Final rises in very short questions

Catherine Lai

Department of Linguistics
University of Pennsylvania, Philadelphia, United States
laic@babel.ling.upenn.edu

Abstract

The meaning of very short questions like what? or do you? necessarily depend on discourse context. Even with context, meaning may be ambiguous without the help of prosody. This paper presents a corpus study investigating the relationship between question use and final rises. Firstly, rhetorical and non-rhetorical wh-questions appear separable by the presence of a final rise. Secondly, the intonation of backchannel questions is more like that of questions than backchannels. These two findings pose challenges for theories that correlate final rises with speaker uncertainty.

1. Introduction

A key factor for question interpretation and detection is the match of form and its intonation. Prototypical question types: wh, yes/no and declarative questions, have been characterized as ending in a final fall, final rise, and a higher final rise respectively [1]. However, the matching problem is much harder if we consider that there is some gradience in what we mean to be a question.

There are clearly more types of questions than the three listed above. Bolinger [2] distinguishes alternative, tag, and reclamatory questions to name a few. Moreover, interrogative forms have many other uses. For example, the broad set of questions can be used to make assertions, clarifications, acknowledge turn control, or to express agreement [3, 2]. We can consider a question to be genuine if the speaker does not know the answer [1, 4, 5]. That is, the speaker is seeking information from the hearer. However, the line between information seeking and non-information seeking questions is vague.

Even if a question seeks information, there can be ambiguity in what type of information is being sought. In such situations, prosodic factors appear to cue the interpretation a question receives and what response is required from the hearer.

This paper considers the connection between final falls/rises and the interpretation of very short questions (two words or less). In general, such short questions must be interpreted with respect to the previous discourse. However, even with previous context, their interpretation and discourse function are often extremely ambiguous when stripped of sound.

We report the results a corpus study of very short questions. Section 4 describes the corpus data and methodology for extracting the relevant pitch slope information from it. Sections 5-7 present results of the study and discusses their relevance to the issues outlined above. The next section presents some examples of ambiguous short questions in the light of theories linking intonation and question meaning discussed in Section 3.

2. Ambiguity in short questions

Wh-question are often ellipted in natural speech. Different uses of these questions often reduce to the same wh-word. For example, ellipted wh-question can be used to seek information (eg. What (did you do?) or to elicit repetitions (eg. What (did you just say)?). Here, the intonation disambiguates what the question is used for.

As mentioned previously, questions are not always used to elicit information. Utterances with yes/no question syntax are often used as backchannels in dialogue [6]. That is, they can be used to express acknowledgement or agreement, as continuers or to mark incipient speakership. Consider the dialogue in (1).

1. B: New Jersey had the stiffest gun laws
   A: oh, do they?
   B: oh yeah
   B: but you know where where are the most crimes committed?

However, the exact same utterance could be used to convey surprise, disbelief, demand confirmation, or more generally question the truth of the previous utterance. Like the ellipted wh-questions, these different uses are not really discernible without considering the prosody of the utterance. In the example above, speaker B’s response is clearly going to be affirmative. So, this sort of question is highly biased. In this case, it is not clear that if these questions are questions at all. We wish to see if how intonation can help categorize these utterances. If so, whether this categorization works with theories connecting intonation and question meaning.

3. Question intonation in discourse

Haan [1] has argued that the less an utterance is syntactically marked as a question, the more intonationally marked it will be (the functional hypothesis). This explains wh-questions falls and declarative questions rises. However, it is clear that the characterization above is not rigid. Attempts have been made to explain intonational variation in questions by looking at the issue in a broader discourse framework.

Hudson [4] proposed sincerity conditions for questions. To be genuine question, the speaker should defer to the hearers opinion on the current proposition does and wish the hearer divulge this. Similarly, Steedman [7] argues that high boundary tones imply hearer responsibility for the current proposition. Low boundary tones imply speaker responsibility. These approaches predict a high rises when the speaker is uncertain of truth of the proposition and falls when they are certain. In a wh-question like Who ate the cake? the speaker has responsibility for the proposition Somebody ate the cake. Hence, wh-questions do not require a final rise.

This proposed link between speaker confidence and question intonation predicts final falls/rises should help in determin-
ing question status and subtype. We would expect ellipted wh-questions to retain final falls and genuine yes/no questions to rise. True backchannel questions are highly biased, so we would expect them to fall. However, backchannel questions that have a final rise may convey something more than acknowledgement. This is mostly likely speaker uncertainty of the previous utterance. These predictions are tested in the following sections.

4. Data

This study used the Conversational Telephone Speech component of the the MDE RT-04 corpus (LDC2005S16). This comprises of approximately 40 hours of speech from the Switchboard-1 Release 2 corpus. The annotation provided by the MDE project (LDC2005T24) provides a range of discourse metadata. In particular, it identifies question and backchannel type turns. Switchboard also includes demographic information.

We located 315 questions turns containing two or less words. The timing data of the final word in each of these question was recorded. F0 contours were extracted from final words using praat. F0 values were then normalized as follows. Some speakers participated in more than one conversation, so the F0 data for each speaker was combined. Outliers were then removed for each individual question. These were defined as as F0 values outside the range \((q_1 - 1.5 \times (q_3 - q_1), q_3 + 1.5 \times (q_3 - q_1))\), where \(q_1, q_3\) are the first and third quartiles of combined F0 data of each individual speaker. The F0 data was further normalized to a log scale between 0-10.

Given the short nature of the utterances, it was hoped that enough information could be encoded in the overall slope of the final word to be of use. The final word pitch slopes were calculated from the normalized F0 data using the linear regression function \(\text{lm}\) from the R language. The pitch extraction and normalization process eliminated five questions from the data set. Two more question were eliminated due to lack of speaker information, leaving 308 questions.

We manually coded the remaining questions from the transcripts alone. Each question was placed into one of the following categories. Genuine yes/no question (Y) and wh-questions (W) were identified along with a number of other question types. Reclamatory (R) questions elicit repetitions. They were identified by the fact that they caused the hearer to repeat the immediate previous utterance. Confirmation (C) questions clarify the current topic of discourse. Incomplete (I) questions were used to elicit non-specific speech from the hearer (eg. Are you there? Hello?). In tag questions (T), speakers questioned their own prior statement. Backchannel questions were utterances were the speaker questioned an immediately prior statement of the hearer. Moreover, an acknowledgement only interpretation could be derived from the transcript. Suggestion (S) questions arose from speakers offering possible but indefinite options to the hearer. For example, the speaker may be finishing the hearer’s own sentence. Complementary (Cp) questions elicit the same responses as wh-questions but without the wh-word. Alternative (A) questions present question in the form of a list of alternatives.

One more utterance was removed from the data set as its transcription did not clearly match the audio recording. This left 307 questions. The distribution of these questions is presented in the following section.

5. Final word slope in questions

The proportions of utterances with positive and negative slope across question types is shown in the Figure 1. Pearson’s \(\chi^2\) test was used to see if there was any difference between positive and negative slope counts across gender, age (10 year groups) and accent. The test did not show a significant differences for the data as a whole. Similarly, no significant difference in age or sex was found when looking at specific question types using the Fisher exact test (due to the small amount of data). However, there was a significant different in comparing accent type and reclamatory questions (p-value = 0.03).

It is important to keep in mind the small amount of data available for each of these question types. We clearly cannot make any generalizations about the distribution of complementary questions from two examples. Also, we cannot draw conclusions about the alternative questions. These comprise of single components of a larger question. Since the position of the disjunct was not encoded the data is not informative.

However, we can make some general observations at this point about the other types of questions. Yes/no questions do appear to have final rises and wh-questions have final falls, the majority of the time. The results show that confirmation, incomplete, and suggestion questions tend to end in a rise. These question are really
declarative questions so this result is in line with Haan’s functional hypothesis. The ambivalence associated with these types of questions also agrees with the association of rising and speaker uncertainty.

However, it appears that wh-reclamatories are distinguished from genuine wh-questions by their final rise. The large number of final rises in backchannel questions also requires further attention.

6. Wh-questions and reclamatory questions

Figure 2 shows a striking difference in the distribution of slope data in Wh (W) and reclamatory (R) questions. Approximately 86% of the R data had positive slope, compared to 26% for the W data. Reclamatory wh-questions (RW) were further identified. We will refer to reclamatory questions without wh-words by R-W.

Pearson’s $\chi^2$ test was applied compare W and R counts of samples with positive and negative slope. The differences were enough to reject the null hypothesis that the counts were independent of row and column groupings (p-value < 0.0001). That is, the type of question makes a significant difference to the final direction of pitch contour. This difference is true of R data and the RW subset (p-value < 0.001). However, this does not explain why RW and W questions differ in intonation in this way.

The large number of rises agrees with the description of R questions in [2]. The R-W questions are easily explained. The data appear to be set phrases such as Pardon me? Sorry? Excuse me?. However, they can have an apologetic meaning without the final rise. Moreover, they appear to be genuine question with respect to speaker uncertainty. In this context a final rise is expected. As noted above, accent difference was found to be significant. In fact, "SOUTH" accented speakers contributed three negative sloped R-W questions and no positive ones. Given data sparsity, this warrants further study. However, we will focus on RW questions and so not address this here.

RW questions are not so easy to explain. They are certainly lexically marked as questions. In many ways RW questions are just W questions requiring a very specific response. We can consider reclamatory what? is elipted what did you say?. This can also be interpreted as a non-reclamatory wh-question with appropriate intonation. Intuitively, the speaker should be equally confident that the hearer did say something on both interpretation. Confidence in the hearer’s ability to answer the question should also be equivalent. Differentiating the intonation of these questions via speaker confidence differences appears to be a very hard task.

An easier explanation may be that the intonation pattern of RW occurs simply to differentiate it from W questions. The form of this intonation difference may have arisen due to factors independent from speaker confidence. For example, the rise may be motivated by the fact that reclamatories often occur as responses to turn interruptions. So, for example, some element of surprise may be involved.

7. Backchannel Questions

Most of the utterances in this category are syntactically marked as questions with subj-aux inversion. The slope data showed that most of these questions have final rises. In fact, the $\chi^2$ test over positive and negative slope counts did not distinguish these backchannel questions from unbiased yes/no questions (p-value = 0.1949). This differs from in [5] who found final falls and rises to be evenly distributed in backchannel questions. This suggest that these questions may actually be questions, or at least something other than pure backchannels.

These questions elicited responses in 62% of cases (55/89) with (first quartile, third quartile) = (0.020, 1.255). Questions that did not get a response occurred in a narrower range: (first quartile, third quartile) = (0.000, 0.548). However, the $\chi^2$ test did not distinguish between backchannel questions that elicited a response and those that did not.

Backchannel questions were also compared to backchannels identified by the MDE annotation. Besides the yes/no form questions, really? and yeah? questions were also categorized as backchannel questions. These words appeared labelled annotated as pure backchannels (247, 6212 counts respectively) in the corpus far more frequently than as questions (17, 1 respectively). Slope data for these backchannels was obtained as described in Section 4.

The majority (76%) of backchannel yeah’s had negative slope. On the other hand, the backchannel really’s had negative slope was found 39% of the time. Negative slope was found in question-really in 11% of cases. Our backchannel questions as a whole were found to be significantly different from backchannel really (p-value = 0.017) and yeah (p-value < 0.0001). The difference between backchannel really and question really also appears to be significant (p-value = 0.04). Boxplots of backchannel and question really are shown in Figure 3.

Thus, our backchannel question category appears lie somewhere between genuine backchannels and genuine questions. In
fact, these questions were used to convey a range of signals from acknowledgment to disbelief and shock, depending on prosody. This difference cannot really be captured in the current final rise/fall approximation without a clearer categorization of the data by use (eg. acknowledgement, incredulity). This requires a more controlled perception study. However, we can make some qualitative observations about the different instantiation of the class of questions.

To start we can try to infer the use of *really* from its effect in dialogue. In (2) *really* seems to be expressing surprise and disbelief. We can see from the text that speaker B justifies his question in the last line.

2. B : You like Lubbock better than Dallas  
   A : Yeah  
   B : Why?  
   A : Uh, because people are so much nicer  
   B : Really?  
   A : Yes  
   B : Well people are nice here in Dallas

In (3), the use of *really* appears to be a backchannel. Even though it apparently required a response (*Yeah*), there did not appear to be any need for either speaker to justify their statements any further.

3. B: I kind of enjoyed that boat  
   I looked at today  
   A: Really?  
   B: It wasn’t  
   B: Yeah  
   A: Did it have a cabin?

The pitch contours for these instances of *really* can be found in Figure 4. According to the theory outlined above we would expect backchannel *really* to remain relatively flat. This is indeed the case for dialogue (2). Surprise, on the other hand, probably requires a more exclamative intonation. For example, a steep rise. This is what is found in Dialogue (2). In fact, ‘surprised’ *really* can also occur with steep falls, as it did in dialogue (4).

4. A : Um I major in psychology in family studies  
   B : Really?  
   A : Yeah  
   A : I like it a lot

So, the steepness of the final pitch movement is probably a good indicator of deviation from acknowledgment status. However, this must be considered alongside a number of other prosodic factors such as intensity and timing effects. These sorts of factors can also be applied to Wh-questions to get disbelief. However, in order determine this scale of acknowledgement to shock, we need to agree whether or not a question actually expresses this.

### 8. Conclusion

Prosody has an important role to play in the interpretation of very short questions. The shortness of the questions can lead to ambiguity that is not always resolvable from the previous context.

The final pitch movement of many of the question types examined here followed concurred with theories linking speaker uncertainty to final rises. However, the idea of speaker uncertainty does not appear sharp enough to explain why declaratory and non-declaratory wh-questions differ in this respect. We can say, however, that these two types of questions are interpreted very differently depending on the presence of final rise.

### 9. References


