

Prosody, focus, speaker variation and focus perception in Hindi

Frank Kügler

Institute of Linguistics, Goethe-University Frankfurt, kuegler@em.uni-frankfurt.de

In this talk I will discuss research I have done on the prosodic marking of focus in Hindi and the perception of focus by native listeners. It is agreed on that roughly every prosodic word is associated with a rising pitch contour except the last one, and the rises within an intonation phrase are downstepped to each other (Moore 1964, Harnsberger & Judge 1996, Patil et al. 2008). If a particular constituent is focused there seems to be no prosodic difference in pitch and duration of the focused constituent (Patil et al. 2008). In contrastive focus, there is an indication of prosodic marking by means of an increase in pitch span (Genzel & Kügler 2010). In all cases, however, the pitch register of the post-focal constituent is compressed; nevertheless the rising contours are still realized (Patil et al. 2008; Genzel & Kügler 2010). Given that there is no clear prosodic distinction of the focused constituent, the question arises whether post-focal pitch register compression is a prosodic cue that is functionally used by listeners to perceive the focus.

This is tested considering the case of contrastive ellipsis as in (1) that are ambiguous up to the conjunction *naa ki*, ‘rather than’. The predictions based on the findings in Patil et al. (2008) are that in speech production post-focal compression should occur after the object in the matrix clause that is contrasted with the object in the remnant. In (1a) the register compression would affect the verb, in (1b) it would affect the direct object. 30 native Hindi speakers were recorded reading twelve sentences of the kind in (1). The corresponding constituents were labeled, and for each one the low and high tones of the rises were measured in Hertz (Table 1). Time-normalized mean F0 contours across sentences and speakers show that (1a) and (1b) only differ in the realization of the rise on the direct object (Figure 1). Thus, the prediction for (1b) is borne out. The register is compressed by 20 Hz on average after the indirect object contrast. However, for (1a) post-focal compression on the verb has not been produced. Yet, (1a) and (1b) are prosodically differentiated depending on the contrast to be expressed. Considering speaker variation however we observe that speakers vary in the degree of post-focal compression they realize. Some speakers do not realize post-focal compression at all, while others compress their register up to 30 Hz.

To test the functional load of the feature ‘post-focal compression’ a sentence completion experiment was carried out. Data from six different speakers taken from the production study were selected. Two of these speakers did not produce post-focal compression, two of them produced a mean of 10 Hz, and the remaining two produced a mean of about 30 Hz. These three groups were chosen to test for speaker variation. The sentences in (1) were cut after the conjunction *naa ki*, and 18 native Hindi speakers were auditorily presented these sentence fragments. In a forced-choice completion experiment listeners were asked to decide which of the two possible objects contrasts (direct or indirect object) correctly will complete the sentence. The results show that for sentences like (1a) containing no prosodic cue related to contrastiveness listeners were unable to choose the intended sentence completion (Figure 2). The decisions were about chance level (C01 in Figure 2). The same holds for sentences like (1b) when speakers do not produce post-focal compression (C02 in upper and lower left panels in Figure 2). Yet, correct sentence completion judgments increased significantly if listeners were provided with the prosodic information of post-focal compression (upper and lower mid and right panels in Figure 2). The results further indicate that a higher degree of post-focal compression yields a higher rate of correct completion decisions.

My conclusion would be that the prosodic contour of a sentence in Hindi guides sentence disambiguation at an early stage in the parsing process, and thus post-focal compression can be viewed as a cue for focus perception in Hindi. Why speakers vary in their production of prosodic focus however remains an open issue.

References

- Moore, R.R. (1965). *A study of Hindi intonation*. Ph.D. thesis, University of Michigan.
- Harnsberger, J. D., Judge, J. (1996). Pitch range and focus in Hindi. *JASA* 99:2493.
- Genzel, S. & Kügler, F. (2010). The prosodic expression of contrast in Hindi. *Proc. Speech Prosody 2010, Chicago*.
- Patil, U., Kentner, G., Gollrad, A., Kügler, F., Féry, C., Vasishth, S. (2008). Focus, word order, and intonation in Hindi. *Journal of South Asian Linguistics* 1, 53-70.

Materials

- (1) a. raahul ne ma ko davaaii dii naa ki ghaddii
 RaahulERG mother DAT medicine give NEG that car
 Raahul gave the medicine to the mother and not the car.
- b. raahul ne ma ko davaaii dii naa ki naanii ko
 RaahulERG mother DAT medicine give NEG that granny DAT
 Raahul gave the medicine to the mother and not to granny.

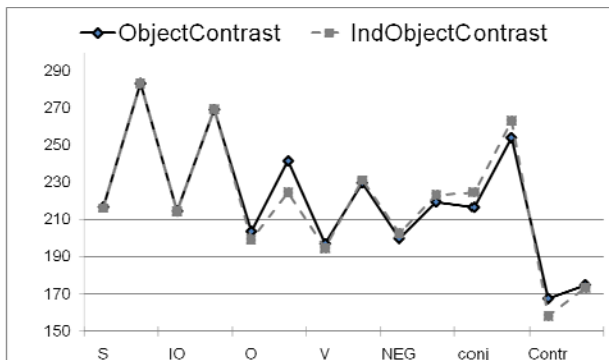


Figure 1. F0-minimum and F0-maximum per constituent, showing the mean across all speakers (n=30) comparing condition ((1a) – solid black line) with condition ((1b) – dashed grey line).

Table 1. F0-maximum, F0-minimum, and duration on the Direct Object and following Verb in conditions (1a) and (1b) displayed by token.

	F0max [Hz]		F0min [Hz]		Duration [s]	
	Object	Verb	Object	Verb	Object	Verb
Condition (1a)						
1	248	231	201	195	0,3980	0,2034
2	229	215	202	192	0,3113	0,3088
3	235	241	198	203	0,4491	0,2704
4	247	227	206	200	0,4177	0,3927
5	245	245	211	203	0,3312	0,3809
6	247	218	205	191	0,6056	0,4036
mean	242	230	204	197	0,4188	0,3266
Condition (1b)						
1	230	228	190	191	0,3548	0,1814
2	210	230	197	191	0,2653	0,3235
3	211	232	199	199	0,3975	0,2599
4	241	234	205	198	0,3699	0,3599
5	217	239	204	198	0,2668	0,3594
6	238	226	198	191	0,5656	0,4121
mean	225	231	199	195	0,3706	0,3160

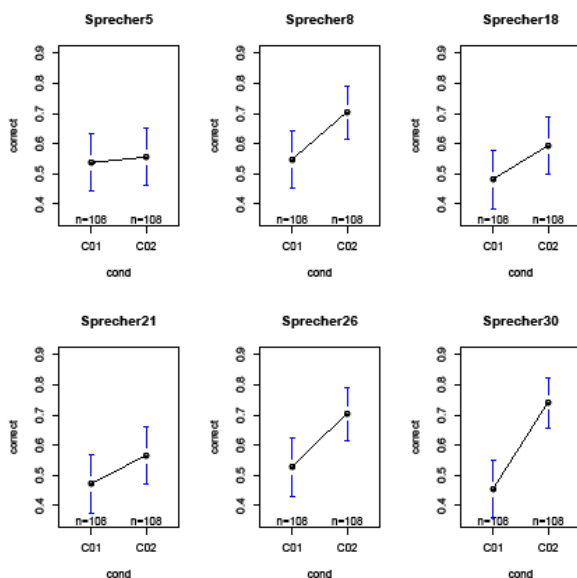


Figure 2. Degree of correct sentence completion of items spoken by six speakers based on 18 listeners's judgments. Speakers 5 and 21 (upper left and lower left panel) realized no post-focal compression, speakers 8 and 18 (upper mid and right panel) realized a mean of 10 Hz post-focal compression, and speakers 26 and 30 (lower mid and right panel) a mean of 30 Hz.