10th International Conference on Speech Prosody 2020
25-28 May 2020, Tokyo, Japan

Learning to Anticipate Contrast with Prosody: A Visual World Study with L2 Learners

Chie Nakamura[1], *Jesse A. Harris*[2], *Sun-Ah Jun*[2]

[1]Waseda University, Japan
[2]University of California Los Angeles, USA
chienak@aoni.waseda.jp, jharris@humnet.ucla.edu, jun@humnet.ucla.edu

Abstract

This study tested how L2 learners use contrastive accent to anticipate upcoming information during instructed visual search. In two visual-world eye-tracking experiments, we compared the processing patterns between native English speakers and Japanese learners of English. Participants' eye-movements were recorded and analyzed to investigate whether listeners reached the correct target object faster when the sentence carried contrastive accent (L+H*) on the adjective in an adjective-noun pair (e.g., *First, find the red cat. Next, find the PURPLE*$_{\text{L+H*}}$ *cat*) compared to the condition in which the adjective carried new information accent (H*). The results showed that the use of contrastive L+H* accent led anticipatory looks to the target object both with native speakers and with L2 learners. In addition, the difference between the two types of prosody was increased in the final block of the experiment for L2 learners. This indicates that L2 learners learned to use the contrastive function of prosody in processing more as the experiment advanced by associating the phonetic features of L+H* with a contrastive interpretation with increased exposure.

Index Terms: Prosody, contrastive accent, second language processing, learning, visual-world eye-tracking

1. Introduction

Past research on sentence processing in a second language (L2) suggests that L2 learners, like native speakers, exploit various types of linguistic information to generate incremental representations in their L2 during real-time sentence processing [1-3]. To date, relatively little is known about how prosodic cues influence the interpretation and comprehension in L2 processing. For example, several studies reported that L2 learners are sensitive to the alignment of prosodic boundary and syntax [4], much like native speakers [5]. Other studies have reported a difference between native speakers and L2 learners in the perception of intonational meaning [6,7]. While these results are perhaps hard to reconcile because they considered different types of prosody (e.g., prosodic boundary and pitch accents) and participants with different L1s at different proficiency levels, there is a general consensus that L2 learners' use of prosody is generally slower and less accurate compared to L1 processing [8].

Broadly speaking, processing deficits might originate from two possible sources: integration or prediction [9]. Integration refers to the bottom up, reactive process of incorporating an element into a structure incrementally. Prediction, or anticipation, refer to the top down, proactive process of anticipating words or structure ahead of the input already received. When successful, predictive processing could ease future integrative processes by pre-activating likely content, possibly even pre-updating representations prior to encountering the predicted stimulus (see [10]).

While integration is understood as a conceptual necessity in the sentence processing literature, many recent studies have probed the extent, and conditions under which, language comprehenders engage in predictive processing [11]. Although a complex picture is now emerging, there is already good deal of evidence that language users make predictive inferences at multiple levels of representations, and that they adapt such strategies in response to various contextual and cognitive demands required by the processing task or situation. Despite its putative advantages, predictive processing may, however, come at a further cost for some populations – e.g., older adults [9,12] or non-native speakers [13-16].

In particular, the RAGE (Reduced Ability to Generate Expectations) approach to L2 processing argues that L2 learners are so overtaxed by integrative processes that little resources remain for predicting upcoming information as effectively as native speakers [16]. The strongest, and a perhaps unlikely, version of the RAGE hypothesis would predict that L2 learners are categorically unable to engage in any processing beyond simply integrating words into structure. Under such a view, L2 processing would be purely reactive. However, results from recent studies report that L2 learners use grammatical information such as the lexical-semantics of verbs and gender-marked adjectives to restrict the selection of upcoming information [18,19]. These studies indicate that prediction in L2 is not an all-or-nothing process and that predictive or anticipatory processing in L2 could indeed be selectively limited.

A weaker, and more plausible, version of the RAGE hypothesis would allow for considerable variance in terms of multiple factors such as population (e.g., the L2 learners' native language and their level of L2 proficiency), task (e.g., how demanding a particular task is), and linguistic factors (e.g., whether the prediction can be generated from purely grammatical information, how constraining the context is, and familiarities and frequency information of the cue). Accordingly, if the limitations of L2 learners to make predictions in sentence processing is due to reduced access to processing resources, they might nonetheless be able to anticipate upcoming information when integration is relatively simple and does not demand extensive cognitive effort. Under this view, L2 processing is selectively proactive, depending on the extent to which speakers have (i) learned when content is predictable, (ii) sufficient cognitive or attentional reserves to make the predictions, and (iii) a possibly implicit belief that predictive processing will be beneficial given the task.

The current study further investigates anticipation in L2 processing by employing a simple instructed visual search task similar to the ones used in [19,20]. Previous research finds that

10.21437/SpeechProsody.2020-177

contrastive L+H* accent on a contrastive adjective leads to increased anticipatory fixations to the target referent for native speakers of English, compared to non-contrastive H* accent on the adjective [20]. We tested whether contrastive L+H* accent on the adjective in an adjective-noun pair such as (1) would lead L2 learners of English to correctly anticipate the referent. If L2 learners' difficulty predicting upcoming information is due to increased attentional demand required by the complex integration of multiple sources of information, they might be able to use prosodic information in anticipatory processing in a relatively simple visual search task.

In the study, we tested Japanese speakers learning English as a second language. In Japanese, focus is realized by pitch range expansion on the focused word, which is phonetically similar to how focus is realized in English. However, the pitch range expansion in Japanese does not change the phonological tonal categories unlike in English [21,22]. Therefore, Japanese speakers might not interpret the difference in rising slope in English H* vs. L+H* phonologically. If Japanese speakers use pitch range difference to distinguish between H* and L+H*, however, they might use the contrastive meaning of L+H* to correctly anticipate upcoming information.

(1) *First, find the red cat.*
a. *Next, find the purple*$_{H*}$ *cat.*
b. *Next, find the PURPLE*$_{L+H*}$ *cat.*

In addition to L2 learner's anticipatory use of prosody, we also investigated whether L2 learners' processing patterns in the use of prosody would change over the course of the experiment. Some recent studies in sentence processing have shown that language users generate predictions about upcoming information in response to language input they receive. These studies suggest that language users have abilities to cope with linguistic variability by updating distributional statistics based on the exposure they receive [23,24], and generate expectations for the information that is most likely to come next [25,26]. Interestingly, such learning effects, demonstrated as a structural priming effect in some studies, are shown to be larger with less preferred or less frequent utterances than generally preferred and frequent utterances, possibly because processing of low-frequent irregular sentences benefits almost exclusively from specific experience of the exact irregular sentence types [27].

This points to an interesting set of possibilities in which the size of learning effects might differ for native speakers and L2 learners, as the two groups have different access to grammatical information and frequency distributions for the language. One possibility is that native speakers would not exhibit learning over the course of the experiment, because they can already fully distinguish between H* and L+H* on the basis of the rate at which f0 rises, and understand the conventional interpretations as new information and contrastive information, respectively (even if they do not always use such distinctions consistently [28,29]). Thus, there would be no change in the pattern of anticipatory eye-movements to the contrastive object with native speakers. Another possibility is that L2 learners may have relatively weak associations between a sharply rising f0 and the contrastive meaning in English, and they would learn the contrastive meaning of L+H* as they receive more input. If this is the case, L2 learners would show learning effects as they are exposed to more sentences with contrastive prosody.

2. Experiments

2.1. Experiment 1 (Native speakers of English)

In Experiment 1, we tested whether native speakers, as a control group, would use contrastive accent to predict an upcoming referent during visual search. We expected to observe anticipatory eye movements to the target object (*purple cat*) prior to the onset of the target noun (*cat*) when the sentence has contrastive accent on a contrastive adjective (1b) compared to the control condition with new information (H*) accent (1a).

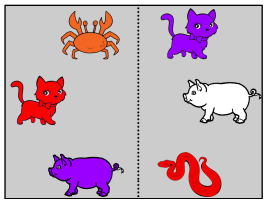

Figure 1: *Visual array presented with (1)*

2.1.1. Participants

Thirty-six native speakers of English with unimpaired vision and hearing participated in the experiment for course credit.

2.1.2. Stimuli

Thirty-six experimental items were created. Each item consisted of a sound file of a sentence and a corresponding visual scene (*Figure 1*). The auditory stimuli were recorded by a male native speaker of English who is trained in English ToBI [31,31]. Figure 2 shows the f0 contours of the sentence (1) in each condition. The visual scenes were prepared using clip art images. The position of the objects was counter-balanced across the items. Two experimental lists were created following the Latin square design including 36 target items and 54 filler items. Filler sentences contained nouns without color adjectives (e.g., *First, find the cat. Next, find the giraffe*). The 90 items in each list were presented in pseudo-random order.

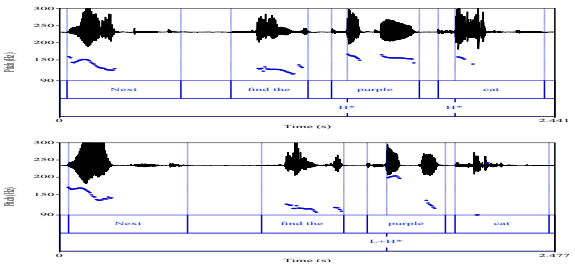

Figure 2: *Waveform, pitch track, and accent type for (1a, top) and (1b, bottom).*

2.1.3. Procedure

Participants were instructed to listen to the sentences carefully while attending to the picture on the computer monitor. As soon as the sentence ended, participants were asked to respond by pressing the left or right bumper on a gamepad that corresponded to the position of the second-mentioned object. The visual array was presented for viewing for 3000ms prior to the onset of the sentence. Participants' eye-movements were recorded with EyeLink 1000 Plus (SR Research) at a sampling rate of 500 Hz. A 5-point calibration was conducted at the beginning of the experiment and as needed. Drift correction was

performed before each trial. Experimental sessions lasted approximately 30 minutes.

2.1.4. Data analysis and results

The mean percentage of correct responses was 99.5%. For the eye-movements analysis, we summed the gazes to each entity in the scene and calculated the logit of looks to each entity out of looks to all the objects in the scene, including the background, for the period of interest. Statistical analyses for the duration of anticipatory time window (from the onset of the color adjective until the minimum onset of the target noun, 400ms) were conducted using linear mixed-effect regression models [32]. Prosodic accent type (L+H* or H*) was included as a fixed factor in the model. In order to investigate whether participants' performance changed over the course of the experiment, the 90 trials were divided into 3 blocks and included as an additional factor in the model (Block). By-participants and by-item effects were modeled as random intercepts. The best fitting model was explored using a backward selection approach. Table 1 summarizes the results from the optimal model in the anticipatory time window. Figure 3 shows the proportion of looks to the target object by native speakers, time-locked to the onset of the color adjective. The analysis for the duration of anticipatory time window showed a main effect of Prosody (p<.05); more looks to the target object were observed with contrastive accent than without it. Crucially, the effect was observed before the onset of the target noun. In a more complex model with Block included as an interactive term, there was no effect of Block (p=0.430), nor an interaction between Block x Prosody (p=0.782).

Table 1: *Analysis of looks to the target object in the anticipatory time window (native speakers)*

	Estimate	**SE**	***t*-value**	***p*-value**
Intercept	-5.91	2.35	-25.19	<.001
Prosody	2.37	1.05	2.25	< .05

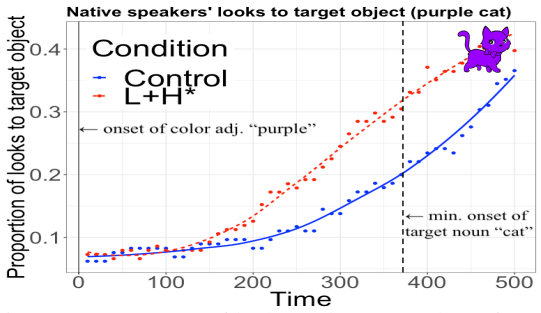

Figure 3: *Proportion of looks to the target object from the onset of the color adjective to 500ms (native speakers).*

Consistent with the findings of the previous research [20], native speakers in our study used contrastive accent to anticipate the upcoming word that was most likely to contrast with the previous word in the visual array.

2.2. Experiment 2 (Japanese learners of English)

2.2.1. Participants

Twenty-nine native Japanese-L1 English-L2 speakers (L2 learners) participated in Experiment 2 for monetary compensation. Participants were all undergraduate students at Waseda University, Japan (Age=18-22; M=19.2, SD=1.04). They had at least six years of English education before enrolling in the university. We obtained English proficiency scores from all the participants. The mean score for our participants corresponded to the proficiency levels of 'Intermediate' to 'Advanced' on the Common European Framework of Reference for Languages (CEFR) [33].

2.2.2. Stimuli and procedure

Stimuli and procedure were identical to those in Experiment 1.

2.2.3. Data analysis and results

The mean percentage of correct responses was 98.1%. Analyses on gazes to each entity were performed in the same manner for the same intervals as in Experiment 1. Each participant's score for a standardized English test was added as an additional factor (Proficiency) in the model. Table 2 summarizes the results from the best-fitting model. Figure 4 shows the proportion of looks to the target object, time-locked to the onset of the color adjective. With L2 learners, the analysis for the duration of the anticipatory time window showed a main effect of Prosody (p<.05). This finding indicates that there were more looks to the target object with contrastive accent on the adjective than without it before the onset of the target noun. This demonstrates that, like native speakers, L2 learners used contrastive meaning of prosody in anticipating the likely referent of the upcoming noun. There was no effect of Proficiency, most likely because the proficiency level of our L2 participants was relatively homogeneous.

Table 2: *Analysis of looks to the target object in the anticipatory time window (L2 learners)*

	Estimate	**SE**	***t*-value**	***p*-value**
Intercept	-4.69	0.64	-7.37	< .001
Prosody	1.35	0.65	2.26	< .05
Block	0.49	0.26	1.90	0.061
Prosody x Block	-0.55	0.30	-1.83	0.067

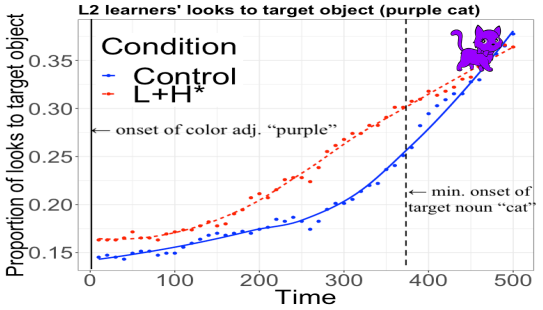

Figure 4: *Proportion of looks to the target object from the onset of the color adjective to 500ms (L2 learners).*

In addition, L2 learners' results also showed that there was a marginally significant effect of Block and a marginally significant interaction between Prosody x Block, suggesting the possibility that L2 learners' performance changed over the course of the experiment by learning the contrastive meaning of the contrastive L+H* accent. L2 learners' anticipatory use of prosody and L2 learner's learning effect in the use of prosodic information will be discussed separately.

2.2.3.1. Anticipatory use of prosody with L2 learners

L2 learners looked more at the target object in the L+H* condition than in the H* condition during the anticipatory time

window, suggesting that not only do L2 learners understand the interpretive difference between H* and L+H* in English intonation, but also they were able to recruit this knowledge to anticipate an upcoming referent. This result provides evidence against the possibility that L2 learners looked at the target object more with L+H* simply because the adjective was phonetically more prominent, because there was no effect of Prosody in looks to the competitor object with the same color (i.e., no increased looks to "purple pig" in Fig. 1 with L+H* on "purple").

In addition, the difference between Figure 3 and 4 indicates that L2 learners looked at the target object with L+H* at the onset of the color adjective, which is even earlier than native speakers did. This suggests another interesting possibility that L2 learners might have used additional acoustic cues that were not used by native speakers in anticipatory processing. In English intonation, L+H* and H* are known to differ phonetically by the f0 rising slope (i.e., sharp vs. smooth rise), but they also differ in the f0 on the syllable before the f0 peak (i.e., lower f0 in L+H*) [28,29]. In our stimuli, an acoustic analysis comparing the pitch peak on the "find the" segment in "Next, find the red cat" confirmed that the L+H* condition had lower f0 during "find the" (*mean pitch peak*=125Hz, *SD*=8, $t(35)=4.79$, $p<.001$) compared to that in the H* condition (*mean pitch peak*=133Hz, *SD*=7). [

Although an effect of Prosody in the looks to the target object for the duration "find the" was not statistically significant, L2 learners' looks to the target object started to increase in the L+H* condition from about 150ms after the onset of the adjective, suggesting that they were more sensitive to the slope of f0 rising to distinguish L+H* from H*. This may imply that L2 learners used the f0 difference to anticipate the target noun on the following adjective. It is possible that L2 learners paid more attention to the phonetic details of the stimuli, instead of parsing the L2 prosody phonologically. As Japanese is a lexical pitch accent language, our L2 participants may have been more sensitive to f0 cues than native English speakers, who did not seem to use f0 height before the adjective in anticipatory processing. Further study is needed to confirm our hypothesis.

2.2.3.2. Learning effect of prosody with L2 learners

In order to investigate whether L2 learners used the contrastive meaning of L+H* prosody more as they experienced more trials in the experiment, we analyzed the change in looks to the target object over the course of the experiment. Looks to the target object from the onset of the noun until the minimum offset of the sentence (e.g., the duration of "cat", henceforth "disambiguating time window") were analyzed. Figure 5 shows the changes in looks to the target object in the disambiguating time window.

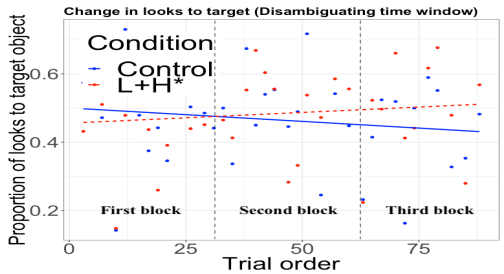

Figure 5: *Change in looks to the target object in the disambiguating time window.*

Further analyses on an effect of Prosody in each block found no effect of Prosody in the first and second blocks (p's>.1). In the third block, there was a marginal effect of Prosody (p=.077), which indicates that L2 learners only learned to look to the target object more in the final part of the experiment. The influence of the contrastive meaning of L+H* prosody became stronger as L2 learners experienced more trials in the experiment. Crucially, Figure 5 also shows that L2 learners looked less at the target referent in the H* condition in later trials of the experiment. This suggests that L2 learners were not simply getting better at anticipating the noun given the adjective, but adapted to the particular meaning of contrastive accent.

3. Discussion

Our results showed an anticipatory effect of contrastive L+H* accent with both native speakers and L2 learners, demonstrating a faster increase in fixation to the target noun than with H*. The results provide evidence against the strongest view of the RAGE hypothesis, which assumes that L2 learners are incapable of making predictions about upcoming information in online sentence processing. Instead, our results support a weaker version, in that L2 learners appeared to learn to make anticipatory eye movements to a predicted target noun on the basis of prosodic information on the immediately preceding adjective.

Our results also suggest the possibility that L2 learners were sensitive to the pitch difference even before the onset of the color adjective, and used the low f0 information before the peak as a contrastive marking cue. We hypothesized that our L2 participants, whose L1 is Japanese, were sensitive to pitch changes over the syllables, and the low f0 before the color adjective triggered their anticipation of contrast. In contrast, English native speakers did not use the f0 difference before the color adjective, which might suggest that the f0 cue before the peak has a lower weight compared to the f0 rising slope cue in distinguishing between L+H* and H* in English intonation.

In addition, we also observed that L2 learners' ability to correctly anticipate the upcoming referent increased with exposure. The difference between the two types of the prosody was found to be the largest in the final part of the experiment, suggesting that sufficient exposure to the mapping between prosody and meaning may be necessary for L2 learners to begin anticipating upcoming speech during online sentence comprehension using prosodic cues. For native speakers, however, no such learning effect was observed, most likely because native speakers have established the conventional meaning of a L+H* pitch accent as contrastive. Thus, their interpretation of the contrastive meaning with L+H* accent did not change over the course of the experiment.

We speculate that adequate exposure is necessary in language learning to be able to generate the appropriate conceptual mappings between the H* and L+H* tonal patterns and their conventional interpretive uses. Additionally, gaining familiarity with the task and prosodic categories over the course of the experiment may also reduce the overall processing load on L2 learners, which in turn increases cognitive or attentional resources to engage in anticipatory strategies.

4. Acknowledgements

We thank Canaan Breiss and the research assistants at the UCLA Language Processing Lab and the Waseda BLIT lab. This work was supported by JSPS grant number 19H01279.

5. References

[1] Soares, C., and Grosjean, F. “Bilinguals in a monolingual and a bilingual speech mode: the effect on lexical access.” *Memory & Cognition*, vol 12, pp. 380–386, 1984.

[2] Hahne, A. “What’s different in second-language processing? Evidence from event-related brain potentials.” *Journal of Psycholinguistic Research,.* vol 30, 251–266, 2001.

[3] Frenck-Mestre, C. "An on-line look at sentence processing in the second 1469 language," in Advances in Psychology, Vol. 134, eds R. R. Heredia, and J. 1470 Altarriba, (Amsterdam: Elsevier Science Publishers), 217–236, 2002.

[4] Nakamura, C., Harris, J. A., & Jun, S. A. “L2 adaptation to unreliable prosody during structural analysis: A visual world study.” *BUCLD 43 Proceedings.* pp. 454-468, 2019.

[5] Snedeker, J. & Trueswell, J. “Using prosody to avoid ambiguity: Effects of speaker awareness and referential context.” *Journal of Memory and Language,* vol 48, 103-130, 2003.

[6] Akker, E., & Cutler, A. “ Prosodic cues to semantic structure in native and nonnative listening.” *Bilingualism,* vol 6, pp. 81-96, 2003.

[7] Braun, B. & Tagliapietra, L. “On-line interpretation of intonational meaning in L2.” *Language, Cognition, and Neuroscience,* vol 26, pp. 224-235, 2011.

[8] Cutler, A. *Native Listening: Language Experience and the Recognition of Spoken Words*. MIT Press, 2012.

[9] Federmeier K. D. “Thinking ahead: the role and roots of prediction in language comprehension.” *Psychophysiology*, vol 44, 491–505, 2007.

[10] Lau, E.F., Holcomb, P., & Kuperberg, G. “Dissociating N400 effect of prediction from association in single-word contexts.” *Journal of Cognitive Neuroscience.*, vol 25, pp.484-502, 2013.

[11] Kuperberg, G. R., & Jaeger, T. F. “What do we mean by prediction in language comprehension?” *Language, Cognition and Neuroscience*, vol 31, pp. 32–5, 2016.

[12] Federmeier, K. D., Kutas, M., & Schul, R. “Age-related and individual differences in the use of prediction during language comprehension.” *Brain and Language*, vol 115, pp. 149–161 , 2010.

[13] Kaan, E. “Predictive sentence processing in L2 and L1: What is different?” *Linguistic Approaches to Bilingualism, vol 4, pp. 257-282, 2014*.

[14] Lew-Williams, C. & Fernald, A. “Real-time processing of gender-marked articles by native and non-native Spanish speakers.” *Journal of Memory and Language*, vol 63, pp. 446-464.

[15] Martin, C., Thierry, G., Kuipers, J. R., Boutonnet, B., Foucart, A., & Costa, A. “Bilinguals reading in their second language do not predict upcoming words as native readers do.” *Journal of Memory and Language*, vol. 69, 574–588, 2013.

[16] Grüter, T., & Rohde, H., & Schafer, A. J. “The role of discourse-level expectations in non-native speakers’ referential choices.” *BUCLD 38 Proceedings*, 2014.

[17] Ito, A., Pickering, M. J., & Corley, M. “Investigating the time-course of phonological prediction in native and non-native speakers of English: A visual world eye-tracking study.” *Journal of Memory and Language,* vol 98, pp. 1–11, 2018.

[18] Hopp, H., & Lemmerth, N. “Lexical and syntactic congruency in L2 predictive gender processing.” *Studies in Second Language Acquisition*, vol 40, pp. 171-199, 2018.

[19] Dahan D., Tanenhaus, M. K., & Chambers, C. G. “Accent and reference resolution in spoken-language comprehension.” *Journal of Memory and Language*, vol 47, pp. 292–314, 2002.

[20] Ito, K. & Speer, S. R. “Anticipatory effects of intonation: Eye movements during instructed visual search.” *Journal of Memory and Language*, vol 58, pp. 541-573, 2008.

[21] Pierrehumbert, J. & MBeckman, M. “Japanese Tone Structure.”, Cambridge, Mass.: MIT Press, 1988.

[22] Ishihara, S. “Japanese focus prosody revisited: freeing focus from prosodic phrasing.” *Lingua*, vol 121, pp. 1870-1888, 2011.

[23] Norris, D., McQueen, J. M., Cutler, A. “Prediction, Bayesian inference and feedback in speech recognition.” *Language, Cognition and Neuroscience,* vol 31, pp. 4–18, 2015.

[24] Kraljic, T., Samuel, A. G. “ Generalization in perceptual learning for speech.” *Psychonomic Bulletin and Review*, vol 13, pp. 262–268, 2006.

[25] Kaschak, M. P., Glenberg, A. “This construction needs learned.” *Journal of Experimental Psychology: General*, vol 133, pp. 450–467, 2004.

[26] Fine, A. B., Jaeger, T. F., Farmer, T., Qian, T. “Rapid expectation adaptation during syntactic comprehension.” *PLoS ONE*, vol 8, 2013.

[27] Pearlmutter, N. J. & MacDonald, M. C. “Individual differences and probabilistic constraints in syntactic ambiguity resolution.” Journal of Memory and Language, vol 34, pp. 521-542, 1995.

[28] Ito, K., Speer, S.R., & Beckman, M. “Informational status and pitch accent distribution in spontaneous dialogues in English.” *Proceedings of Speech Prosody 2004*, pp. 279–282, 2004.

[29] Royer, A & Jun, S-A. “Surface categorical variation in the prosodic marking of English focus.” *JASA 140*, Pt.2 p.3397, 2016.

[30] Beckman, M., Hirschberg, J., Shattuck-Hufnagel, S. “The original ToBI system and the evolution of the ToBI framework.” In: Jun, S.-A. (ed), *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press, pp. 9-54, 2005.

[31] Beckman, M. and Ayers, G. “Guidelines for ToBI labelling”, ms. Ohio State Univ., 1994.

[32] Baayen, R. H., Davidson, D. J., and Bates, D. M. “Mixed-effects modeling with crossed random effects for subjects and items.” *Journal of Memory and Language*, vol 59, pp. 390–412, 2008.

[33] Council of Europe. “Common European framework of reference for languages: Learning, teaching, assessment.” Cambridge, U.K: Press Syndicate of the University of Cambridge, 2001.