Focus, word order, and intonation in Hindi

Umesh Patil, Gerrit Kentner, Anja Gollrad, Frank Kügler, Caroline Féry, Shravan Vasishth, *Institute for Linguistics, University of Potsdam, Germany*

Submitted July 15, 2008; DO NOT CITE OR DISTRIBUTE

Abstract

A production study is presented that investigates the effects of word order and information structural context on the prosodic realization of declarative sentences in Hindi. Previous work on Hindi intonation has shown that: (i) non-final content words bear rising pitch accents (Moore 1965, Dyrud 2001, Nair 1999); (ii) focused constituents show greater pitch excursion and longer duration and that post-focal material undergoes prosodic compression (Moore 1965, Harnsberger 1994, Harnsberger and Judge 1996); and (iii) focused constituents may be followed by a phrase break (Moore 1965). By means of a controlled experiment, we investigated the effect of focus in relation to word order variation using 1200 utterances produced by 20 speakers. Fundamental frequency (F0) and duration of constituents were measured in Subject-Object-Verb (SOV) and Object-Subject-Verb (OSV) sentences in different information structural conditions (wide focus, subject focus and object focus). The analyses indicate that (i) regardless of word order and focus, the constituents are in a strict downstep relationship; (ii) focus is mainly characterized by post-focal compression rather than raising of the element in focus; (iii) given expressions that occur prefocally appear to undergo no compression; (iv) pitch excursion and duration of the constituents is higher in OSV compared to SOV sentences. A phonological analysis suggests that focus affects pitch scaling and that word order influences prosodic phrasing of the constituents.

JSAL Volume 1, Issue 1, January 2007.
Focus, word order, and intonation in Hindi.

1.1 Introduction

Prosody is an integral component of language, and so it is only natural that online sentence comprehension and production critically depend on the information-structuring cues provided by prosody. The role of prosody in comprehension and production is especially relevant for languages such as German and Hindi that involve relatively free word order; word order flexibility increases the number of options available for expressing information structure, significantly complicating the means by which the incoming signal can be decoded. Any extra-syntactic cue, such as prosody, would a priori be expected to provide an important cue for facilitating comprehension.

European languages such as English and German (e.g. Gussenhoven 2008, Ladd 1996, Selkirk 2007, Truckenbrodt 1995, Féry and Kügler to appear) have received a great deal of attention concerning the prosodic marking of information structure. However, not much is known about Hindi in this respect. In an attempt to fill this gap, we investigate the effect of word order and intonation on prosody. We carried out a production study of Delhi Hindi (20 participants) which showed that the intonation of Hindi and its interaction with focus and word order differs from well-studied intonational languages such as English and German. First, the primary prosodic cue accompanying focus on a constituent is post-focal compression, rather than the raising of F0 as observed in intonational languages such as English; this is consistent with previous research on Hindi intonation (Moore 1965, Harnsberger 1994). Second, in sentence-initial focus, canonical (SOV) word order shows a greater post-focal compression than non-canonical (OSV) order. Third, when sentence-medial elements are focused, (i) the sentence-initial object in non-canonical (OSV) order has a higher F0 peak as well as a greater F0 range than the sentence-initial subject in canonical order (SOV) order, (ii) the duration of the medial (focused) element is longer in the non-canonical order compared to canonical order, and (iii) no evidence is found of pre-focal compression for given (previously mentioned) elements.

We propose an analysis of phrasing in Hindi according to which each content word in Hindi is phrased separately as a prosodic-phrase (p-phrase). Each p-phrase receives a low pitch accent and a high phrase boundary associated with the right edge of the prosodic word. Regarding pitch range effects, tones are scaled relative to abstract reference lines, and we assume the all-new sentence pattern to represent the neutral baseline. Focus, then, compresses the post-focal register in Hindi.

1.2 Background

Hindi belongs to the Indo-European branch of languages and is an Indo-Iranian language (Kachru 1987) genetically related to European intonation languages such as English, German and Russian. It is natively spoken by approximately 366 million speakers (source: Ethnologue, www.ethnologue.com), mainly in the central and northern part of India, but also in Bangladesh, Nepal, the United Kingdom and many other countries. In addition, Hindi is also used as a second language or a lingua franca by many Indians in non-Hindi speaking regions. The syntax of Hindi and Urdu (which is spoken in Pakistan as well as India) are virtually identical, although the scripts and choice of content words differ. We use the term 'Hindi' in this paper but the conclusions presented are expected to extend to Urdu as well.

1.2.1 Hindi intonation

The work of Nair (2001) and Dyrud (2001) suggests that Hindi has lexical stress, in the sense that every word has a designated syllable on which prominence is realized (see Hayes 1995, Moore 1965, Ohala 1986, who notice contradictions on this view). Nair (1999) and Dyrud (2001) find acoustic correlates of prominent syllables, like higher pitch and longer duration. Position of lexical stress is dependent on syllable weight, the heavier syllables attracting stress first. Hussain (1997) shows how the position of stress can be predicted by syllable weight (but see Ohala 1986, who finds differences in stress position depending on whether a word is uttered in isolation or in a sentence).

All researchers on Hindi intonation appear to agree that each content word except the final one is associated with a rising contour.[1] According to Moore (1965, 68), every 'foot' contains a pitch accent, where foot is defined as "one to several syllables in length, which normally is uttered with a continuously rising pitch from beginning to end". Harnsberger (1994) makes a similar observation, and proposes that the low part of the rising contour is a low pitch accent, annotated as L* in an autosegmental-metrical notation system (Pierrehumbert 1980). The high part of the rising contour is either a trailing tone H-[2], or a boundary tone H_P. The subscript 'P' represents a phrase boundary smaller than the intonation phrase (see Hayes and Lahiri 1991, for this annotation convention). In the analysis presented below, we treat the high

[1] Moore's data suggests that wh-questions are realised with a different intonation pattern, though Moore himself does not emphasize this fact and a systematic analysis has yet to be done.

[2] A trailing tone is the part of a bitonal pitch accent that follows the starred tone.

part as a boundary tone.

Moore's account of Hindi intonation comprises an analysis of pitch, intensity and duration for three prosodic phenomena: the expression of emphasis (or focus), the expression of speaker attitudes, and phrasing (Moore 1965, 62). Although Moore distinguishes different melodic contours on a single syllable, i.e. level, rise, fall, rise-fall, fall-rise (Moore 1965, 65), the underlying pitch accent is considered a rising one. Deviations from this underlying pattern on the surface are the result of speaker's attitude, or, alternatively, the result of a pure phonetic effect, namely tonal transition from a very high rising pitch to the following low tone (1965, 68, 75).

As for phrasing, Moore assumes two prosodic levels below the intonation phrase: the foot (as defined above), and a higher prosodic level called 'the measure' (1965, 68ff). This latter level of phrasing separates a focused element from the rest of the phrase. Thus, according to Moore, focus has a phrasing effect. The same effect of focus has also been observed in the closely related language Bengali (Hayes and Lahiri 1991). In addition, a pause indicates a measure break, for example in continuation contexts (Moore 1965, 80). Moore (1965, 81, 84) also assumes that an intonation phrase contains at least one measure, but every word in an intonation phrase may be phrased separately on this level. Furthermore, each measure contains at least one foot indicated by a rising pitch accent. Harnsberger (1994) also assumes two levels of phrasing and two kinds of boundary tone, low and high. In contrast to Hayes and Lahiri (1991) for Bengali, Harnsberger's tonal analysis does not assume bitonal boundary tones.

Focus has been claimed to have three prosodic effects (see Harnsberger 1994, 1999, Moore 1965, Dyrud 2001, among others). First, the rising pitch pattern may show a higher excursion, a greater intensity and longer duration.[3] Second, after the focused element, a phrase break may occur. Third, post-focally the pitch range may be compressed or even completely flat and deaccented (Harnsberger and Judge 1996), although rising pitch accents are still realized in compressed pitch range.

Hindi shares some properties with intonation languages. It has lexical stress, and accordingly has pitch accents on designated syllables. It uses melodic changes for pragmatic functions, such as signaling questions and speaker attitudes, and it uses phrasing intensively for articulating information structure (see Jun 2005, for a more fine-grained

[3] Note, however, that according to Moore (1965) a focus needs not necessarily be realized by means of all three phonetic correlates but any combination of one or two of these may suffice to phonetically express focus. Thus, F0 as a correlate may be absent.

typological classification of prosodic features). There is a striking similarity between the intonation of Hindi and detailed descriptions of other South-Asian languages like Bengali and Tamil: Hayes and Lahiri (1991) and Khan (2007) assume several similar prosodic properties for Bengali, such as lexical stress, pitch accents associated with stressed syllables, and intonational phrasing on at least two levels. The Dravidian language Tamil, although genetically unrelated, also shows similarities with Hindi (Keane 2007a,b); this could be a consequence of language contact.

1.2.2 Information structure and Hindi syntax

Hindi is a head-final (Subject-Object-Verb) language, with relatively free word order. Constituents may be scrambled to express different information structural configurations, or for stylistic reasons. The first syntactic constituent in a sentence is usually the aboutness topic (Gambhir 1981, Butt and King 1996), which may under certain conditions be marked by the clitic -*to*, similar in some respects to Japanese -*wa* (Kuno 1981, Kidwai 2000). The term 'aboutness topic' is understood here as a referent which the remainder of the sentence is about, possibly contrasting with other referents, and followed by a focused constituent (see Reinhart 1981, Jacobs 2001, among others).

In Hindi, a focused constituent typically occupies the immediately preverbal position, and wh-markers also tend to occur preverbally (Kidwai 2000, 116). Nominal clitics can serve to mark focus (similar to English focus particles like 'only', 'even' or 'also', Sharma (2003)). However, focused constituents need not be morphologically marked. In this paper, 'focus' is used rather traditionally as the part of the sentence which introduces alternatives (Rooth 1985, 1992). The term 'focus' is applied to constituents which are informationally more important than other backgrounded parts of the same sentence. In the general case, an all-new sentence does not trigger a set of alternatives, though the possibility of focusing a whole sentence cannot be excluded in principle. Below, we call an all-new sentence a 'wide-focused' sentence. According to Butt and King (1996), in situ focusing of a phrase in Hindi is possible with multiple foci and results in contrastive focus readings. Kidwai (2000, 114-137) presents detailed arguments that focus is responsible for scrambling operations such as preposing (as XP adjunction operations).

Butt and King also provide evidence that background information occurs postverbally, and completive information – which is information of secondary importance to the information structure of the discourse – occurs in the preverbal region preceding the focus position.

1.2.3 Aim of the present study

Thus, although much is known or hypothesized about word order and focus in Hindi, very few controlled experimental studies exist that explore the interaction with prosody. In order to remedy this situation, we took this previous work as a theoretical starting point and designed a production experiment that investigates the intonational realization of focus and its interaction with different word orders.

1.3 Production experiment

1.3.1 Method

Design and Materials

The experiment involved a 3x2 factorial design with two factors: focus (subject, object and wide focus) and word order (SOV and OSV[4])

Each trial consisted of a question-answer pair: a question and a response to the question. The question set up either a subject, object or wide focus for the response utterance; see examples (1), (2), (3). In (1-b) and (1-c) the question involves subject focus, in (2-b) and (2-c) object focus, and in (3) wide focus. In the examples, a bracketed segment with a subscripted F stands for the focused element relative to the preceding question.

(1) Subject question

a. kis ne davaaii ko khariidaa?
who ERG medicine ACC buy.PAST?
Who bought the medicine?

b. [graahak ne]$_F$ davaaii ko khariidaa
customer ERG medicine ACC buy.PAST
(The) customer bought the medicine

c. davaaii ko [graahak ne]$_F$ khariidaa
medicine ACC customer ERG buy.PAST
(The) customer bought the medicine

(2) Object question

a. graahak ne kyaa khariidaa?
customer ERG what buy.PAST?
What did the customer buy?

b. graahak ne [davaaii ko]$_F$ khariidaa

c. [davaaii ko]$_F$ graahak ne khariidaa

[4]Only two word orders (SOV and OSV) are considered here in order to keep the number of experiment conditions tractable, and because not all word orders are possible in Hindi and the constraints on word order variation are far from clear.

(3) Wide focus question

a. kyaa huaa?
   what happen.PAST?
   What happened?

b. [graahak ne davaaii ko khariidaa]$_F$

c. [davaaii ko graahak ne khariidaa]$_F$

The questions were always in canonical (SOV) word order, and the answers always contained a transitive verb and two arguments, with subject arguments in ergative case and objects in accusative case. Subjects were nouns referring to humans and objects referred to inanimate referents. Past tense and perfective aspect was used in all sentences. Half of the subject and object nouns were bi-syllabic with initial stress and the other half were tri-syllabic with stress on the second syllable. The complete set of target items is shown in the Appendix.

A note on terminology: Since at most one argument (subject or object) is focused in the question utterance, the non-focused argument will be designated as given in the response utterance (previously mentioned in the question; Allerton 1978, Lambrecht 1994). For example, in the subject question (1), *davaaii*, 'medicine', is mentioned, and so in the response utterances (1-b) and (1-c) the referent denoted by *davaaii* is given. By contrast, in the wide focus condition (3), since none of the arguments are mentioned, in the corresponding response utterances (3-b) and (3-c) none of the noun phrases refer to given elements. This distinction between the focused and given element becomes relevant when the results of the experiment are presented.

A total of 18 unique question-answer pairs were constructed and each pair was realized in the 6 conditions, resulting in 18×6=108 sentences per speaker. All the 108 sentence-pairs were presented to each speaker in a pseudo-randomized manner; items from four other unrelated experiments were interspersed as fillers. Four pseudo-randomized lists were prepared to minimize order effects.

The questions were recorded in a speech recording laboratory in the University of Potsdam in preparation for presentation of stimuli to participants; the presentation procedure is described below.

Participants

30 native speakers of Hindi participated in the experiment. All were female students at the University of Delhi, India and were residents of Delhi and surrounding areas. Each speaker was paid 150 Indian Rupees for participation and took approximately 45 minutes to complete the experiment.

Procedure

The experiment was carried out using presentation software. First, participants were equipped with a set of headphones and a microphone head-set, and familiarized with the task through written and verbal instructions, followed by two practice trials. Each trial consisted of a presentation of the question and its answer on the computer screen, written in Devanagari. Participants heard the pre-recorded question over headphones, spoken by a male voice. At the same time the target sentence was presented on the screen. Participants were instructed to speak out the answer displayed on the screen as a response to the question. If the question was answered without any hesitations or false starts, the next trial was presented. If there were hesitations, participants were asked to repeat the answer. A total of 48 items (4%) had to be repeated because of false starts or hesitations. Presentation flow was controlled by the experimenter, and participants were allowed to take a break whenever they wanted. The sentences produced by participants were recorded at the University of Delhi on a DAT tape recorder.

Data pre-processing and statistical analysis

Due to limited resources we analyzed a subset of the data. Of the 18 items, five items from each syllabic and stress pattern, i.e., a total of 10 items, were selected for annotation and analysis (the first five items of each syllabic pattern) Of the 30 speakers, utterances from the last 20 speakers was used in the data analysis. This resulted in a total of 1200 utterances (20 speakers $\times$ 10 items $\times$ 6 conditions). The above criteria for subsetting the data were decided upon arbitrarily.

The recordings were re-digitized from DAT at a sampling frequency of 44.1 kHz and 16 bit resolution. Data were labeled by hand at the level of the constituent, as shown in (4). The vertical lines mark constituent boundaries.

(4) | graahak ne | davaaii ko | khariidaa |
customer ERG medicine ACC buy.PAST
(The) customer bought the medicine.

The pitch analysis was conducted using a Hanning window of 0.4 seconds length with a default 10 ms analysis frame. The pitch contour was smoothed using the Praat (Boersma and Weenink 2005) smoothing algorithm (frequency band 10 Hz) to diminish microprosodic perturbations. Stylized pitch tracks were calculated. For this purpose, each constituent in (4) was divided into five equal intervals, and the mean pitch was aggregated over the 20 speakers and 10 sentences for each interval. The resultant values were interpolated separately for each condition.

Dependent variable	*Where measured*
F0-maximum (Hz)	at the right edge of preverbal constituents (This is where H boundary tones are expected.)
F0-range (Hz)	preverbal constituents
Duration (ms)	preverbal constituents

TABLE 1 Dependent variables.

For each constituent in (4), the maximum F0, the minimum F0 and the duration were detected using a Praat script. In the second constituent, only those F0 maxima were measured that followed the F0 minimum in that constituent; this was done in order to exclude maxima due to transitions from preceding H tones. The maximum after the low tonal target represents the high tone in the LH gesture. Based on the measurements of F0-maximum and F0-minimum the F0-range was calculated (F0-max minus F0-min).

The statistical analysis relied on three dependent variables, F0-maximum, F0-range, and duration; the loci of these measurements are shown in Table 1. All dependent measures were log-transformed to meet the assumption of the regression model.

A multilevel model (Gelman and Hill 2007, Bates and Sarkar 2007, Pinheiro and Bates 2000) was fit, using crossed random factors speaker and item, and focus status of constituent (wide focus, narrow focus, given), and word order of sentence (SOV vs. OSV) as fixed factors.

1.3.2 Results and Discussion

Effect of Focus

The contours in Figures 1-3 show time-normalized mean pitch tracks for each focus condition averaged over all 20 speakers. The contours show rising tonal patterns on the non-final constituents and falling patterns on the final verb. Table 2 in the Appendix summarizes the results of the statistical analyses.

As shown in Figure 1, for SOV structures, in the subject, object and wide focus conditions a rising pitch gesture is realized on both pre-verbal constituents. Object focus and wide focus do not show a significant difference, but subject focus is realized with a significantly higher F0 excursion compared to the wide focus counterpart (although the magnitude of the difference is small). In the subject focus condition, the rising gesture on the object is realized in a clearly lower and compressed range compared to the other conditions.

A similar pattern is seen in OSV structures. Here, the pitch tracks

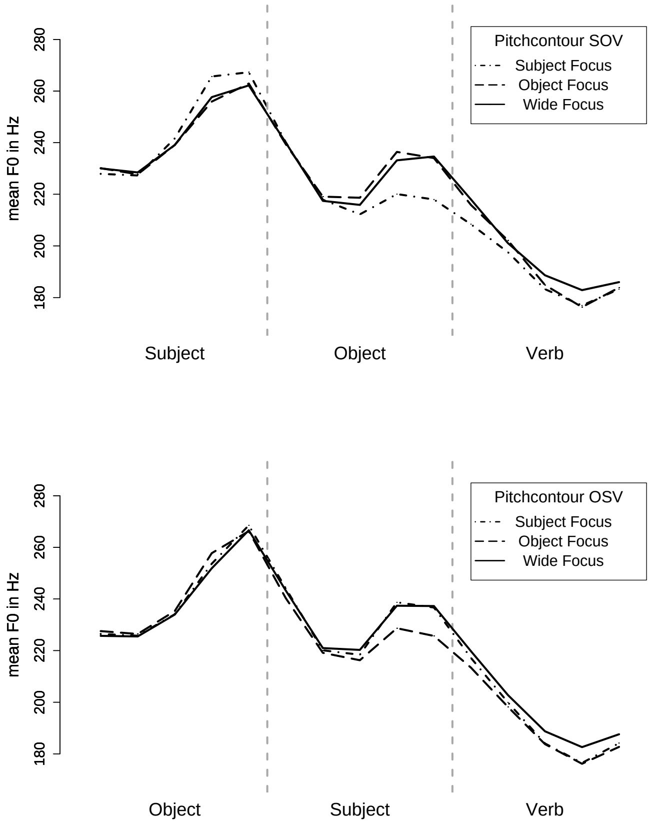

FIGURE 1 Time-normalized pitch tracks, based on five measuring points per constituent, showing the mean across all speakers. The upper plot shows SOV order and the lower plot OSV order. The comparisons of interest in each plot are subject focus (dotted line) and object focus (dashed line) with respect to wide focus (solid line).

of subject focus and wide focus are nearly identical. This absence of a difference between the two types of focus may due to the fact that in OSV word order, the subject is in the default preverbal focus position (Section 1.2.2) in both the conditions. In OSV structures, the realization of (sentence-initial) object focus shows a divergent pattern, although not on the focused constituent itself. Only post-focal compression is visible: the rising pitch gesture in the post-focal constituent is significantly lower than the one for wide focus.

Focus thus induces post-focal compression of the pitch range, which confirms the results of Moore (1965), and Harnsberger and Judge (1996). However, the expected effect of greater pitch excursion on the focused constituent itself is only borne out for initial focus in SOV sentences but not for medial focus or OSV sentences.

We now turn to the results of the statistical analyses on the three dependent variables. Subject focus in SOV order had a higher F0-maximum ($t=4.26$), a greater pitch range ($t=4.97$), and longer duration ($t=2.62$) on the focused subject, compared to the wide focus baseline. Compared to the baseline, the given object showed a significantly lower F0-maximum ($t=-9.06$), a smaller F0-range ($t=-9.94$), and shorter duration ($t=-6.24$).

Although no effect of focus was found on the object in OSV sentences, post-focal compression on the medial constituent was seen when the initial object was in focus (lower F0-maximum ($t=-8.23$), lower pitch range (t=-6.7) and shorter duration ($t=-3.62$) compared to the wide focus baseline) much as in SOV order. When the medial subject was focused, it had a slightly but significantly higher F0 range ($t=2.34$) compared to the baseline. F0-maximum and duration did not yield significant effects here.

Pre-focally given constituents do not show any clear difference compared to wide focus baseline: an initial given subject is nearly identical in F0 maximum (initial subject and baseline: 274 Hz), F0 range (initial subject 58 Hz, baseline 59 Hz), and duration (initial subject 556 ms, baseline 558 ms). The same result holds for an initial given object (F0-max: initial object 276 Hz, baseline 279 Hz; F0-range: initial object 67 Hz, baseline 68 Hz; and duration: initial object 551 ms, baseline 558 ms).

Effects of word order

Wide focus. Figure 2 shows SOV and OSV word orders in the wide focus condition, and Table 3 in the Appendix summarizes the results of the statistical analyses.

Almost no difference is seen in the time-normalized pitch tracks. The

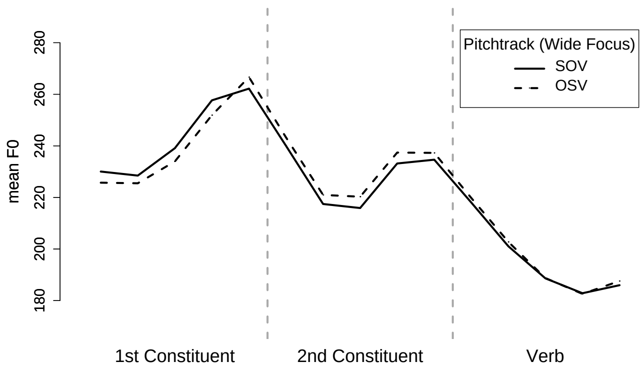

FIGURE 2 Time-normalized pitch track based on five measuring points per constituent averaged across all speakers for SOV (solid line) and OSV (dotted line) in the wide focus condition.

marked word order (OSV) is on average slightly lower in the rising part of the first constituent, higher on its peak, and it is slightly higher on both the rising part and on the peak of the second constituent. Regarding the dependent variables, the F0-range on the initial constituent, but not the F0-maximum, is significantly larger (9 Hz) in OSV than in SOV, t =4.59. This suggests that the low pitch accent is realized lower in OSV order. Additionally, the duration of the preverbal constituent is on average 17 ms longer in OSV compared to SOV. This difference is significant (t=2.51).

First or second constituent in focus. We turn next to the word order comparison when either the first (Figure 3a) or second constituent is in focus (Figure 3b). In sentence-initial focus (Figure 3a), no difference was found in the realisation of the focused constituent, but in the post-focal constituent the amount of post-focal compression is larger for the unmarked word order. In other words, the post-focal pitch range is higher for OSV word order. When the first constituent is the focus, the medial subject (OSV) displays a higher F0-maximum (t=4.55), a greater F0-range (t=2.65), and longer duration (t=3.76) than the medial object (SOV). Although the contour plot (Figure 3a) suggests an earlier and steeper rise in pitch on the first constituent for SOV structures, as compared to OSV sentences, no significant differences in the dependent variables were found on the initial focused constituents.

In case of second-constituent focus, a small difference appears on

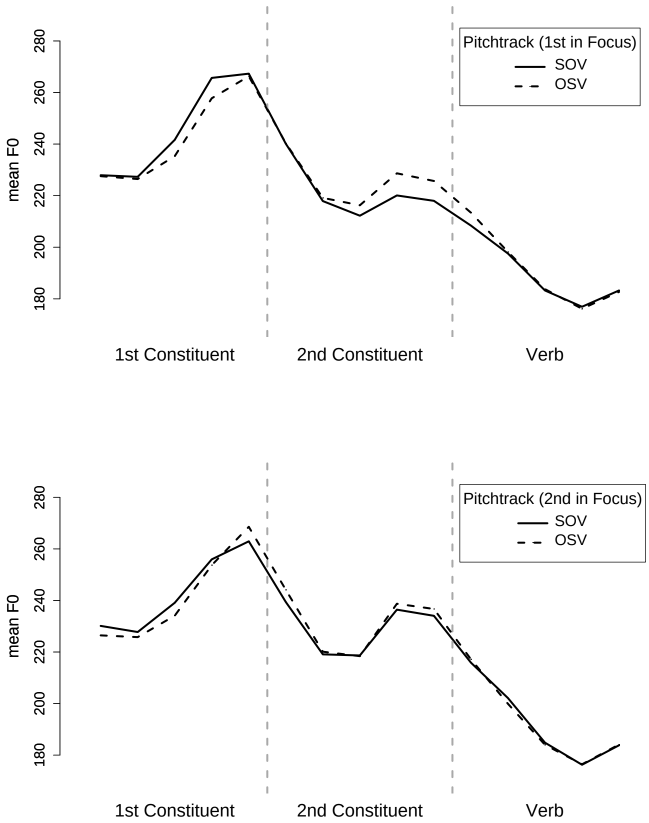

FIGURE 3 Time-normalized pitch tracks based on five measuring points per constituent averaged across all speakers for SOV (solid line) and OSV (dotted line) word order; in the upper plot the first constituent, and in the lower plot the second constituent is in focus.

the pre-focal given constituent. The marked word order (OSV) shows a higher F0 peak on the initial given object. The F0-maximum (t=3.44), and correspondingly, the F0-range (t=4.77) on the initial constituent is significantly higher in OSV, as compared to SOV structures. The duration of the medial (focused) constituent itself is 20 ms longer in OSV than SOV sequences (t=3.56). In the duration measure, the same pattern is seen for these constituents when they are given (22 ms, t=3.76), i.e., when the focus is on the first constituent.

1.3.3 Summary of the results

The results can be summarized as follows. Pitch contours of SOV and OSV sentences display the same basic pattern: both arguments have a rising tonal structure, and the verb has a falling structure. This result is in line with previously observed pitch patterns in Hindi (Moore 1965, Harnsberger 1994). The highest part of all three constituents are in a very clear downstep relationship to each other.[5]

With the exception of initial subject focus in SOV structures, focus was not found to affect the pitch excursion and duration of the focused elements. However, it does affect the post-focal constituent, when the initial element is in focus: in both word orders, the medial, post-focal constituent has a lower F0-maximum, a smaller F0-range and duration is shorter than in the baseline wide focus condition. When focus is on the second, preverbal constituent, no reliable difference was found compared to the wide focus baseline in F0-maximum, F0-range and duration on either of the preverbal constituents. Thus, we found no evidence of pre-focal compression due to givenness, and no raising of F0 as a consequence of focus as reported for many intonation languages (e.g. Bartels and Kingston 1994, Baumann et al. 2006, Cooper et al. 1985, Féry and Kügler to appear). The absence of any prosodic effect when the preverbal constituent is focused might be due to the fact that the preverbal position is the syntactic default position for focus (Kidwai 2000). Prosodic marking of focus in this position might therefore be redundant.

Word order has an effect on prosody: significant differences were found between SOV and OSV word orders, appearing most clearly on the given constituents. First, in sentence-initial focus, the amount of post-focal compression was larger in SOV than in OSV sentences, as reflected by a lower F0-maximum, a smaller F0 range and shorter duration of the medial constituent in SOV sentences. Moreover, in sentences with focus on the second constituent, the F0-maximum, as well as the

[5] Although Moore does not mention the effect of downstep it appears to be visible in his data, e.g., his examples (11) or (14) (Moore 1965, 80, 101).

F0-range on the initial given constituent are slightly but significantly higher in OSV, as compared to SOV structures. Third, the duration of the medial focus constituent is longer in OSV than SOV sequences.

Based on these results, we turn now to a phonological analysis of Hindi intonation.

1.4 A phonological interpretation

A phonological analysis (including phrasing and pitch scaling) is presented next, based on the production data. Phrasing is a direct consequence of syntactic structure and is thus very simple, as the syntactic structure of the sentences investigated is quite simple and all constituents are phrased individually. The tonal realization is, on the one hand, dependent on the syntactic structure (downstep pattern), as well as finality or non-finality of the smaller prosodic phrases in a larger intonation phrase, and, on the other hand, the result of information structure. The variation observed in pitch scaling comes from information structure and is discussed in more detail below.

1.4.1 Phrasing

The general pattern of intonation described by Moore (1965), Harnsberger (1994, 1999) and others was confirmed in our data. Every content word (here every argument) except for the final one (the verb) has a rising contour, which can be analyzed as a prosodic domain. This implies that the phrasal contour is clearly realized, especially the final boundary tone of a prosodic-phrase or p-phrase. Every constituent forms its own p-phrase, and all three constituents form an intonation phrase or i-phrase. The final verb and the preceding object are more tightly phrased together than the initial argument, a fact that we can express as recursive phrasing (see for instance Ito and Mester 2007, for recursive phrasing), as illustrated in (5): The p-phrases of the object and of the verb are grouped together in a single p-phrase that comprises them both.[6] By contrast, and as shown in (6), the order OSV does not have recursive phrasing. A subject and a following verb are separated by a stronger syntactic boundary than an object and a following verb. As a result, they are not grouped together in a common p-phrase.

(5) a. $[[\text{ S }]_{\text{P}}\ [[\text{ O }]_{\text{P}}\ [\text{ V }]_{\text{P}}]_{\text{P}}]_{\text{I}}$
b. $[[\text{graahak ne}]_{\text{P}}\ [[\text{davaaii ko}]_{\text{P}}\ [\text{khariidaa}]_{\text{P}}]_{\text{P}}]_{\text{I}}$
customer ERG medicine ACC buy.PAST
The customer bought the medicine.

[6] Alternatively, three levels of phrasing may be assumed, as proposed by Khan (2007) for Bengali.

(6) a. $[[\text{ O }]_P\ [\text{ S }]_P\ [\text{ V }]_P]_I$
b. $[[$davaaii ko$]_P$ $[$graahak ne$]_P$ $[$khariidaa$]_P]_I$
medicine ACC customer ERG buy.PAST
The customer bought the medicine.

The difference between the tonal realization of the p-phrase of the subject and the object on the one hand and verb on the other, comes from the non-finality of the former, and finality of the latter. The tonal structure of the verb is determined by the final low boundary tone (see below), while the tonal structure of the arguments is influenced by the non-final high boundary tone.

Importantly for the phonological analysis, phrasing is unchanged by narrow focus, since the phrasal boundaries are always realized, albeit sometimes only weakly. This points to an absence of an effect of focus for phrasing: focus does not insert a prosodic boundary (see similar results for Bengali in Khan 2007, 39-40), though this result needs confirmation from a larger and more varied set of data. This analysis would contradict that of Moore (1965), who assumes a phrase break after a focused constituent.

1.4.2 Pitch and pitch scaling

Following Nair (1999) and Dyrud (2001) among others, we assume that Hindi has lexical stress, which means that the low part of the rising pattern observed on all non-final constituents can be analyzed as a starred low tone L* for a pitch accent. Harnsberger (1999) proposed that, phonologically, the high part of the rising gesture may be analysed in two different ways: as a high trailing tone (+H-) or as a high phrase tone (H_P). Because of the clear phrasing found in our data, we analyze the rising pitch gesture as a low pitch accent L* and a high phrase boundary tone H_P, see (7). The final verb has a falling contour, which we assume is coming from a high pitch accent H* and a low boundary tone at the level of the intonation phrase (L_I).[7]

L* H_P L* H_P H* L_I
(7) $[[$graahak ne$]_P$ $[[$davaai ko$]_P$ $[$khariidaa$]_P]_P]_I$
customer ERG medicine ACC buy.PAST

Based on the time-normalized pitch contours in Figure 1, we assume that H_P is associated with the right edge of the constituent, i.e. the

[7]The last p-phrase can optionally end with a rising contour which does not necessarily strike Hindi speakers as a list intonation. Some of our speakers regularly realized a rising final intonation, others only occasionally.

target noun plus case marker.[8] Further evidence for this claim comes from an unpublished study by Genzel (2007) on Hindi. Genzel manipulated the number of syllables of a target word from one to five in order to investigate the behavior of the high tone. Her data suggest that the high tone is associated with the right edge of a prosodic word (see Khan 2007, for the same conclusion for Bengali).

The wide focus pattern can serve as a baseline for the remaining contours: All three high tones are subject to downstep. The downstep pattern affects the high tones of prosodic domains. Each high tone is lower than the preceding high tones in the same level of prosodic structure. Figure 4-A shows the metrical structure of sentences of both word orders plus the top lines of the prosodic domains to illustrate pitch scaling. Every constituent is the head of its own prosodic phrase, and for this reason, each constituent has the same metrical level. The downstepped lines above the metrical structure illustrate the top lines of the prosodic phrases. They show the highest point that the speaker's voice can reach at this moment, and define the maximal height of the high tones. We take the downstep pattern of the p-phrases as an obligatory feature of Hindi intonation. A sequence of prosodic phrases of the same level is organized in downstepped p-phrases.

Narrow focus on one constituent is accompanied by givenness of the other constituents. We represent prominence due to focus and givenness with the help of abstract metrical grid positions. The constituent with narrow focus acquires an additional grid mark, and the following ones become, in relation, less prominent. This difference may change the scaling of tones, as is illustrated with an $\mathrm{S_F}$OV configuration, see Figure 4-B. It is important to realize that the difference in prominence can in principle be realized in two ways implying pitch scaling: either by a rise on the focused constituent, or by a lowering on the given constituents. In Hindi, a change in the focus relationship is expressed by compression of the given constituents, but only of the post-focal ones. The first focused constituent does not change its level.

When the second constituent is narrowly focused, no difference in scaling appears (Figure 4-C). There is neither raising of the focused constituent nor lowering of the given constituent. We assume that the reason for the total absence of prosodic effects in such a configuration is that the top lines associated with prosodic domains cannot be changed in such a way that downstep inside of an i-phrase is cancelled. This means that the change in the metrical structure as a consequence of

[8] Moore (1965) reported a similar observation, the rising pitch gesture may continue throughout the noun and any following grammatical morpheme.

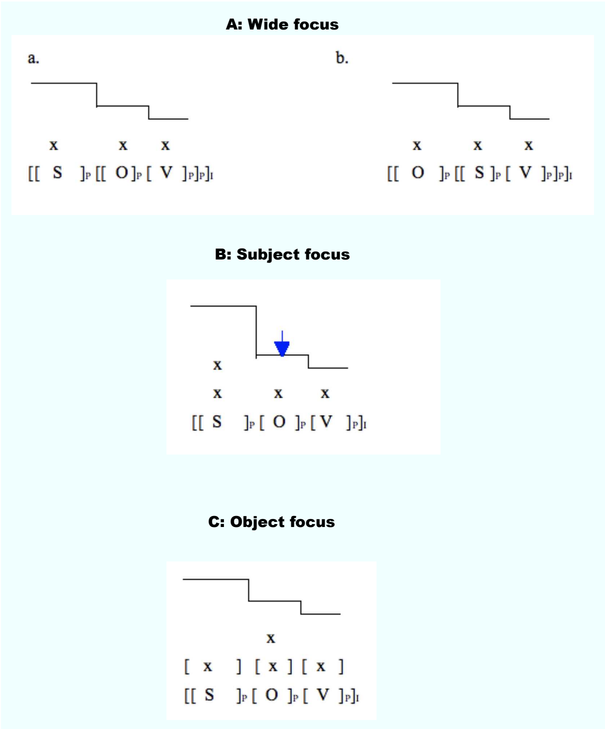

FIGURE 4 Figure A shows the phonological interpretation of $[\mathrm{SOV}]_{\mathrm{F}}$ (a) and $[\mathrm{OSV}]_{\mathrm{F}}$ (b) in the wide focus context; the prosodic phrasing, metrical grid and top lines relationship are displayed. Figure B shows the phonological interpretation of an $\mathrm{S_FOV}$ sentence in subject focus context (initial focus). The prosodic phrasing, metrical grid and top lines relationship are displayed. The arrow indicates a lowering of the post-focal pitch range. Figure C shows the phonological interpretation of an $\mathrm{SO_FV}$ sentence in object focus context (preverbal focus). The prosodic phrasing, metrical grid and top lines relationship are displayed.

the change in information structure has no effect on the relative height of the top lines, and thus, also on the height of the individual tones which are scaled according to these top lines. Raising of the medial constituent would result in suppressing the difference between the first and the second constituent. The same result would appear if the first constituent were compressed. Since both operations would cancel the downstep relation, none of them is performed, and pitch scaling is not affected by focus on the preverbal constituent.

We turn next to word order considerations. First, post-focal compression is larger in SOV than in OSV order. This effect correlates with the difference in phrasing shown in (5) and (6), which correlates with a difference of syntactic boundary strength between a preverbal argument and a verb: it is weak in SOV and strong in OSV sentences. In other words, an object and a following verb are more tightly connected than a subject and a following verb. The difference in post-focal compression reflects this difference. The stronger boundary between a subject and a following verb is also reflected in the significantly longer duration that a focused subject has, as compared to a focused object in the same position (548 ms vs. 528 ms, t=3.56).

Second, in case of preverbal focus, the initial argument displays a greater pitch range and higher F0-maximum in the non-canonical OSV than in the canonical SOV order. This may be the consequence of the scrambled word order in case the object is preposed. As discussed in Section 1.2, a preposed constituent is generally interpreted as a topic (Gambhir 1981, Butt and King 1996). It may be the case that the speakers, or some of them, have realized the given object as a topic, which would explain the extra high boundary tone. Nevertheless we refrain from analyzing the prosodic phrasing of the initial given object differently from an initial subject, for instance as a separate intonation phrase. There are two reasons for this. First, the context did not give any indication for the speakers to interpret this constituent as a topic; and second, the effects obtained were rather weak. We point to the fact that boundaries of prosodic phrases are subject to gradience anyway, but do not provide a deeper explanation for this effect of word order.

1.5 Concluding remarks

Based on the Hindi production study, we have proposed that each constituent forms its own prosodic domain, of the size of a prosodic phrase. Non-final p-phrases have a rising pattern (L^*H_P) and the final ones have a falling pattern (H^*L_I). This structure is not changed by focus. Thus, focus does not introduce a different pattern of phrasing; the prosodic phrases are in a strict downstep relationship which cannot be disturbed.

Hindi differs with respect to the expression of focus and its interaction with the downstep pattern from languages like English or German. In these languages, a sequence of downstepped accents is always interrupted by focus, and focus is realized with an upstep or a raising of the high tone on the focused word (e.g. Bartels and Kingston 1994, Baumann et al. 2006, Cooper et al. 1985, Féry 1993, Féry and Kügler to appear). By contrast, in Hindi focus prominence appears to be expressed after the focused item, by means of post-focal compression. In our material, sentences with focus on the preverbal constituent do not reveal any prosodic difference compared to the wide focus sentences. Since the preverbal position is the syntactic default position for focus (Kidwai 2000), prosodic marking of focus in this position might be redundant and therefore remains unrealized.

Given the analysis above, the global downstep pattern is more important than local register changes induced by focus.

Second, although Hindi's use of post-focal compression is similar to other languages that reduce the prominence of given material to enhance the salience of focused material (Cruttenden 2006), a rising pitch accent on content words is required even in post-focal position. This contrasts with other intonational languages, which commonly employ complete deaccentuation of post-focal material.

Third, in contrast to German (and to Bengali, Hayes and Lahiri 1991), pre-focal given elements in Hindi appear to not undergo compression. Given the downstep pattern, it follows that any register compression before a focus would disturb the downstep pattern. If the pitch range of a pre-focal constituent would be compressed, the dissimilative tonal effect of downstep would be blocked, minimizing the prosodic difference between a pre-focal and focal constituent. Downstep seems to be compulsory; we do not find any reduction of the pre-focal pitch register in case of medial focus. It may be that the downstep pattern facilitates sentence comprehension in that it clearly marks constituent boundaries.

It remains to be determined whether the prosodic structure and cues identified here are used by comprehenders to parse sentences more efficiently. Another important open question is whether the constraints identified here are valid for more complex utterances. Answering these questions lies outside the scope of the present study and must be left for future work.

Acknowledgments

We are very grateful to Professor Ramakant Agnihotri for logistical support in running the experiments at Delhi University; without his assistance, none of the work reported here would be possible. Dinah Rottmann and Verena Thiessen assisted with processing the raw data. The data analysis was carried out using the statistical computing language R (R Development Core Team 2006) and the R packages lme4 (Bates and Sarkar 2007). Furthermore, we thank two anonymous reviewers for helpful comments and suggestions. The research presented here was funded by the Deutsche Forschungsgemeinschaft (DFG), through the project *Prosody in Parsing* (DFG Schwerpunktprogramm 1234 (2006-2009), FE 292/5-1), Principal Investigators Caroline Féry, Frank Kügler, and Shravan Vasishth.

References

Allerton, David J. 1978. The notion of "givenness" and its relations to presupposition and to theme. *Lingua Amsterdam* 44(2-3):133–168.

Bartels, C. and J. Kingston. 1994. Salient pitch cues in the perception of contrastive focus. In P. Bosch and R. van der Sandt, eds., *Focus and natural language processing, vol 1: Intonation and syntax*, pages 11–28. IBM working papers of the Institute for Logic and Linguistics.

Bates, Douglas and Deepayan Sarkar. 2007. *lme4: Linear mixed-effects models using S4 classes*. R package version 0.9975-11.

Baumann, Stefan, Martine Grice, and Susanne Steindamm. 2006. Prosodic marking of focus domains categorical or gradient? In *Proceedings of Speech Prosody, Dresden, Germany*, pages 301–304.

Boersma, Paul and David Weenink. 2005. Praat: doing phonetics by computer (Version 4.5).

Butt, Miriam and Tracy H. King. 1996. Structural Topic and Focus without Movement. In M. Butt and T. H. King, eds., *The First LFG Conference*. CSLI Publications.

Cooper, William E., Stephen J. Eady, and Pamela R. Mueller. 1985. Acoustical Aspects of Contrastive Stress in Question-answer Contexts. *Journal of the Acoustical Society of America* 77:2142–2156.

Cruttenden, Alan. 2006. The deaccenting of given information: a cognitive universal? In G. Bernini and M. L. Schwartz, eds., *The pragmatic organization of discourse in the languages of Europe*, pages 311–356. The Hague: Mouton.

Dyrud, Lars O. 2001. *Hindi-Urdu: Stress Accent or Non-Stress Accent?*. Ph.D. thesis, University of North Dakota.

Féry, C. 1993. *German Intonational Patterns*. Max Niemeyer Verlag.

Féry, Caroline and Frank Kügler. to appear. Pitch accent scaling in German. *Journal of Phonetics* .

Gambhir, Vijay. 1981. *Syntactic restrictions and discourse functions of word order in standard Hindi*. Ph.D. thesis, University of Pennsylvania, Philadelphia.

Gelman, Andrew and Jennifer Hill. 2007. *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

Genzel, Susanne. 2007. Phonetische Realisierung von Fokus auf Adjektiven im Hindi. Unpublished MA thesis, Potsdam University.

Gussenhoven, Carlos. 2008. Notions and subnotions of information structure. *Acta Linguistica Hungarica* 55:231ff.

Harnsberger, James D. 1994. Towards an intonational phonology of Hindi. Unpublished manuscript.

Harnsberger, James D. 1999. The role of metrical structure in Hindi intonation. In *South Asian Language Analysis Roundtable, University of Illinois, Urbana-Champaign*.

Harnsberger, James D. and Jasmeet Judge. 1996. Pitch range and focus in Hindi. *The Journal of the Acoustical Society of America* 99:2493.

Hayes, Bruce. 1995. *Metrical Stress Theory. Principles and Case Studies*. Chicago: University of Chicago Press.

Hayes, Bruce and Aditi Lahiri. 1991. Bengali intonational phonology. *Natural Language and Linguistic Theory* 9(1):47–96.

Hussain, Sarmad. 1997. *Phonetic Correlates of Lexical Stress in Urdu*. Northwestern University.

Ito, Junko and Armin Mester. 2007. Categories and projections in prosodic structure. Paper presented at Old World Conference in Phonology-4, Rhodes, Greece.

Jacobs, Joachim. 2001. The dimensions of topic–comment. *Linguistics* 39(4):641–681.

Jun, Sun-Ah. 2005. Prosodic Typology. In S.-A. Jun, ed., *Prosodic Typology. The Phonology of Intonation and Phrasing*, pages 430–458. Oxford: Oxford University Press.

Kachru, Yamuna. 1987. Hindi-Urdu. In B. Comrie, ed., *The World's Major Languages*, pages 470–489. London: Routledge.

Keane, Elinor. 2007a. Distribution and Alignment of F0 Contours in Tamil. In *Proceedings of the 16th International Congress of Phonetic Sciences, Saarbruecken, Germany*.

Keane, Elinor. 2007b. Investigating the Intonational Phonology of Tamil. Poster presented at satellite meeting of ICPhS, Saarbruecken, Germany, August 5, 2007.

Khan, Sameer. 2007. Phrasing and Focus in Bengali. Poster presented at satellite meeting of ICPhS, Saarbruecken, Germany, August 5, 2007.

Kidwai, Ayesha. 2000. *XP-Adjunction in Universal Grammar: Scrambling and Binding in Hindi-Urdu*. New York: OUP.

Kuno, Susumo. 1981. The syntax of comparative clauses. Proceedings of CLS 17.

Ladd, D. Robert. 1996. *Intonational Phonology*. Cambridge: CUP.

Lambrecht, Knud. 1994. *Information structure and sentence form*. Cambridge Univ. Press.

Moore, Robert R. 1965. *A study of Hindi intonation*. Ph.D. thesis, University of Michigan.

Nair, Rami. 1999. Syllables and Word Edges. Northwestern University. Unpublished Dissertation.

Nair, Rami. 2001. Acoustic correlates of lexical stress in Hindi. In A. Abbi, R. Gupta, and A. Kidwai, eds., *Linguistic Structure and Language Dynamics in South Asia - papers from the proceedings of SALA XVIII roundtable*.

Ohala, Manjari. 1986. A search for the phonetic correlates of hindi stress. In B. Krishnamurti, ed., *South Asian Languages: structure, convergence, and diglossia*, pages 81–92. Delhi: Motilal Banarsidass.

Pierrehumbert, Janet. 1980. *The phonology and phonetics of English intonation*. Ph.D. thesis, MIT. Distributed by Indiana University Linguistics Club, Bloomington.

Pinheiro, José C. and Douglas M. Bates. 2000. *Mixed-Effects Models in S and S-PLUS*. New York: Springer-Verlag.

R Development Core Team. 2006. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Reinhart, Tanja. 1981. Pragmatics and linguistics: An analysis of sentence topics. *Philosophica* 27:53–94.

Rooth, Mats. 1985. *Association with Focus*. Ph.D. thesis, University of Massachusetts, Amherst.

Rooth, Mats. 1992. A theory of focus interpretation. *Natural Language Semantics* 1(1):75–116.

Selkirk, Elisabeth. 2007. Contrastive Focus, Givenness and the Unmarked Status of Discourse-New. In C. Féry, G. Fanselow, and M. Krifka, eds., *Interdisciplinary Studies on Information Structure (ISIS)*, vol. 6, pages 125–145.

Sharma, Devyani. 2003. Nominal clitics and constructive morphology in Hindi. In M. Butt and T. King, eds., *Nominals: Inside and Out*, pages 59–84. CSLI publications.

Truckenbrodt, Hubert. 1995. *Phonological phrases: their relation to syntax, focus and prominence*. Ph.D. dissertation, MIT.

1.6 Appendix

1.6.1 Tables accompanying the Results section

	Comparison	F0 max (Hz)			F0 range (Hz)			Duration (ms)		
		Means		t-score	Means		t-score	Means		t-score
First constituent										
SOV	Wide vs Narrow	274	281	4.26*	58	68	4.97*	556	567	2.62*
	Wide vs Given	274	274	0.15	58	59	<1	556	558	<1
	Given vs Narrow	274	281	4.15*	59	68	4.23*	558	567	2.05*
OSV	Wide vs Narrow	276	279	1.29	67	69	<1	558	559	<1
	Wide vs Given	276	279	1.56	67	68	1.01	558	551	-1.09
	Given vs Narrow	279	279	-0.09	68	69	<1	551	559	1.85
Second constituent										
SOV	Wide vs Narrow	247	250	1.49	44	47	1.32	526	528	<1
	Wide vs Given	247	230	-9.06*	44	30	-9.94*	526	504	-6.24*
	Given vs Narrow	230	250	9.97*	30	47	10.85*	504	528	6.77*
OSV	Wide vs Narrow	249	250	<1	42	48	2.34*	543	548	1.64
	Wide vs Given	249	237	-8.23*	42	32	-6.7*	543	526	-3.62*
	Given vs Narrow	237	250	8.65*	32	48	8.26*	526	548	5.65*

TABLE 2 F0-maximum, F0-range, and duration on the first and second constituent for SOV and OSV order in different focus conditions as well as their statistical comparisons by means of t-tests; absolute t-values above 2 are significant and are marked with an asterisk.

First constituent									
	F0 max			F0 range			Duration		
Focus	SOV	OSV	t-score	SOV	OSV	t-score	SOV	OSV	t-score
Wide	274	276	1.5	58	67	4.59*	556	558	<1
Narrow	281	279	<1	68	69	<1	567	559	-1.37
Given	274	279	3.44*	59	68	4.77*	558	551	-1.18
Second constituent									
Wide	247	249	1.31	44	42	1.23	526	543	2.51*
Narrow	250	250	<1	47	48	<1	528	548	3.56*
Given	230	237	4.55*	30	32	2.65*	504	526	3.76*

TABLE 3 Maximum F0, duration, and F0-range on the first and second constituent for different focus conditions comparing SOV and OSV word order as well as their statistical comparisons by means of t-tests; absolute t-values above 2 are significant.

1.6.2 Stimuli

Stressed syllables are capitalized.

Pattern-1 (2 syllabic Subject, 3 syllabic Object)

(1) GRAAhak ne daVAAii ko khariidaa
(The) customer bought the medicine

(2) BAAlak ne suRAAhii ko chhupaayaa
(The) kid hid the jar

(3) GAAyak ne darVAAje ko dhakelaa
(The) singer pushed the door

(4) NAUkar ne kaTOre ko hataayaa
(The) servant took away the bowl

(5) AADmii ne gaVAAhii ko sudhaaraa
(The) man corrected the statement

(6) MAAlik ne kiRAAye ko badhaayaa
(The) landlord increased the rent

(7) BRAAHman ne cheTAAVnii ko sunaayaa
(The) brahman announced the warning

(8) SAANsad ne jaanKAArii ko failaayaa
(The) parliamentarian spread the awareness

(9) CHHAAtra ne kamPYUter ko ghumaayaa
(The) student turned the computer

Pattern-2 (3syllabic Subject, 2syllabic Object)

(10) maNUshya ne CHAAdar ko jalaayaa
(The) man burnt the bedcover

(11) saVAArii ne JOOte ko utaaraa
(The) passenger took off the shoe

(12) shiKAArii ne PAUdhe ko ukhaadaa
(The) hunter uprooted the plant

(13) khiLAAdii ne GHOde ko bhagaayaa
(The) sportsman made the horse run (faster)

(14) kanDAKtar ne GAAdii ko rukaayaa
(The) conductor stopped the vehicle

(15) shaRAAbii ne BOtal ko bajaayaa
(The) drunkard made sound with the bottle

(16) maiKEnik ne TAxi ko chalaayaa
(The) mechanic drove the taxi

(17) adHYAApak ne MOORti ko banaayaa
(The) teacher made the sculpture

(18) adHYAKsha ne PYAAle ko uthaaya
(The) chairperson picked up the glass