosody, participants in Experiment 2 should have been less tempted to use prosodic information to resolve structural ambiguity compared to those in Experiment 1. As expected, boundary locations indicating an instrument interpretation did not result in anticipatory eye movements to the instrument-noun target. However, the reduction of anticipatory processing did not extend to conditions with modifier prosody, which instead appeared to elicit more modifier interpretations of the PP. A post-hoc analysis of trial order revealed that participants looked at the modifier object more often as they progressed in the experiment with both types of prosody. Participants apparently learned to ignore certain kinds of prosodic boundary information the more they were presented with fillers with unconventional boundary locations, opting to look to nouns compatible with modifier interpretations rather than instrument interpretations. The modifier object in the visual array was always a combination of an animal and an object, and participants appeared to have chosen the element that was compatible with both interpretations (a picture of the tiger that has a pair of binoculars can be interpreted both as "the tiger that has the binoculars" and "the tiger was seen by the binoculars").

An alternative interpretation of the asymmetric effect of boundary location on predictive processing in Experiment 2 is that modifier prosody is simply a stronger or more reliable indicator of a PP modifier. In the previous literature, attaching the ambiguous PP to the verb (i.e. the instrument interpretation) has sometimes said to be the preferred interpretation in English (Frazier 1987; Schafer, 1997; Spivey-Knowlton & Sedivy, 1995; Rayner, et al., 1983; Rayner, et al., 1992; but see also Schütze, 1995; Schütze and Gibson, 1999). It has been reported in previous studies on implicit error-based learning accounts that an effect of syntactic priming is larger with non-default, usually dispreferred structures because they generate greater prediction error (Chang et al., 2006; Dell & Chang, 2013; Jaeger & Snider, 2013). Assuming that comprehenders tend to prefer instrument interpretations of ambiguous PPs, it is possible that the influence of prosodic boundary information was stronger with the non-default, modifier interpretation.

In addition, it is also possible that modifier prosody had a stronger influence than instrument prosody because prosodic boundary was located at an earlier point in the sentence stimuli for modifier prosody compared to instrument prosody. Since the prosodic boundary was placed before the patient NP in modifier prosody whereas it was placed after the patient NP in instrument prosody, participants had more time to incorporate boundary location information into anticipatory processing, resulting in a delayed effect for modifier prosody.

Further, the lack of anticipatory commitment to the instrument interpretation in Experiment 2 was also reflected in the results in the sentence-final word time window. Unlike the results of Experiment 1, no interaction between Prosody and Noun Type was observed in Experiment 2. This finding supports the claim that participants in Experiment 2 did not generate as strong of an expectation for an instrument noun when sentences had instrument prosody, thus showing reduced processing costs upon hearing an implausible instrument noun compared to participants in Experiment 1. The combined analysis of Experiment 1 and 2, as well as the permutation analysis, support this interpretation.

The findings are highly compatible with the claim that listeners use prosodic information to anticipate the structure of upcoming speech, and consequently experience processing difficulty when they receive input that does not match their expectations. Listeners relied less on prosodic information when prosodic boundaries did not consistently align with syntactic constituency. As suggested by many previous studies, listeners were found to be highly attuned to how a particular speaker used prosodic information, and adjusted the extent to which they utilized this information to resolve structural ambiguity.

Most previous research on adaptation has examined whether exposure to specific contexts can shift judgments towards another possible interpretation, such as categorization shift in the perception of phonetic and prosodic information (Norris et al., 2016; Kurumada, Brown, & Tanenhaus, 2018), or, as with our study, how structural ambiguity is resolved (e.g. high and low attachment structure in Kamide, 2012, or main verb and relative clause structure in Fine et al., 2013). However, in these studies, participants may have learned the unconventional use of linguistic information as a speaker-specific property, whereas the present study manipulated the overall reliability of the information. In other words, whereas previous results can be explained by a rapid re-mapping between a linguistic form and its category or meaning, our findings suggest that listeners evaluate the overall reliability of prosodic information and adjust the extent to which prosody serves as an indicator of upcoming structure.

We now consider the implications of our results on sentence processing models that account for linguistic adaptation. In recent years, highly influential approaches have cast the adaptive nature of the language processing system within a framework of hierarchical generative models to explain how language users successfully accommodate linguistic variation among speakers (Clark, 2013). These accounts assume that language users generate top-down predictions from probabilistic models derived from bottom-up cues, models which are continuously updated as additional input is received. As discussed in Introduction, Bayesian belief-updating variants of this general model propose that adaptation is driven by attempting to minimize the prediction error – i.e. the mismatch between prediction and observation. More specifically, the error-based implicit learning account suggests that when comprehenders receive information that is different from what was expected, an error signal is generated. This error signal is used to update probabilistic distributions of how linguistic cues are used, allowing comprehenders to adapt to the statistics of the input (e.g. Chang et al., 2006; Foltz, 2020; Jaeger & Snider, 2013; Kleinschmidt & Jaeger, 2015; Qian, Jaeger, & Aslin, 2012; Tanenhaus, et al., 2015; Roettger & Franke, 2019). Such an approach could plausibly account for prosodic adaptation, as reliance on prosodic information in resolving structural ambiguity is reduced in order to minimize the error generated by a mismatch between prosody-based expectations and the input the listener actually receives.

Foltz (2021) used a Bayesian belief-updating model to account for previous findings in which L1 German and L2 English listeners appear to adapt to infelicitous contrastive L+H* accent (e.g. *Click on the red duck. Click on the GREEN*$_{L+H*}$ *banana*) in German (Foltz, 2020). The use of contrastive accent on the adjective *green* would normally lead listeners to expect a color contrast within the same noun category (*duck*), instead of a new noun (e.g. Ito & Speer, 2006; Sudhoff, 2010). She argued that the extent to which listeners rely on contrastive L+H* accent in predictive processing is relative to the extent to which the accent is supported as a valid, predictive cue within the experiment. She showed how a Bayesian adaptation model could account for inhibition of an unexpected referent as the informativity of contrastive accent progressively decreased over the course of the experiment in response to prediction error. Inhibition of an unexpected target word was also observed after native German listeners were exposed to an intervening training phase with only felicitous pitch accent from the same speaker, suggesting that listeners are able to rapidly adjust their expectations given relatively local statistics of the data (see also Nakamura et al., 2019).

Foltz's (2021) findings may also be consistent with the distinction between predictive pre-activation and predictive pre-updating proposed in some recent theories of language comprehension (Lau, Holcomb, Kuperberg, 2013; Ness & Meltzer-Asscher, 2018). Predictive pre-activation occurs when comprehenders activate linguistic representations before they receive an actual input (e.g. prediction of a repeated noun on hearing the contrastive L+H* in the above example), and predictive pre-updating generates even stronger commitment to the predicted material (e.g. inhibition of unpredicted referents in the above example). The results of the present study are also consistent with this view in the sense that the degree of predictive structural commitment listeners made based on prosody may have been relative to the validity of prosodic boundary information as a predictive cue.

In our study, the experimental materials were identical in the two experiments, which differed only in whether the filler items contained conventional or unconventional uses of prosody in ways that were orthogonal to the experimental materials. If comprehenders do not track fine-grained usage statistics regarding how prosody aligns with structure, then the unconventional prosody of filler items could, in principle, have had little to no effect on how listeners resolve the structural PP attachment ambiguities in the experimental items. However, the prosody of the filler items disrupted anticipatory processing in the experimental items, as the effect of boundary location was reduced in Experiment 2 compared to Experiment 1. It seems likely that listeners in Experiment 2 learned that prosodic boundary information, in general, had a limited utility for predicting upcoming structure.

It has been suggested from previous studies that listeners continuously model many aspects of particular speakers (e.g. Clark & Marshall, 1978) and rapidly adjust their assessment of potential cues in the signal such as foreign accents (Bradlow & Bent, 2008; Clarke & Garrett, 2004; Clarke & Luce, 2005) and emotional states of the speaker (Sobin & Alpert, 1999; Williams & Stevens, 1972; Chin & Pisoni, 1997; Johnson, Pisoni, & Bernacki, 1990; Pisoni and Martin, 1989). These studies show that listeners use information about the source of speech production to accommodate non-standard speech, and are seemingly biased towards understanding variation in the speech signal as a characteristic which is specific to the speaker. For example, listeners interpret a non-standard production of /s/ as a speaker-specific characteristic, unless they are shown or told that the speaker has a pen in her mouth while speaking (Kraljic et al, 2008; see also Kraljic and Samuel, 2011). Subsequent studies have proposed that experiences of atypical or ambiguous percepts may be maintained within the perceptual system, even when no causal explanation of the non-standard segment is explicitly provided (Liu & Jaeger, 2018). Prior visual world studies further show that explicitly informing participants that the speaker has reduced verbal fluency (Arnold et al., 2007) or social ability (Grodner & Sedivy, 2011) reduces discourse and pragmatic inferences. Even linguistic indicators of age, gender, and socio-economic class rapidly generate early stereotypical inferences about a speaker's likely viewpoint, yielding electrophysiological disruption when the message clashes with the stereotype (e.g. Van Berkum, 2009; Van Berkum et al., 2008).

Our results contribute to this literature by noting that listeners can adapt to speaker-specific productions even when deviance in speech production is not explained by properties or characteristics that could be attributed to the speaker or to the speech event. In other words, listeners use various sources of information in speech processing including consistency/inconsistency in the use of particular information. This in turn suggests that Experiment 2 may have yielded different results if prosody was used to signal unconventional meaning consistently throughout the experiment and participants were able to explicitly provided with a reason for the use of unconventional prosody.

A recent visual world study tested this prediction. Nakamura, Harris, & Jun (2019) showed that listeners anticipate referents on the basis of contrastive L+H* accent when a speaker consistently uses it in conventional ways, i.e. to mark contrastive focus in English. Once the use of contrastive accent deviated from convention, listeners failed to generate anticipatory looks to the referent, even though the speaker's use remained highly systematic (their Experiment 2). However, importantly, they further demonstrated that when listeners were told that the speaker would try to trick them, listeners quickly learned the unconventional use of contrastive accent, even anticipating which referent would be mentioned next (their Experiment 3).

In all, the results support models in which listeners rapidly adjust to speaker-specific patterns in production, and that they may do so with or without any explicit explanation for the variation. However, assuming that anticipatory processing is a potentially risky processing strategy, listeners may be less willing to engage in language prediction when not only is speaker variation high, but the source of speaker variation remains unexplained.

5. Conclusions

Two visual world studies investigated the role of prosodic boundary placement in anticipatory language processing. The results suggest two major conclusions. First, prosody drives structural prediction prior to disambiguating information, lending support to models of language comprehension in which multiple sources of information, including prosody, generate predictions about upcoming speech. Second, we argued that the degree to which prosodic cues were used in structural analysis was modulated by the listeners' evaluation of the overall informativity of how prosodic cues are used by the speaker. Listeners seemed to have adapted to the reduced utility of prosodic boundary information when prosodic and syntactic information were not consistently aligned. In general, the results support the view that the predictive mechanisms in language comprehension are finely tuned not only to speaker-specific variation, but also to the presumed quality of information driving predictive processing. If erroneous predictions pose additional processing risks for comprehension, listeners may strategically avoid engaging in predictive processes without having sufficient confidence in the source of speaker variation.

Notes

1. In Schafer (1997), the prosodic boundary was an Intermediate Phrase (ip) boundary, not an Intonational Phrase (IP) boundary. An IP is a larger prosodic unit than an ip. Both types of prosodic boundary are marked by phrase-final lengthening and a boundary tone, but the degree of final lengthening is weaker in an ip than an IP and the boundary tone contour is simpler in ip than IP. Also, a pause can follow an IP boundary, but not an ip boundary (Beckman & Pierrehumbert, 1986).
2. Acoustic analysis confirmed that the duration of the pause after the subject NP was significantly longer in the instrument prosody than in the modifier prosody ($p<0.001$).
3. The empirical logit of looks to the modifier object was calculated using the function $\eta' = \ln\left(\frac{y+0.5}{n-y+0.5}\right)$ where y is the number of looks to the modifier object and n is the total number of looks to all the objects in the scene including background (Barr, 2008). The same calculation was used in the analyses of the looks to the instrument object.
4. In order to verify that the effects found in the anticipatory time window reflect anticipatory processing, we conducted the same analysis by time-locking to the onset of the sentence-final noun in each item and going back to the mean onset of the ambiguous PP (henceforth, backward analysis). This analysis ensured that the time window of analysis included only eye-movement patterns that were made prior to the onset of the sentence-final noun in all items. The results of the backward analysis provided different levels of significance, but found similar results except for the analysis in looks to the modifier object in the anticipatory time window in Experiment 2. See the footnote in each section reporting the results of the backward analysis.
5. We again conducted the backward analysis for the looks to the instrument object by time-locking to the onset of the sentence-final noun in each item and going back to the mean onset of the ambiguous PP. The backward analysis showed the same results as those reported in the main text.
6. We also conducted a backward analysis for the looks to the modifier object in Experiment 2 by time-locking to the onset of the sentence-final noun in each item and going back to the mean onset of the ambiguous PP. In addition to the effects reported in the main text, there was an interaction between Prosody and Trial Block ($p < .001$). A separate analysis on the simple effect of Prosody in Trial Block 1 and Trial Block 2 showed that the difference between the two types of prosody was significant only in Trial Block 1 ($p < .001$) and there was no significant effect of Prosody in Trial Block 2 ($p = 0.11$). The results are consistent with the findings reported in the sentence-final word time window.
7. The backward analysis for the looks to the instrument object in Experiment 2 showed the same results as those reported in the main text.

Acknowledgements

This word was supported by JSPS KAKENHI Grant-in-Aid for Scientific Research (B) Grant Number 19H01279. Portions of the research were previously presented at the 31st Annual CUNY Conference on Human Sentence Processing in March 2018, and at the 2nd California Meeting on Psycholinguistics (CAMP) in December 2018. We thank Stephanie Rich for her assistance as lab manager and Lalitha Balachandran, Katie Barnett, Anthony Chadwick, Janice Chen, Arjun Gananathan, Joonhwa Kim, Durgesh Rajandiran, Delaney Warren, Chenchen Wang, Pranav Singh, and Livia Witteveen for their help administering the experiments in the UCLA Language Processing Lab. We also thank Franklin Chang for his help with the permutation analysis.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by JSPS (JSPS KAKENHI Grant Number 19H01279).

ORCID

Chie Nakamura http://orcid.org/0000-0003-4534-7385

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov, & B. F. Csaki (Eds.), *Second International symposium on information theory* (pp. 267–281). Academiai Kiado.

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264. doi:10.1016/S0010-0277(99)00059-1

Altmann, G. T. M., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, *57*(4), 502–518. doi:10.1016/j.jml.2006.12.004

Arnold, J. E., Kam, C. L. H., & Tanenhaus, M. K. (2007). If you say thee uh you are describing something hard: The on-line attribution of disfluency during reference comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(5), 914–930. doi:10.1037/0278-7393.33.5.914

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. doi:10.1016/j.jml.2007.12.005

Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, *59*(4), 457–474. doi:10.1016/j.jml.2007.09.002

Bates, D. M., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *arXiv: Methodology*. arXiv:1506.04967.

Beach, C. M. (1991). The interpretation of prosodic patterns at points of syntactic structure ambiguity: Evidence for cue trading relations. *Journal of Memory and Language*, *30*(6), 644–663. doi:10.1016/0749-596X(91)90030-N

Beckman, M. (1996). The parsing of prosody. *Language and Cognitive Processes*, *11*(1/2), 17–68. doi:10.1080/016909696387213

Beckman, M., Hirschberg, J., & Shattuck-Hufnagel, S. (2005). The original ToBI system and the evolution of the ToBI framework. In S.-A. Jun (Ed.), *Prosodic Typology: The Phonology of Intonation and phrasing* (pp. 9–54). Oxford Univ. Press.

Beckman, M., & Pierrehumbert, J. (1986). Intonational structure in Japanese and English. *Phonology Yearbook*, *3*, 255–309. doi:10.1017/S095267570000066X

Bögels, S., Schriefers, H., Vonk, W., & Chwilla, D. (2011). Prosodic breaks in sentence processing investigated by event-related potentials. *Language and Linguistics Compass*, *5*(7), 424–440. doi:10.1111/j.1749-818X.2011.00291.x

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, *106*(2), 707–729. doi:10.1016/j.cognition.2007.04.005

Breen, M., Watson, D. G., & Gibson, E. (2011). Intonational phrasing is constrained by meaning, not balance. *Language and Cognitive Processes*, *26*(10), 1532–1562. doi:10.1080/01690965.2010.508878

Brennan, S. E., & Hanna, J. E. (2009). Partner-specific adaptation in dialog. *Topics in Cognitive Science*, *1*(2), 274–291. doi:10.1111/j.1756-8765.2009.01019.x

Carlson, K., Clifton, C.Jr., & Frazier, L. (2001). Prosodic boundaries in adjunct attachment. *Journal of Memory and Language*, *45*(1), 58–81. doi:10.1006/jmla.2000.2762

Chan, A., Yang, W., Chang, F., & Kidd, E. (2018). Four-year-old cantonese-speaking children's online processing of relative clauses: A permutation analysis. *Journal of Child Language*, *45*(1), 174–203. doi:10.1017/S0305000917000198

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, *113*(2), 234–272. doi:10.1037/0033-295X.113.2.234

Chin, S. B., & Pisoni, D. B. (1997). *Alcohol and speech*. Academic Press.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. doi:10.1017/S0140525X12000477

Clark, H. H., & Marshall, C. R. (1978). Reference diaries. In D. L. Waltz (Ed.), *Theoretical issues in natural language processing* (pp. 57–63). Association for Computing Machinery.

Clarke, C., & Luce, P. (2005). Perceptual adaptation to speaker characteristics: VOT boundaries in stop voicing categorization. Paper presented at the *ISCA Workshop on Plasticity in Speech Perception*.

Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *Journal of the Acoustical Society of America*, *116*(6), 3647–3658. doi:10.1121/1.1815131

Clayards, M., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition*, *108*(3), 804–809. doi:10.1016/j.cognition.2008.04.004

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. doi:10.1177/002383099704000203

Dahan, D., Tanenhaus, M. K., & Chambers, C. G. (2002). Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language*, *47*(2), 292–314. doi:10.1016/S0749-596X(02)00001-3

Dautriche, I., Swingley, D., & Christophe, A. (2015). Learning novel phonological neighbors: Syntactic category matters. *Cognition*, *143*, 77–86. doi:10.1016/j.cognition.2015.06.003

De Jong, N. H., & Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods*, *41*(2), 385–390. doi:10.3758/BRM.41.2.385

Dell, G. S., & Chang, F. (2013). The P-chain: Relating sentence production and its disorders to comprehension and acquisition. Philosophical transactions of the royal Society of London. Series B. *Biological Sciences*, *369*(1634), 20120394. doi:10.1098/rstb.2012.0394

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

Fine, A. B., Jaeger, T. F., Farmer, T. A., & Quan, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PloS ONE*, *8*(10), e77661. doi:10.1371/journal.pone.0077661

Foltz, A. (2020). Using prosody to predict upcoming referents in the L1 and the L2: The role of recent exposure. *Studies in Second Language Acquisition*, *43*, 1–28. doi:10.1017/S0272263120000509

Foltz, A. (2021). Adaptation in predictive prosodic processing in bilinguals. *Frontiers in Psychology*, *12*, 661236. doi:10.3389/fpsyg.2021.661236

Frazier, L. (1987). Theories of sentence processing. In J. L. Garfield (Ed.), *Modularity in knowledge representation and natural-language understanding* (pp. 291–307). The MIT Press.

Grodner, D., & Sedivy, J. C. (2011). The effect of speaker-specific information on pragmatic inferences. In E. Gibson, & N. J. Pearlmutter (Eds.), *The processing and acquisition of reference* (pp. 239–272). The MIT Press.

Groppe, D. M., Urbach, T. P., & Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields II: Simulation studies. *Psychophysiology*, *48*(12), 1726–1737. doi:10.1111/j.1469-8986.2011.01272.x

Hayes, B. (1989). The prosodic hierarchy in meter. In P. Kiparsky, & G. Youmans (Eds.), *Rhythm and meter* (pp. 201–260). Academic Press.

Ito, K., & Speer, S. R. (2006). Using interactive tasks to elicit natural dialogue. In S. Sudhoff, D. Lenertova, R. Meyer, P.

Augurzky, S. Pappert, & I. Mleinek (Eds.), *Language, context, and cognition: Vol. 3. Methods in empirical prosody research* (pp. 229–257). Walter de Gruyter.

Ito, K., & Speer, S. R. (2008). Anticipatory effects of intonation: Eye movements during instructed visual search. *Journal of Memory and Language*, *58*(2), 541–573. doi:10.1016/j.jml.2007.06.013

Jaeger, T. F., & Snider, N. E. (2013). Alignment as a consequence of expectation adaptation: Syntactic priming is affected by the prime's prediction error given both prior and recent experience. *Cognition*, *127*(1), 57–83. doi:10.1016/j.cognition.2012.10.013

Johnson, K., Pisoni, D. B., & Bernacki, R. H. (1990). Do voice recordings reveal whether a person is intoxicated? A case study. *Phonetica*, *47*(3-4), 215–237. doi:10.1159/000261863

Jun, S.-A. (1998). The accentual phrase in the Korean prosodic hierarchy. *Phonology*, *15*(2), 189–226. doi:10.1017/S0952675798003571

Jun, S.-A. (2005). Prosodic typology. In S.-A. Jun (Ed.), *Prosodic Typology: The Phonology of Intonation and phrasing* (pp. 430–458). Oxford University Press.

Jun, S.-A. (2014). Prosodic Typology: by prominence type, word prosody, and macro-rhythm. In S.-A. Jun (Ed.), *Prosodic Typology II: The Phonology of Intonation and phrasing* (pp. 520–539). Oxford University Press.

Kamide, Y. (2012). Learning individual talkers' structural preferences. *Cognition*, *124*(1), 66–71. doi:10.1016/j.cognition.2012.03.001

Kaschak, M., & Glenberg, A. (2004). This construction needs learned. *Journal of Experimental Psychology: General*, *133*(3), 450–467. doi:10.1037/0096-3445.133.3.450

Kaschak, M. P., & Borreggine, K. L. (2008). Is long-term structural priming affected by patterns of experience with individual verbs? *Journal of Memory and Language*, *58*(3), 862–878. doi:10.1016/j.jml.2006.12.002

Kjelgaard, M. M., & Speer, S. R. (1999). Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language*, *40*(2), 153–194. doi:10.1006/jmla.1998.2620

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), 148–203. doi:10.1037/a0038695

Kraljic, T., & Samuel, A. G. (2011). Perceptual learning evidence for contextually-specific representations. *Cognition*, *121*(3), 459–465. doi:10.1016/j.cognition.2011.08.015

Kraljic, T., Samuel, A. G., & Brennan, S. E. (2008). First impressions and last resorts how listeners adjust to speaker variability. *Psychological Science*, *19*(4), 332–338. doi:10.1111/j.1467-9280.2008.02090.x

Kurumada, C., Brown, M., Bibyk, S., Pontillo, D. F., & Tanenhaus, M. K. (2014). Is it or isn't it: Listeners make rapid use of prosody to infer speaker meanings. *Cognition*, *133*(2), 335–342. doi:10.1016/j.cognition.2014.05.017

Kurumada, C., Brown, M., & Tanenhaus, M. K. (2018). Effects of distributional information on categorization of prosodic contours. *Psychonomic Bulletin & Review*, *25*(3), 1153–1160. doi:10.3758/s13423-017-1332-6

Ladd, D. R. (1996/2008). *Intonational phonology*. Cambridge University Press.

Lau, E. F., Holcomb, P. J., & Kuperberg, G. R. (2013). Dissociating N400 effects of prediction from association in single-word contexts. *Journal of Cognitive Neuroscience*, *25*(3), 484–502. doi:10.1162/jocn_a_00328

Liu, L., & Jaeger, T. F. (2018). Inferring causes during speech perception. *Cognition*, *174*, 55–70. doi:10.1016/j.cognition.2018.01.003

Maris, E. (2012). Statistical testing in electrophysiological studies. *Psychophysiology*, *49*(4), 549–565. doi:10.1111/j.1469-8986.2011.01320.x

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. doi:10.1016/j.jneumeth.2007.03.024

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. doi:10.1080/03640210802035357

Nakamura, C., Arai, M., & Mazuka, R. (2012). Immediate use of prosody and context in predicting a syntactic structure. *Cognition*, *125*(2), 317–323. doi:10.1016/j.cognition.2012.07.016

Nakamura, C., Harris, J. A., & Jun, S.-A. (2019). Listener's belief influences prosodic adaptation: Anticipatory use of contrastive accent during visual search. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of International congress of phonetic sciences, melbourne 2019* (pp. 447–451). Australasian Speech Science and Technology Association Inc.

Nakamura, C., Harris, J. A., & Jun, S.-A. (2021). Adaptation to atypical prosody: The L2 advantage. In D. Danielle, & V. C. L22-Ann (Eds.), *Proceedings of the 45th Boston University Conference on language development* (pp. 568–583). Cascadilla Press.

Nespor, M., & Vogel, I. (1986/2007). *Prosodic phonology*. Foris Publications.

Ness, T., & Meltzer-Asscher, A. (2018). Predictive pre-updating and working memory capacity: Evidence from event-related potentials. *Journal of Cognitive Neuroscience*, *30*(12), 1916–1938. doi:10.1162/jocn_a_01322

Norris, D., McQueen, J., & Cutler, A. (2016). Prediction, Bayesian inference and feedback in speech recognition. *Language, Cognition and Neuroscience*, *31*(1), 4–18. doi:10.1080/23273798.2015.1081703

Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences*, *11*(3), 105–110. doi:10.1016/j.tics.2006.12.002

Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. Cohen, J. Morgan, & M. Pollack (Eds.), *Intentions in communication* (pp. 271–311). MIT Press.

Pisoni, D. B., & Martin, C. S. (1989). Effects of alcohol on the acoustic-phonetic properties of speech: *Perceptual and acoustic analyses. Alcoholism, Clinical and Experimental Research*, *13*(4), 577–587. doi:10.1111/j.1530-0277.1989.tb00381.x

Price, P., Ostendorf, M., Shattuck-Hufnagel, S., & Fong, C. (1991). The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America*, *90*(6), 2956–2970. doi:10.1121/1.401770

Qian, T., Jaeger, T. F., & Aslin, R. N. (2012). Learning to represent a multi-context environment: More than detecting changes. *Frontiers in Psychology*, *228*, 1–20. doi:10.3389/fpsyg.2012.00228

Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning & Verbal Behavior*, *22*(3), 358–374. doi:10.1016/S0022-5371(83)90236-0

Rayner, K., Garrod, S., & Perfetti, C. A. (1992). Discourse influences during parsing are delayed. *Cognition*, *45*(2), 109–139. doi:10.1016/0010-0277(92)90026-E

Roettger, T. B., & Franke, M. (2019). Evidential strength of intonational cues and rational adaptation to (un-)reliable intonation. *Cognitive Science*, *43*(7), e12745. doi:10.1111/cogs.12745

Roettger, T. B., & Rimland, K. (2020). Listeners' adaptation to unreliable intonation is speaker-sensitive. *Cognition*, *204*, e104372. doi:10.1016/j.cognition.2020.104372

Ryskin, R., Shukhan, N., Mimnaugh, K., Brown-Schmidt, S., & Federmeier, K. D. (2020). Talker-specific predictions during language processing. *Language, Cognition and Neuroscience*, *35*(6), 797–812. doi:10.1080/23273798.2019.1630654

Schafer, A., Carter, J., Clifton, Jr. C., & Frazier, L. (1996). Focus in relative clause construal. *Language and Cognitive Processes*, *11*(1-2), 135–164. doi:10.1080/016909696387240

Schafer, A., Speer, S., Warren, P., & White, S. (2000). Intonational disambiguation in sentence production and comprehension. *Journal of Psycholinguistic Research*, *29*(2), 169–182. doi:10.1023/A:1005192911512

Schafer, A. J. (1997). *Prosodic parsing: The role of prosody in sentence comprehension*. University of Massachusetts Doctoral Dissertation.

Schuster, S., & Degen, J. (2020). I know what you're probably going to say: Listener adaptation to variable use of uncertainty expressions. *Cognition*, *203*, 104285. doi:10.1016/j.cognition.2020.104285

Schütze, C. T. (1995). PP attachment and argumenthood. In T. S. Carson, J. B. Ganger, & K. Broihier (Eds.), *Papers on language processing and acquisition* (pp. 95–152). 26 MIT Working Papers in Linguistics.

Schütze, C. T., & Gibson, E. (1999). Argumenthood and English prepositional phrase attachment. *Journal of Memory and Language*, *40*(3), 409–443. doi:10.1006/jmla.1998.2619

Selkirk, E. (1986). On derived domains in sentence phonology. *Phonology Yearbook*, *3*, 371–405. doi:10.1017/S0952675700000695

Selkirk, E. (2000). The interaction of constraints on prosodic phrasing. In M. Horne (Ed.), *Prosody: Theory and experiment* (pp. 231–261). Kluwer Academic Publishers.

Selkirk, E. (2007). Focus prominence drives focus phrasing: Bengali intonation revisited. In C.-m. Lee, M. Gordon, & D. Büring (Eds.), *Topic and focus: A cross-linguistic perspective* (pp. 217–246). Kluwer.

Selkirk, E. (2011). The syntax–phonology interface. In J. Goldsmith, J. Riggle, & A. Yu (Eds.), *The handbook of phonological theory* (2nd ed., pp. 435–484). Blackwell.

Shattuck-Hufnagel, S., & Turk, A. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research*, *25*(2), 193–247. doi:10.1007/BF01708572

Snedeker, J., & Casserly, E. (2010). It is all relative? Effects of prosodic boundaries on the comprehension and production of attachment ambiguities. *Language and Cognitive Processes*, *25*(7-9), 1234–1264. doi:10.1080/01690960903525499

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. doi:10.1016/S0749-596X(02)00519-3

Sobin, C., & Alpert, M. (1999). Emotion in speech: The acoustic attributes of fear, anger, sadness, and joy. *Journal of Psycholinguistic Research*, *28*(4), 347–365. doi:10.1023/A:1023237014909

Speer, S. R., Kjelgaard, M. M., & Dobroth, K. M. (1996). The influence of prosodic structure on the resolution of temporary syntactic closure ambiguities. *Journal of Psycholinguistic Research*, *25*(2), 249–271. doi:10.1007/BF01708573

Spivey-Knowlton, M., & Sedivy, J. C. (1995). Resolving attachment ambiguities with multiple constraints. *Cognition*, *55*(3), 227–267. doi:10.1016/0010-0277(94)00647-4

Sudhoff, S. (2010). Focus particles and contrast in German. *Lingua*, *120*(6), 1458–1475. doi:10.1016/j.lingua.2009.02.011

Tanenhaus, M. K., Kurumada, C., & Brown, M. (2015). Prosody and intention recognition. In L. Frazier, & E. Gibson (Eds.), *Explicit and implicit prosody in sentence processing: Studies in Theoretical psycholinguistics* (pp. 99–118). Springer.

Taraban, R., & McClelland, J. L. (1988). Constituent attachment and thematic role assignment in sentence processing: Influences of content-based expectations. *Journal of Memory and Language*, *27*(6), 597–632. doi:10.1016/0749-596X(88)90011-3

Trouvain, J. (2004). *Tempo variation in speech production. Implications for speech synthesis. PHONUS 8*. Institute of Phonetics, Saarland University.

Truckenbrodt, H. (1999). On the relation between phonological phrases and syntactic phrases. *Linguistic Inquiry*, *30*(2), 219–255. doi:10.1162/002438999554048

Van Berkum, J. J. (2009). The neuropragmatics of 'simple' utterance comprehension: An ERP review. In U. Sauerland, & K. Yatsushiro (Eds.), *Semantics and pragmatics: From Experiment to theory* (pp. 276–316). Palgrave Macmillan.

Van Berkum, J. J., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(3), 443–467. doi:10.1037/0278-7393.31.3.443

Van Berkum, J. J., Van den Brink, D., Tesink, C. M., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience*, *20*(4), 580–591. doi:10.1162/jocn.2008.20054

Von Holzen, K., & Mani, N. (2012). Language nonselective lexical access in bilingual toddlers. *Journal of Experimental Child Psychology*, *113*(4), 569–586. doi:10.1016/j.jecp.2012.08.001

Wagner, M., & Watson, D. G. (2010). Experimental and Theoretical advances in prosody: A review. *Language and Cognitive Processes*, *25*(7-9), 905–945. doi:10.1080/01690961003589492

Watson, D., Breen, M., & Gibson, E. (2006). The role of syntactic obligatoriness in the production of intonational boundaries. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *32*(5), 1045–1056. doi:10.1037/0278-7393.32.5.1045

Watson, D., & Gibson, E. (2004). The relationship between intonational phrasing and syntactic structure in language production. *Language and Cognitive Processes*, *19*(6), 713–755. doi:10.1080/01690960444000070

Weber, A., Braun, B., & Crocker, M. W. (2006). Finding referents in time: Eye-tracking evidence for the role of contrastive accents. *Language and Speech*, *49*(3), 367–392. doi:10.1177/00238309060490030301

Williams, C. E., & Stevens, K. N. (1972). Emotions and speech: Some acoustical correlates. *Journal of the Acoustical Society of America*, *52*(4B), 1238–1250. doi:10.1121/1.1913238

Appendix

The 24 experimental items used in Experiment 1 and 2. The sentences had a prosodic boundary either after the verb (Modifier prosody) or between the noun and the PP (Instrument prosody). The sentence-final noun was either a plausible-instrument noun or an implausible-instrument noun.

1. The boy % will write to (%) the pig (%) with the pen/cup.
2. The boy % will write to (%) the bear (%) with the chalk/glass.
3. The boy % will write to (%) the panda (%) with the crayon/jelly.
4. The girl % will look at (%) the cat (%) with the glasses/coffee.
5. The girl % will look at (%) the tiger (%) with the telescope/steak.
6. The girl % will look at (%) the mouse (%) with the lens/juice.
7. The girl % will see (%) the turtle (%) with the goggles/egg.
8. The boy % will see (%) the tiger (%) with the binoculars/popcorn.
9. The boy % will see (%) the frog (%) with the microscope/milk.
10. The boy % will sing to (%) the frog (%) with the guitar/necktie.
11. The boy % will sing to (%) the dog (%) with the microphone/tea.
12. The boy % will sing to (%) the cat (%) with the piano/pudding.
13. The girl % will hide (%) the turtle (%) with the towel/jam.
14. The girl % will hide (%) the sheep (%) with the blanket/cake.
15. The girl % will hide (%) the bear (%) with the scarf/soup.
16. The girl % will feed (%) the rabbit (%) with the spoon/headphones.
17. The boy % will feed (%) the panda (%) with the chopsticks/sunglasses.
18. The boy % will feed (%) the pig (%) with the fork/watch.
19. The boy % will scratch (%) the elephant (%) with the shovel/water.
20. The boy % will scratch (%) the mouse (%) with the stick/wine.
21. The girl % will scratch (%) the sheep (%) with the hairbrush/yogurt.
22. The girl % will clean (%) the elephant (%) with the brush/cereal.
23. The girl % will clean (%) the dog (%) with the comb/collar.
24. The girl % will clean (%) the rabbit (%) with the sponge/necklace.