THE PHONOLOGICAL AND PHONETIC ENCODING OF INFORMATION STRUCTURE IN AMERICAN ENGLISH NUCLEAR ACCENTS

Eleanor Chodroff and Jennifer Cole

Department of Linguistics, Northwestern University
eleanor.chodroff@northwestern.edu, jennifer.cole1@northwestern.edu

ABSTRACT

Information structure is said to play an important role in determining phrasal prominence and the assignment of nuclear pitch accents in English. Early accounts claim that discourse-new or focused words receive a prominence-lending high/rising pitch accent, while given words are unaccented, with reduced prominence. Empirical findings are varied, but paint a more complex picture of the prosodic encoding of information structure. The present study investigated the phonological and phonetic encoding of information status and contrastive focus in nuclear position in American English, from speech read under neutral and lively affect. Given information was associated with decreased phonological and phonetic prominence, contrastive information with enhanced prominence, while new information corresponded to increased phonological, but not phonetic prominence, as assessed in pitch accent type, duration, intensity, and voice quality. The findings indicate a probabilistic relationship between information structure and nuclear pitch accent type, and gradient expression of information structure in acoustic prominence.

Keywords: prosody, information structure, givenness, focus, nuclear pitch accents

1. INTRODUCTION

The prosodic realization of a phrase reflects a myriad of linguistic and extra-linguistic functions. Understanding the relation between form and function is central to prosodic research. While these relations are quite complex, one that has received considerable attention from phonologists and semanticists alike is that between information structure and prosodic realization [4, 7, 9, 11, 14]. Information structure (IS), including both focus status and information status (or degree of givenness) have been argued to constrain prosodic realization, particularly for nuclear pitch accents, though empirical analysis of this relationship has yielded highly variable results.

Information status has traditionally been characterized as a binary distinction between given and new information [11, 14]. In its phonological expression in English, given information tends to be deaccented or receive a low pitch accent, whereas new information, as well as narrow or contrastive focus, is marked with high or rising pitch accents [4, 7, 9, 11, 14]. Empirical studies reveal a more complex and probabilistic relation between information structure and pitch accent type: given information is not always deaccented [1, 13, 18], and high and rising tones are often, but not always used to signal contrastive focus or newness [13]. Similar findings are reported for German [10, 15].

Several studies have examined the phonetic encoding of IS: for American English, Breen et al. [3] report effects of focus location (subject, verb, object) in duration, intensity and F0, with further distinction of focus type (contrastive/non-contrastive) for objects. Calhoun [6] observed lower, but more delayed peaks on given information (theme) relative to new information (rheme) in Australian English, though these relationships were notably weak. Röhr & Baumann [15] also report gradient effects of givenness on f0 in German.

While previous studies show a probabilistic or gradient relationship between information structure and prosodic prominence, no study has yet examined the influence of information structure on both phonological (pitch accent) and phonetic correlates of prominence in American English, and specifically in the position of the default nuclear prominence in the intonational phrase. The nuclear pitch accent is the final pitch accent of an intonational phrase, and though it may not manifest acoustic prominence, it is nevertheless associated with increased perceptual prominence [4, 12]. Critically, this accent has been argued to convey meaning related to information structure [4, 5], whereas the relationship between IS and *prenuclear* accents is minimal at best [5, 8]. The present study therefore provides a thorough investigation of the phonological and phonetic encoding of information structure in nuclear prominence position in American English.

This research importantly addresses several limitations in previous production studies of the IS-prosody relationship. In addition to confining the analysis to the nuclear region of the phrase, we also obtain a high degree of phonological and semantic control in our stimulus for a systematic comparison of prosodic effects between IS conditions. We adopt a three-level representation of information status with levels not only for given and new, but also accessible information, which has been implicated in perceptual processing of the IS-prosody relation [2, 7]. Finally,

we implemented a dual phonological and phonetic analysis of the nuclear region. From previous studies, we expect informativity to be positively correlated with prominence of the nuclear pitch accent in its phonological status (pitch accent) and phonetic realization (duration, intensity, and voice quality). If IS is categorically encoded in phonology, we expect a direct relation between IS and pitch accent type. If IS influences prominence in a gradient manner, we expect a positive but weak correlation between the degree of newness and phonological prominence (pitch accent), as well as phonetic prominence (duration, intensity, voice quality).

2. METHODS

2.1. Participants

Data were collected from 32 native speakers of American English (23 female, 9 male), recruited from undergraduate Linguistics courses. An additional 11 participants completed the experiment but as non-native speakers of English, they were excluded from analysis. Participants were retained who reported that English was their first language or learned simultaneously with another (Mandarin: 2, Spanish: 1, Urdu: 1). All participants were compensated with course credit.

2.2. Stimuli

Twenty sets of mini-stories were created in which the IS of the final object noun phrase was contextually defined as discourse-given, -accessible, -new, or contrastive. Each story contained three sentences: two context sentences followed by one target sentence. Context sentence 1 and the target sentence were the same within each set of stories, while context sentence 2 modulated the IS of the final noun phrase in the target sentence. The target word was *given* when previously mentioned; *accessible* when in a hypernym–hyponym relation with a non-coreferential word in sentence 2; *new* when not mentioned or inferable from prior context. *Contrastive* focus was implemented in a double focus construction that paralleled sentence 2. All target words were trisyllabic with dictionary-defined initial stress. The syntactic and metrical structure of the target sentence was identical across stories, and voiceless obstruents were avoided. See Table 1 for an example story.

A yes-no question targeting information in the second sentence was created for each individual story to encourage full comprehension of the story content. The correct answer to half of the comprehension questions was 'yes' for every participant.

2.3. Procedure

Each participant received one IS condition for each of the 20 stories. Equal numbers of productions per story and condition were obtained by counterbalancing the IS-story pairings among every four participants. The order of the stories was randomized for each participant, and then repeated with the same order for four blocks. Participants were asked to modulate their speaking style between blocks: odd blocks were to be read in a casual, neutral manner, and even blocks in a lively, expressive manner. This manipulation was introduced primarily to promote variety in the pitch accent type. In total, there were 80 productions per participant, with 20 productions per IS condition. The comprehension question was presented after each story. All participants achieved at least 85% accuracy with a median accuracy of 97%.

Recordings were digitized at a sampling rate of 22050 Hz (16-bit format). The audio was force-aligned to the story transcript using FAVE-align [16], and the alignment of the final critical word was manually corrected at the word level. The final word was then re-aligned using FAVE for more precise phone boundaries.

Table 1: An example story structure with each of the four IS types in Context Sentence 2. The critical word in Sentence 3 is bolded.

Story Structure
Context Sentence 1
Our sister Jamie spent all day Saturday in the kitchen.
Context Sentence 2
Given: She knew it would take hours to make the marmalade.
Accessible: She especially enjoyed making homemade preserves.
New: She likes to make everything from scratch.
Contrastive: Our father loved the strawberry jam.
Target Sentence
Our nana loved the **marmalade.**

2.3. Pitch accent labeling & acoustic measures

A ToBI label was assigned to each critical word based on auditory impression and visual inspection of the f0 contour. Three trained annotators completed the task. Annotators first assigned one of the following ToBI labels to each word: H*, L*, L+H*, L*+H, or unaccented (UA). Because of annotator uncertainty in the distinction between H* and L+H*, as well as L* and UA, the labels were ultimately binned into one of two categories: (L+)H* and L*/UA. While some pitch accents were clearly high-flat or shallow rises (H*) and others sharply rising (L+H*), many accents had intermediate rises that were difficult to categorize. The distinction between L* and UA was also impaired by pervasive utterance-final creak,

which is analyzed below. Words labeled L*+H were few in number and excluded from analysis.

Duration (sec) and RMS intensity (dB) were measured in the initial trochee of the critical word. Intensity was normalized to the preceding subject word via subtraction (relative intensity). Intervals of modal and creaky voice were also manually marked throughout the critical word to explore interactions between IS and voice quality.

3. RESULTS

A total of 2560 items (32 participants × 80 productions) were collected. 154 items were excluded due to production error (143 instances) or a perceived primary stress on the third syllable of the critical word (11 instances). The number of errors per participant ranged from 0 to 17 (median = 4 errors). As mentioned, items labeled as L*+H were also removed from analysis (51 instances), resulting in 2355 items.

3.1. Pitch accent type

The count of pitch accent types per IS condition and affect are presented in Figure 1. A logistic mixed effects analysis was used to predict the likelihood of the critical word receiving an H* pitch accent as opposed to L* or no pitch accent. The model had fixed effects of IS condition, affect (lively vs. neutral), their interactions, and a random intercept for speaker and word. Models with more complex random effects failed to converge, even after decorrelating slopes. Condition and affect were sum-coded with weights of ±1 and respective base levels of accessible and neutral affect.

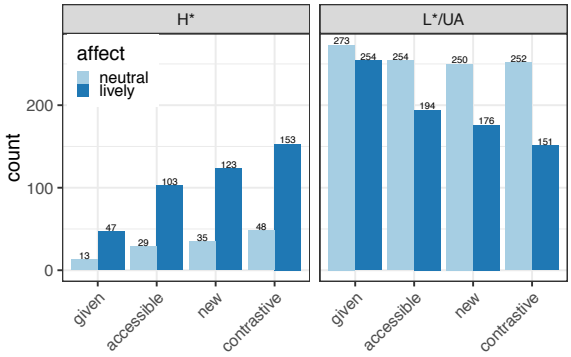
The model yielded significant effects of IS condition, as well as affect, but no significant interactions between factors. H* was significantly less likely in the given condition, more likely in the new and contrastive conditions, and with a lively speaking style, all relative to the grand mean (β_{given} = -1.06, $p < 0.001$; $\beta_{\text{new}} = 0.33$, $p < 0.01$; $\beta_{\text{contr}} = 0.71$, $p < 0.001$; $\beta_{\text{lively}} = 0.89$, $p < 0.001$).

3.2. Duration

The relation between duration and IS condition was assessed with a linear mixed effects model predicting the duration of the trochee (seconds) based on the pitch accent type (H* or L*/UA), IS condition, affect, and the full interactions between those effects. A random intercept, slope for condition, and slope for affect was included for participant, as well as a random intercept for the critical word. Main effects were sum-coded with weights of ±1 and respective base levels of L*/UA, accessible, and neutral.

The model revealed significant effects of given and affect. Relative to the average production, given words were approximately 8 ms shorter (β_{given} = -0.008, $p < 0.01$), contrastive 5 ms longer (β_{contr} = 0.005, $p < 0.05$), and lively 9 ms longer (β_{lively} = 0.008, $p < 0.001$). New and H* items were approximately 3 ms longer on average ($\beta_{\text{new}\times\text{H*}}$ = 0.003, $p < 0.05$). No other main effects or interactions reached significance.

Figure 1: Counts of pitch accent type by IS and affect.



3.3. Relative intensity

A linear mixed effects model was also used to assess the relative intensity of the trochee (dB) based on pitch accent type, IS condition, affect, and the interactions between those factors. The model was identical in structure to that for duration, though the random by-participant slope for condition failed to converge and was therefore excluded.

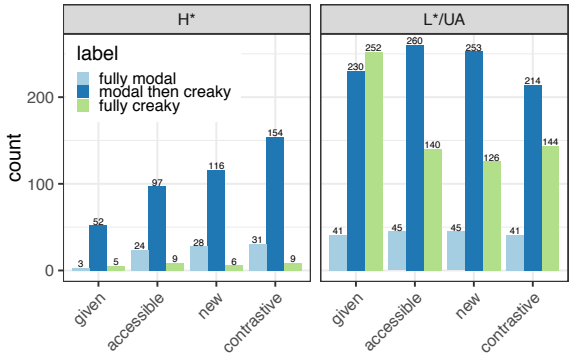
We observed significant effects of pitch accent type, given and contrastive conditions, and an interaction between pitch accent and affect. Critical words with an H* accent were approximately 1.4 dB louder on average ($\beta_{\text{H*}} = 1.36$, $p < 0.001$), given items 0.9 dB quieter (β_{given} = -0.88, $p < 0.001$), and contrastive items 1.4 dB louder (β_{contr} = 1.37, $p < 0.001$). Lively, H* items were approximately 0.2 dB louder than average ($\beta_{\text{lively}\times\text{H*}}$ = 0.20, $p < 0.05$). A main effect of affect was likely not observed due to normalization with respect to the subject word. No other main effects or interactions reached significance.

3.4. Voice quality

Voice quality in the critical word was often dynamic in nature. Though a handful of tokens had more than two voice quality changes, we defined the following primary voice quality categories: fully modal (258 items), modal followed by creaky (1376 items), and fully creaky (691 items). Because there were relatively few instances of creaky followed by modal voice, these were excluded from statistical analysis (30 items). The counts for each voice quality type

within each pitch accent category are presented in Figure 2.

Figure 2: Count of voice quality type by pitch accent type and IS.



A multinomial logistic regression was implemented to analyze the likelihood of voice quality sequence based on IS condition, pitch accent type, affect, and their interactions. The contrast coding of these factors was the same as for the linear models above, and the base category was 'fully modal'. There were significant effects of given, new, pitch accent type, and affect, as well as a significant interaction between newness and pitch accent. Given items were more likely to be fully creaky and new items were less likely to be modal followed by creaky as opposed to fully modal ($\beta_{\text{fullCr}\sim\text{given}} = 0.89$, $p < 0.01$; $\beta_{\text{modalCr}\sim\text{new}} = -0.33$, $p < 0.05$). Lively items were more likely to be fully modal than fully creaky ($\beta_{\text{fullCr}\sim\text{lively}} = -0.39$, $p < 0.01$), and H* items were also more likely to occur as fully modal than fully creaky ($\beta_{\text{fullCr}\sim\text{H*}} = -0.97$, $p < 0.001$). The interaction between new and H* indicated that these tokens were more likely to be realized with modal followed by creaky voice ($\beta_{\text{modalCr}\sim\text{new}\times\text{H*}} = -0.33$, $p < 0.05$). The remaining effects and interactions were not statistically significant, though numerically the number of modal then creaky H* items appeared to increase with informativity. Note that no numerical difference was observed between male and female speakers in the rate of voice quality patterns (*fully creaky*: F – 30%, M – 29%; *modal then creaky*: F – 58%, M – 59%; *fully modal*: F – 11%, M – 11%; *creaky then modal:* F – 1%, M – 1%).

4. DISCUSSION

The present analysis found both phonological and phonetic encoding of information structure for words in nuclear prominence position in American English, though the relationship was highly probabilistic [see also 10, 13]. Given information was more likely to be conveyed by a low pitch accent or be unaccented, whereas new and contrastive information was more likely to receive a high or rising accent; however, this relationship was not one-to-one. High or rising accents were also found on given items, and low accents or no accent also occurred on new and contrastive items. Acoustic correlates of IS were found primarily for words with given status: given information was more likely to be shorter, quieter, and realized with creak throughout the duration of the word when compared against all productions. Contrastive information was also relatively louder compared to average. Speaker affect or style also influenced prosodic prominence in its phonological and phonetic encoding in that a lively speaking style yielded increased usage of high or rising accents, along with an increase in duration and intensity. These effects did not significantly interact with IS.

Overall, these findings indicate a measurable but probabilistic relation between IS and prosodic prominence. While the usage of a high or rising pitch accent increased with informativity, the majority of phonetic effects were observed in the realization of given information in nuclear position. This may relate to previous claims that givenness is of primary relevance in its relation to prosodic implementation [17]. Nevertheless, the presence of phrase-final creak in American English may have obscured further effects of IS on prosodic prominence. Because the nuclear position often coincides with the edge of a phrase, any prosodic implementation of IS at the edge will be confounded by phrase-final effects. Between the high degree of creakiness and L*/UA realizations, the requirement to mark phrase finality may have overpowered effects of IS that require increased prominence. Research is currently underway to disentangle these relations by examining nuclear accents that are non-final and therefore less likely to contain influences of phrase-final creak.

5. ACKNOWLEDGEMENTS

We are grateful to Alaina Arthurs, Katherine Glew, Priya Kurian, and Jonah Pazol for assistance in data collection and processing. We also thank Stefan Baumann, Suyeon Im, José Hualde, and Jane Mertens for their role in the early phase of this project. Finally, we thank Timo Roettger, Daniel Turner, and three anonymous reviewers for their insightful feedback.

6. REFERENCES

[1] Bard, E., & Aylett, M. (1999). The dissociation of deaccenting, givenness, and syntactic role in spontaneous speech. In J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A. C. Bailey (Eds.), *Proc. 14th ICPhS* San Francisco, 1753–1756.

[2] Baumann, S., & Riester, A. (2012). Referential and lexical givenness: semantic, prosodic and cognitive aspects. In G. Elordieta & P. Prieto (Eds.), *Prosody and Meaning* (pp. 1–34). Mouton de Gruyter.

[3] Breen, M., Fedorenko, E., Wagner, M., & Gibson, E. (2010). Acoustic correlates of information structure. *Language and Cognitive Processes*, *25*(7), 1044–1098.
[4] Büring, D. (2007). Semantics, Intonation and Information Structure. In W. Ramchand & C. Reiss (Eds.), *The Oxford Handbook of Linguistic Interfaces* (pp. 1–36). New York.
[5] Calhoun, S. (2010). How does informativeness affect prosodic prominence? *Language and Cognitive Processes*, *25*(7), 1099–1140.
[6] Calhoun, S. (2012). The theme/rheme distinction: Accent type or relative prominence? *J. Phon.*, *40*(2), 329–349.
[7] Chafe, W. L. (1974). Language and Consciousness. *Language*, *50*(1), 111–133.
[8] Chodroff, E., & Cole, J. (2018). Information structure, affect, and prenuclear prominence in American English. *Proc. Interspeech* Hyderabad, 1849-1852.
[9] Cruttenden, A. (2006). The de-accenting of given information: A cognitive universal. *Pragmatic organization of discourse in the languages of Europe*, 311-355.
[10] de Ruiter, L. E. (2015). Information status marking in spontaneous vs. read speech in story-telling tasks - Evidence from intonation analysis using GToBI. *J. Phon.*, *48*, 29–44.
[11] Halliday, M. A. K. (1967). Notes on transitivity and theme in English: Part 2. *J. Ling.*, *3*(2), 199–244.
[12] Hualde, J. I., Cole, J., Smith, C. L., Eager, C. D., Mahrt, T., & Souza, R. N. De. (2016). The perception of phrasal prominence in English, Spanish and French conversational speech. *Proc. Speech Prosody 2016*, *50*, 459–463.
[13] Ito, K., Speer, S., & Beckman, M. (2004). Informational status and pitch accent distribution in spontaneous dialogues in English. In *Proc. Speech Prosody 2004*, 279–282.
[14] Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. R. Cohen, J. Morgan, & M. E. Pollack (Eds.), *Intentions in Communication* (pp. 271–311). Cambridge: MIT Press.
[15] Röhr, C. T., & Baumann, S. (2010). Prosodic marking of information status in picture story descriptions. In *Proc. Speech Prosody 2010*, 1–4.
[16] Rosenfelder, I., Fruehwald, J., Evanini, K., Seyfarth, S., Gorman, K., Prichard, H., & Yuan, J. (2014). FAVE Program Suite v1.2.2.
[17] Schwarzschild, R. (1999). Givenness, AvoidF, and other constraints on the placement of accent. *Natural Language Semantics*, *7*, 141–177.
[18] Terken, J., & Hirschberg, J. (1994). Deaccentuation of words representing "given" information: Effects of persistence of grammatical function and surface position. *Language and Speech*, *37*(2), 125–145.