Tonal Constituents and Meanings of *Yes-No* Questions in American English

Nancy Hedberg, Juan M. Sosa & Lorna Fadden

Department of Linguistics
Simon Fraser University, Burnaby, B.C., Canada
{hedberg; sosa; fadden}@sfu.ca

Abstract

We analyzed the different meanings associated with the tonal contours of 104 positive *yes-no* questions from the CallHome Corpus of American English. We take into consideration such broad constituents as the head, nucleus and tail of intonational phrases, as well as ToBI sequences of pitch accents, phrase accents and boundary tones. The meaning of a question as unmarked or marked in a variety of ways is shown to depend upon the intonational contours associated with these broad constituents, and even with the contour associated with the question as a whole.

1. Introduction

We first outline our method of reliably arriving at a ToBI notation for American English *yes-no* questions, and its correlates in terms of higher constituents such as head, nucleus, and tail. We then discuss and exemplify the meanings we have found to correlate with these categories. Finally we consider whether it is the intonational pattern stretching across the entire question that is most relevant for characterizing basic aspects of the meaning of the question, such as whether it is marked or unmarked in particular ways.

2. Methods

2.1. Reliability Study

We selected 48 positive *yes-no* questions for the training phase of the study from the CallHome corpus [1], [2], a corpus of 30-minute recorded telephone calls between people who know each other. The purpose of the training phase was to learn to reliably code ToBI categories [3]. Together we listened to each example and examined pitch tracks and came to a consensus on the appropriate tonal notation for each. We took notes on how decisions were made so that we could refer back to them later.

The second phase involved each of the three members' independent coding of 56 positive *yes-no* questions. We calculated transcriber-pair-word agreement by comparing the labels of each transcriber against the labels of every other transcriber. If all three pairs agreed, we gave the word a score of 1. If two out of three pairs agreed we gave it a score of .33. If none agreed, we gave it a score of 0. Agreement was calculated three different ways. First, we calculated whether we agreed on pitch accents and terminal contours, and we agreed in 72% of the cases. Second, we calculated the presence and type of pitch accent, and agreement was 75.7%. Third, we calculated the agreement of the presence or absence of pitch accent, and agreement was 85.3%.

Our results compare about equally with previous studies reported in the literature. Using a variety of read and spontaneous speech, [4] report 68% agreement on particular pitch accent or no pitch accent, and 80.6% agreement on presence or absence of pitch accent. In a study on German ToBI, [5] report 71% agreement on particular pitch accent or no pitch accent, and 87% on presence or absence of pitch accent. [6] calculated agreement for one professional male and one professional female speaker. They reached 71% and 72% agreement for the female and male speaker, respectively, for particular pitch accent or no pitch accent. They report 92% and 91% agreement for the female and male speaker for presence or absence of pitch accent. In a study using ToBI-Lite on samples from the Switchboard Corpus [7], [8] report agreement of 85.6% for particular pitch accent or no pitch accent, and 86.6% for presence or absence of pitch accent. In [9], reliability on declarative questions in the Santa Barbara Corpus of American English [10], [11] was 47% agreement for particular pitch accents using full ToBI, and 87% using ToBI-Lite.

Once pairwise agreement had been calculated, we worked collectively to arrive at a consensus for the final coding of the test questions. After that, we recoded the training data for inclusion in our study. By that time, we felt confident about the reliability and validity of our coding categories.

2.2. Classification of ToBI Patterns into Constituent Types

We found it productive to classify our questions according to broad constituents (head-nucleus-tail), see [12], because of the different patterns and meanings that we found to be associated with the different types of final and non-final contours. For prior literature on this strategy, we can point to [13] and [14], which discuss the relationship between ToBI and such broader categories.

Ladd [14] (p. 82) discusses the mapping of ToBI final tunes onto British-style nuclei, and also discusses viewing ToBI pre-nuclear pitch accent sequences as constituents corresponding to heads (pp. 210-211). We classified all the pre-nuclear pitch accent sequences in our question data into heads of different shapes. To do this we had to generalize across some instances of ToBI categories, e.g., an initial L+H* or L*+H or H* followed by a !H* or sequence of !H*'s was classified as a falling head. In a few instances, we also identified a pre-head (%H in ToBI notation).

3. Results

In this section we present the results of our mapping of the ToBI sequences of high and low pitch accents, phrase accents and boundary tones onto nuclei, tails and heads. Our goal in this categorization was to identify stable meanings associated

with particular tonal patterns across the three primary broad constituents.

3.1. Nuclei

Table 1 presents the distribution of nuclei in our data, classified according to both British tradition and ToBI categories, some of which we have defined ourselves.

Table 1: *Distribution of Nuclei*

Nucleus	ToBI Category	Number
Low rise	L*HH%	63
High rise	H*HH%	18
Low low rise	L*LH%	5
Extra high low rise	L*+HHH%	6
Level	H*HL%	4
Low fall	L*LL%	4
High fall	H*LL%	3
Extra high fall	L+H*LL%	1
Total:		104

For terminal contours, we found that the low rise is by far the most common nucleus, occurring in 63 cases or 60.5% of the time. We concluded from this frequency data and from considering the meaning of the questions so marked that the low rise is the unmarked nucleus of *yes-no* questions in American English. Variants of the low rise were low-low rise, where the rise was delayed and did not extend very high and a category that we labeled extra-high low rise, where the rise extended to an extra-high height. Together these three categories account for 71% of the data.

The next most frequent nucleus is the high rise, which occurred 18 times, or 17% of the time. Identification of a specific meaning associated with the high rise was not clear cut. One pattern that we did find (5 times) was that high rises occurred when a word that would be normally be an unaccented part of a tail due to being anaphoric or deictic is exceptionally accented, as in (1).

(1) Is **anybody coming** before **that**?
L* L* H*HH%

High rises also occurred (5 times) when the speaker is reminding the hearer of the content of the question, so that the propositional information is familiar to the hearer but not activated in the terms of [15], rather than 'out of the blue' or related only to the current conversation. An example is shown in (2), where it can be assumed that the speaker and hearer already share the information that Adriana had been looking for a job.

(2) Did **Adriana get** a **job**?
H* !H* H*HH%

Finally they occurred (3 times) when a complex pitch-accent occurred contrastively immediately prior in the head, perhaps motivating a high starting point for the nucleus due to the height of the immediately preceding pitch accent in the head:

(3) Is that **really** such an **awful job**?
L*+H L*+H H*HH%

Climbing heads (step-pattern rising pre-nuclear contours), such as the one in (3), and rising heads are associated with high-rise nuclei in that the nucleus continues the overall rising pattern begun by the head, as will be discussed in 4.2 below.

We found that falls (high falls, low falls, and extra-high falls) occurred 8 times (7.6%), and were associated with non-genuine questions, as is consistent with the conclusions of [16], which analyzed the discourse functions of 249 falling *yes-no* questions out of 3,789 *yes-no* questions in the Switchboard Corpus, as coded in [17]. First, it is noteworthy that the proportion of their falling compared to rising questions was quite comparable to our much smaller sample (for them, 6.6%). Second, the functions were comparable. In speech act terms, we found that falling nuclei expressed requests for action (1 time)–(4), as opposed to requests for information, announcements (2 times)–(5), and suggestions (1 time). Other falling nucleus questions sounded like aggressive demands for information (3 times), as in (6).

(4) Can we **talk** about the **job** things now?
H* L*LL%

(5) Did **I** **tell** you that **I** have a new **job**?
L+H* !H* !H* L*LL%

(6) **Do** you **agree**?
H* H*LL%

In our data, we found no requests for action with rising or level nucleus, and we found one announcement produced with a level nucleus. Two other level nuclei were meta-conversational, such as the one shown in (7).

(7) Did I **tell** you **that**?
H* H*HL%

As suggested in [18], level nuclear contours in *yes-no* questions thus seem to pattern with falling nuclei in having marked meanings.

3.2. Tails

Table 2 presents the distribution of tails (unaccented words after the nucleus) in our data. All but two occur after rising nuclei and continue the rise to end of the question. Rising tails are not often mentioned in the literature, but [14] cites [19] as maintaining that "the shape of the tail is largely or entirely dictated by the choice of nuclear tone" (p. 209).

Table 2: *Distribution of Tails*

Tails	Number
Function words	10
Nonactivated, semantically light content words.	4
Pronouns and deictic words	11
Other activated information	7
Total:	32

As the table shows, the types of words that occur as components of tails in questions are the same types of words

that have been reported in the literature to occur as tails in statements. A clear summary of the factors that go into determining accentability (and its converse, ability to be unaccented) is presented in [20]. Part of speech (i.e. function words); 'semantic weight' (e.g. 'empty' content words like *thing* or *stuff*), and givenness are relevant factors.

Seven examples occurred in our data with the unaccented final, function word *yet*, as in (8):

(8) Did **she have** her **baby** yet?
H* !H* L*HH%

Eleven examples occurred with a post-nuclear anaphoric or deictic pronoun or adverb, such as that in (9):

(9) But does he **always** want to **do** it?
L*+H L*HH%

Example (10) shows activated, contentful information occurring in a tail. It had just been mentioned that a certain person would probably not be able to come. The question asks contrastively if *they* are going to still come, with the 'X is/is not going to still come' information activated, and therefore realized by means of a long, unaccented tail in the question, and a narrow focus nuclear accent on *they*.

(10) A: So th- anyway they just found something and they're going to start treating it like this week so now they're not sure if he can come but mom thought it was so sweet that he wanted to come.
B: Yeah. But are **they** gonna still come?
H*HH%

In (11), *other* is contrastive in that some of the children had just been discussed, and thus the head noun *children* is deaccented.

(11) And **did** they **mention** their **other** children?
H* H*L- L*HH%

We conclude that the information structure of tails in questions corresponds to that of statements.

3.3. Heads

Table 3 presents the distribution of heads, showing that the most frequent, unmarked pattern (68%) is the no head, single H* head, or falling head (when followed by the low rise).

Table 3: *Distribution of Heads*

Pre-Nuclear Contours	Number
No head	18
H* head	23
Falling head	30
Low head	14
High level head	6
Rising head	11
Climbing head	2
Total:	104

The examples in (12) – (14) illustrate the unmarked patterns, and a pitch track for (14) is shown in Figure 1.

(12) Can you **swim**?
L*HH%

(13) Is **Matt** through **school**?
H* L*HH%

(14) **Have** you **seen** a **doctor** yet?
H* !H* L*HH%

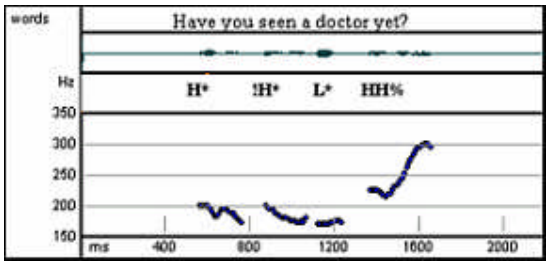

Figure 1: *Falling head with a low-rise nucleus*

The examples in (15) and (16) show low heads before low rise nuclei.

(15) A: I used to nurse her four times a day
B: Does she take, does she **take** a **bottle**?
L* L*HH%

(16) **S'are** you **happy**?
L* L*HH%

We found that questions with low heads would not be used to ask 'out of the blue' questions. Instead there is an expectation in the conversation that the question is a natural continuation of previous discourse (note the *so* in (16) linking the question to preceding context). In only two out of 14 cases does the question introduce a new discourse topic. This may relate to [21]'s claim that L* accented words are salient but do not "form part of what S is predicating in the utterance" (p.291).

The following examples show high level heads preceding different marked nuclear tune types. In our data, none is followed by the unmarked low-rise nucleus.

(17) Had you **been there before**?
H* H* L+H*LL%

(18) Have you **seen** them at **all lately**?
H* H* H*HH%

(19) Is **she close** by **you**?
H* H* H*HL%

4. Discussion

4.1. Identity of pre-nuclear accents.

It has been claimed [14] that pre-nuclear pitch accents have to be identical, with some exceptions. We found that this is in general true for unmarked questions. The primary exception was with complex pitch accents that highlight words for contrast or emphasis. Examples are shown in (20)-(21). The contrastive, 'scalar' semantic behavior of L*+H and L+H* fits [21]'s claims about these pitch accents.

(20) Did the **outdoor pictures come** out **fuzzy**?
L+H* !H* !H* L*HH%

(21) **Did** she **ever show** up?
H* L*+H L*HH%

4.2. Restricted combinations of head and nucleus.

We found that eight out of 13 (61.5%) high-rise nuclei occurred with rising or climbing heads, showing a strong correlation between these two constituent shapes. Examples are shown in (22)-(23), and a pitch track for (23) is shown in Figure 2.

(22) **Are** they **selling well**?
L* H* H*HH%

(23) **Are** you gonna **be** in New **York**?
L*+H H* H*HH%

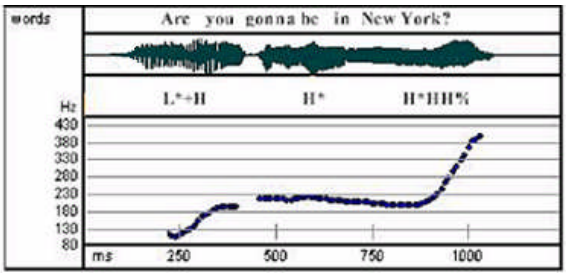

Figure 2: Rising head with high-rise nucleus

As discussed above, our results show that the unmarked question intonation is no head, H* head or falling head followed by a low-rise nucleus. The choice between these three patterns depends upon the amount of pre-nuclear material available to receive pitch accents, and this doesn't affect the meaning of the question as unmarked.

The results discussed in this section show that the choice of head, given a particular nucleus, is not completely free, and thus that overall contour shape must be taken into account in characterizing the intonation of *yes-no* questions.

5. Conclusion

We conclude that it is illuminating to group pitch accent and edge tone sequences into the larger constituents of head, nucleus and tail, enabling overall contour shape to be examined. We next intend to pursue perception experiments to further test our conclusions about the markedness and meanings of questions of different shapes.

The research reported in this paper was supported by a SSHRC Small Grant and an SFU Discovery Park Grant.

6. References

[1] Carnavan, A.; Graff, D.; Zippener, G. 1997. CALLHOME American English Speech. University of Pennsylvania: Linguistic Data Consortium.

[2] Kingsbury, P.; Strassel, S.; McLemore, C.; McIntyre, R., 1997. CALLHOME American English Transcripts. University of Pennsylvania: Linguistic Data Consortium.

[3] Beckman, M.; Ayers Elam, G. 1997. Guidelines for ToBI labeling, Version 3, Ohio State University.

[4] Pitrelli, J.F.; Beckman, M. E.; Hirschberg, J., 1994. Evaluation of prosodic transcription labeling reliability in the ToBI framework. *Proceedings of the International Conference on Spoken Language Processing*. Yokohama, Japan, 123-126.

[5] Grice, M.; Reyelt, R.; Benzmüller, R.; Mayer, J., and Batiner, A. 1996. Consistency in transcription and labeling of German intonation with GToBI. *Proceedings of the International Conference on Spoken Language Processing*. Philadelphia, USA, 1716-1719.

[6] Syrdal A.; McGory, A. 2000. Inter-transcriber Reliability of ToBI prosodic labeling. *Proceedings of the International Conference on Spoken Language Processing*. Beijing, China, 235-238.

[7] Godfrey, J. J.; Holliman, E.C.; & McDaniel, J., 1992. SWITCHBOARD: Telephone speech corpus for research and development. *Proceedings of the International Conference on Audio, Speech and Signal Processing*, 517-520.

[8] Yoon, T.; Chavarria, S.; Cole, J.; Hasegawa-Johnson, M. 2004. Intertranscriber reliability of prosodic labeling on telephone conversation using ToBI. *ICSA International Conference on Spoken Language Processing*, Interspeech 2004, Jeju, Korea, Oct. 2004, 2729-2732.

[9] Safarova, M.; Swerts, M. 2004. On recognition of declarative questions in English. *Proceedings of Speech Prosody 2004*. 313-316.

[10] Du Bois, J. W.; Chafe, W. L.; Meyer, C.; Thompson, S. A. 2000. Santa Barbara Corpus of Spoken American English. Part I. University of Pennsylvania: Linguistic Data Consortium.

[11] Du Bois, J. W.; Chafe, W. L.; Meyer, C.; Thompson, S.A.; Martey, N. 2003. Santa Barbara Corpus of Spoken American English. Part II. University of Pennsylvania: Linguistic Data Consortium.

[12] O'Connor, J. D.; Arnold, G. F. 1973. *Intonation of Colloquial English* (2nd edition).

[13] Ladd, D. R. 1986. Intonational phrasing: the case for recursive prosodic structure. *Phonology Yearbook* 3, 311-40.

[14] Ladd, D. R. 1996. *Intonational Phonology*. Cambridge University Press.

[15] Gundel, J. K.; Hedberg N; Zacharski, R. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69, 274-307.

[16] Banuazizi, A.; Creswell, C. 1999. Is that a real question?: Final rises, final falls and discourse function in question intonation. *Proceedings of the Chicago Linguistic Society*. *Papers from the Main Session*, 1-13.

[17] Jurafsky, D.; Shriberg, E.; Biasca, D. 1997. *Switchboard SWBD-DAMSL Shallow-Discourse-Function Annotation Coder's Manual. Draft 13.* University of Colorado.

[18] Hedberg, N.; Sosa, J. M.; Fadden, L. 2004. Meanings and configurations of questions in English. *Proceedings of Speech Prosody 2004*, Nara, Japan, 309-312.

[19] Palmer, H. 1922. *English intonation, with systematic exercises.* Cambridge: Heffer.

[20] Zacharski, R.; Monaghan, A. I. C; Ladd, D. R.; Delin, J. 1995. *BRIDGE: Basic Research on Intonation for Dialogue Generation.* University of Edinburgh.

[21] Pierrehumbert, J.; Hirschberg, J. 1990. The meaning of intonation contours in the interpretation of discourse. In Cohen, P.R.; Morgan, J.; Pollack, M. E. (eds.) *Intentions in Communication*. MIT Press, 271-311.