F0 cues for the discourse functions of "*hã*" in Hindi

Kalika Bali

[1] Microsoft Research Labs India, Bangalore, India
kalikab@microsoft.com

Abstract

Affirmative particles are often employed in conversational speech to convey more than their literal semantic meaning. The discourse information conveyed by such particles can have consequences in both Speech Understanding and Speech Production for a Spoken Dialogue System. This paper analyses the different discourse functions of the affirmative particle *hã* ("yes") in Hindi and in explores the role of fundamental frequency (f0) as a cue to disambiguating these functions.

Index Terms: affirmative particles, discourse, dialogue, f0, Hindi

1. Introduction

Affirmative particles are often employed in conversational speech to convey more than their literal semantic meaning. Earlier work in English [1, 2, 3] has shown that turn-taking in a dialogue is often cued by affirmative particles such as *okay, alright, right, yeah,* etc. The linguistic context as well as the acoustic-prosodic properties of these particles aid listeners to interpret whether the speaker is continuing, ending or starting an utterance. This helps the listener to provide an appropriate response by either taking up his turn (in the case of turn-yielding cues), or providing feedback to indicate his/her participation in the conversation (in the case of turn-continuing cues). In telephonic conversation, the role of acoustic and prosodic features like fundamental frequency modulations, and pause and its duration, as well as contextual information assumes more importance in the absence of any other facial or visual clues.

As spoken dialogue systems (SDS) focus more and more on naturalness and user-initiated dialogues, the role of discourse markers becomes crucial in completing a task. The right interpretation by the system of such a particle in the user's speech can avoid interruptions, and misunderstandings [4,5]. On the other hand, the use of appropriate discourse markers by the system can provide a better more natural experience for the user.

In India, with 18 official languages, and low literacy rates, the use of SDS for information access and dissemination, as well as transactional purposes like mobile banking, can help overcome a major barrier for a large percentage of the population. While there are some deployments of SDS in local languages in the Indian sub-continent [6,7,8], these still remain in the constrained domain of system-directed dialogues. There has been some anecdotal evidence that a more natural speech interface may lead to a better human-computer interaction [9] and a better understanding of discourse markers can only help improve the quality of such a system.

In this paper, we explore the role of fundamental frequency (f0) in identifying the discourse functions of the Hindi. Following previous work in English [2], we differentiate between the literal or "sentential" use of *hã* as an affirmative particle meaning "yes" , and its different discourse functions especially as a backchannel and turn-taking cue.

In the next section we deal briefly with Hindi language and discourse markers. Section 3 describes the data used and annotation methodology. The results are discussed in section 4 and we conclude with future directions in section 5.

2. Discourse markers in Hindi

Hindi is the official language of India and is spoken by 337 million Indians (1991 census), mainly originating from the North Indian states of Bihar, Chhattisgarh, Delhi, Haryana, Himachal Pradesh, Jharkhand, Madhya Pradesh, Rajasthan, Uttar Pradesh, and Uttarakhand. The standard Hindi dialect called Hindustani or *Khari Boli* is the mother tongue of about 180 millions of Indians as per 1991 census. It is in fact the lingua franca of India and due to the high popularity of Hindi cinema; it serves as a link between the numerous language groups in India.

There has been very little work done on the discourse markers in spoken Hindi. [10] delineates the clitics *hi*, *bhi* and *to* as the main discourse markers in Hindi and proceeds to show how these topic-focus markers carry clause-level information that helps identify the discourse at the clause-level. The focus on a purely grammatical function of discourse markers is the stated objective of the study. According to [11], a number of discourse markers exist in Hindi which carry no grammatical meaning but provide "other" information about the utterance. These discourse markers may mark attitude, topic switch, turn-taking, repair, intimacy, hedging or hesitation, among other things. [11] lists more than 20 such particles that are used in Hindi to express agreement, signal sarcasm, challenge a statement etc. She notes that *hã* ("yes") along with *bIkul* ("absolutely"), *əc^h:a* ("good" used as English "okay"), *thik* ("right") and *ji* (honorific marker, that can be used on its own as well as with other particles) is used in discourse to mark "agreement".

None of the above studies looks into the use of such discourse markers in Spoken Language, especially dialogues where these discourse markers can assume a more varied role and importance. In section 3.3 we will illustrate the different functions of *hã* as found in natural telephonic conversation, where only one of them is that of agreement.

3. Data and Annotation

3.1. Data

The data presented in this paper consists of recordings of telephonic conversations in colloquial Hindi on both landline as well as mobile phone. The full databsase consists of over 400 conversations by more than 900 speakers, with equal distribution of gender as well as coverage of a number of Hindi dialects including Hindustani, Awadhi etc. However, the current study presents initial results for only 12 conversations from 24 male speakers (12 pairs) of the Hindustani dialect of

Hindi on mobile phone. The Hindustani dialect is primarily spoken in regions in and around Delhi as well as parts of Western Uttar Pradesh.

The conversations were spontaneous and not scripted. The subjects were asked to call friends or relatives and conduct normal conversations on topics of their choice. The topics of conversation in these recordings covered a wide range including examinations taken, upcoming marriages, welfare of families, weather, trips, visits home, etc.

All recordings are channel separated with the caller on the in-line and the receiver on the outline. The recordings were made at 8 KHz sampling rate, and 8-bit μ-law. Each conversation lasts between 3-4 minutes and a total of 41 minutes of speech data was analysed.

3.2. Annotation

A total of 1134 tokens were annotated by a trained linguist using Praat. The annotator was asked to mark the discourse functions according to the labels discussed in the next section on the first tier. The second tier was used to mark the pitch. Only the f0 movement on the word *hã* was considered, and marked as FLAT for no change, FALL for a sharp fall or valley, RISE for a sharp rise or peak, RISE-FALL for a rise followed by a short fall, FALL-RISE for a fall followed by a short rise. All the annotated data was rechecked by another linguist and all discrepancies were also marked as "Others". In the future, it is planned that the data would be simultaneously annotated by two annotators to better capture inter-annotator agreement.

Discourse Labels	Discourse Function	Description of the Function
A	Agreement/ Acknowledgement	Denotes agreement with the other interlocutor
B	Backchannel	Feedback to denote the presence of the interlocutor in the conversation
Y	Yes	The literal meaning of "Yes" as an answer to a question.
CB	Cue-Beginning	Denotes the start of a new topic
CE	Cue-End	Denotes the end of current topic
PB	Pivot-Beginning	Denotes agreement and start of new topic (A + CB)
PE	Pivot-End	Denotes agreement and end of current topic (A+CE)
I	Interjection	Denotes interruption or request for repeat
O	Others	Undecided

Table 1. *Discourse function labels for hã*

3.3. Discourse Labels of *hã*

We modified the classification in [2] for Hindi and Table 1 lists the labels used for annotating the different discourse functions of *hã* and their description. A total of nine labels were used as follows:

Agreement/Acknowledgement (A) indicates the only discourse function of *hã* discussed in the literature. Here *hã* is used by Speaker 2 to agree with Speaker 1's proposal.

Speaker 1: उसे कानपुर में चालू होने देर में है

"(It) will take sometime to start in Kanpur"

Speaker 2: **हाँ** कानपुर में देख लो

"*hã,* check it out in Kanpur"

Backchannel (B) use of *hã* provides feedback to Speaker 1 that Speaker 2 is still online and is attentive to what Speaker 1 is saying.

Speaker 1: हम तो दिखाए डॉक्टर को भी # मैडिकल स्टोर…

"I showed it to doctor # the medical store…"

Speaker 2: **हाँ**

"*hã"*

Speaker 1: उसने यही कहा है कि इनफ़ैक्शन है

"He (emph) said that it is an infection"

Speaker 2: **हाँ**

"*hã"*

Speaker 1: और कुछ नहीं

"nothing else"

Speaker 2: **हाँ**

"*hã"*

Literal Meaning "yes" (Y) of *hã* is used to answer the Yes/No question by Speaker 1

Speaker 1: हैलो # छोटू सब ठीक हैं#

"Hello# (name) is everyone okay?"

Speaker 2: हैलो # **हाँ** ठीक है सब #

"Hello# Yes, everyone is okay"

Cue-Beginning *hã* is used by Speaker 1 to introduce a completely new topic in the conversation

Speaker 1: **हाँ**, वे पापा अपना गन्ना-वन्ना सब लगा रहे है

"*hã,* father is planting sugarcane (etc)"

Speaker 2: ओ अच्छा

"Oh, okay"

Cue-End *hã* is used to end the topic in this example.

Speaker 1: **हाँ** सौ दो सौ क्विंटल बचा है

"*hã,* 100-200 quintal is left,*"*

Pivot-Beginning marks not only the start of a new topic (in this case by Speaker 2) but *hã* is also used to acknowledge the request of Speaker 1.

Speaker 1: जीजाजी को कहे हमको फ़ोन कर दें

"Ask brother-in-law to call me"

Speaker 2: हाँ आप आना ना इधर

"*hã*, you come over (sometime)"

Pivot-End *hã* marks not only the end of the topic by Speaker 2 but also to agree with Speaker 1's previous statement

Speaker 1: तो फिर आप लखनऊ जा रहे है

"So, you are going to Lucknow"

Speaker 1: **हाँ**, जा रहे है हम

"*hã* I am going,"

Interjection *hã* are mainly employed to request repeats in case of misunderstanding or disbelief or surprise. In this case, the second speaker did not understand Speaker 1's first utterance.

Speaker 1: लखनऊ…
"Lucknow…"
Speaker 2: हाँ
"*hã* (?)"
Speaker 1: लखनऊ में चालू है
"(It) has started in Lucknow"

Other

These include instances of *hã* where the annotator was not sure of the discourse function or where there was a disagreement between the annotator and the linguist checking the annotations.

4. Results

Table 2 summarizes the total number of *hã* tokens by their discourse function. Of the 1134 tokens, the highest (22%) were those functioning as regular affirmative particle "yes". These were followed by Pivot Beginning (21.96%), Backchannel (14.99), Agreement (12.69%), and Cue-Beginning (11.9%). A small percentage of the tokens were Interjections (5.9%) and around 10% of the tokens were undecided (Others). What was surprising was that not one token was marked as Cue End or Pivot End. This might be interpreted to mean that *hã* is less used as a marker of utterance ends than for the start of utterances.

Discourse Function	Total
A	144
B	170
Y	258
CB	135
CE	0
PB	249
PE	0
I	67
O	111

Table 2. *Discourse function labels for hã*

The high number of Y is expected as the conversations did include a number of direct Y/N questions. What is interesting to note is that the number though individually the largest is very small when compared to that of all the discourse functions put together. The high number of Pivot Beginning is also not surprising because in telephone conversation you start a topic by first agreeing to or acknowledging the previous statement. The actual numbers for Cue-Beginning and Agreement are comparable and half those of PB.

Backchannel has the third highest score and again this is to be expected as in a non face-to-face communication, the need for constant feedback is required for the interaction to be successful. These numbers might have been considerably higher if other markers such as "hmm", *əc*^{h}*:a* ("okay"), *thik* ("right") and *ji* (hon. marker) were included.

Interjections in this data were mainly request for repetitions because of mishearing. Some view interjections as backchannel as well (especially interruptions) however, in this case not only were they functionally distinct but prosodically these tokens were like a wh-question.

If we consider the distribution of f0 contours across discourse functions of *hã* given in Table 3, this becomes clear. While backchannel is mostly characterized by a flat f0, interjections are always associated with a sharp rising f0.

Discourse Labels	Flat	Fall	Rise	Rise-Fall	Fall-Rise
A	24	20	100	0	0
B	125	23		22	0
Y	0	0	130	128	0
CB	12	0	48	19	56
PB	34	0	137	78	0
I	0	0	67	0	0
O	29	71	11	0	0

Table 3. *Discourse function labels for hã*

The results in Table 3 distinctly support the hypothesis that f0 patterns may play a big role in cueing different discourse functions of *hã*. In addition to B and I mentioned above, Figure 1 shows that Agreement, and PB are mostly cued by rising f0. Cue-Beginnings are the only ones to employ a fall-rise f0 but even though their results seem to be more spread across, they still show a preference for rising f0. Y seems evenly split between rise and rise-fall. One possible explanation for this might be that it is usually a single word utterance and there could be some end-of-utterance effect. However, it is possible that in case of the literal meaning, the linguistic context might be more decisive than the pitch cue alone. This merits deeper investigation.

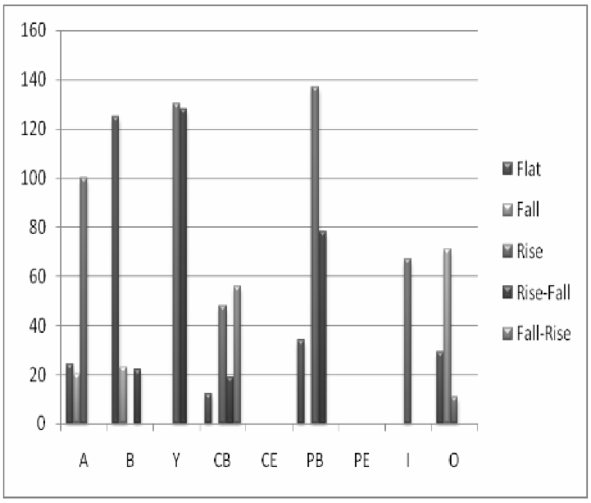

Figure 1: *Number of occurrence of different discourse functions of hã split by all f0 contour*

5. Discussion and Conclusions

The data discussed above underlines the importance of *hã* as a discourse marker in spoken Hindi. The results indicate that while *hã* in its literal linguistic form is more frequent than any single discourse function, the combined frequency of *hã* in all its discourse forms is significantly larger. Another interesting aspect of the data is the absence of *hã* to mark the end of an utterance. While it would be wrong to say that *hã* is never used for ending utterances on the basis of such a small sample (and at least a couple of examples were found elsewhere) it would be safe to assume that this is not one of the primary functions of *hã*.

Based on f0 alone, it might be difficult to disambiguate PB and Agreement as both are marked by sharp rises, and here again it would require investigating other acoustic-prosodic features to determine if at all these can be differentiated based on acoustic features alone.

While acknowledging these discrepancies, the current data nevertheless points to f0 as a possible cue to the different discourse functions of *hã*. As it is used as "yes" in its regular linguistic function it is extremely important that a Spoken Dialogue System interprets it correctly to avoid confusion and interruptions. This poses further problems as ASR accuracies for a small word like *hã* are notoriously low. Pitch may be employed more effectively in such cases.

Though initial results presented do indicate that f0 is a good indicator of discourse functions this needs to be validated by more data across dialects and gender as well as perceptual studies using both natural and synthetic speech.

As more annotated data becomes available, future work plans to focus on an in-depth study exploring the relative importance of both lexical and acoustic-prosodic features of *hã* perception and automatic classification of spoken Hindi.

6. Acknowledgements

The author would like to thank Mr. Avijit Chakraborty for data annotation as well as initial discussions on the classification scheme.

7. References

Gravano, A., Benus, S, Hirschberg, J, Mitchell, S, Voysha, I. , "Classification of discourse functions of affirmative words in spoken dialogue", Proceedings of Interspeech 2007, 1613-1616. Antwerp, Belgium, August 2007.

Gravano, A, Benus, S., Chavez, H, Hirschberg, J., Wilcox, L., One the role of context and prosody in the interpretation of *okay*", Proceedings of ACL 2007, 800-807, Prague, Czech Republic, June 2007.

Benus, S., Gravano, A., Hirschberg, J., "The role of prosody backchannels in American English", Proceedings of ICPhS 2007, 1065-1068, Saarbrucken, Germany, August 2007.

Higashinaka, R., Sudoh, K., Nakano, M., "Incorporating discourse features into confidence scoring of Intention Recognition Results in Spoken Dialogue Systems", *ICASSP,* vol.1, pp.25-28, 2005

Ishi, C.T., Ishiguro, H., Hagita, N. "Analysis of prosodic and linguistic cues of phrase finals for turn-taking and dialog acts," Proceedings of The Ninth International Conference of Speech and Language Processing 2006 (Interspeech'2006 - ICSLP), 2006-2009

Grisedale, S., Graves, M and Grünsteidl, A. "Designing a graphical user interface for healthcare workers in rural India" Proc. SIGCHI conference on Human factors in computing systems, Atlanta, USA, (1997), 471-478.

Plauche, M., *et al.,* "Speech Recognition for Illiterate Access to Information and Technology". Proc. Information & Communications Technologies and Development, Berkeley, USA, May 2006.

Sherwani, J and Rosenfeld, R., "The Case for Speech Technology for Developing Regions", Proc HCI for Community and International Development, Florence, Italy, April, 2008

Jindal, R., Kumar, R., Sahajpal, R., Sofat, S., Singh, S., "Implementing a Natural Language Conversational Interface for Indian Language Computing", IETE Journal of Technical Review, July - August 2004 Edition

Sharma, D., "Nominal clitics and constructive morphology of Hindi", in Proceedings of the LFG 99 Conference, Manchester, 1999.

Kachru, Yamuna, "Hindi" Publihed by John Benjamins Publishing Company, 2006.