Exploring Variation in English Intonational Acoustic Phonetics from Grammatical Perspectives

Byron Ahn, Alejna Brugos, Sunwoo Jeong, Stefanie Shattuck-Hufnagel, Nanette Veilleux

1 Overview

1.1 Proposed Research: Goals and Products

It is well-established that prosody can communicate important aspects of the meaning of an utterance, but we do not yet have a comprehensive understanding of all of the systematic ways in which it does so, even in a language as intensively studied as English. Of particular interest to many is the nature of the relationship between semantic/pragmatic (henceforth **"sem/prag"**) meanings and intonational contours; however, formal analysis of this relationship and the structures involved is still in the early stages. This project has as its goal to make substantial contributions to understanding in this area, focusing on which particular units of intonation align with which particular units of sem/prag meaning, and how to model these alignments. We consider these broad goals through the lens of intonational variation —i.e., where we find differences in intonational form both within/across speakers and within/across contexts— and address three specific and deeply intertwined questions that will be probed with a series of experiments.

Q1: What sorts of intonational variation are found in the speech signal (as differences in *acoustic cues*)?

Q2: What variation is *conditioned by sem/prag variables* (i.e., phonemic differences)?

Q3: What variation is found *within particular sem/prag contexts* (i.e., non-phonemic differences)?

The proposed experiments are centered around three sem/prag domains whose relations to intonation have been the focus of recent interest —common ground management (Expt Set 1), information structure (Expt Set 2), and scope ambiguity (Expt Set 3)— and will examine how these concepts relate to intonational variation. We employ a multi-layered paradigm of paired production and perception tasks, whose designs employ both new and existing tools in new ways. The experiments have been crafted with an eye towards determining what part of the variation we find is underwritten by grammatical structures (phonological/phonetic, semantic/pragmatic, and syntactic/morphological), as opposed to social and non-grammatical factors.

While it is well understood that these sorts of investigations are desirable, progress has been hindered by (at least) three types of challenges: (A) difficulties in analyzing detailed but generalizable aspects of the acoustic signal that are relevant to intonational phonology, (B) difficulties in reliably situating experimental participants in particular sem/prag contexts, and (C) the siloing of relevant linguistic expertise in intonation vs semantics/pragmatics. These challenges are addressed here in the following ways: (A) utterances elicited in the production experiments will be annotated using PoLaR (see §1.2), a recently-developed system designed to capture potentially meaningful variation of the acoustic signal that is often lost in purely phonological annotation, (B) experiments will use cartoon-based contexts with scenarios that contrast minimally with respect to sem/prag variables (see §1.3), to elicit and evaluate intonation contours, and (C) our grant team includes researchers with expertise in intonation (all), its relationships with phonetics/phonology (Ahn, Shattuck-Hufnagel, Brugos), psycholinguistics (Shattuck-Hufnagel), computational methods (Veilleux), experimental design and statistics (Scotina), morphosyntax (Ahn), and sem/prag (Jeong), so that direct interactions among these subfields of linguistics will occur naturally.

1.2 PoLaR: The challenge of annotating potentially meaningful intonational acoustic detail

Much recent progress in the domain of intonational meaning (in English) has relied on using intonational annotation systems that are phonological in nature (at least to some extent) and within the Autosegmental-Metrical tradition (e.g., ToBI [Beckman & Ayers 1997], IViE [Grabe et al. 2001], and RaP [Dilley & Brown 2005]). Any such phonological annotations require particular theoretical commitments to the number of phonological categories, as well as commitments on how those categories are realized in the intonational signal. However, recent work has noted the role of phonetic variation within/across these categories that may also be relevant to meaning (e.g., Dilley 2010, Breen et al. 2010, Ouyang & Kaiser 2015, Cole 2019, Zhou & Ahn 2019). The PIs have made similar observations in analyzing results from a pilot experiment on intonational meaning (detailed in §1.4). Taken together, these findings suggest a need to go beyond existing purely phonological annotation (e.g., IPrA [Hualde & Prieto 2016], IViE [Grabe et al. 2001], and DIMA [Kügler et al. 2015], each with somewhat different goals and premises).

To highlight the need to capture more acoustic information, consider that, by nature, a phonological label bundles together a range of acoustic cue patterns which vary in, e.g., alignment and relative scaling. This is profitable when all annotators share the same understanding of how to do this, but potentially catastrophic when they do not. Such bundling is the essence of phonological annotation, even in cases where the acoustic cues are ambiguous (cf. Fig.1) or where the number of phonemic contrasts is uncertain (cf. Fig.2).

Figure 1. Two Pitch Accents, Five Pitch Contours: How Many Phonological Representations?

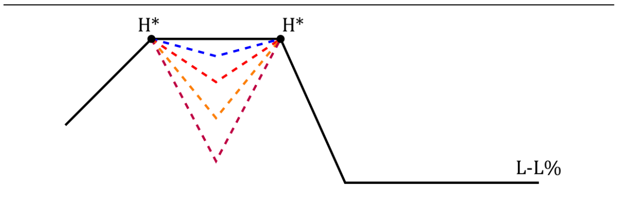
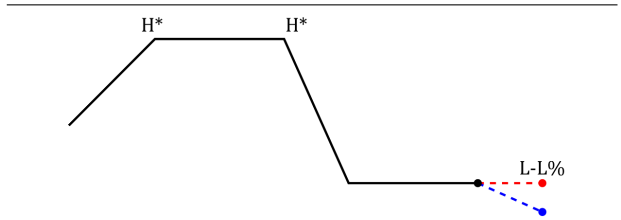


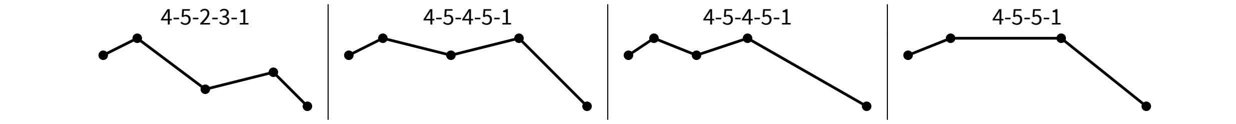
Figure 2. Two Pitch Contours: How Many Phonological Representations?



In Fig.1, while the contours with the deepest valley (bottom dotted line) vs. no valley (topmost) are often considered phonemically contrastive, some of the intermediate cases are not as clearly assigned to category labels (e.g., some may be instances of phonetic “sagging” [Ladd & Schepman 2003], but which?). Such cases can lead to interlabeller disagreement and labeller uncertainty. A forced choice for the second pitch accent’s label (e.g., ToBI H* vs. L+H*) may obscure possible systematic variation in the intermediate cases, including the possibility of (continuous) interpretive differences or phonological environments that coincide with these variants. As for contours like those in Fig.2, they are often treated as phonemically the same, but there may be distributional differences. Exploring these questions requires overt annotation of both acoustic differences and information about the surrounding contexts.

PoLaR (Ahn et al. 2018, 2019, 2020) is a prosodic annotation system recently developed to address issues like these. PoLaR is a decompositional system whose particular labels map onto individual (non-bundled) acoustic correlates of intonational characteristics (more) transparently than in a phonological annotation system, because the cues are labelled independently of one another, on separate tiers; this independence allows the labeller to remain agnostic about many aspects of prosodic phonology. (Even phonetically-oriented annotation frameworks [e.g., IPrA, IViE] bundle together disparate features like prominence, scaling, and timing, as they relate to specific phonemic categories.) At the same time, PoLaR annotation relates a richer set of acoustic-phonetic cues (i.e., turning points in the pitch contour on the Points tier, relative pitch heights on the Levels tier, and local pitch ranges on the Ranges tier) to some uncontroversial aspects of the phonology (i.e., the location of prominences and boundaries, on the Prosodic Structure tier), to capture interdependencies among cues that signal a particular event. Anno-

Figure 3. Four abstract intonational contours with PoLaR Levels annotations above.



tating these relationships requires human judgment (e.g., ignoring pitch halving/doubling errors in estimating the size of the pitch range), facilitating further automatic measures (e.g., segmental alignment of pitch movements). Lastly, PoLaR labels can be useful for the future development of automatic systems, which cannot yet distinguish between acoustic characteristics that should be labeled (and how) from those that are part of a noisy signal. (Figures 4–6 of §1.4 provide example sets of PoLaR labels, produced in Praat.)

A benefit of PoLaR is it can capture generalizations by extracting patterns of interest that are similar but not identical over multiple tokens. For example, while it is easy to visually generalize over the four stylized contours in Fig.3 as having a similar rise-{steady high/fall-rise}-fall shapes, PoLaR annotation's scaled pitch "Levels" annotations (relative to the speaker's labeled Range) can be used to find such patterns computationally, and can generate the sort of straight line stylizations shown in these figures. This can facilitate the identification of patterns of interest in order to investigate, e.g., whether these four contours distribute similarly or differently, with regard to some other annotated variable (e.g., sem/prag function/meaning).

PoLaR can be used alongside labels from other annotation systems (e.g., ToBI, RaP, IViE, IPrA), which the team will use to complement the acoustic analyses that derive from PoLaR, where appropriate. To further address our research questions about intonational acoustics, PoLaR will be used in conjunction with directly extractable acoustic measures (e.g., duration, amplitude, etc.). Lastly, PoLaR is conceived as an extendable framework; plans are underway to notate cue strength (e.g., duration, intensity, pauses, and voice quality; cf. Brugos et al. 2018) as they relate to labels on various PoLaR tiers (e.g., PrStr tier labels; see Ahn et al. 2019, 2020). In sum, PoLaR was designed to supplement existing transcription systems by providing a greater degree of acoustic-phonetic granularity that is nevertheless informed by intonational intuitions.

1.3 Context Presentation with Comics: Experimental Methodology

In experiments designed to elicit speech, task-oriented paradigms (e.g., a map task [cf. Anderson et al. 1991], or a game [e.g., Ito et al. 2004, Zhou & Ahn 2019]) are advantageous in that they elicit speech that is quite naturalistic, while also allowing for control or manipulation of sem/prag context by constraining possible situations and conversational moves. However, in such tasks, the text of the speech can be quite variable, complicating comparisons across speakers and across scenarios.

Other paradigms aim to keep the text of the speech constant, and manipulate sem/prag variables via text-based contexts (e.g., through written scripts or instructions). However, at least anecdotally, this sort of text-based context manipulation may not result in consistent prosodic behaviors, both within and across speakers. Such inconsistent results suggest that participants may not have internalized the context as intended by the experimenters. One reason could be that the explicitly provided context is vague enough to allow speakers to "fill in the gaps" differently. Another reason could be that the presentation of context as text is difficult for a speaker to internalize, remember, and then reflect in their speech. In both of these cases, speakers in the experimental task may vary in their internal beliefs about the context, resulting in variable prosodic realizations (cf. Zubizaretta & Vergnaud's [2006:523] discussion of the idea of "context-neutral").

To address this, we make use of a methodology in which contexts are presented visually, as a series of pictures, which provide a reader with a naturalistic way to experience a rich sem/prag context. We have three reasons for using this methodology. First, research has shown that pictures help in encoding information in memory (Glenberg & Langston 1992), which is important when participants will need to produce/perceive intonation with context in mind. Second, cartoons have been shown to be effective in establishing sem/prag context in a number of domains, such as in elicitation tasks with underdocumented languages (cf. Burton & Matthewson 2015), in first-language acquisition research (e.g., Truth-Value Judgment tasks; Crain & McKee 1986), and in evaluation of L2 pragmatic capabilities (e.g., Cartoon Oral Production Tasks; Rose 2000). Finally, a pilot experiment (described more in depth in §1.4 below) suggests that participants reliably internalize the sorts of complex contexts that we are testing, on the basis of the comic stimuli. For these reasons, this method holds great promise for delving deep into the study of intonation, and will be used throughout this project in ways that will be further described below, for each proposed experimental task.

1.4 Pilot Experiment: Example of Comics-elicitation and PoLaR analysis

Our experience in a pilot experiment (Ahn et al. 2016, *in prep*) suggests comics can effectively communicate subtle sem/prag contextual information, since intonation was observed to vary according to sem/prag variables that defined the comics. In this case, the variables concerned the type of evidence (direct, indirect, none) that the speaker and hearer each have about a proposition. The 2016 findings were cast in terms of which ToBI pitch accents and boundary tones were found. At the same time, the authors observed intonational patterns that distinguished the different conditions that the categorical ToBI labels were not adequately capturing.

As an example, consider the three comics in upper portions of Fig.s 4–6, which are a subset of 6 different evidential conditions from the task. Each comic presents a context establishing who has what evidence about the proposition, p, _the copier is out of order again_. Below each comic are annotated pitch tracks of productions by a single speaker (Participant 20) for the target (last) utterance of the comic – note that the contour differs based on context encoded in the comic.

Figure 4. Speaker (left) has only indirect evidence for p; Hearer (right) has direct evidence. Note the final rise in the labeled pitch track below (ToBI: H-H%).

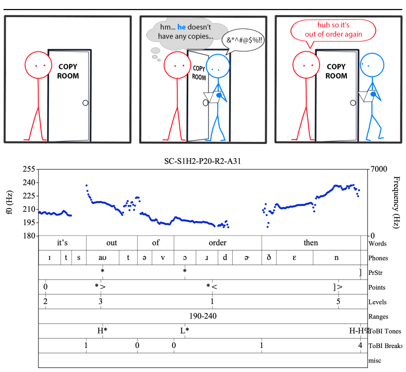


Figure 5. Speaker (right) has direct evidence for p; Hearer (left) has no evidence. Note the final low, flat pitch (ToBI: L-L%).

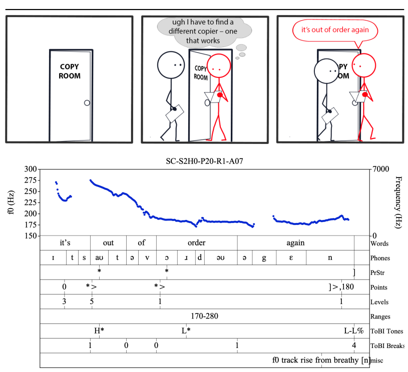
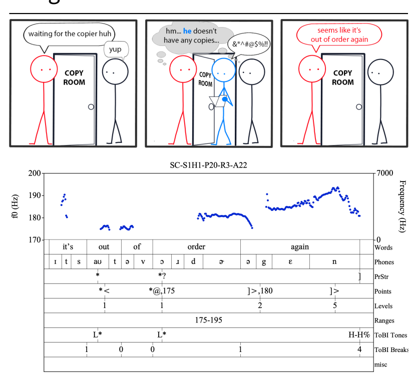


Figure 6. Speaker (left) and Hearer (rightmost) both have indirect evidence for p. Note the same H-H% final rise as in Figure 3, but with differences in the Ranges tier.



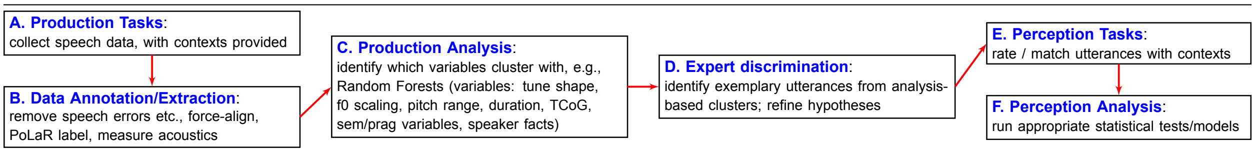
The contrast between Fig.4 and Fig.5 exemplifies that boundary tone type, as encoded phonologically, might be encoding some of these sem/prag differences. However, the phonological differences between Fig.4 and Fig.6 are not as obvious: both end in a /L* H-H%/ ToBI sequence. However, the former production has a much wider local pitch range, 190Hz-240Hz, as compared to the latter, 170Hz-195Hz (encoded in the PoLaR Ranges tier). Considering these sorts of acoustic labels alongside the careful manipulation of the context of these comics (and by proxy, sem/prag variables in the mind of the speaker) can help to reveal which, if any, acoustic differences are cues to meaningfully distinct (i.e., phonemic) categories.

The pilot study described here is a production task, but these comics-based representations of context will be useful tools for perception tasks as well, as described in the proposed experiments below.

1.5 Three Experimental Sets

Returning to our organizing questions (§1.1) about the relationship between intonational variation and semantic/pragmatic meaning, we propose to employ the methodological tools outlined above (comics-based context manipulation, PoLaR annotation) to execute the three planned sets of experiments (common ground management, information structure, and quantifier scope) in order to explore the relation between intonational form and meaning. Each of these experimental sets will involve the same general paradigm, visualized in Fig.7, with some variation in the details, to be determined by the specific goals of each experiment.

Figure 7. A flowchart exhibiting some general logic of the workflow for the proposed sets of experiments. Individual experimental sets are variations on this theme, with respect to specifics.



The first step (Fig.7A) for all studies involves a production task (eliciting speech using context-specific comics), with the goal of identifying the basic intonational patterns in the acoustics (addressing Q1 in §1.1). These production tasks will be conducted either in person, using standard lab protocols, or remotely, as has been successfully undertaken in recent works (Buz et al. 2016, Heidenreich 2019, Jennifer Cole *p.c.* 2020). Financial incentives will be offered as compensation, following standard IRB practices. In order to partially control for inter-speaker variation, we will recruit a gender-balanced sample of younger adults (aged 25-50) who are monolingual native speakers of English, born and raised in the United States. Previous research suggests that 25 participants per production task will be adequate (see, e.g., Ferreira 1993, Beaver et al. 2007, Katz & Selkirk 2011, *inter alia*) for documenting patterns, for testing initial hypotheses, and for generating more refined hypotheses about Mainstream US English (MUSE; Lippi-Green 2000). Focusing on MUSE speakers, as the intonationally best-investigated form of American English, will enable comparisons with existing literature. That said, no subject population is monolithic. Demographic information about each participant (e.g., age, hometown, racial and gender identities, educational background, and socio-economic background) will be collected, which may lead to identifying and understanding patterns that are outside of MUSE varieties, as well as other sociological sources of intonational variation, in the analysis process.

The next step (Fig.7B) is to build a corpus from these recordings by removing utterances with speech errors and annotating each utterance file with speaker information, sem/prag variables, PoLaR labels, and extracted acoustic cues (e.g., scaling and alignment). Using this corpus, we will begin the process of addressing Q2/Q3 (§1.1) by first (Fig.7C) using data mining/machine learning techniques. Depending on the specific experiment, candidate algorithms include Random Forests (cf. Arnhold & Kyröläinen 2017's success in identifying multiple acoustic cues to focus in Finnish and Villarreal et al. 2020's classification of /r/ variation; also Hain et al. *in prep*), regression models (e.g., Zhou & Ahn 2019 to determine the connection between spurious peaks and sem/prag context), Principal Component Analysis (e.g., for acoustic feature analysis, as in Ward 2019), Functional Principal Component Analysis (e.g., for intonational curve fitting in categorical contexts, as in Lohfink et al. 2019) or clustering techniques (e.g., k-means, as in Rosenberg 2013) to identify which acoustic features best predict the sem/prag context in which it was produced. Once these acoustic patterns are identified through this analysis, the research team will then (Fig.7D) identify exemplary utterances from identified patterns, and generate (newly refined) hypotheses for which aspects of intonational variation have grammatical underpinnings, especially related to sem/prag factors.

This enables the next experimental task (Fig.7E), in which participants listen to these exemplary utterances, and judge the appropriateness of intonation in different sem/prag contexts (to address Q2/Q3 of §1.1). The goal for these perceptual tasks is to investigate which intonational patterns used by a speaker in these sem/prag contexts match those used by a listener in understanding the relevant sem/prag meanings. We will recruit adult listeners with unimpaired hearing from the same demographic pool, again offering IRB-approved compensation and collecting demographic information. The perception tasks will be conducted via Mechanical Turk and Prolific (using best practices to verify continued participant engagement; cf. Mason & Suri 2012.). Since judgment tasks differ over our three experiments (and may depend on the production test results), they are described below in more detail for each experiment. For each new task, the research leads and consulting statistician will identify the appropriate participant pool sizes based on estimated effect size. Recruiting large numbers of participants will be facilitated by using online platforms, and work on intonational perception has been successfully carried out over such platforms (e.g., Buxó-Lugo & Watson 2016, Burdin & Tyler 2018, Jeong 2018a,b, *inter alia*).

Finally (Fig.7F), the perception task results will be analyzed, to identify the relationship between the phonetic/acoustic features and the sem/prag variables, as well as to test whether any sociological variables are predictive of differences in measured behaviors. In preparing for this stage, the research leads and statistician will again collaborate to determine the appropriate statistical models for analysis.

More specifics for each production task (Fig.7A), perception task (Fig.7E), and the relevant hypotheses, statistical tools, and analytical models (Fig.7C/F) can be found in the following sections, which outline the specifics of the three studies, each involving distinct sem/prag domains linked to the kinds of intonational variation outlined in our organizing research questions (§1.1). While probing the domain of intonational meaning (i.e., the relationship between intonational forms and sem/prag interpretation) in Stalnakerian/Lewisian models of discourse structure (Stalnaker 1978, Farkas & Bruce 2010), these experiments will yield new insight into phonetic variation in intonation and its relationship to phonological models/inventories in an A-M phonological framework (Pierrehumbert 1980, Dilley & Breen 2018), as well as insight into the representation of such intonational chunks of meaning (i.e., morphemes) in the morphosyntax in a Distributed Morphology framework (Halle & Marantz 1993, Bobaljik 2017).

2 Experiment Set 1: Common Ground Management and Intonational Meaning

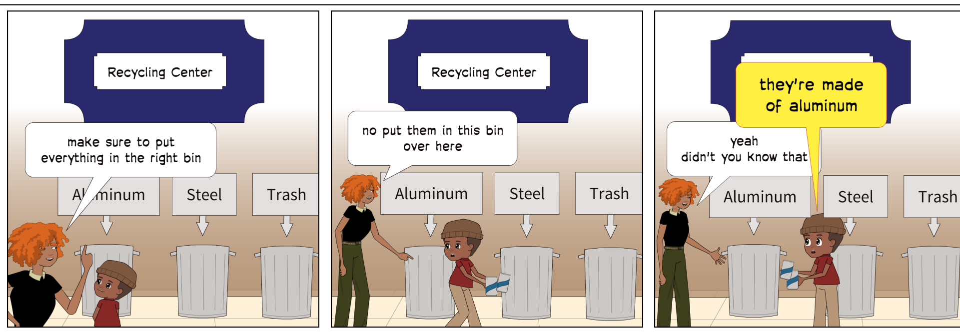
2.1 Background

One function of intonational tunes relates to what Repp (2013) calls ‘common ground (CG) managing’ functions. Ideas in this vein were already present in Pierrehumbert & Hirschberg 1990, but advancements in Stalnakerian/Lewisian approaches to discourse and dynamic semantics have enabled development in a variety of directions more directly grounded in semantic theories. In this domain, a primary function of intonation is to signal how the context should be updated with the content/proposition under discussion: Is the speaker conveying intent to update the CG with the content?, Is she inviting the hearer to resolve the issue raised by the content?, etc. For instance, Gunlogson (2004) has claimed that a final-rise tune attributes commitment to the truth of the proposition to the *hearer*, whereas the final-fall tune attributes the commitment to the *speaker* (see also: Truckenbrodt 2012, Malamud & Stephenson 2015, Westera 2017, Jeong 2018b, Rudin 2018, a.o.). Others have argued that intonation can even operate at the level of truth-conditional semantic composition to determine the type of content that will update the context: for instance, final-rise tunes may transform the assertive propositional content of a declarative into inquisitive content (e.g., Farkas & Roelofsen 2017, Jeong 2018b), relating to the sentential force / mood of the utterance (generally accepted to be represented in the syntax). The cited works have focused on CG-managing aspects of intonation, and results suggest that tunes may signal different discourse moves that work towards updating the CG with the at-issue content. However, only a few forms (often discussed

Table 1. Experimental design for three dimensions of contextual features: 1) Degree of evidence of a proposition validity held by the Speaker compared to the Addressee 2) Prior bias of the Speaker (True or False) of the proposition and 3) Whether the contexts indicates that the speaker is seeking information from or providing information to the addressee

Cond.#	**Evidence of p**	**Prior bias**	**Force**	**Cond.#**	**Evidence of p**	**Prior bias**	Force
1	Spkr $=$ Hearer	$\neg p$	Inquisitive	5	Spkr $<$ Hearer	$\neg p$	Inquisitive
2	Spkr $=$ Hearer	$\neg p$	Assertive	6	Spkr $>$ Hearer	$\neg p$	Assertive
3	Spkr $=$ Hearer	p	Inquisitive	7	Spkr $<$ Hearer	p	Inquisitive
4	Spkr $=$ Hearer	p	Assertive	8	Spkr $>$ Hearer	p	Assertive

Figure 8. A three panel cartoon stimulus that situates the participant in the experimental context (Spkr < Hearer, $\neg p$, Inquisitive).



as idealized contours) have been studied: e.g., the binary distinction between final-rise vs. final-fall tunes (or between H- and L-). Additionally, little attention has been paid to what types of additional enriched inferences may emerge in context from the abstract, conventionalized meanings of varieties of final-rise vs. final-fall tunes.

To address these gaps, we propose a pair of tasks, described below. The findings will enrich our understanding of the form domain by using insights gained from analyzing PoLaR-annotated utterances. Specifically, we aim to expand the range of investigated prosodic phenomena that have close connections with CG-managing functions, and determine whether potential differences in meaning are generated by different varieties of nuclear pitch accents (see Ahn et al. 2016 for approaches in this vein), as well as differences in slopes of the F0 contour (Jeong 2018a,b). Moreover, the contexts used are manipulated to vary systematically along salient pragmatic dimensions known to affect linguistic form (evidentiality, prior epistemic bias of interlocutors, and intended speech acts), which are likely candidates for enriched second-order inferences that can be derived from core meanings of the tunes (whatever they may be).

2.2 Production: Hypotheses and Task 1.1 Design

Our first experiment will involve the following contextual manipulations: (i) how much evidence the speaker believes they have about a proposition p, relative to the hearer, (ii) the speaker's epistemic prior bias about p (positive or negative) and (iii) whether the speaker intends to ask/inquire "whether p" or assert "that p". By investigating the three types of inferences in tandem and in multiple combinations (see Table 1)), we seek to clarify the bounds and landscapes of inferences that can arise from the core meaning of a tune.

This experiment and the other production experiments proposed below will follow a general paradigm, by presenting contextual conditions as cartoon strips and instructing the participant to read each panel, taking on the role of the rightmost figure and speaking that figure's final utterance, highlighted in yellow. An example is shown in Fig.8 (where participants are instructed that punctuation has been purposefully omitted). We have the following hypotheses about the results.

Hyp.1.1. When the Speaker has less evidence than the Hearer, the most prominent word will correspond to low values in Levels tier annotations (i.e., a low pitch accent, L* in ToBI; Ahn et al. 2016)

Hyp.1.2. Holding a prior bias of $\neg p$ will lead a Speaker to utter p with intonation that marks surprise (e.g., the surprise-redundancy contour, Sag & Liberman 1975)

Hyp.1.3. Inquisitive force will result in steeper upward slopes in the boundary tone (Jeong 2018b)

2.3 Acoustic Analysis

Our analyses of the production data described below will be informed by the broader semantic/pragmatic literature in constructing a derivational story where the empirically observed contextual inferences arise from a systematic interaction between the input context, the core meaning of the tunes, and pragmatics. Rather than assuming that the intonation will directly encode contextual aspects, we will be careful in interpreting correlations between prosodic cues and contexts.

The production data will be annotated using PoLaR (§1.2), and segmented automatically using a forced aligner (e.g., the Montreal Forced Aligner; McAuliffe et al. 2019). This will allow us to extract a set of features related to our hypotheses: raw and normalized F0 at PoLaR points, F0 scaling, and F0-scale-defined tunes (sequences such as “1-4-2” for a Rise-Fall), segment durations, Tonal Center of Gravity (TCoG), etc. Using statistical tools (potentially PCA to explore feature combinations followed by Random Forests) to determine the most relevant features, we will determine which factors co-occur under which conditions and these findings will be used to manipulate re-synthesized speech to use as stimuli in the perception experiment described in the next section.

2.4 Perception: Hypotheses and Task 1.2 & Task 1.3 Design

As described generally in §1.5, production results will provide data on which prosodic variations occur in which context; these data will be used to evaluate and refine hypotheses 1.1-1.3, and potentially generate new ones. We will test these refined hypotheses by way of a perceptual experiment designed to further vet our results. In task 1.2, perception stimuli will consist of productions from task 1.1 that are typical of the patterns found in the analysis (§2.3). Participants will be presented with a recording and asked to identify the appropriate comic-context for that utterance, from a set of three comics in a forced choice design. The set of choices will contain a matched choice (the comic previously used to elicit the recording) and two distractors. Task 1.3 is similar, but stimuli will be constructed through resynthesis, using neutral utterances (produced in a frame sentence to dictate minimal prosodic features), and re-synthesized to include different numbers of prosodic cues identified in the production analysis (§2.3). The degree of anticipated match in task 1.3 will enable further evaluation of which intonation cues generate consistent meaning differences. Perception experiments adopting similar key features of this experimental paradigm were conducted by Jeong & Condoravdi (2017) (see also Jeong 2018a,b).

2.5 Perceptual Judgment Analysis

The results from the perceptual experiments are, in general, more straight-forward to analyze than those of the production experiment. On the most general level, the results will indicate if the prosodic features used in the re-synthesis are correlated with participants’ judgments of the “correct” semantic-pragmatic context. We will rate each context-match judgment by its closeness to the target context (e.g., preliminarily, a judgment that matches in 2 of 3 dimensions is rated a 2; this will be refined if some dimensions are found in the production analysis to independently carry more weight than others). For this experiment, the distance metric will take into account all three of the separate dimensions: evidence, bias and force.

3 Experiment Set 2: Information Structure and Phonemic Inventory of MAE

3.1 Background

The connection between intonational contours and information structure (I-S) is perhaps one of the best studied relationships between intonation and meaning, in particular with regard to the meaning of diverse pitch accents (e.g., /H*/, /L+H*/, /L*/; Prince 1981, Pierrehumbert & Hirschberg 1990, Schwarzschild 1999, Büring 2007). These approaches posit that choice of pitch accents contributes to conveying different I-S properties/statuses, such as given-ness, newness, contrastive topic, etc. One area that has received special attention in the intonation literature is the two-way distinction within the umbrella category of I-S 'focus': '(discourse) new' versus 'contrastive (focus)'. In their often-cited work on the meaning of ToBI intonational atoms, Pierrehumbert & Hirschberg (1990) say "the items made salient by the H* are to be treated as 'new' in the discourse" while L+H* is used "to mark a correction or contrast". More recently, there have been investigations into the acoustics of these 'new'/'contrastive' I-S categories (e.g., Watson et al. 2008, Breen et al. 2010, Calhoun 2012, Chodroff & Cole 2019, Roettger et al. 2019).

Because identifying elements as given/new/contrastive in spontaneous speech can be challenging, this work on spontaneous speech has been complemented by work on laboratory speech that uses question prompts to establish a particular I-S status on various elements in a sentence. For example, it is well established that in an answer to a WH-question, the element that corresponds to the WH-phrase in the question will have focus marking (Question-Answer Congruence, QAC; cf. Halliday 1967, Krifka 2004). This QAC generalization is well established and holds quite robustly (e.g., Breen et al. 2010).

Many recent investigations in the sem/prag literature have aimed to model interactions in terms of questions under discussion (QUDs) and their responses (Roberts 1996, 2004, Büring 2003, Beaver & Clark 2008, Ginzburg 2012, Ciardelli et al. 2013, Farkas & Roelofsen 2017, among many others). Not all QUDs are explicit, though, which may result in apparent violations of QAC. Consider the following examples ((1) is modified from Büring 2003), in which the question that B answers is not the precise question A asked:

(1) A: What did the pop stars wear?
B: The FEMALE pop stars wore CAFTANS.

(2) A: Who made those cookies?
B: MARIANNA made them.

For (1), the explicit question in is an object WH-question, but B's (felicitous) response has focus marking within the subject as well as the object. Spuriously adding focus marking that doesn't adhere to QAC is disallowed (Schwarzschild 1999); however, a QAC violation is avoided if B is answering an (implicit) subquestion (à la Roberts 1996) of A's question: "Which pop stars wore what?". In other words, the prosody of the response indicates that B is answering an *implicit* question related to the *explicit* one. Such implicit subquestions can introduce salient focus alternatives, allowing a response to contain elements with contrastive focus marking, despite the fact that the explicit question would not otherwise license such alternatives.

In a similar way, in (2), the focus in B's answer, *Marianna*, can have an I-S status that is not licensed by the question A asks explicitly. Because A's question lacks salient focus alternatives, *Marianna* is predicted to be a 'new' focus. However, *Marianna* can indeed be a 'contrastive' focus if B is addressing an implicit question like "Did Marianna or Luther make these cookies?", which can introduce *Luther* as an alternative. Thus text-based question-answer pairs, in the absence of richer context, are open to interpretation by the experimental participant such that they could be crossing categorical boundaries in their answers.

Our goal of the proposed experiments here is to connect intonational research on I-S to research done on I-S in other domains. In particular, we aim to build on recent advances in the understanding of QUDs and discourse structure (as just described) to construct naturalistic visual stimuli to signal whether an answer responds to an explicit question or some other subquestion. Additionally, the analysis of production data and construction of perception stimuli will target particular acoustic cues that are relevant to intonation with PoLaR annotation. Finally, this work aims to avoid syntactic confounds in the stimuli/analysis, based on previous findings related to the relationship between syntax and I-S (e.g., Selkirk 1996, Zubizarreta 1998, Wagner 2006, Ahn 2015, Kratzer & Selkirk 2020).

We propose production and perception tasks, to better understand the acoustic (and phonological, after some analysis) intonational counterparts to I-S statuses. Specifically, we use question-answer pairs and comics to set up rich sem/prag contexts in the mind of the participant, and then measure (i) the acoustic changes with production tasks and (ii) interpretation of acoustic changes with perception tasks. Our selection of sem/prag contexts draws on current research in the burgeoning area of discourse as it relates to the sem/prag of spoken questions and answers.

3.2 Production: Hypotheses and Task 2.1 Design

Like previous experiments on the intonational acoustics of I-S, this production task will present question-answer contexts, manipulated to affect the answer I-S. We will present three conditions of salient alternatives to the focused element: the comic-based context provides (i) some salient alternative(s) made explicit by the preceding question (Fig.9), (ii) some salient alternative(s) left implicit by the preceding question (Fig.10), or (iii) no salient alternatives provided at all (Fig.11). Thus these conditions differ on the salience of alternatives, which can determine the choice 'contrastive' and 'new' I-S status, leading to Hyp.2.1.

Figure 9. Explicit salient alternatives in the question.

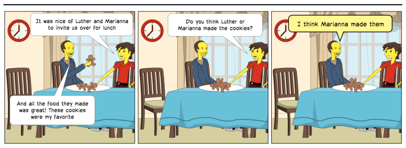


Figure 10. Implicit salient alternatives in the context.

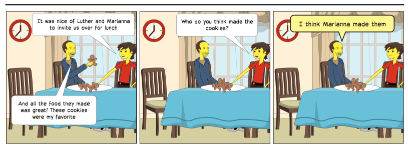


Figure 11. No salient alternatives in the question or in the context.

Hyp.2.1. Contexts that do have salient alternatives (e.g., Fig.9–10) are predicted to result in 'contrastive' I-S focus status in the answer, while contexts that do not (e.g., Fig.11) are predicted to result in 'new' I-S focus status in the answer.

Based on findings in the literature (especially Breen et al. 2010 and Ouyang & Kaiser 2015), we expect that conditions with explicit salient alternatives ('contrastive' ones) will lead to productions that acoustically differ from productions from conditions with no available salient alternatives ('new' ones) . In particular, we expect 'new' foci to be realized with more gradual, subtler rises, as opposed to 'contrastive' foci being realized more prominently in various ways. Using PoLaR-labels to implement the Tonal Center of Gravity (TCoG) model (which has been shown to be relevant to perceptual discrimination of different types of rises (Barnes et al. 2010, 2012), Hyp.2.2's acoustic predictions can be stated in more general terms as Hyp.2.3.

Hyp.2.2. In the automatically extracted acoustic measures (many facilitated by PoLaR labels), 'contrastive' focus will, compared to 'new' focus, manifest as: **(a)** louder relative to surrounding speech, **(b)** longer, **(c)** with a steeper slope (as measured between pitch turning points on the Points tier), **(d)** as a greater number of pitch turning points associated with the relevant pitch accented word, as annotated on the Points tier, **(e)** with pitch turning points that are more at the extreme edges of the pitch range, as annotated on the Levels tier, and/or **(f)** in bigger pitch ranges, as annotated on the Ranges tier.

Hyp.2.3. 'New' foci will have early TCoGs, and 'contrastive' foci will have late TCoGs.

We anticipate that better-controlled context (encouraging/discouraging implicit subquestions) will result in more participant consistency in I-S parsing, so the acoustic differences between these I-S categories may be more distinctive. But since other (implicit) subquestions can be created by the speaker even where discouraged, we expect that I-S 'contrastive' focus marking will be found with some frequency in all contexts. On the other hand, we do not expect to find marking of I-S 'new' focus marking in the answers to questions that directly license salient alternatives creating an asymmetric distribution, leading us to Hyp. 2.4:

Hyp.2.4. There will be distributional overlap between intonational forms and these QUD-defined contexts, in that **(a)** 'contrastive' focus will occur with some regularity in all contexts, and/or **(b)** 'new' focus will occur infrequently in response to a QUD that directly licenses salient alternatives.

The phonological contrasts proposed for marking focus have been thought to be categorical (e.g., H* vs. L+H* in Pierrehumbert & Hirschberg 1990), and so are the meaning contrasts involved ('new' and 'contrastive'). However, the acoustics that signal the putative phonological contrasts are gradient and the interpretive properties that signal the putative meaning contrasts may also be gradient. This leaves us with two competing hypotheses about the phonological analysis of the findings:

Hyp.2.5. Two competing hypotheses about phonological categories:

- **(a)** There are indeed two phonemic categories, 'new' and 'contrastive' (potentially /H*/ vs. /L+H*/), with phonetic variation such that there is overlap/neutralization in some contexts
- **(b)** There is a single phonemic category, with gradient intonational variation for 'new' vs. 'contrastive'

A priori, we do not know which of these two hypotheses does a better job for predicting intonational production. Moreover, as stated in Hyp.2.4, we predict distributional overlap from this production study. Thus we will return to this hypothesis with a perception task (Task 2.2), which we will describe in §3.4.

3.3 Acoustic Analysis

As in the previous experiment, production data will be annotated using PoLaR (§1.2), and segmented using the automatic segmentation of a forced aligner (McAuliffe et al. 2019). This will allow us to extract the set of features related to our hypotheses: F0 slope, intensity, TCoG, etc. Using statistical tools such as those described generally above (e.g., K-means to explore if features align with clusters/categories), we will determine whether there are, in fact, predictable distinct distributions corresponding to our QUD contexts and therefore whether the acoustic distributions plausibly map to H* vs. L+H* categories. If we are able to find two distributions, we will determine which factors contribute most significantly (e.g., with Random Forests) and use this information to manipulate re-synthesized speech that will serve as stimuli in the perception experiment described in the next section.

3.4 Perception: Hypotheses and Task 2.2 Design

As in perception task 1.2 on CG management, we predict overlapping distributions (cf. Hyp.2.5), such that particular intonations may be compatible with different types of QUDs. To probe the relationship between the (potentially ambiguous) intonation with meaning, we will run a task that involves perception and discourse completion. This task (again like task 1.2) will use the insights gained from the paired production study to create a set of resynthesized stimuli from a neutral base utterance. The specifics of these manipulations will be determined by the acoustic parameters that present as significant in the production analysis. For example, we might find a categorical distinction that corresponds to notions of H* and L+H* and these two categories are produced under specific QUD contexts. Of course, the non-categorical outcome is also possible. In this case we would present manipulations based on gradient steps from endpoints found in the production data.

In this variant of a Discourse Completion Task (cf. Brown 2001), participants will read/listen to an incomplete comic that is missing a key utterance in the penultimate panel. They will be instructed to listen to a recording of the final cell, and, on the basis of that recording, provide (in writing) a free-response answer to the prompt: "What question was asked right before the answer in the final panel?".

We hypothesize that the I-S of the questions participants provide will differ depending on intonation, and further that there may be a relationship between responses and gradient acoustic information in the stimuli.

Hyp.2.6. The questions provided as written responses will be ones that evoke **(a)** multiple focus alternatives when presented with production stimuli from the multiple-alternative contexts, and **(b)** either multiple or no alternatives when presented with production stimuli based on no-alternative contexts.

Hyp.2.7. The strength of cues to different I-S categories will have an effect on the frequency of question-type provided as a response in this free-form response task.

Figure 12. Participants are given a comic-context and an auditory stimulus, corresponding to the final speech bubble (yellow). They then asked to fill-in-the-blank with a question that could prompt the final (yellow) speech bubble as produced in the auditory stimulus.

3.5 Perceptual Judgment Analysis

The freeform responses will be coded (e.g., for I-S status) using methods common in qualitative research. Using Grounded Theory, we can induce the set of QUD categories/features to create tags to annotate each fill-in-the-blank response, q, based on existing sem/prag and syntactic taxonomies of questions (e.g., Ciardelli et al. 2013, Farkas & Roelofsen 2017). For example, tags may include: (i) whether q presupposes salient alternatives, (ii) q's alternatives are a closed vs. open set, and (iii) q's syntactic form (e.g., WH Q vs. polar Q vs. alternative Q). Regression models will be used to see which tags significantly map to which intonational production, allowing us to postulate a mapping between aspects of the meaning and intonation.

4 Experiment Set 3: Conventionalized Intonational Meaning and Scope Ambiguity

4.1 Background

A third way in which intonation can signal sem/prag meaning is the association of certain stylized contours with conventionalized meanings; this mechanism has been argued to contribute a separate track of meaning ("not-at-issue content") calculated in parallel to the "at-issue" content (characterized as conventional implicature). For example, the calling contour (/H* !H-L%/; Ladd 1978, Jeong & Condoravdi 2017) does not affect the at-issue content of the clause (and can therefore be viewed as an implicature) and the particulars of its meanings are non-compositional (i.e., are conventionalized). Another tune that has been ascribed a conventionalized meaning is the Rise-Fall-Rise (henceforth "RFR") contour. The RR contour has been argued to convey the speaker's understanding that she cannot safely claim the (Roothian) focus alternatives (Constant 2012). Here we focus on aspects of the RFR contour's form and meaning.

Starting with the form, the designation of "RFR" refers not to just any sequence of rise, fall, and rise, but to a specific contour. The literature (e.g., Ward & Hirschberg 1985, Büring 2007, Constant 2012) breaks breaks RFR down into two key components: (i) a low intonational target on the main sentence stress with a trailing high (ToBI: L*+H), and (ii) a phrase-final low into a final rise at the phrase boundary (L-H%).

Interpretively, the RFR contour is known to contribute conventionalized meanings, affecting relative scope between semantic operators; e.g., between a universal quantifier ($\forall$) and negation ($\neg$). Consider a sentence like (3), which is ambiguous between two possible readings, paraphrased in (a) and (b) below.

(3) All of us didn't order a cheeseburger.

a. <u>All</u> of us are such that we didn't [order a cheeseburger]. ($\forall > \neg$; surface scope reading)

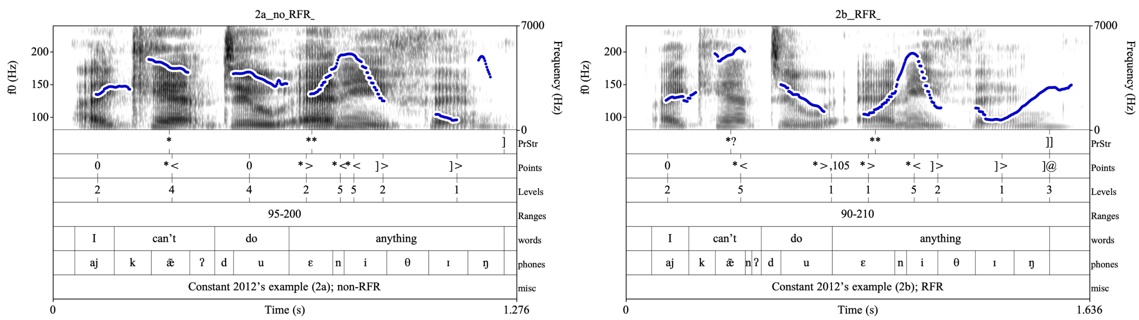
b. It isn't [that <u>all</u> of us ordered a cheeseburger]. ($\neg > \forall$; inverse scope reading)

Above, the scope of negation *n't* is marked with brackets. In (3a), the universal quantifier *all* is outside of it ($\forall > \neg$), and *all* is also outside the syntactic constituent with *n't*; i.e., the scope and surface syntax match: the "surface scope" reading. On the other hand, in (3b), *all* is interpreted within the scope of negation ($\neg > \forall$), in spite of the surface syntax: the "inverse scope" reading. While both readings are available with

Table 2. Inverse scope contexts support both default and RFR contours; surface scope contexts do not support RFR contours.

	Non-RFR Contour	RFR Contour
Surface-Scope Context	✔	✘
Inverse-Scope Context	✔	✔

Figure 13. PoLaR annotation of Constant 2012's recordings of the same string, with a surface scope interpretation (left, non-RFR contour) and an inverse scope interpretation (right, RFR contour).



"default" intonation (ToBI /H* L-L%/), it has been argued that certain tunes can force one reading (Ward & Hirschberg 1985, 1988, Hirschberg & Ward 1992, Constant 2012, 2014). In particular, Constant (2012) argues that the (conventionalized) interpretive contribution of RFR contour (alongside the semantics of these operators) is such that it "unambiguously convey[s]" (*ibid.* p.408) inverse scope readings. This is laid out in Table 2.

This interpretive contribution of the RFR contour ("RFR interpretation", henceforth) makes it an ideal domain to investigate intonational variation in well-defined sem/prag contexts.

We investigate two sets of questions concerning RFR contours (form) and RFR interpretation (meaning), and their interdependence: (i) what phonological/phonetic variation exists within the "RFR contour" (§4.2) and (ii) which components of the RFR contour are necessary for the availability of an RFR interpretation and whether its availability is gradiently dependent on gradient phonetic variability (§4.4).

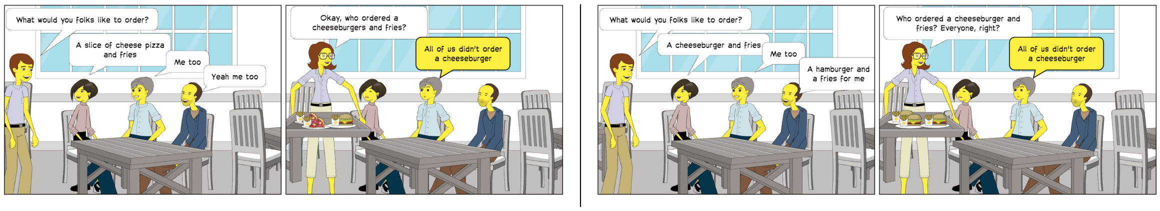
4.2 Production and Perception: Hypotheses and Tasks 3.1 & 3.2 Design

Consider the examples of RFR and non-RFR tokens (Fig.13) from Constant 2012 (cf. *ibid.*, footnote 1): impressionistic similarities in shape and timing of the initial rises in the examples in Fig.13 may lead to interlabeller disagreement, labelling difficulty (especially for novices), or outright ambiguity – even for experienced labellers who might generally be confident in their abilities to distinguish different phonemic (e.g., L+H* vs. L*+H) rises (Barnes et al. 2012, p.c.). This may in part be due to shared acoustic properties between these two cases, and the tendency for different labellers to attend to acoustic cues differently. Thus our goal is to investigate the acoustics that distinguish RFR and non-RFR contours, as well as those that are shared.

Thus data should not be categorized as RFR and non-RFR not the basis of phonemic labels. Instead, this experimental set will use a production and a perception task to uncover what defines an RFR contour intonationally. To do so, we exploit the previous findings in Table 2 in a multi-step process, in which we elicit speech (task 3.1), then run a perception task with that speech (task 3.2), and finally analyze the intonational differences between RFR and non-RFR. The initial production task uses the methodology described above in tasks 1.1 and 2.1. Comic-strip stimuli are designed to evoke either the surface scope meaning (Fig.14, left) or the inverse scope meaning (Fig.14, right).

Before we can analyze the phonetic variation in "the RFR contour" elicited in this production task, we must identify which of the gathered utterances instantiate the RFR contour. (This step is necessary

Figure 14. Both comics (left and right) have QUDs that pick out alternatives on the subject; while the comic on the left unambiguously picks out a surface scope reading, the comic on the right unambiguously picks out an inverse scope reading.



because elicitations from inverse-scope contexts could be either RFR contours or not; Table 2.) Rather than circularly using preconceptions about an intonational definition of RFR, we will identify contours as RFR by (again) leaning on the interpretive properties of RFR, and running a perception experiment (task 3.2). Task participants will be shown a comic depicting a surface-scope interpretation (i.e., the context claimed to uniquely distinguish the acceptability of RFR contours from non-RFR, cf. Table 2) and will be asked to rate the felicity of an auditory stimulus drawn from the production task. These previous findings inform our hypotheses about the perception task results:

Hyp.3.1a. When listeners perceive a contour as being an "RFR contour" in surface-scope contexts, they will give very low felicity ratings, as the "RFR contour" category unambiguously maps onto inverse-scope readings.

Hyp.3.1b. If listeners rate a contour highly in surface-scope contexts, it is not an RFR contour.

We will use felicity ratings to bifurcate our production data into RFR and non-RFR contours and conduct phonetic analyses (described below) to explore the following hypotheses:

Hyp.3.2. RFR (vs. non-RFR) contours will have a distinct phonetic profile in our analysis:

- **(a)** RFR contours will occur with wider pitch range domains, and with pitch turning points at more extreme edges of these ranges (as reflected through Points and Ranges tier annotations).
- **(b)** The low part of the pitch accent (i.e., preceding the initial rise) in RFR contours will be reflected in a greater number of pitch turning Points (to reflect the scoopier quality), and the high part of the pitch accent will have a later alignment in terms of Tonal Center of Gravity.

4.3 Analysis of Perceptual Data and Acoustic Analysis

We will use a percent judgment threshold to identify the RFR and non-RFR contour categories, on the basis of the unavailability of surface-scope readings in task 3.2. We will explore the variation across these functionally defined sets through the analysis methodology described above in §2.3 and §3.3 for the production data, and on this basis both the surface and inverse scope recordings will be examined for intonational patterns. Having isolated those utterances that are more likely to exhibit the RFR pattern, we can then use our hypotheses as a guide to compare the acoustic patterns of those RFR contours with the non-RFR contours. Moreover, in the event that hypothesis 3.2 is not eliminated, we will be able to quantify the extent of the variation (e.g., that the pitch range domain is 40% wider). Machine learning techniques (Functional PCA may be particularly useful in modeling the RFR contour's shape), using perception judgments as the independent variable and the acoustic features as dependent variables, will further identify the acoustic form, and underlying intonational structure, of the RFR contour.

4.4 Perception: Hypotheses and Tasks 3.3 & 3.4 Design

We now return to our second set of questions, regarding how "RFR interpretation" relates to the intonational components of the "RFR contour". In extant discussions of this contour, there is disagreement as to whether the RFR interpretation is composed from the meaning of the intonational components (i.e., the pitch accent and boundary tone), or whether their meaning is conventionalized and/or non-compositional. Pike (1945: 55) treats the RFR interpretation as the result of combining the meanings of a "Contrastive Attention" pitch accent and an "Incomplete Deliberation" boundary tone. Ward & Hirschberg (1985: 752), however, argue that the interpretation of the RFR contour "cannot be seen as the mere sum of its parts".

Again taking as the premise that RFR contours are incompatible with surface scope reading (cf. Table 2), we can test these competing analyses of what triggers RFR interpretations by varying aspects of the intonational contour and seeing where the surface scope reading is (un)available, via perception tasks.

Hyp.3.3. A surface-scope interpretation is less available (i.e., fewer participants report it) when: **(a)** all RFR contour components are present in the signal, **(b)** the boundary tone is (the right kind of) fall-rise, **(c)** the pitch accent is a rise (of the right kind), and/or **(d)** the number and strength of cues to these categorical differences is greater, at a global level.

As with task 3.2, participants in two perception tasks (3.3 & 3.4) will see a visual context showing a surface-scope interpretation, and will provide a felicity rating for a particular audio stimulus. Tasks 3.3 & 3.4 differ from each other in which intonational properties are manipulated in the the audio stimuli.

In task 3.3, audio stimuli will be created by resynthesizing a neutral base utterance, according to the acoustic parameters that characterize RFR contours (as identified in tasks 3.1 & 3.2), targeting two categorical dimensions of the contour: (i) which pitch accent, if any, occurs on the focus (determined by QAC) of the sentence, and (ii) which boundary tone is used. In particular, the pitch accent will vary along the following categories: scoopy rise (L*+H), gradual rise (H*), and none; and the boundary tone will vary between fall-rise (L-H%) or fall (L-L%). These manipulations have been chosen on the basis of the literature that identifies RFR as /L*+H L-H%/ and 'default' intonation as /H* L-L%/. Other aspects of the contour (prosodic phrasing, surrounding pitch accents, pitch range, lengthening effects, intensity changes) will be controlled for. This will allow us to address Hypotheses 3.3a-c, concerning whether RFR interpretations are compositional ("a sum of its parts", *ibid.*) or not.

The audio stimuli for task 3.4 will again be created by resynthesizing a neutral base, this time along two continuous dimensions for pitch movements: scaling and timing. We have chosen these variables on the basis of work on TCoG (Barnes et al. 2012); the scaling dimension will be manipulated in steps from flat F0 (no accent) to a high in the speaker's range value. The time-alignment dimension will be manipulated by moving the peak from the vowel onset in steps through the vowel and coda consonant. Moreover, changes to these variables will cause changes to the slopes of the various pitch rises/falls, which have been argued to affect interpretation (e.g., Jeong 2018a). Again, other aspects of the contour will be controlled for. This will allow us to address Hyp.3.3d, i.e., whether the RFR interpretive effects are categorical differences, or whether the interpretation is gradient such that it changes as a function of (continuous) acoustic variables.

4.5 Acoustic Analysis and Analysis of Perceptual Data

As with tasks 3.1 and 3.2, we will use a percent judgment threshold to identify the RFR and non-RFR contour categories, on the basis of the unavailability of surface-scope readings in task 3.3 and 3.4. We will use logistic and linear regression analyses to model which manipulated RFR contour components predict the percent judgments. We are especially interested in two questions: (i) whether the RFR interpretation is compositional on the basis of sub-parts of the intonational contour, and (ii) whether the mapping between form and meaning for RFR is categorical or continuous. For the former question, we consider whether a contour with only the final rise (L-H%) will have felicity ratings that are closer to non-RFR contours that are closer to the canonical RFR contour, or to non-RFR contours that fall somewhere in between. This will help deepen our understanding of the intonational components that meaningfully contribute to the RFR contour. For example, if we find that both intonational phonemes are necessary for RFR interpretive effects, this would not entail that (unlike meaning composed of segmental morphemes) intonational meaning is never composed of meaningful parts; in fact, such a result for the RFR contour is completely consistent with our understanding of idioms (wherein the meaning of morphemes becomes opaque through contextual allosemy [Marantz 2013], which requires syntactic locality.) As for the latter question, we wish to understand whether this sort of intonational meaning is primarily categorical in nature (i.e., semantic) such that perception ratings suddenly flip (cf. perception tasks for VOT; e.g., Repp 1984), or whether it is continuous and gradient (i.e., as a pragmatic inference) such that perception ratings shift gradually with continuous changes to the acoustics. For example, does reducing the strength of the various components (e.g., the Level of the Rise, the TCoG alignment of the accent) reduce the return of inverse scope readings? And if there is a gradient decline, which components have which sizes of effects?

5 Intellectual Merit

Broadly speaking, the proposed experiments address rigorously the connections between sem/prag meanings and prosodic structures in English, an issue that has long tantalized the field, The proposed work builds on recent studies in which connections between these domains have begun to be seriously addressed, establishing these connections by close examination of patterns of acoustic variability that relate to our initial guiding questions (§1.1). These three deeply intertwined questions —about differences in acoustic cues (Q1), how sem/prag variables condition phonemic aspects of this variation (Q2), and what (non-phonemic) differences occur within a sem/prag context (Q3)— are addressed in an integrated way in each of the proposed sets of experiments, and results will have significant implications for linguistic theory. For example, discoveries of variation conditioned by particular sem/prag variables may lead to expansion of the set of contrastive phonological categories in (varieties of) US English. More specifically, where some languages use segmental morphemes to convey evidentiality, Expt Set 1 tests whether (spoken) English regularly employs intonation to do so. Expt Set 2 investigates whether or not a particular set of acoustic/intonational cues cluster around a particular meaning; i.e., whether these are cues to a single phonological category of intonation. Finally, Expt Set 3 examines the intonational variation in a conventionalized meaning and its relationship to a sequence of tonemes, and whether the intonational cues that are present signal the relevant meaning compositionally and/or gradiently. In all three sets of experiments, we will consider the possibility that a given meaning may be cued by more than one set of intonational patterns, including the possibility that there are intonational morphemes with multiple allomorphs or that meanings are represented through a variable set of intonational cues. Additionally, we expect that some variation will be arbitrary or idiosyncratic to a speaker (or group of speakers), helping us to uncover non-phonemic intonational variation that is idiosyncratic to a speaker or representative of a community of speakers.

As an interface-oriented investigation, this project has significant implications for linguistic theory at the architectural level. It addresses theoretical questions at the heart of the intonational enterprise: the nature of the phonetics-phonology interface, the inventory of phonological categories, the interface between intonation and sem/prag meanings, and the morphological/syntactic representations. This requires a concrete framework with specific architectural principles, for which we appeal to Distributed Morphology.

In approaching intonation this way, this work will be able to meaningfully contribute towards conversations of how much intonation is (un)like segmental morphology. (See Bocci 2013, Constant 2014, or Wakefield 2020 for sophisticated analyses that treat them as deeply similar.) For example, it might be that for a particular sem/prag context, both H* and L+H* (or even H* and L*) can be used, and the alternation is a case of allomorphy, governed by morphological principles that refer to morphological structure. (For discussions of allomorphic principles, see Embick 2010 or Bobaljik 2012.) In addition, intonational meanings may be sometimes compositional and sometimes opaque/idiomatic – and it might be that a particular intonational morpheme has multiple interpretations, including opaque/idiomatic ones that depend on syntactic context. (For discussions of such allosemic constraints, see Borer 2013 or Marantz 2013.)

6 Broader Impacts

Advancing our understanding of the subtle, complex and varied ways in which meaning is realized by intonation has the potential to positively impact society via several avenues. First, findings from this research can be used to improve increasingly ubiquitous computer-human speech interfaces, to reflect the meanings investigated here (e.g., speaker attitudes or truth-conditional ambiguity) which are effortlessly used by human speakers but as yet are little utilized in automatic speech recognition applications. Likewise, synthesis of natural-sounding context-appropriate contours requires an understanding of the range of acoustic possibilities for a contour and the factors that govern the speaker's choice among these possibilities. Findings can similarly be used to improve machine-translation of spoken utterances, a critical application as the world globalizes. Such advances in automatic speech processing can also decrease barriers for those with speech difficulties (e.g., dysarthria, ALS, TBI) who rely on speech technologies.

Second, consideration of individual cue patterns in the segmental domain has proven valuable in characterizing the speech of certain clinical populations (e.g., Parkinson's disease [Godino-Llorente et al. 2017], child language development [Talkar et al. 2017], and decompensated heart failure [Murton et al. 2018]); and there is increasing interest in analyzing individual cues to prosodic prominence and phrasing, in both typical (Brugos et al. 2018, Cole & Shattuck-Hufnagel 2018) and atypical (Jun & Bishop 2015) speakers. By extending and enriching our understanding of form-meaning relationships to the level of individual prosodic cues in typical speakers, we lay the groundwork for annotating, analyzing and modeling fine-grained differences that arise in atypical speakers. This work will be of particular value in developing new methods for early diagnosis of disordered speech, a problem which can be socially devastating, so the earliest possible intervention can be critical. For example, currently it is difficult to reach a definitive diagnosis of Parkinson's disease within the time window from onset during which treatment can be effective; early indications that speech cues can track the onset of the disease may help address this issue (Godino-Llorente et al. 2017).

Finally, our methodological approaches have great potential for the future investigation of variation across different speech communities/contexts within a language, a subject of great importance and intense recent interest for both theoretical and practical reasons. Researchers who study aspects of intonational variation across language varieties (e.g., Burdin et al. 2018) have found that there is substantial and systematic phonetic variation across intonational events labelled as the same phonological category. PoLaR annotation offers the means to capture and analyze such systematic variation. Further, there has been little research on the variation in how intonation maps to meaning across language varieties. Such variation is likely to hinder communication between members of different groups, both directly and by way of speech technology intermediaries, which can have a negative impact on speakers of non-standard language varieties in areas as diverse as healthcare and courtroom testimony. (As a concrete example, a court case about the meaning of a contract may hinge on a single comma (Victor 2017); comma-related and other intonationally-based misunderstandings may emerge as speech-based contractual interactions become more ubiquitous.) The development and refinement of this methodology to carefully examine the interface between intonation and semantic/pragmatic meanings paves the way for future work involving different speech communities.

To work toward these goals, the team will communicate with a broad audience, including both specialists (in a variety of language-related and linguistic fields) and non-specialists. We will engage with researchers in a broad range of linguistic subdisciplines, such as sociolinguistics and language acquisition, as well as with speech pathologists, ASD behavioral therapists, language teachers, and NLP specialists. We will foster dialog through dissemination of our findings at (satellite) workshops, at meetings (e.g., at BUCLD, INTERSPEECH, LSA, ASA, ASHA) and at prosody-focused meetings such as Speech Prosody and ETAP. We hope that engagement with researchers across subdisciplines will lead to future collaborations and dialog, such as through a future cross-disciplinary workshop. To make products of the research broadly accessible to a wider audience, we will develop and publicize a freely accessible website that presents tools and overviews of findings in non-technical language.

Beyond the expected societal benefits of the findings and methods of the study, an additional direct positive impact of this project is the inclusion of women students of computer science. Women continue to be under-represented in computational, data intensive and machine learning activities. Much of the data mining and machine learning work in this study will be conducted by undergraduate women from Simmons University, giving them opportunities for professional development.

References Cited

Ahn, Byron. 2015. There's nothing exceptional about the phrasal stress rule. lingBuzz/002458.

Ahn, Byron, Stefanie Shattuck-Hufnagel & Nanette Veilleux. 2016. Evidence and intonational contours: An experimental approach to meaning in intonation. *Proceedings of the Sixteenth Australasian International Conference on Speech Science and Technology* 189–192.

Ahn, Byron, Stefanie Shattuck-Hufnagel & Nanette Veilleux. 2018. PoLaR annotation conventions: A tool for annotating prosodic variation. Presented at Experimental and Theoretical Advances in Prosody 4. `http://doi.org/10.17605/OSF.IO/RQ7FB`.

Ahn, Byron, Stefanie Shattuck-Hufnagel, Nanette Veilleux & Alejna Brugos. In preparation. Evidentiality and influence on intonational acoustics.

Ahn, Byron, Nanette Veilleux & Stefanie Shattuck-Hufnagel. 2019. Annotating prosody with PoLaR: Conventions for a decompositional annotation system. In Sasha Calhoun, Paola Escudero, Marija Tabain & Paul Warren (eds.), *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019*, 1302–1306. Canberra, Australia: Australasian Speech Science and Technology Association Inc.

Ahn, Byron, Nanette Veilleux, Stefanie Shattuck-Hufnagel & Alejna Brugos. 2020. Guidelines for PoLaR annotation. Available at http://www.polarlabels.com. doi:10.17605/OSF.IO/RQ7FB.

Anderson, Anne H., Miles Badera dn Ellen Gurman Bard, Elizabeth Boyle, Gwyneth Doherty, Simon Garrod, Stephen Isard, Jacqueline Kowtko, Jan McAllister, Jim Miller, Catherine Sotillo, Henry S. Thompson & Regina Weinert. 1991. The HCRC map task corpus. *Language and Speech* 34(4). 351–366. doi: 10.1177/002383099103400404.

Arnhold, Anja & Aki-Juhani Kyröläinen. 2017. Modelling the interplay of multiple cues in prosodic focus marking. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 8(1). 4. doi: 10.5334/labphon.78.

Barnes, Jonathan, Nanette Veilleux, Alejna Brugos & Stefanie Shattuck-Hufnagel. 2010. The effect of global F0 contour shape on the perception of tonal timing contrasts in American English intonation. In *Speech Prosody 2010*, paper 445. ISCA Speech.

Barnes, Jonathan, Nanette Veilleux, Alejna Brugos & Stefanie Shattuck-Hufnagel. 2012. Tonal center of gravity: A global approach to tonal implementation in a level-based intonational phonology. *Laboratory Phonology* 3(2). 337–383. doi:10.1515/lp-2012-0017.

Beaver, David & Brady Clark. 2008. *Sense and sensitivity: How focus determines meaning*. Chichester: Wiley-Blackwell.

Beaver, David, Brady Zack Clark, Edward Flemming, T. Florian Jaeger & Maria Wolters. 2007. When semantics meets phonetics: Acoustical studies of second-occurrence focus. *Language* 83(2). 245–276.

Beckman, Mary E. & Gayle M. Ayers. 1997. Guidelines for ToBI labelling. *The OSU Research Foundation* 3. Available at http://www.ling.ohio-state.edu/~tobi/ame_tobi/labelling_guide_v3.pdf.

Bobaljik, Jonathan. 2012. *Universals in comparative morphology: Suppletion, superlatives and the structure of words*. Cambridge, MA: MIT Press.

Bobaljik, Jonathan. 2017. Distributed morphology. In *Oxford research encyclopedia of linguistics*, Oxford University Press. doi:10.1093/acrefore/9780199384655.013.131.

Bocci, Giuliano. 2013. *The syntax-prosody interface*. John Benjamins. doi:10.1075/la.204.

Borer, Hagit. 2013. *Taking form*, vol. III (Structuring Sense). Oxford: Oxford University Press.

Breen, Mara, Evelina Fedorenko, Michael Wagner & Edward Gibson. 2010. Acoustic correlates of information structure. *Language and Cognitive Processes* 25. 1044–1098. doi:10.1080/01690965.2010.504378.

Brown, James D. 2001. Pragmatics tests: Different purposes, different tests. In Kenneth R. Rose & Gabriele Kasper (eds.), *Pragmatics in language teaching*, 301–325. Cambridge: Cambridge University Press.

Brugos, Alejna, Mara Breen, Nanette Veilleux, Jonathan Barnes & Stefanie Shattuck-Hufnagel. 2018. Cue-based annotation and analysis of prosodic boundary events. In *Speech Prosody 2018*, 245–249. doi:10.21437/SpeechProsody.2018-50.

Burdin, Rachel, Nicole Holliday & Paul Reed. 2018. Rising above the standard: Variation in L+H* contour use across 5 varieties of American English. In *Speech Prosody 2018*, 354–358. doi:10.21437/SpeechProsody.2018-72.

Burdin, Rachel Steindel & Joseph Tyler. 2018. Rises inform, and plateaus remind: Exploring the epistemic meanings of “list intonation” in american english. *Journal of Pragmatics* 136. 97–114.

Büring, Daniel. 2007. Semantics, intonation and information structure. In Gillian Ramchand & Charles Reiss (eds.), *The Oxford handbook of linguistic interfaces*, 445–473. New York: Oxford University Press.

Burton, Strang & Lisa Matthewson. 2015. Targeted construction storyboards in semantic fieldwork. In M. Ryan Bochnak & Lisa Matthewson (eds.), *Methodologies in semantic fieldwork*, chap. 5. New York: Oxford University Press. doi:10.1093/acprof:oso/9780190212339.001.0001.

Buxó-Lugo, Andrés & Duane G. Watson. 2016. Evidence for the influence of syntax on prosodic parsing. *Journal of Memory and Language* 90. 1–13.

Buz, Esteban, Michael K. Tanenhaus & T. Florian Jaeger. 2016. Dynamically adapted context-specific hyper-articulation: Feedback from interlocutors affects speakers’ subsequent pronunciations. *Journal of Memory and Language* 89. 68–86. doi:10.1016/j.jml.2015.12.009.

Büring, Daniel. 2003. On D-trees, beans, and B-accents. *Linguistics and Philosophy* 26(5). 511–545. doi:10.1023/A:1025887707652.

Calhoun, Sasha. 2012. The theme/rheme distinction: Accent type or relative prominence? *Journal of Phonetics* 40(2). 329–349. doi:10.1016/j.wocn.2011.12.001.

Chodroff, Eleanor & Jennifer Cole. 2019. The phonological and phonetic encoding of information structure in American English nuclear accents. *The Proceedings of 19th International Congress of Phonetic Sciences* .

Ciardelli, Ivano, Jeroen Groenendijk & Floris Roelofsen. 2013. Inquisitive semantics: A new notion of meaning. *Language and Linguistics Compass* 7(9). 459–476. doi:10.1111/lnc3.12037.

Cole, Jennifer. 2019. An information-based approach to explaining variation in intonation: Reconciling theory and empirical data. Plenary talk at the Annual Meeting of the LSA.

Cole, Jennifer & Stefanie Shattuck-Hufnagel. 2018. Quantifying phonetic variation: Landmark labelling of imitated utterances. In Francesco Cangemi, Meghan Clayards, Oliver Niebuhr, Barbara Schuppler & Margaret Zellers (eds.), *Rethinking reduction*, 164–204. Berlin: De Gruyter.

Constant, Noah. 2012. English rise-fall-rise: A study in the semantics and pragmatics of intonation. *Linguistics and Philosophy* 35(5). 407–442.

Constant, Noah. 2014. *Contrastive topic: Meanings and realizations*: University of Massachusetts-Amherst dissertation.

Crain, Stephen & Cecil McKee. 1986. The acquisition of structural restrictions on anaphora. In Stephen Berman, Jae-Woong Choe & Joyce McDonough (eds.), *Proceedings of North East Linguistics Society 16*. 94–110. Amherst, Massachusetts: Graduate Linguistic Student Association.

Dilley, Laura. 2010. Pitch range variation in English tonal contrasts: Continuous or categorical? *Phonetica* 67. 63–81. doi:10.1159/000319379.

Dilley, Laura & Mara Breen. 2018. An enhanced autosegmental-metrical theory (AM^+) facilitates phonetically transparent prosodic annotation. *Tonal Aspects of Language 2018* 77–81. doi:10.21437/TAL.2018-14.

Dilley, Laura & Meredith Brown. 2005. The RaP (rhythm and pitch) labeling system. Available at http://tedlab.mit.edu/tedlab_website/RaPHome.html.

Embick, David. 2010. *Localism versus globalism in morphology and phonology*. MIT Press. doi:10.7551/mitpress/9780262014229.001.0001.

Farkas, Donka & Kim Bruce. 2010. On reacting to assertions and polar questions. *Journal of Semantics* 27. 81–118.

Farkas, Donka & Floris Roelofsen. 2017. Division of labor in the interpretation of declaratives and interrogatives. *Journal of Semantics* 34(2). 237–289. doi:10.1093/jos/ffw012.

Ferreira, Fernanda. 1993. Creation of prosody during sentence production. *Psychological Review* 100(2). 233–253.

Ginzburg, Jonathan. 2012. *The interactive stance: Meaning for conversation*. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199697922.001.0001.

Glenberg, Arthur M. & William E. Langston. 1992. Comprehension of illustrated text: Pictures help to build mental models. *Journal of Memory and Language* 31. 129–151.

Godino-Llorente, Juan Ignacio, Stefanie Shattuck-Hufnagel, J. Y. Choi, Laureano Moro-Velázquez & Jorge Gómez-García. 2017. Towards the identification of idiopathic Parkinson's disease from the speech: New articulatory kinetic biomarkers. *PLOS ONE* 12(12). 1–35. doi:10.1371/journal.pone.0189583.

Grabe, Esther, Brechtje Post & Francis Nolan. 2001. Modelling intonational variation in English: the IViE system. In *Proceedings of Prosody 2000*, 51–58. Poznan, Poland: Adam Mickiewicz University.

Gunlogson, Christine. 2004. *True to form: Rising and falling declaratives as questions in English*. New York: Routledge.

Hain, Cristina, Nanette Veilleux & Byron Ahn. In preparation. Modelling the mapping between intonation and meaning of questions in spoken American English using Random Forests.

Halle, Morris & Alec Marantz. 1993. Distributed morphology and the pieces of inflection. In Kenneth Hale & Samuel Jay Keyser (eds.), *The view from building 20*, 111–176. Cambridge, MA: MIT Press.

Halliday, M.A.K. 1967. Notes on transitivity and theme in English: Part 2. *Journal of Linguistics* 3(2). 199–244.

Heidenreich, Samantha. 2019. *Disjunctive questions: Experimental evidence in prosody and discourse*: Northwestern University dissertation.

Hirschberg, Julia & Gregory Ward. 1992. The influence of pitch range, duration, amplitude and spectral features on the interpretation of the rise-fall-rise intonation contour in English. *Journal of Phonetics* 20(2). 241–251.

Hualde, José I. & Pilar Prieto. 2016. Towards an international prosodic alphabet (IPrA). *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 7(1). 5. doi:http://doi.org/10.5334/labphon.11.

Ito, Kiwako, Shari R. Speer & Mary E. Beckman. 2004. Informational status and pitch accent distribution in spontaneous dialogues in English. In *Speech Prosody 2004*, 279–282. ISCA Speech.

Jeong, Sunwoo. 2018a. Intonation and sentence-type conventions: Two types of rising declaratives. *Journal of Semantics* 35(2). 305–356.

Jeong, Sunwoo. 2018b. *Melodies in context: The semantics and pragmatics of English rising declaratives*. Stanford, CA: Stanford University dissertation.

Jeong, Sunwoo & Cleo Condoravdi. 2017. Imperatives with the calling contour. In *Proceedings of the Berkeley Linguistic Society 43 (BLS 43)*, vol. 1, 185–210.

Jun, Sun-Ah & Jason Bishop. 2015. Priming implicit prosody: Prosodic boundaries and individual differences. *Language and Speech* 58(4). 459–473. doi:10.1177/0023830914563368.

Katz, Jonah & Elisabeth Selkirk. 2011. Contrastive focus vs. discourse-new: Evidence from phonetic prominence in English. *Language* 87(4).

Kratzer, Angelika & Elisabeth Selkirk. 2020. Deconstructing information structure. ms., lingbuzz/004201.

Krifka, Manfred. 2004. The semantics of questions and the focusation of answers. In Chungmin Lee & Matthew Gordon (eds.), *Topic and focus: A cross-linguistic perspective*, 139–151. Kluwer Academic Publishers.

Kügler, Frank, Bernadett Smolibocki, Denis Arnold, Stefan Baumann, Bettina Braun, Martine Grice, Stefanie Jannedy, Jan Michalsky, Oliver Niebuhr, Jörg Peters, Simon Ritter, Christine T. Röhr, Antje Schweitzer, Katrin Schweitzer & Petra Wagner. 2015. DIMA – annotation guidelines for German intonation. In The Scottish Consortium for ICPhS 2015 (ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*, The University of Glasgow.

Ladd, D. Robert. 1978. Stylized intonation. *Language* 54. 517–539.

Ladd, D. Robert & Astrid Schepman. 2003. "sagging transitions" between high pitch accents in English: experimental evidence. *Journal of Phonetics* 31(1). 81–112. doi:10.1016/S0095-4470(02)00073-6.

Lippi-Green, Rosina. 2000. Reprint *the standard language myth*. *Voice and Speech Review* 1(1). 23–30. doi:10.1080/23268263.2000.10761384.

Lohfink, Georg, Argyro Katsika & Amalia Arvaniti. 2019. Variability and category overlap in the realization of intonation. In Sasha Calhoun, Paola Escudero, Marija Tabain & Paul Warren (eds.), *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019*, 701–705. Canberra, Australia: Australasian Speech Science and Technology Association Inc.

Malamud, Sophia A. & Tamina Stephenson. 2015. Three ways to avoid commitments: Declarative force modifiers in the conversational scoreboard. *Journal of Semantics* 32(2). 275–311. doi:10.1093/jos/ffu002.

Marantz, Alec. 2013. Locality domains for contextual allomorphy across the interfaces. In Ora Matushansky & Alec Marantz (eds.), *Distributed morphology today: Morphemes for Morris Halle*, 95–115. MIT Press.

McAuliffe, Michael, Michaela Socolof, Sarah Mihuc, Michael Wagner & Morgan Sonderegger. 2019. Montreal forced aligner [computer program] (version 1.0.1). Retrieved from `http://montrealcorpustools.github.io/Montreal-Forced-Aligner/`. doi: http://doi.org/10.5281/zenodo.2630943.

Murton, Olivia, Jeung-Yoon Choi, Daryush Mehta & Stefanie Shattuck-Hufnagel. 2018. Toward the automatic detection of manually labeled irregular pitch periods. *The Journal of the Acoustical Society of America* 143(3). 1869–1869. doi:10.1121/1.5036134.

Ouyang, Iris & Elsi Kaiser. 2015. Prosody marks different kinds of informativity: Interactions between frequency, probability and focus. *University of Pennsylvania Working Papers in Linguistics* 21(1).

Pierrehumbert, Janet. 1980. *The phonology and phonetics of English intonation*. Cambridge, MA: MIT dissertation.

Pierrehumbert, Janet & Julia Hirschberg. 1990. The meaning of intonational contours in the interpretation of discourse. In Philip R. Cohen, Jerry L. Morgan & Martha E. Pollack (eds.), *Intentions in communication*, 271–311. Cambridge, MA: MIT Press.

Pike, Kenneth L. 1945. *The intonation of American English*. Ann Arbor, MI: University of Michigan Press.

Prince, Ellen F. 1981. Toward a taxonomy of given-new information. In Peter Cole (ed.), *Radical pragmatics*, New York: Academic Press.

Repp, Bruno. 1984. Categorical perception: Issues, methods, findings. *Speech and Language* 10. 243–335. doi:10.1016/B978-0-12-608610-2.50012-1.

Repp, Sophie. 2013. Common ground management: Modal particles, illocutionary negation and verum. In Daniel Gutzmann & Hans-Martin Gärtner (eds.), *Beyond expressives: Explorations in use-conditional meaning*, 231–274. Leiden, The Netherlands: Brill. doi:10.1163/9789004183988_008.

Roberts, Craige. 1996. Information structure: Towards an integrated formal theory of pragmatics. In Jae-

Hak Yoon & Andreas Kathol (eds.), *Papers in semantics (Working Papers in Linguistics 49)*, The Ohio State University. doi:http://linguistics.osu.edu/files/linguistics/workingpapers/osu_wpl_49.pdf.

Roberts, Craige. 2004. Context in dynamic interpretation. In Laurence R. Horn & Gregory Ward (eds.), *The handbook of pragmatics*, 197–220. Malden, MA: Blackwell. doi:10.1002/9780470756959.ch9.

Roettger, Timo, Tim Mahrt & Jennifer Cole. 2019. Mapping prosody onto meaning – the case of information structure in American English. *Language, Cognition and Neuroscience* 34(7). 841–860. doi: 10.1080/23273798.2019.1587482.

Rose, Kenneth R. 2000. An exploratory cross-sectional study of interlanguage pragmatic development. *Studies in Second Language Acquisition* 22(1). 27–67.

Rosenberg, Andrew. 2013. Modeling prosodic sequences with k-means and Dirichlet process GMMs. In Frédéric Bimbot, Christophe Cerisara, Cécile Fougeron, Guillaume Gravier, Lori Lamel, François Pellegrino & Pascal Perrier (eds.), *Proceedings of the 14th annual conference of the international speech communication association (interspeech 2013)*, 520–524. ISCA.

Rudin, Deniz. 2018. *Rising above commitment*. Santa Cruz, CA: UC Santa Cruz dissertation.

Sag, Ivan & Mark Liberman. 1975. The intonational disambiguation of indirect speech acts. In *Papers from the eleventh regional meeting Chicago Linguistics Society*, 487–497.

Schwarzschild, Roger. 1999. Givenness, AvoidF and other constraints on the placement of accent. *Natural Language Semantics* 7(2). 141–177. doi:10.1023/A:1008370902407.

Selkirk, Elisabeth. 1996. Sentence prosody: Intonation, stress, and phrasing. In John A. Goldsmith (ed.), *The handbook of phonological theory*, 550–569. Blackwell Publishing.

Stalnaker, Robert. 1978. Assertion. In P Cole (ed.), *Syntax and semantics*. 315–332. New York: Academic Press.

Talkar, Tanya, Jennifer Zuk, Maria X. Guerrero, Jeung-Yoon Choi & Stefanie Shattuck-Hufnagel. 2017. Acoustic cues to distinctive features are modified in the speech of typically-developing versus atypically developing children. *The Journal of the Acoustical Society of America* 141(5). 3837–3837. doi: 10.1121/1.4988535.

Truckenbrodt, Hubert. 2012. Semantics of intonation. In Klaus Von Heusinger, Claudia Maienborn & Paul Portner (eds.), *Semantics: An international handbook of natural language meaning*, vol. 3, 2039–2069. De Gruyter. doi:10.1515/9783110253382.2039.

Victor, Daniel. 2017. Lack of Oxford comma could cost Maine company millions in overtime dispute. New York Times, March 16 2017.

Villarreal, Dan, Lynn Clark, Jennifer Hay & Kevin Watson. 2020. From categories to gradience: Auto-coding sociophonetic variation with random forests. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 11(1). 6. doi:10.5334/labphon.216.

Wagner, Michael. 2006. Givenness and locality. In Masayuki Gibson & Jonathan Howell (eds.), *Proceedings of SALT XVI*, 295–312. CLC Publications.

Wakefield, John. 2020. *Intonational morphology*. Springer: Springer. doi:10.1007/978-981-15-2265-9.

Ward, Gregory & Julia Hirschberg. 1985. Implicating uncertainty: The pragmatics of fall-rise intonation. *Language* 61. 747–776.

Ward, Gregory & Julie Hirschberg. 1988. Intonation and propositional attitude: The pragmatics of L*+H L H In J. Powers & K. de Jong (eds.), *Proceedings of eastern states conference on linguistics (escol) 5*, 512–522. Columbus, OH: Ohio State University.

Ward, Nigel G. 2019. *Prosodic patterns in English conversation*. Cambridge, UK: Cambridge University Press. doi:10.1017/9781316848265.

Watson, Duane, Michael Tanenhaus & Christine Gunlogson. 2008. Interpreting pitch accents in online comprehension: H* vs. L+H*. *Cognitive Science* 32. 1232–1244.

Westera, Matthijs. 2017. *Exhaustivity and intonation: a unified theory*. Amsterdam: University of Amsterdam dissertation.

Zhou, Z.L. & Byron Ahn. 2019. Is this in the phonology? examining the intonational phonetics-phonology interface with American English polar questions. *The Proceedings of 19th International Congress of Phonetic Sciences* .

Zubizarreta, María Luisa. 1998. *Prosody, focus, and word order*. Cambridge, MA: The MIT Press.

Zubizarreta, María Luisa & Jean-Roger Vergnaud. 2006. Phrasal stress and syntax. In Martin Everaert & Henk van Riemsdijk (eds.), *The Blackwell Companion to Syntax*, vol. 3, chap. 49. Blackwell Publishing.