Marked Initial Pitch in Questions Signals Marked Communicative Function

Mark A Sicoli
Georgetown University, USA

Tanya Stivers
University of California, USA

NJ Enfield
University of Sydney, Australia

Stephen C Levinson
Max Planck Institute for Psycholinguistics, The Netherlands

Abstract

In conversation, the initial pitch of an utterance can provide an early phonetic cue of the communicative function, the speech act, or the social action being implemented. We conducted quantitative acoustic measurements and statistical analyses of pitch in over 10,000 utterances, including 2512 questions, their responses, and about 5000 other utterances by 180 total speakers from a corpus of 70 natural conversations in 10 languages. We measured pitch at first prominence in a speaker's utterance and discriminated utterances by language, speaker, gender, question form, and what social action is achieved by the speaker's turn. Through applying multivariate logistic regression we found that initial pitch that significantly deviated from the speaker's median pitch level was predictive of the social action of the question. In questions designed to solicit agreement with an evaluation rather than information, pitch was divergent from a speaker's median predictably in the top 10% of a speakers range. This latter finding reveals a kind of iconicity in the relationship between prosody and social action in which a marked pitch correlates with a marked social action. Thus, we argue that speakers rely on pitch to provide an early signal for recipients that the question is not to be interpreted through its literal semantics but rather through an inference.

Corresponding author:
Mark A Sicoli, Department of Linguistics, Georgetown University, 1421 37th Street NW, Box 571051, Poulton Hall 240, Washington, DC 20057-1051, USA.
Email: mas498@georgetown.edu

Introduction

It is well known that we cannot tell the communicative function or speech act of an utterance directly from its surface form (Austin, 1962; Gordon & Lakoff, 1971; Heritage, 1984; Searle, 1969, among many others). For instance, the question *Where are you going dressed like that?* has all of the surface indications of a request for information. However, in context, it is likely to be better analyzed as a way of admonishing someone to change her clothes (Cole & Morgan, 1975; Gordon & Lakoff, 1971; Grice, 1975; Sadock, 1975; Searle, 1975). Despite the fact that languages typically lack any direct mapping of form to function, addressees generally show no difficulty ascribing actions to speakers' utterances (Levinson, 2013). Recipients also make these judgments very quickly. A comparative study of question–response sequences in 10 languages found that the most common time taken between the question and the response was zero with an overall mean of only 208 ms (Stivers et al., 2009). Moreover, questioners relatively rarely treat recipient answers as revealing any misunderstanding, which suggests that question recipients are generally accurate in their action ascription. The problem remains: how do people successfully and quickly ascribe social actions to each other's utterances? Is the question a common request for information or does it request an object? Does the question initiate repair, or is it used as a vehicle to evaluate the conduct or morality of another social actor?

Schegloff notes that "one of the most intuitively inviting prospects in incorporating prosody in the study of interaction has often been that it is the intonation of some utterance which has it 'do' some action or produce some outcome" (1998, p. 247). Prosodic features of utterances have been shown to have a high functional load in discourse function recognition. Shriberg et al. (1998), for example, showed that including prosodic information in statements, questions, and greetings used for machine-based classification tasks resulted in significantly better performance than word information alone. The work of Gumperz and his collaborators incorporated prosody into the study of spoken discourse, arguing that prosodic features can act as *contextualization cues* (Gumperz, 1992; Levinson, 2003), providing a semiotic ground for analyzing culturally and linguistically based understanding and misunderstanding (Gumperz, 1982).

Our aim in this paper is to examine initial pitch in conversations with a focus on utterances that perform a range of actions but that may nonetheless be classified as *questions*. We ask whether the initial pitch of a question may help participants recognize the action that a given question is being used to implement. In this study we address this issue in an examination of speakers' turns-at-talk from a corpus of video recordings of everyday conversation in 10 languages (data described in more detail elsewhere: Stivers et al., 2009; Stivers, Enfield, & Levinson, 2010). In this cross-linguistic sample, we examined communicative but non-referential functions of pitch, focusing our efforts on early parts in the utterance. Our study provides evidence for a kind of iconic mapping from form to function: if a question shows out-of-the-ordinary initial-pitch properties, then it is most likely being used to implement an out-of-the-ordinary action. For questions, this refers to, say, complaining or expressing surprise as opposed to requesting information. More precisely, our data suggest that when questions are formulated with divergent initial pitch, this is an early signal that the addressee should make an additional inference in order to ascribe an indirect speech act function to the utterance. Our analyses show that pitch is an important tool for recipient design (Schegloff, 2002, p. 79) and a resource for action ascription (Schegloff, 1998, p. 237). Initial pitch frames the early content of an utterance, signaling an aspect of question design. In our study initial

pitch contributes to a contrast between questions that are being used to *request information* and evaluative questions that are rather being used to *seek agreement.* We also found that initial pitch was to a lesser degree predictive of question form in the corpus, particularly for content questions (Wh-). This is something we might expect among languages that front Wh-words to a focal position, but as the tendency to show higher than median pitch was independent of whether a language fronted Wh-words, it is possible that both sets of divergent pitch facts have communicative functions.

2 Background

2.1 Question–response sequences

Psycholinguistic research suggests that it takes close to 600 ms to initiate and produce an utterance (Levelt, 1989). In conversation a substantial amount of talk is turn-by-turn talk in which sequences such as the basic two-part question–response sequence (Schegloff, 2007) require an alternation between speakers. The norm in conversation is to achieve no overlap and no gap between turns (Sacks, Schegloff, & Jefferson, 1974), and this has been empirically documented across multiple languages (de Ruiter, Mitterer, & Enfield, 2006; Stivers et al., 2009). Taken together, this suggests that question recipients must begin preparing their response well before the end of the question if they are to produce an on-time response. Final intonation, long thought to be an important cue in signaling that a given declarative utterance is a question, is likely to be, in fact, redundant due to coming so late in the questioner's turn. Rather, the most critical signals of an utterance's action type are likely to be early indicators (Levinson, 2013). This would make something like syntactic inversion and Wh-fronting good cues. Yet, neither inversion nor fronting are universals of question form. Even a language such as English, which does have syntactic inversion, relies more on declarative forms than interrogative forms (Stivers, 2010). Could prosody provide such an early cue in question design?

2.2 Early pitch in utterances and its role in questions

There is evidence that prosody early in the utterance is used by speakers to discriminate between actions. For example, high pitch can signal new topics in both conversation and narrative (Beckman & Pierrehumbert, 1986; Bolinger, 1989; Chafe, 1998; Gussenhoven, 2002, p. 51; Sicoli, 2007). In data from radio phone-in programs, Couper-Kuhlen (2001) demonstrates that onset pitch was deployed as a contextualization cue to index the topic at issue. She shows that turn-constructional units that announce the "reason-for-the-call" are routinely designed with high pitch onset. She further showed that when such a sequence lacks high pitch onset, interactional difficulties were more abundant, and the radio moderator often treated these turns as preliminary to the reason for the call rather than constitutive of it.

Relatedly, intonation has been reported to be effective in the disambiguation of indirect speech acts from their direct readings (Sag & Liberman, 1975). Brazil, Coulthard, and Johns (1980) and Brazil (1997) developed the notion of "key accent", which is (most often) a pre-nuclear pitch accent that is realized on the first stressed syllable of an utterance or a phrase unit. Brazil's is an initializing model of pitch range in which an initial pitch is interpreted on the basis of the previous unit, and sets the pitch space of the current utterance. This is not the only way to interpret a pitch peak. A listener is somehow able to accurately assess how high a pitch peak is in a speaker's range even in the absence of other peaks. We model this second interpretation by comparing initial pitch with a speaker's median pitch established for that speaker through a representative number of utterances.

While many languages have been shown to have rising final pitch contours in questions, not only is this not universal but it is also not the only part of an utterance that shows a tendency toward pitch perturbation in questions. Henriksen (2012), for example, has suggested early rising pitch contours may have communicative value signaling question intent in Manchego Spanish dialogs. Higher *initial* pitch has been documented in questions for Finnish (Iivonen, 1998), Mandarin (Shen, 1990), Zapotec (Sicoli, 2007), Portuguese (Cruz Ferreira, 1998; de Moraes, 1998; Fónagy, 1998), Spanish (Sosa, 1999), and Dutch (Haan, 2002). Just from this list, we see that higher initial pitch can occur in both tonal and non-tonal languages, and both in languages that front Wh-words and those that leave them in situ.

Many tonal languages, such as Mandarin (Shen, 1990) and Zapotec (Sicoli, 2007), are said to show no final rising intonation contour for questions. These languages derive most final pitch contours from lexical tone specifications. Yet, both Zapotec and Mandarin do show higher pitch in interrogatives (Peng et al., 2005; Sicoli, 2007). This higher pitch is a register phenomenon, detectable at the onset of the utterance on the first prominent syllable, and cuing the interrogative frame. Similarly, Finnish is also described as lacking final rise but showing higher initial pitch in questions (Iivonen, 1998, p. 318), as is Moroccan Arabic (Benkirane, 1998, p. 353). Based on experimental studies of read Dutch, Haan (2002) argues that Dutch shows both final rise in questions as well as higher onset pitch. For English, Bolinger (1989) argued that Wh-questions canonically show an initial rise with a subsequent fall and thus differ from yes/no questions.

2.3 Social action

The prototypical question performs the social action or speech act of requesting information (e.g., *What time is it?* or *Is Anne home yet?*). Yet, in everyday conversation, questions involve a wide variety of social actions. For instance, requests for information comprised only 43% of all questions asked in an American English corpus of face-to-face conversation (Stivers, 2010). Other social actions accounted for the majority of questions including relatively common social actions, such as initiating repair on the prior utterance (e.g., *What did you say?*), requesting confirmation (e.g., *She's coming tonight isn't she?*) offering (e.g., *Do you want me to bring dinner over tonight?*), or requesting (e.g., *Could you pick up the prescription on your way home?*). Less common varieties include pre-announcements (e.g., *Do you know what I heard tonight?*) as well as evaluative questions (e.g., *Isn't that a horrible color?* or *The weather's just gorgeous isn't it?*). With such variety of social actions implemented through questions and the fact that the actions typically implemented by questions make relevant a response, these sequences offer a ripe area in which to investigate how participants identify social actions (see also Heritage, 2012).

As this background shows, prosody is functional both in creating the form of a question and for the recognition of social action through potentially disambiguating between indirect and direct speech acts. In examining a cross-cultural corpus of questions and responses, we ask: Is initial pitch a cue for the social action that a question implements?

3 Data, variables, and methods

3.1 Data

In order to address our question, we drew on a corpus of question–response pairs in video recordings of everyday informal conversation in 10 languages (Stivers et al., 2010). This sample allowed us to vary the affordances of the languages' grammars to examine whether pitch functioned as a cue across the languages. Table 1 gives an overview of languages, and the contributors to whom we are grateful for access to the data.

Table 1. Languages included in the project.

Language name	Language group	Data collected by
ǂĀkhoe Haiǁom (Namibia)	Khoisan	Hoymann
Danish	Germanic	Heinemann
Dutch	Germanic	Englert
English (US)	Germanic	Stivers
Italian	Romance	Rossano
Japanese	Isolate	Hayashi
Korean	Ural-Altaic	Yoon
Lao	Tai	Enfield
Tzeltal (Mexico)	Mayan	Brown
Yélî Dnye (Papua New Guinea)	Isolate	Levinson

Each interaction in this corpus involved between two and six consenting participants (see Stivers et al., 2009, 2010 for overviews of the data). Participants were not instructed to speak about any particular topic and were simply recorded doing the activity they were otherwise engaged in. Participants were often engaged in additional activities (e.g., eating, drinking, stringing beads). As long as the task was not determining the overall direction or structure of the conversation this was considered acceptable. Thus, institutionalized talk as found in service encounters (Merritt, 1976) and rituals (Kuipers, 1990) were excluded. Each contributor to the original corpus worked on a separate language. The sample was intentionally diverse, but since participants in the project had to obtain substantial video corpora of natural conversation, it was to some extent based on availability. Data were obtained from 10 languages from five continents: Europe, the USA, Southeast Asia, Mexico, Namibia, and Papua New Guinea.

Each contributor identified 350 consecutive questions across 5–17 separate interactions (101 conversations in the total dataset). For the present study all 10 languages were studied using a subset of 70 conversations involving 180 total speakers. A total of 2512 questions are included in the analyses for this study. There were selected from among the recordings with the best audio quality and least background noise. Sometimes microphone distance, noise from the recording environment, or speaker overlaps made measuring the initial pitch unreliable or impossible. For this reason some questions within an otherwise utilized corpus were dropped from the dataset. Our primary unit of analysis is the question utterance.

Differences in linguistic typology raise challenges for a cross-linguistic comparison of prosody. For example, questions are formally marked through different means in these languages. Questions may or may not have phrase-final intonation patterns. Content questions may front Wh-words or leave them in situ. Polar questions may or may not show inversion, or may have final particles that indicate interrogative function. However, initial pitch appears to be a resource that can operate across all of these languages.[1]

3.2 Coding

3.2.1 Questions. Each researcher who contributed data to the questions project coded for question type and social action type. A question was identified on formal grounds (the presence of interrogative morpho-syntax or prosody) and/or functional grounds (the utterance sought information, confirmation, or agreement, whether this was accomplished with or without interrogative morpho-syntax or prosody). For further details, see Stivers and Enfield (2010).

Three types of questions were coded in these data: content, polar, and alternative questions. Content questions were coded as such if there was a content question morpheme ("what", "who", "where", etc.), regardless of its placement in situ or, if the language allowed, fronted (Stivers et al., 2010). Polar questions were coded as such if the answering system for the question was based on polarity ("yes"/"no" or functional equivalent, such as "uh-huh", "nope", head nod, head shake, etc.). Alternative questions were coded when two or more choices were offered without a break (i.e., no pause and a continuous prosodic unit). When two choices were offered in succession (e.g., *Would you like coffee? (0.5) or tea?*) then these were coded as two polar questions.

3.2.2 Social action type.
All questions were coded for the sort of social action (or speech act) they performed (Stivers et al., 2010). For this study a dichotomous variable was created from the various action types. Specifically, *evaluative questions* were defined as questions coded originally as either "assessments" or "rhetorical questions". Assessments were coded as questions in the study if they sought agreement with an evaluation (whether positive or negative) through the use of, for instance, an interrogative morpheme, negative interrogative syntax, or the use of a turn-final question marker such as a tag (Stivers et al., 2010). Examples would be "Isn't this just delicious?" and "It was a bit boring, wasn't it?". Rhetorical questions were coded as such if they were formally marked as questions but were judged not to be seeking an answer to the question. Examples would include "Why do I even bother?" and "Are you nuts?" (Schegloff, 1984; see also Koshik, 2005). Both Schegloff and Koshik argue that rhetorical questions work to assert the speaker's opinion and thus make relevant agreement or disagreement rather than an answer.

Both assessments and rhetorical questions are social actions that depart from the "information request" action that is canonically associated with questions. Assessments and rhetorical questions are both evaluative and seeking agreement (and affiliation) by their interlocutor. For this study we differentiated evaluative questions from all other questions.

3.2.3 Pitch.
The critical variables added to this study were pitch measurements. We measured raw Hz values for each question in the study as well as those for each response to a question. We also measured each speaker's turn prior to their question or response. The first prominent syllable was identified. Then the peak of intensity of the vowel nucleus was measured for its fundamental frequency using PRAAT software (Boersma & Weenink, 2011). One author (MS) and a student assistant conducted all coding. To test inter-rater reliability we calculated the intra-class correlation coefficient (ICC). The ICC was high: $R = .94$ for question-pitch measures ($n = 124$ questions across three languages). Across the languages of the corpus, and across turns within each language, the first prominence could fall on a number of positions in the utterance, but we found it to be most commonly within the first three syllables. For example, the two English questions "Where is it?" and "To where was it sent?" would both be measured on the vowel nucleus of /where/ underlined above. Figure 1 shows the point of measurement for a Dutch question *Hoe kwam ik daar nou zo op?* "How did I get on this topic". There, our pitch measurement is the stressed nucleus of the second syllable. This is located at the peak of prominence shown at the cross-hairs on the amplitude track (smooth light line). In Figure 1, pitch is read off the pitch track (dotted dark line) with the measure shown at the right margin as 166.5 Hz.

In addition to these measurements of pitch, we calculated a median pitch for each of the 180 speakers represented in these data using all pitch measurements (questions, responses, and turns before a speaker's question or response) for each speaker. This measure was preferred over the mean for being more robust to outliers. To examine whether initial pitch of a question is predicted by such factors as the question's action and the question type, controlling for speaker sex, we assessed whether the initial pitch of the target question was near the top of a given speaker's range
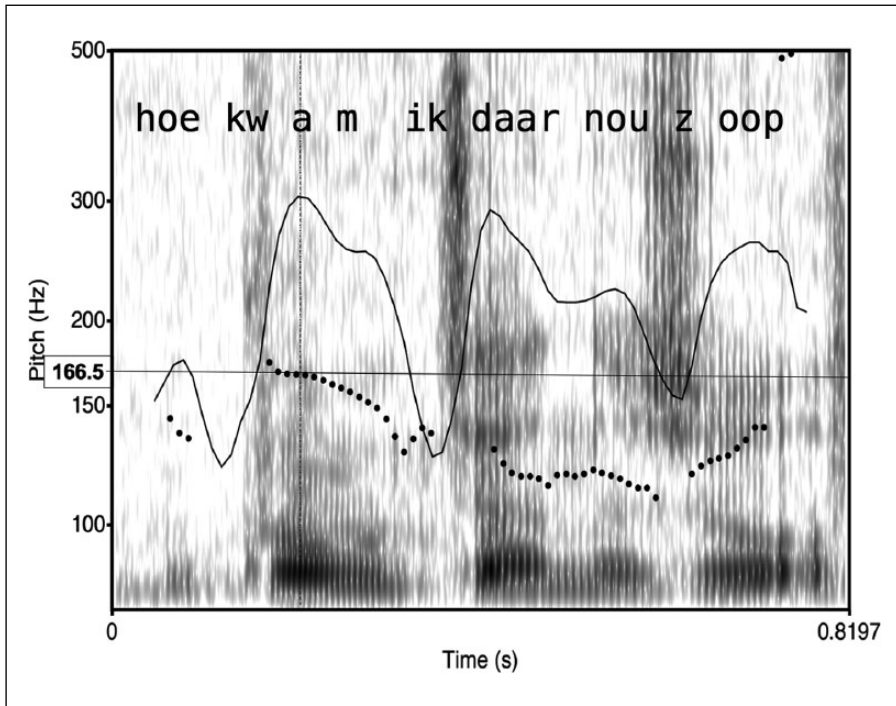
Figure 1. Pitch at first prominence.

of sample utterances. Independent of language, pitch in the top 15% of a speaker's range was predictive of Content (Wh-) and Evaluative questions together and Evaluative questions alone were clustered among turns in the top 10% of a speaker's range.

Using initial pitch as a phonetic indicator raises the question of whether F0 scaling is affecting initial pitch. Reviewing pitch-onset models, Ladd (2008) contrasts the initializing model with normalizing models, in which initial pitch sets the "tonal space" of an utterance given its length or accentual complexity. The normalization of pitch across longer utterances has been hypothesized to produce higher initial-pitch values that provide the tonal space needed for the declination of pitch across the length of the utterance (so that a speaker does not hit the floor of their pitch range before the end). We checked our sample to see if initial pitch values were being influenced by utterance length and found rather a random distribution of utterance lengths. For example, it was typical for a speaker's evaluative questions and Wh-questions to range from one-word minimal responses to multi-phrase utterances. Thus there was no dependency between utterance length and the initial pitch of questions and other utterances we used in the calculation of speaker medians. As an extra precaution, we removed from the analysis any speakers with fewer than five utterances, which could be more easily skewed. The range of utterances used to calculate median initial pitch by language and the upper and lower quartiles for each language is shown in the boxplot in Figure 2.

3.2.4 Covariates. Speaker sex was included as a covariate because females have higher pitch on average cross-linguistically (although how different this is from men's averages varies dramatically across our sample, as it is exaggerated in some societies, and minimized in others, as seen
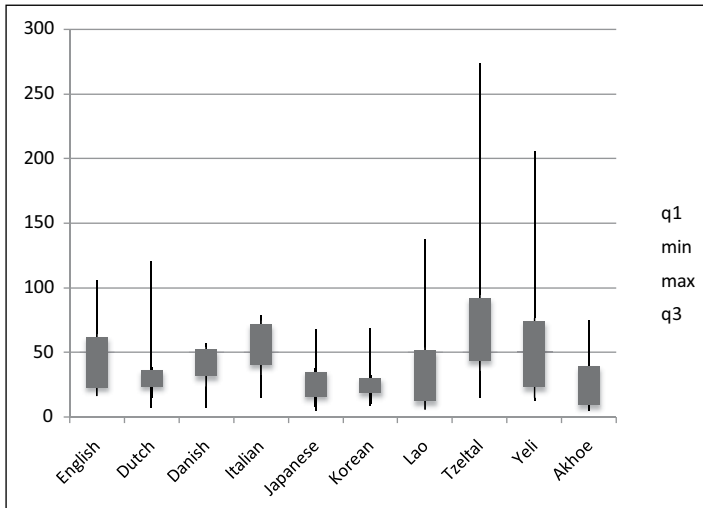
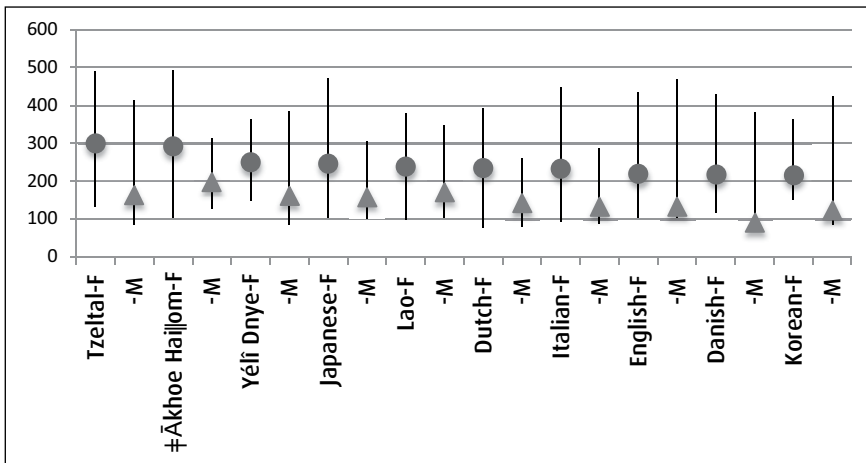**Figure 2.** Range of utterances used to calculate median initial pitch.

Figure 3. Pitch mean and range by language and speaker sex.

in Figure 3). Bivariate results suggested that women's and men's pitch levels varied with respect to Wh-questions in particular, so we also included an interaction effect for speaker sex and question type.

3.3 Analytic methods

Our primary goal was to identify whether there was an association between initial pitch and social action. We used univariate and bivariate analyses to examine relationships that might be better examined multivariately as well as to ascertain in what form we should analyze a variable. For instance, Hz was coded as a continuous variable but ultimately analyzed as dichotomous variable (in the top 10% of a speaker's utterances). Social action was originally a categorical variable with

six values, but we determined that for this study it was best analyzed as a binary variable—either an evaluative question or a "typical" information question.

We used this information to build a multivariate model predicting high pitch onset in a question. These data were clustered within one of ten languages. We therefore used a Generalized Linear Latent and Mixed Model in STATA software (Rabe-Hesketh & Skrondal, 2005). This is one model from a class of multilevel or hierarchical statistical models, which takes into consideration that there is clustering in the data. The dependent variable was constructed to be binary (whether or not the question's first prominent syllable was uttered at a pitch level that was in the top 10% of the speaker's utterances), so a logit model was used to fit the data. This model tested as a two-level model to account for the clustering of the data by language.

4 Results

4.1 Univariate and bivariate results

Pitch levels measured on the first prominent syllable in questions varied somewhat by language and by sex. The overall sample median was 244 Hz for women and 147 Hz for men with similar ranges of 78–493 Hz and 76–468 Hz, respectively. These ranges include speakers' uses of non-modal phonations to achieve pitches higher or lower than their natural modal range. Uses of the non-modal phonations falsetto voice and creaky voice to extend the speech range have been discussed by Sicoli (2007), Podesva (2007), and Sicoli (2010). These results are summarized in Table 2 and presented in a histogram form in Figure 3, which shows the median pitch and the pitch range for males and females of each language. These pairs are ordered from the highest female median (Tzeltal) to the lowest female median (Korean). The circles indicate female median pitch and the triangles male median pitch.

Tzeltal and ǂĀkhoe Hai‖om women both made use of very high pitch with medians close to 300 Hz. Tzeltal women showed a median pitch of 299 Hz and a range of 134–490. ǂĀkhoe Hai‖om women had a median pitch of 291 Hz and a range of 103–493. ǂĀkhoe Hai‖om men had the highest median initial pitch (198 Hz). This may reflect that the ǂĀkhoe Hai‖om videos include children. The next highest median is the all-adult Lao corpus. The difference between male and female medians was greatest for Tzeltal and Danish at around 130 Hz and smallest for Lao (<70 Hz), with the other languages all clustered around 90 Hz.

Initial pitch floor (defined here as the lowest of the range) is also a useful measure to compare and we report here these results. Korean and Yelî-Dnye women had the highest pitch floors at 151 and 147 Hz, respectively. Between those only Yelî-Dnye had a high median (249 Hz). So while not showing peaks as high as Tzeltal and ǂĀkhoe Hai‖om, Yelî-Dnye women were the most consistently high in the corpus. Dutch women, well known for their low pitch in speech (van Bezooijen, 1995) showed the lowest pitch floor, 78 Hz on average, in their initial pitch of questions with a median of 235 Hz. Dutch men had a narrow and low range (79–261 Hz). Lao stood out as different in that women also showed a low pitch floor of 97 Hz. Danish men had the lowest median of 90 Hz. English, Korean, and Tzeltal men all reached the highest extremes of over 400 Hz, but English and Korean men were among the lowest medians, so Tzeltal men spoke higher more frequently (this even while one Tzeltal man had one of the lowest pitched voices the corpus).

There are masculine and feminine biological correlates identified for pitch in which the thicker vocal folds and generally larger larynxes of men, produce tendencies toward lower fundamental frequencies in males (Fitch, 1994; see also Gussenhoven, 2002, 2004; Ohala, 1994; Simpson, 2009). While this is reflected in the overall medians of the speakers in our corpus, we observe that speakers of different languages tend to, as members of distinct ethnolinguistic groups, utilize

Table 2. Pitch median and range by language and speaker sex.

Language name	Female speakers median Hz in Q (min–max)	Male speakers median Hz in Q (min–max)	Median difference (male–female)
Tzeltal (Mexico)	299 (134–490)	162 (84–413)	137
Danish	218 (117–430)	90 (76–382)	128
Italian	233 (93–449)	133 (87–287)	100
Dutch	235 (78–393)	141 (79–261)	94
‡Ākhoe Hai‖om (Namibia)	291 (103–493)	198 (128–312)	93
Korean	216 (151–363)	123 (85–424)	93
Yélî Dnye (Papua New Guinea)	249 (147–363)	160 (86–386)	89
Japanese	246 (105–471)	158 (101–306)	88
English (US)	219 (102–433)	132 (102–468)	87
Lao	238 (97–377)	170 (103–348)	68
Average median (Range)	244 (78–493)	147 (76–468)	97

different portions of the pitch range in their questions, with some exploiting higher targets and wider ranges than others. Because of these pitch facts, the biological affordances need to be seen only as first-order differences that can be appropriated to minimize or maximize a phonetic index of the social distinction. These findings suggest it would be valuable to query this corpus and similar ones to contribute to developing a comparative perspective on speech and gender. Our primary interest, however, was the extent to which pitch was used as an interactional resource to convey to an interlocutor that something communicatively special was being done with the particular utterance.

In preliminary analyses, pitch was examined as both a continuous variable and as a four-step variable, which grouped quartiles together. A relationship between question actions was already visible in these analyses, but given that we believe that what is salient to hearers is not a slight contrast but more of a binary contrast between a high onset or a mid-range onset, we explored using the dichotomous variant. Our view is consistent with literature describing initial pitch in questions as high or not high (Cruz Ferreira, 1998; de Moraes, 1998; Fónagy, 1998; Haan, 2002; Iivonen, 1998; Shen, 1990; Sicoli, 2007; Sosa, 1999). Hz values were ranked by speaker for their percentile within a speaker's range.

In addition to pitch levels varying by sex of speaker, bivariate results showed that the pitch varied in the sample as a whole by social action type, whether the action was a typical question or an evaluative one ($t[2516] = -5.37$, $p = 0.000$). Further tests of each language showed that for nearly all languages, Hz levels for evaluative questions were higher than for questions implementing other sorts of social actions (e.g., information requests, confirmation requests, initiations of repair). These results are summarized in Table 3. However, the *t*-values of these tests were commonly non-significant (in six of our 10 languages) and in Yelî-Dnye and English, speakers were very marginally more likely to use lower pitch in evaluative questions than in other types of questions (but this was a difference of 4–8 Hz and was non-significant). It is possible that not all languages discriminate between evaluative and other sorts of questions using pitch; it is also possible that all languages do mark such questions with pitch different from the norm, but that some prefer higher pitch while others prefer lower pitch (on pitch lowering for pragmatic prominence see Kügler and Genzel (2012)), or we may lack a sufficient number of cases for these languages in the corpus to see the pattern accurately.

Table 3. Summary of results of *t*-tests of Hz values for evaluative questions by language: $p = .06$ §; $p < .05$ *; $p < .01$ **; $p < .001$ ***.

Language name	Difference between non-evaluative versus evaluative questions
ǂĀkhoe Haiǁom (Namibia)	−3.83 Hz, $t(248) = −0.17$
Danish	−14.64 Hz, $t(146) = −0.52$
Dutch	−33.68 Hz, $t(211) = −2.95$**
English (US)	3.66 Hz, $t(180) = 0.17$
Italian	−53.77 Hz, $t(313) = −2.80$**
Japanese	−9.11 Hz, $t(314) = −0.95$
Korean	−26.98 Hz, $t(157) = −1.94$*
Lao	−16.13 Hz, $t(276) = −1.52$§
Tzeltal (Mexico)	−24.19 Hz, $t(335) = −2.05$*
Yélî Dnye (Papua New Guinea)	8.6 Hz, $t(318) = 0.49$

Tests of each language for question type compared Wh-questions to other questions and showed that for Danish, English, Italian, Lao, Tzeltal, and Yelî-Dnye, Wh-questions were higher than other questions. This reached statistical significance in four of the six languages. These results are summarized in Table 4. While there is no grammatical account for the higher pitch on evaluative social actions, with regard to Wh-questions, it is possible that the presence of Wh-fronting in most of these languages provides an explanation for this association. That said, such an account would not preclude the communicative value that high pitch would provide. Indeed, in a language such as Lao, which lacks Wh-fronting, this may be a more valuable communicative resource. Lao showed one of the greatest differences in initial pitch of Wh-questions versus other all other questions as well. Without fronting that would syntactically mark the content question early in the turn, the high pitch may give a clue to the listener that this is a Wh-question.

Conversely though, not all languages that front Wh-pronouns showed higher initial pitch in content questions. Dutch and ǂĀkhoe Haiǁom, both languages that front Wh-words, did not show higher pitch and but were moderately more likely to be lower (~10 Hz). It is notable that Japanese and Korean, both languages that do not front Wh-words, did not show higher initial pitch in Wh-questions but were rather marginally more likely to use lower pitch, with Korean reaching significantly lower pitch in the context of Wh-questions. To the extent that fronting explains some of these facts, the patterns suggest that the fronted position may use pitch differently: higher in some languages and lower in others.

Because both Wh-questions and evaluative questions are associated with the use of high pitch in questions, a multivariate model that considers the effect of each of these factors as well as the independent role of speaker sex, social action, and the interaction between speaker sex and question type, was fit to the data.

4.2 Multivariate results

A multivariate logistic regression model was used to identify predictors of high pitch on the first prominent syllable of the question. The results of the multivariate logistic regression are shown in Table 5 as odds ratios with 95% confidence intervals. The variable *Language* did not contribute significantly to the variance of initial pitch in questions independent of other predictors. Thus,

Table 4. Summary of results of *t*-tests of Hz values for Wh-questions by language: *p* = .06 §; *p* < .05 *; *p* < .01 **; *p* < .001 ***. .

Language name	Difference between Wh- versus other questions
‡Ākhoe Hai‖om (Namibia)	10.20 Hz, *t*(248) = 0.87
Danish	−28.11 Hz, *t*(146) = −2.10*
Dutch	8.33 Hz, *t*(211) = 0.87
English (US)	−28.86 Hz, *t*(180) = −2.40**
Italian	−8.47 Hz, *t*(313) = −1.01
Japanese	7.32 Hz, *t*(314) = 0.64
Korean	22.60 Hz, *t*(157) = 2.13*
Lao	−20.52 Hz, *t*(276) = −2.45**
Tzeltal (Mexico)	−19.36 Hz, *t*(335) = −1.90*
Yélî Dnye (Papua New Guinea)	−3.42 Hz, *t*(318) = −0.54

Table 5. Results of multivariate logistic regression predicting pitch in the top 10% of a speaker's range.

Level 1 variables	Odds ratio	95% CI
Question type		
Content question	1.33	0.91, 1.93
Speech act type		
Evaluative question	1.46*	1.00, 2.14
Co-variates		
Male speaker	1.01	0.72, 1.42
Content Questions × Male speaker	0.83	0.49, 1.38
Context variables	Estimate	Std. error
Level 2: Variance at language level	0.06	0.04
* denotes *p* < 0.05		

although we saw variation in languages, this may have been driven by other factors. We found only one predictor of high pitch (i.e., in the top 10% of the speaker's range) in the first prominent syllable of the question: the social action type. Thus, when a speaker begins a question very high, the likelihood is that the recipient will have to move beyond the literal meaning to ascribe an action and formulate their response. This result is independent of the language spoken, whether the question was a Wh-question or not; whether the speaker was male or female; and an interaction between speaker sex and question type.

The questioner's use of high initial pitch was significantly associated with evaluative questions. Independent of the question type (polar versus Wh), the odds of an evaluative question being asked with raised pitch were 1.46 times that of a question performing another type of social action. This was independent of the language being spoken, and controlling for speaker sex. Furthermore, women were no more likely than men to use high pitch relative to their median in asking questions, nor was there any interaction effect between sex and question type.

In contrast to the significant association of evaluative questions, although bivariately associated, high pitch at question beginning was not significantly associated with content (Wh-) questions relative to polar (yes/no) and alternative questions. Following the presentation in Section 4.1 that showed pitch in content questions to possibly follow several patterns involving focus and

syntactic structure, the most interesting finding to discuss further is that pitch is predictive of social action type.

5 Discussion

When initiating our study of initial pitch we considered that initial pitch could acquire its function either in a local contrast with an immediately prior utterance (an initializing model; perhaps how topic shifts are recognized), or that a current utterance could be held in contrast to an addressee's notion of the speaker's median pitch (a normalizing model). Brazil et al. (1980) argue for the local contrast explanation in which the key pitch of utterances is defined as being in the same key, a higher key, or lower key than the previous utterance. Our findings, however, are more compatible with a normalizing model since in our conversational corpus initial pitch derives its function from a contrast between the current utterance and a median abstracted from a speech sample (see also Gussenhoven, Repp, Rietveld, Rump, and Terken, 1997). Addressing the nature of sociolinguistic variables generally, Labov came to understand variables as showing a statistically reckoned character where participants can show a sensitivity to small deviations from a norm reckoned over a short number of exemplars in interaction (Labov, 1972). Such statistically based reckoning is also apparent in such phenomena as phonological and linguistic accommodation (Babel, 2009; Giles, Coupland, & Coupland, 1991) and in statistical learning theories of child language acquisition (for reviews see Romberg & Saffran, 2010; Saffran, Werker, & Werner, 2006).

While not arguing against the value of interpreting pitch with reference to adjacent pitch values, our data suggest that for initial pitch in questions, participants show sensitivity to a deviation from a median initial pitch calculated from some number of exemplar utterances and that this sensitivity is used to mark questions functioning as indirect speech acts with evaluative function. This finding lends itself as a hypothesis that could be further tested with laboratory methods and compared to other conversational corpora. Interestingly, major theories of prosody, such as Firthean Prosodic Analysis (Firth, 1948; Ogden, 2006; Odgen & Local, 1994) or Autosegmental Phonology (Goldsmith, 1990), which were developed to account for local contrasts, are not well suited to incorporate these results, which emerge from a statistical accounting of the data across a corpus and are not specific to a local sequence or specifically part of an intonational grammar. The signifying process we have discovered to be at play in initial pitch in evaluative questions is better accounted for in terms of semiotics and interactional sociolinguistics.

5.1 Iconicity, indexicality, and markedness

We asked at the beginning of this paper whether the initial pitch of a question may help participants recognize the action that a given question is being used to implement. As our findings support that this is the case for evaluative questions, we suggest that the semiotic concepts of iconicity, indexicality, and markedness are important to understanding how. Iconicity is a pervasive organizing feature of language and talk, as attested by a long list of publications in patterns of *syntax* and *morphology* (Brinton, 1987; Fónagy, 1995; Givón, 1985; Haiman, 1985; Hampe & Schönefeld, 2003; Jakobson, 1966; Posner, 1982), *discourse* (Becker, 1982; Enkvist, 1990; Friedrich, 1979; Givón, 1983; Ishikawa, 1991; Maschler, 1993), and *sound symbolism* (Childs, 1994; Hinton, Nichols, & Ohala, 1994; Sadowski, 2001; Tanz, 1971). These works reinforce the notion of iconicity as a basic sign building process in multiple areas of language and speech. Peirce (1955, p. 102) described the icon from the perspective of the sign–object relationship as grounded in a formal likeness between a sign and its object—the commonplace understanding of iconicity in linguistics since Jakobson (1965) introduced it in "Quest for the essence of language". It is important to point out that

Peirce's understanding of the icon was broader than this to also include signs that do not objectively resemble their object but give rise to iconic interpretations. Such rhematic signs take an icon as their interpretant. Thus, marked initial pitch has its iconicity not insofar as it formally resembles a meaning or speech act type but in it being interpretable as an icon that relates out-of-the-ordinary pitch with out-of-the-ordinary action. Pitch that is markedly different than a speaker's median pitch acts like a vendor cry standing out from the din of background noise at a street market: in the first order it is a rhematic index that directs attention to an utterance as a kind of object. In the second order it involves a diagrammatic iconicity between the divergent pitch and divergent social action.

The iconic interpretation of marked initial pitch is also interpretable in a parallelism between the addition of speaker effort (physically doing something out-of-the-ordinary) to produce a marked initial pitch and the out-of-the ordinary social action to which it corresponds. Here the iconicity is interpretable in a qualitative likeness between form and function where more coding corresponds with more complex semantic representations. In a polysynthetic language, such as Inuit for example, this type of iconicity can be seen in the presence of more affixes corresponding to longer word forms for more complex semantic representations. Alternatively, a pervasively prosodic language such as Chinantec (Otomanguean language family of Mexico) shows us that such complexity can be prosodic. Adding tone specifications, nasalization, or phonation produces a greater prosodic complexity for a verb that parallels its semantic complexity. The process we are pointing to for initial pitch is simpler than these morphological (segmental or prosodic) processes because it does not include the functions of symbolic morphemes. For initial pitch in questions we are not dealing with the presence of formal pitch morphemes but rather with a feature of embodied performance in which a speaker takes the trouble to deviate from his or her median or unmarked initial pitch. Our use of median pitch is a statistically motivated parallel to the idea of a "baseline" in the earliest work on "paralanguage" from which meaningful voice features such as pitch, loudness, and speech rate could diverge (see Birdwhistell, 1961, for a review). A recipient can interpret the addition of the mark on the basis of Grice's maxim of quantity: that one be as informative as needed, but not more, and also Grice's maxim of manner: that if you signal in an unusual way this invites an unusual interpretation (cf. Levinson, 2000, on the M-heuristic in presumptive meanings). Thus, if extra work is being done to mark pitch, such deviation should prompt an increase in the inferential activity through which the social action of a question is ascribed.

Gussenhoven (2002) described "effort" as one of three biological codes related to universal functions of intonation in languages. The "effort code" is related to the amount of energy expended on speech production. Effort can not only lead to higher or lower pitch targets but also to more numerous pitch movements and greater pitch excursions. He related greater pitch excursions to marking contrastive focus, and to functions such as expressing "surprise" and "helpfulness" (p. 50). Our findings on initial pitch in questions are in general agreement with Gussenhoven's interpretations on universal functions for prosodic effort. However, we are able to add that the functions of greater prosodic effort are not simply grounded in biology, but are grounded in semiotic processes available to the speaker in designing an utterance appropriately for an interpreter, and to the interpreter in ascribing a communicative function or social action to the utterance. The added effort in creating a marked pitch signifies a marked frame for interpretation, by means of an iconic sign relationship diagrammatically relating a marked form and a marked action (cf. Brown & Levinson, 1987, pp. 267–268; Jakobson & Pomorska, 1990, p. 134).

5.2 Pitch frames

A recipient's ascription of an action to a speaker's utterance is shaped by the lexical and morphosyntactic aspects of the speaker's turn design, as well as by its sequential context and its prosody. In

our data, prosody that marks off a question as different from its "prima facie" use "invites a search" for its function (Schegloff, 1998, p. 250). Bateson (1955) argued that all natural communication must take place within metacommunicative frames, as only within a delimiting frame does a communicative act become interpretable. The framing function of initial pitch in evaluative questions is akin to what Gumperz (1992) called a contextualization cue, "a[n often] prosodic trigger that in conjunction with lexical material will invoke frames and scenarios within which the current utterance is to be interpreted as an interactional move" (Levinson, 2003, p. 33). Gumperz suggests that contextualization cues "serve to highlight, foreground or make salient certain phonological or lexical strings *vis-à-vis* other similar units, that is, they function relationally and cannot be assigned context-independent, stable, core lexical meanings" (Gumperz, 1992, p. 232). Our findings add that the relational functioning of marked initial pitch in questions relies on general semiotic principles of iconicity, indexicality, and markedness. Pitch, and more specifically pitch at first prominence, is especially well suited to carry information that contrasts one social action type from another. As a framing device, prosodic coding may have an advantage over lexical means in that it is not necessarily adding additional propositional or other semantic content. Since there is a clear difference between the meaning of the words of a conversational turn and what that turn is doing, contextualization cues carry a message about the context for interpretation (Levinson, 2003), thus having pragmatic function. Pitch—early in an utterance—establishes a frame within which a listener can ascribe the action of an utterance. The fact that the mark of difference is very early in the utterance may help explain the great speed at which recipients are able to ascribe an action to an utterance and respond appropriately with little to no gap between turns in conversation (Levinson, 2013).

6 Conclusion

The literature on question intonation has tended to look at redundancy between prosodic and morphosyntactic signaling where parallel signals in multiple channels communicate the same message, or where a non-question form is used to request information. Rather than question forms for which there were only indeterminate results involving Wh-questions and none for other question forms, our findings support the idea that initial pitch is important for distinguishing the social action type ascribed for a question. Pitch provides a means of signaling that a question is not serving a canonical information-requesting function but that something else is going on—that a responder needs to look beyond the unmarked interpretation, as a special interpretation is needed. There is much to be gained by further investigation of prosodic framing in conversation. Is the function we have found for deviant initial pitch in evaluative questions more generalizable to other indirect speech acts?

In investigating the ascription of communicative functions or actions to utterances, a primary question is: How much is explicitly expressed in an utterance and how much can be ascribed to sequential position and other aspects of its environment? The formulation of the utterance must consider both the internal linguistic relations of the proposition and the framing of that proposition through metalinguistic means such as prosody among other co-present features such as gesture and gaze. Bringing together methods from the interactional study of social action with methods of acoustic analysis developed in the experimental study of phonetics and phonology we have developed an understanding that initial pitch is important to question design. For questions with evaluative function that seek agreement rather than request information or confirmation, deviant initial pitch invites the question recipient to begin a search for the question's indirect function.

Acknowledgments

Note

1. Some of the typological variability of the languages is presented in the Appendix.

References

Allan, R., Holmes, H., & Lundskaer-Nielsen, T. (2000). *Danish: An essential grammar*. London, UK & New York, NY: Routledge.

Austin, J. L. (1962). *How to do things with words*. Cambridge, MA: Harvard University Press.

Babel, M. (2009). *Phonetic and social selectivity in speech accommodation*. Berkeley, CA: University of California, Berkeley Ph.D. Dissertation.

Bateson, G. (1955). A theory of play and fantasy. Psychiatric research reports, 2, 39–51. Reprinted in Bateson, G. (1972). *Steps to an ecology of mind. Collected essays in anthropology, psychiatry, evolution, and epistemology*. Chicago, IL: University of Chicago Press.

Becker, A. L. (1982). Beyond translation: Esthetics and language description. In H. Byrnes (Ed.), *GURT 1982: Contemporary perceptions of language: interdisciplinary dimensions* (pp. 124–138). Washington DC: Georgetown University Press.

Beckman, M. E., & Pierrehumbert, J. (1986). Intonational structure in Japanese and English. *Phonology Yearbook*, *3*, 255–310.

Benkirane, T. (1998). Intonation in Western Arabic (Morocco). In D. Hirst & A. Di Cristo (Eds.), *Intonation systems. A survey of twenty languages* (pp. 345–359). Cambridge, UK: Cambridge University Press.

Birdwhistell, R. (1961). *Paralanguage: 25 Years after Sapir,* In H. W. Brosin (Ed.), *Lectures on experimental psychiatry*, (pp. 43–64). Pittsburgh, PA: University of Pittsburgh Press.

Boersma, P., & Weenink, D. (2011). Praat: Doing phonetics by computer [Computer program]. Version 5.*3*. Retrieved from http://www.praat.org/

Bolinger, D. (1989). *Intonation and its Uses*. Stanford, CA: Stanford University Press.

Brazil, D. (1997). *The communicative value of intonation in English*. Cambridge, UK: Cambridge University Press.

Brazil, D., Coulthard, M., & Johns, C. (1980). *Discourse intonation and language teaching*. London: Longman.

Brinton, L. J. (1987). Diagrammatic iconicity in English syntax. *Recherches Semiotiques-Semiotic Inquiry*, *7*, 55–72.

Brown, P. (1996). Isolating the CVC root in Tzeltal Mayan: A study of children's first verbs. In E. Clark (Ed.), *The proceedings of the 28th annual child language research forum* (pp. 41–52). Stanford, CA: Center for the Study of Language and Information.

Brown, P. (2010). Questions and their responses in Tzeltal. In T. Stivers, N. J. Enfield, & S. C. Levinson (Eds.), Question-response sequences in 10 languages [Special Issue]. *Journal of Pragmatics*, 42, 2647–2648.

Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage*. Cambridge, UK: Cambridge University Press.

Chafe, W. (1998). Language and the flow of thought. In M. Tomasello (Ed.), *The new psychology of language: Cognitive and functional approaches to language structure* (pp. 93–111). Mahwah, NJ: Psychology Press.

Childs, G. T. (1994). Expressiveness in contact situations: The fate of African ideophones. *Journal of Pidgin and Creole Languages*, *9*, 257–82.

Cole, P., & Morgan, J. (Eds.) (1975). *Syntax and semantics, 3: Speech acts*. New York, NY: Academic Press.

Couper-Kuhlen, E. (2001). Interactional prosody: High onsets in reason-for-the-call turns. *Language in Society*, *30*, 29–53.

Cruz Ferreira, M. (1998). Intonation in European Portuguese. In D. Hirst, & A. Di Cristo (Eds.), *Intonation systems: A survey of twenty languages* (pp. 167–178). Cambridge, UK: Cambridge University Press.

de Moraes, J. A. (1998). Intonation in Brazilian Portuguese. In D. Hirst, & A. Di Cristo (Eds.), *Intonation systems: A survey of twenty languages* (pp. 179–194). Cambridge, UK: Cambridge University Press.

de Ruiter, J. P., Mitterer, H., & Enfield, N. J. (2006). Projecting the end of a speaker's turn: A cognitive cornerstone of conversation. *Language*, *82*, 515–535.

Enfield, N. J. (2007). *A grammar of Lao*. Berlin, Germany: Mouton de Gruyter.

Enfield, N. J. (2010). Questions and responses in Lao. In T. Stivers, N. J. Enfield, & S.C. Levinson (Eds.), Question-response sequences in 10 languages [Special Issue]. *Journal of Pragmatics*, *42*, 2649–2665.

Englert, C. (2010). Questions and responses in Dutch conversations. In T. Stivers, N. J. Enfield, & S. C. Levinson (Eds.), Question-response sequences in 10 languages [Special Issue]. *Journal of Pragmatics*, 42, 2666–2684.

Enkvist, N. E. (1990). Discourse comprehension, text strategies and style. *Journal of the Australasian Universities Language and Literature Association*, *73*, 166–180.

Firth, J. R. (1948). Sounds and prosodies. *Transactions of the Philological Society*, 127–152.

Fitch, W. T. (1994). *Vocal tract length perception and the evolution of language*. UMI Dissertation Services, Ann Arbor, MI.

Fónagy, I. (1995). *Iconicity of expressive syntactic transformations*. In M. E. Landsberg (Ed.), *Syntactic iconicity and linguistic freezes. The human dimension* (pp. 285–304). Berlin, Germany: Mouton de Gruyter.

Fónagy, I. (1998). Intonation in Hungarian. In D. Hirst & A. Di Cristo (Eds.), *Intonation systems: A survey of twenty languages*, (pp. 179–194). Cambridge, UK: Cambridge University Press.

Friedrich, P. (1979). The relative non-arbitrariness of the linguistic sign. In P. Friedrich (Ed.), *Language, context, and the imagination*, (pp. 1–61). Stanford, CA: Stanford University Press.

Giles, H., Coupland, J., & Coupland, N. (1991). *Contexts of accommodation*. Cambridge, UK: Cambridge University Press.

Givón, T. (1983). Introduction. In T. Givón (Ed.) *Typological studies in language vol 3, topic continuity in discourse: A quantitative cross-language study* (pp. 3–41). Amsterdam, The Netherlands: John Benjamins.

Givón, T. (1985). Iconicity, isomorphism and non-arbitrary coding in syntax. In J. Haiman (Ed.), *Iconicity in syntax* (pp. 187–219). Amsterdam, The Netherlands: Benjamins.

Goldsmith, J. (1990). *Autosegmental and metrical phonology*. Oxford: Basil Blackwell.

Gordon, D., & Lakoff, G. (1971). Conversational postulates. *Papers from the 7th regional meeting, Chicago Linguistic Society* (pp. 63–84). [Reprinted in Cole and Morgan (pp. 83–106).]

Grice, H. P. (1975). *Logic in conversation*. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics, Vol. 3, speech acts* (pp. 41–58). New York, NY: Academic Press.

Gumperz, J. J. (1982). *Discourse strategies*. Cambridge, UK: Cambridge University Press.

Gumperz, J. J. (1992). Contextualization and understanding. In A. Duranti & C. Goodwin (Eds.), *Rethinking context: Language as an interactive phenomenon* (pp. 229–252). New York, NY: Cambridge University Press.

Gussenhoven, C. (2002). Intonation and interpretation: Phonetics and phonology. In *Speech Prosody 2002: Proceedings of the First International Conference on Speech Prosody* (pp. 47–57). Aix-en-Provence: ProSig and Université de Provence Laboratoire Parole et Language.

Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge, UK: Cambridge University Press.

Gussenhoven, C., Repp, B. H., Rietveld, A., Rump, H. H., & Terken, J. (1997). The perceptual prominence of fundamental frequency peaks. *Journal of the Acoustical Society of America*, *102*, 3009–3022.

Haan, J. (2002). Speaking of questions: An exploration of Dutch question intonation. Utrecht NL: LOT. Retrieved from http://www.lotpublications.nl/publish/articles/000097/bookpart.pdf

Haiman, J. (Ed.). (1985). *Iconicity in syntax*. Amsterdam, The Netherlands: Benjamins.

Hampe, B., & Schönefeld, D. (2003). Creative syntax: Iconic principles within the symbolic. In W. G. Müller & O. Fischer (Eds.), *From signs to signing* (pp. 243–261). Amsterdam, The Netherlands: John Benjamins Publishing Company.

Hayashi, M. (2010). An overview of the question-response system in Japanese. In T. Stivers, N. J. Enfield, & S. C. Levinson (Eds.), Question-response sequences in 10 languages [Special Issue]. *Journal of Pragmatics*, 42, 2685–2702.

Heinemann, T. (2010). The question-response system of Danish. In T. Stivers, N. J. Enfield, & S. C. Levinson (Eds.), Question-response sequences in 10 languages [Special Issue]. *Journal of Pragmatics*, 42, 2703–2725.

Henderson, J. E. (1995). *Phonology and grammar of Yele, Papua New Guinea*. Pacific Linguistics B–112. Canberra, Australia: Pacific Linguistics.

Henriksen, N. C. (2012). The intonation and signaling of declarative questions in Manchego peninsular Spanish. *Language and Speech*, *55*, 543–576.

Heritage, J. (1984). A change-of-state token and aspects of its sequential placement. In J. M. Atkinson & J. Heritage (Eds.). *Structures of Social action*, (pp. 299–345). Cambridge, UK: Cambridge University Press.

Heritage, J. (2012). Epistemics in action: Action formation and territories of knowledge. *Research on Language and Social Interaction*, *45*, 1–29.

Hinton, L, Nichols, J., & Ohala, J. (1994). *Sound symbolism*. Cambridge, UK: Cambridge University Press.

Hoymann, G. (2010). *Questions and responses in ǂĀkhoe Hai‖om*. In T. Stivers, N. J. Enfield, & S. C. Levinson (Eds.), Question-response sequences in 10 languages [Special Issue]. *Journal of Pragmatics*, 42, 2726–2740.

Iivonen, A. (1998). *Intonation in Finnish*. In D. Hirst & A. Di Cristo (Eds.), *Intonation systems. A survey of twenty languages* (pp. 311–327). Cambridge, UK: Cambridge University Press.

Ishikawa, M. (1991). Iconicity in discourse: The case of repetition. *Text*, *11*, 552–580.

Jakobson, R. (1965). Quest for the essence of language. *Diogenes*, *13*, 21–37.

Jakobson, R. (1966). Relationship between Russian stem suffixes and verbal aspects. In R. Jakobson (1971). *Selected writings of Roman Jakobson*, Vol II, word and language (pp. 198–202). The Hague, The Netherlands: Mouton.

Jakobson, R., & Pomorska, K. (1990). The concept of the mark. In L. Waugh, & M. Monville-Burston (Eds.), *Roman Jakobson: On Language*, (pp. 134–140). Cambridge MA.: Harvard University Press.

Koshik, I. (2005). Alternative questions used in conversational repair. *Discourse Studies*, *7*, 193–211.

Kügler, F., & Genzel, S. (2012). On the prosodic expression of pragmatic prominence: The case of pitch register lowering in Akan. *Language and Speech*, *55*, 331–359.

Kuipers, J. (1990) *Power in performance: The creation of textual authority in Weyewa ritual speech*. Philadelphia, PA: University of Pennsylvania Press.

Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia, PA: The University of Pennsylvania Press.

Ladd, R. D. (2008). *Intonational phonology*. Cambridge, UK: Cambridge University Press.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Levinson, S. C. (2003). Contextualizing 'contextualization cues'. In S. Eerdmans, C. Prevignano, & P. Thibault (Eds.), *Language and interaction: Discussions with John J. Gumperz* (pp. 31–39). Amsterdam: John Benjamins.

Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. Cambridge, MA: MIT Press.

Levinson, S. C. (2003). Contextualizing 'contextualization cues'. In S. Eerdmans, C. Prevignano, & P. Thibault (Eds.), *Language and interaction: Discussions with John J. Gumperz*, (pp. 31–39). Amsterdam: John Benjamins.

Levinson, S. C. (2010). Questions and responses in Yélî Dnye, the Papuan language of Rossel Island. In T. Stivers, N. J. Enfield, & S. C. Levinson (Eds.), Question-response sequences in 10 languages [Special Issue]. *Journal of Pragmatics*, 42, 2741–2755.

Levinson, S. C. (2013). Action formation and ascription. In T. Stivers & J. Sidnell (Eds.), *The handbook of conversation analysis* (pp. 103–130). Malden, MA: Wiley-Blackwell.

Lim, B. J. (2001). The role of syllable weight and position on prominence in Korean. *Japanese Korean Linguistics*, *9*, 139–150.

Maschler, Y. (1993). Iconicity in discourse: The story of Echo. *Poetics Today*, *14*, 653–689.

Merritt, M. (1976). On questions following questions in service encounters. *Language and Society*, *5*, 315–357.

Ogden, R. (2006). Firthian phonology. In K. Brown (Ed.), *Encyclopedia of language and linguistics*, Second edition. Oxford, UK: Elsevier.

Ogden, R., & Local, J. K. (1994). Disentangling autosegments from prosodies: A note on the misrepresentation of a research tradition in phonology. *Journal of Linguistics*, *30*, 477–498.

Ohala, J. J. (1994). The frequency codes underlies the sound symbolic use of voice pitch. In L. Hinton, J. Nichols, & J. J. Ohala (Eds.), *Sound Symbolism*, (pp. 325–347). Cambridge: Cambridge University Press.

Peirce, C. S. (1955). *Logic as semiotic: The theory of signs*. In J. Bucher (Ed) . *Philosophical writings of Peirce* (pp. 98–104). New York, NY: Dover.

Peng, S., Chan, M. K. M., Tseng, C., Huang, T., Lee, O. J., & Beckman, M.E. (2005). Towards a Pan-Mandarin system for prosodic transcription. In S. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 230–270). Oxford, UK: Oxford University Press.

Podesva, R. J. (2007). Phonation type as a stylistic variable: The use of falsetto in constructing a persona. *Journal of Sociolinguistics*, *11*, 478–504.

Posner, R. (1982). Iconicity in syntax: The natural order of attributes. *Rational discourse* (pp. 49–80). Amsterdam, The Netherlands: Mouton.

Rabe-Hesketh, S., & Skrondal, A. (2005). *Multilevel and longitudinal modeling using Stata*. College Station, TX: Stata Press.

Robinson, S. (2002). Constitutent order in Tenejapa Tzeltal. *International Journal of American Linguistics*, *68*, 51–80.

Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science*, n/a. DOI:10.1002/wcs.78.

Rossano, F. (2010). Questioning and responding in Italian. In T. Stivers, N. J. Enfield, & S. C. Levinson (Eds.) Question-response sequences in 10 languages [Special Issue]. *Journal of Pragmatics*, 42, 2756–2771.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking in conversation. *Language*, *50*, 696–735.

Sadock, J. (1975). *Toward a linguistic theory of speech acts*. New York, NY: Academic Press.

Sadowski, P. (2001). The sound as an echo to the sense: The iconicity of English gl-words. In O. Fischer & M. Nänny (Eds.), *The motivated sign* (pp. 69–88). Amsterdam, The Netherlands: John Benjamins.

Saffran, J. R., Werker, J. F., & Werner, L. A. (2006). The infant's auditory world: Hearing, speech, and the beginnings of language. In D. Kuhn & R. S. Siegler (Eds.), *Handbook of child psychology: Cognition, perception, and language, Vol 2*. (pp. 58–108). Hoboken, NJ: John Wiley & Sons.

Sag, I., & Liberman, M. (1975). The intonational disambiguation of indirect speech acts. In R. E. Grossman, L. James San & T. J. Vance (Eds.), *Papers from the eleventh regional meeting*, (pp. 487–498). Chicago: Chicago Linguistic Society.

Schegloff, E. A. (1984) 'On Some Questions and Ambiguities in Conversation.' In J. M. Atkinson & J. Heritage (Eds), *Structures of Social Action*, (pp. 28–52). Cambridge: Cambridge University Press.

Schegloff, E. A. (1998). Reflections on studying prosody in talk-in-interaction. *Language and Speech*, *41*, 235–263.

Schegloff, E. A. (2007). *Sequence organization in interaction: A primer in conversation analysis*. Cambridge, UK: Cambridge University Press.

Schegloff, E. A. (2002). Accounts of conduct in interaction: Interruption, overlap and turn-taking. In J. H. Turner (Ed) . *Handbook of Sociological Theory*, (pp. 287–321). New York: Plenum.

Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge, UK: Cambridge University Press.

Searle, J. R. (1975). Indirect speech acts. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics* Vol. 3: Speech acts (pp. 59–82). New York, NY: Academic Press.

Shen, X. S. (1990). *The prosody of Mandarin Chinese*. University of California Publications in Linguistics 118. Berkeley, CA: University of California Press.

Shriberg, E., Bates, R., Stolcke, A., Taylor, P., Jurafsky, D., Ries, K., & …van Ess-Dykema, C. (1998). Can prosody aid the automatic classification of dialog acts in conversational speech? *Language and Speech*, *41*, 443–492.

Sicoli, M. A. (2007). *Tono: A linguistic ethnography of tone and voice in a Zapotec region*. Ann Arbor, MI: University of Michigan dissertation.

Sicoli, M. A. (2010). Shifting voices with participant roles: Voice qualities and speech registers in Mesoamerica. *Language in Society*, *39*, 521–553.

Simpson, A. P. (2009). Phonetic differences between male and female speech. *Language and Linguistics Compass*, *3*, 621–640.

Sosa, J. M. (1999). *La Entonación Del Español: Su Estructura Fónica, Variabilidad Y Dialectología*. Madrid, Spain: Cátedra.

Stivers, T. (2010). An overview of the question-response system in American English conversation. In T. Stivers, N. J. Enfield, & S. C. Levinson (Eds.), Question-response sequences in 10 languages [Special Issue]. *Journal of Pragmatics*, 42, 2772–2781.

Stivers, T., & Enfield, N. J. (2010). A coding scheme for question-response sequences in conversation. *Journal of Pragmatics, 42*, 2620–2626.

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., … Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 10587–10592.

Stivers, T., Enfield, N. J., & Levinson, S.C. (2010). Question-response sequences in 10 languages [Special Issue]. *Journal of Pragmatics*, *42*, 2615–2619.

Tanz, C. (1971). Sound symbolism in words relating to proximity and distance. *Language and Speech*, *14*, 266–276.

Van Bezooijen, R. (1995). Sociocultural aspects of pitch differences between Japanese and Dutch Women. *Language and Speech*, *38*, 253–265.

Yoon, K. (2010). Questions and responses in Korean conversation. In T. Stivers, N. J. Enfield, & S. C. Levinson (Eds.), Question-response sequences in 10 languages [Special Issue]. *Journal of Pragmatics*, 42, 2782–2798.

Appendix. Relevant typology: Sketches of corpus languages.

	Tone/phonation	Stress	Word order	TAM	Wh-Front	Polar inversion	Final particles
‡Ākhoe Hai‖om	Tonal	Syll Troch	Free/SOV dominant	Free morph	Yes	No	Yes (rare use)
Danish	Laryngealization	Syll Troch + Contrastive	SVO dominant	Suffixes	Yes	Yes	No
Dutch	No	Syll Troch, Penult, QS	SVO dominant	Suffixes	Yes	Yes	No
English	No	Syll Troch, QS + Contrastive	SVO dominant	Suffixes	Yes	Yes	No
Italian	No	Syll Troch, Penult + Contrastive	(S)VO	Suffixes	Yes	No	No
Japanese	Pitch-accent	Mor Troch, QS	SOV dominant	Suffixes	No	No	Yes
Korean	No	Near Init Accent, QS	SOV dominant	Suffixes	No	No	Yes
Lao	Tonal	Syll Iamb, Word Final	Free/SVO dominant	Free morph	No	No	Yes
Tzeltal	Laryngealization	Syll Final	VO(S)	Suffixes	Yes	No	Yes
Yélî-Dnye	No	Syll Troch, Initial Primary	Free/SOV dominant	Clitics	Yes	No	No

Sources: Allan, Holmes & Lundskaer-Nielsen, 2000 and Heinemann, 2010 for Danish for Danish; Robinson, 2002 and Brown, 1996, 2010 for Tzeltal; Enfield, 2007, 2010 for Lao; Englert, 2010 for Dutch; Hayashi, 2010 for Japanese; Henderson, 1995 and Levinson, 2010 for Yélî-Dnye; Hoymann, 2010 for ‡Ākhoe Hai‖om; Lim, 2001 and Yoon, 2010 for Korean; Rossano, 2010 for Italian.