The prosody of embedded coordinations in German and Hindi

Caroline Féry, Gerrit Kentner

Institut für Kognitive Linguistik, Goethe-Universität Frankfurt am Main, Germany
caroline.fery@googlemail.com, kentner@lingua.uni-frankfurt.de

Abstract

This paper reports a cross-linguistic investigation into the syntax-prosody mapping in German and Hindi. For both languages, comparable speech production experiments were carried out, using ambiguous coordination structures with four names in different syntactic conditions. Pitch contours and duration values were compared at syntactic boundaries of the different coordination structures within and across the two languages. The results show that, in German, the prosodic parameters transparently indicate the syntactic constituent structure while in Hindi, no reliable effect of syntactic structure on prosodic realization could be found. We propose two principles to account for the German results and, at the same time, question the universality of syntax-prosody mapping constraints. The cross-linguistic differences are discussed against the background of the respective intonation systems of the two languages.

Index Terms: syntax-prosody interface, coordinations, speech production, German, Hindi

1. Theoretical background

German is an intonation language, in which pitch accents and boundary tones are varied freely to express pragmatic, semantic and scopal meanings. It has been shown that the rich intonational system of intonation languages is also used to indicate and parse the syntactic structure of utterances (e.g. [1], [2], [3]). Hindi, by contrast, is a 'phrase language' ([4], [5]), in which the melody of sentences is primarily determined by rather invariant phrasal tones. It is thus much more rigid in its use of intonation than German, and this rigidity may hinder the prosody to express the syntactic structure in a transparent way. In order to examine and compare the syntax-prosody mapping in German and Hindi, ambiguous coordination structures were used as a test.

Coordinated names, like *Anna and Bill or Mary*, form an ambiguous structure, in the same way as an arithmetic procedure like *3 minus 2 plus 1*, which can be resolved as 2 or as 0, depending on the order of the operations. Researchers have examined how different groupings of coordinated names or numbers are realized prosodically (as for instance [6], [7], [8] for English, [9] for Hungarian). All authors focus on differences in duration at conjunct boundaries and find a strong dependency between the duration of constituents and their place in the coordination structure.

Wagner [8] compared coordinations with different syntactic depths of embedding such as (1a-c).

(1)

a) Morgan and Joey and Norman and Ronny
b) Morgan or Joey or (Norman and Ronny)
c) ((Morgan or Joey) and Norman) or Ronny

To account for the phrasing differences between the different coordination structures, Wagner [8] proposes the Scopally determined Boundary Rank (SBR) algorithm. The SBR (2) posits that the level of syntactic embedding determines the prosodic boundary rank.

(2) Scopally determined Boundary Rank
If Boundary Rank at a given level of embedding is n, the rank of the boundaries between constituents of the next higher level is n+1.

Although the SBR algorithm successfully predicts effects of syntactic embedding on prosodic realization, it cannot easily account for the finding that boundaries of complex constituents and simplex elements at the same level of embedding differ in strength.

This difference, however, is predicted by Watson and Gibson's [10] LRB algorithm which states that the likelihood of a prosodic boundary increases with the size and complexity of the surrounding constituents. Unlike Wagner's SBR, however, the LRB cannot accommodate effects of constituents that are non-adjacent to the prosodic boundary under examination. Also, the LRB cannot explain the different influences concerning the size of preceding versus upcoming constituents on prosodic boundary strength that were confirmed by Watson and Gibson [10] themselves and also by [11] for German.

Two new principles that account for the various effects of syntax on the prosodic realization of ambiguous coordination structures are proposed below. The principles share features of the SBR and the LRB but avoid shortcomings of these algorithms. First, *Proximity* operates on syntactic constituency, reflecting syntactic boundaries in prosodic structure as expressed by pitch and duration. It requests that adjacent elements grouped together into one syntactic constituent should be realized in close prosodic proximity. A corollary of Proximity is the opposite effect: adjacent elements not grouped together into one constituent should be realized with prosodic distance. This effect is formulated in (1b) as Anti-Proximity.

(3) Proximity

a) An element is realized in close prosodic proximity to a following element in the same constituent.
b) (Anti-Proximity): An element is realized in prosodic distance to a following element in a separate constituent.

Proximity between two elements is achieved by shortening the duration of the left element and flattening its F0 contour, thereby weakening a group internal boundary. Likewise, Anti-Proximity is achieved by group-final lengthening and stronger pitch excursion. Note that Proximity and Anti-Proximity involve a directional asymmetry as shortening only affects the left element of a grouped constituent while lengthening affects the right element of groupings.

The second principle, *Similarity*, is sensitive to the depth of syntactic embedding. Its goal is prosodic balance:

constituents at the same level of embedding resemble each other in prosodic structure.

(4) Similarity
Constituents at the same level of embedding have a similar prosodic realization, irrespective of the constituent's inherent complexity.

Similarity predicts prosodic adjustment of simplex elements as compared to complex constituents at the same level of embedding. More specifically, simplex elements are lengthened to approximate the duration of the complex constituent. This also holds for simplex elements that are non-adjacent to complex constituents.

2. Experiments

Two speech production experiments were set up to test the predictions of Proximity and Similarity in German and Hindi.

2.1. German

2.1.1. Material and Method

The material consisted in coordination structures with four names each, organized in six different syntactic conditions, as exemplified in (5a-f). The coordination *und* 'and' was used within groupings and the coordination *oder* 'or' between groupings. Four lexically unique target sentences were created for each of the six conditions.

(5)
- a) Suse oder Nino oder Mila oder Anna
- b) Suse oder Nino oder (Mila und Anna)
- c) (Suse und Nino) oder Mila oder Anna
- d) Suse oder (Nino und (Mila und Anna))
- e) ((Suse und Nino) und Mila) oder Anna
- f) (Suse und Nino) oder (Mila und Anna)

All names are trochaic disyllables with as many sonorant segments as possible.

For each item, a context question, spoken by a female native speaker of German, had been previously recorded. The contexts were presented together with a target sentence both visually on screen and aurally over headphones. To emphasize the structure of the target sentence, it was displayed with parentheses, as in the above example. The items were presented on a 15'' computer screen. Participants were asked to read and listen to the context and then read aloud the target sentence as an answer to the question. In case of hesitations or slips of the tongue, participants were asked to repeat the sentence. 21 female participants, all native German speakers from the Berlin area (North Germany), read out the complete set of target sentences (n=24) interspersed with numerous fillers. Recordings took place in a sound-proof chamber equipped with an AT4033a audio-technica studio microphone, using the a C-Media Wave soundcard at a sampling rate of 44.1 kHz with 16 bit resolution.

The full set of 504 target sentences was hand-annotated by two phonetically trained students and subjected to phonetic analysis using Praat software [12]. Durations of each name plus the following pause were measured. Prior to F0 analysis, pitch was smoothed (frequency band 10 Hz) to reduce microprosodic perturbations. The maximum F0 in the second half of the names was measured. F0 values were normalized taking the utterance wide mean F0 as the normalizing quotient. For all conditions, time-normalized pitch contours were created by dividing the name into five equal-sized intervals and interpolating the mean F0 (in Hz) of these intervals.

2.1.2. Results and Discussion

The structure without embedding in (5a) is taken as the baseline against which the other conditions are compared. The results are depicted in Figure 1 and 2.

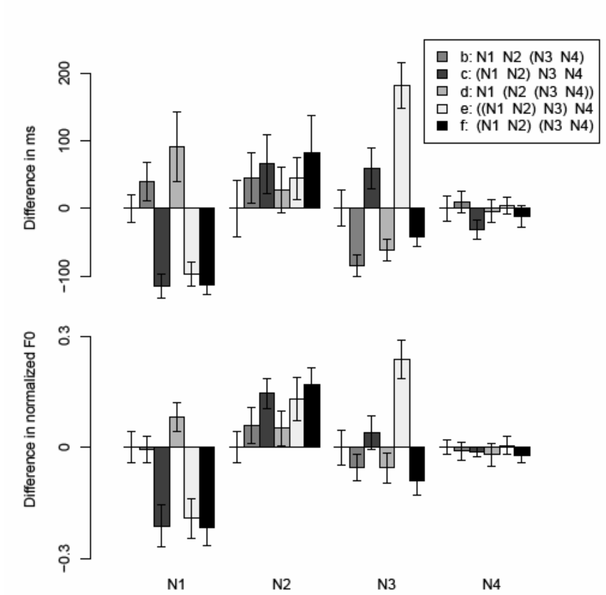

Figure 1: *Durational (upper panel) and F0 differences (lower panel) between baseline (0) and the other conditions on the four names. Error bars depict 95% confidence intervals.*

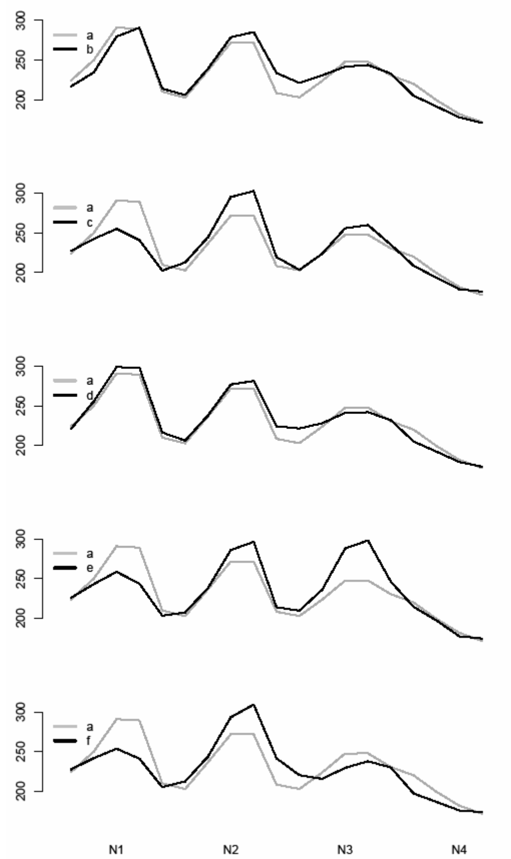

Figure 2: *Time-normalized interpolated F0 tracks (in Hz) of baseline (grey) plotted against other conditions (black).*

The baseline condition follows a regular downstep pattern (cf. Figure 2). As predicted by Proximity, left elements of groupings show less pitch excursion and are shortened to reduce the distance to a following element in the same constituent (conditions c, e, f on N1; b, d, f on N3). Likewise, as predicted by Anti-Proximity, elements preceding a constituent boundary are lengthened and show pitch upstep (b, c, e, f on N2; e on N3). The second principle, Similarity, predicts lengthening of simplex elements if complex constituents appear at the same level of embedding. This prediction is also borne out (b, d on N1, b on N2, c on N3). Note that, in condition (d), Proximity and Similarity make opposite predictions for N2. According to Similarity, it should be lengthened to approximate a complex constituent which appears at the same level of embedding. At the same time, Proximity demands shortening of N2 in this condition, as it is a left element of a grouping. As a result, N2 shows no difference to baseline in this case, i.e. the two principles cancel each other out. On the other hand, if the principles agree in their predictions, the prosodic effect is strengthened: Compare N1 in conditions (b) and (d): While in (b), only Similarity requires lengthening of N1, both Similarity and Anti-Proximity call for the lengthening in (d). As a consequence, N1 is longer in (d) compared to (b).

While Wagner's SBR and Watson and Gibson's LRB can account for some of the effects found here, they fail to predict prosodic effects in critical conditions. The LRB, for example, cannot explain the longer duration of N1 in condition (b) that is predicted by Similarity. The SBR fails to predict the shortening of N1 in conditions (c), (e) and (f), or of N3 in (b), (d), (f) which is straightforwardly derived from Proximity.

2.2. Hindi

In Hindi, as well, four lexically unique coordination structures with usual names were devised. A difference from German was unavoidable: the coordinated names were always followed by the postposition *ke saath* 'with', as the context of the target sentence required a postpositional phrase. The sentences appear in the same six syntactic conditions as in German (see (4)); the coordinations *aur* 'and' and *yaa* 'or' were used within and outside of the groupings respectively. An example is given in (5).

(6) (viral aur vaaman) yaa (yaman aur yogi) ke saath
Viral and Vaaman or Yaman and Yogi with
'With Viral and Vaaman or Yaman and Yogi'

Disambiguating contexts, spoken by a male native speaker of Hindi, had been previously recorded in a speech recording lab in the University of Potsdam. The experiment was carried out in a quiet room at the University of Delhi. Devanagari script was used for contexts and target sentences. Each item was presented as follows. Participants saw a context question and its answer on a 15'' laptop screen and simultaneously heard the question over headphones. Again, the target sentences were presented with parentheses. Participants were instructed to read out the target sentence as a response to the question they heard. In case of hesitation or slips of the tongue, they were asked to repeat the answer. The participants' answers were recorded on a DAT tape recorder using a SM10A head set microphone. 20 female native speakers of Hindi from the Delhi area participated in the experiment.

The full set of 480 target sentences was hand annotated by three phonetically trained students at Potsdam University and subjected to phonetic analysis.

2.2.1. *Results and Discussion*

The results of the Hindi speech production experiment are depicted in Figures 3 and 4.

No reliable differences between the conditions were found with respect to F0 (cf. lower panel of Figure 3). Quite to the contrary, the F0-patterns of the six conditions show a striking similarity with rising pitch on each name irrespective of its syntactic status in the coordination structure (see Figure 4). Also, the durational values lack the differentiation that was found in the German experiment (see upper panel of Figure 3). The only pattern that shows a significant difference is condition (e), in which N2 and N3 are lengthened compared to baseline. We cannot exclude that this effect reflects general processing difficulties with this condition rather than reflection of syntactic structure. Given the lack of further significant effects, neither the predictions derived from the principles Proximity and Similarity nor the SBR or LRB can explain the Hindi prosody which seems to be largely insensitive to the structural differences.

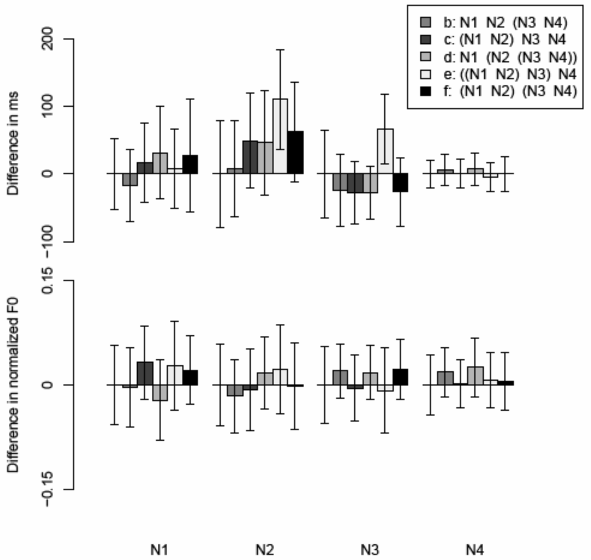

Figure 3: *Durational (upper panel) and F0 differences (lower panel) between baseline (0) and other conditions on the four names. Error bars depict 95% confidence intervals.*

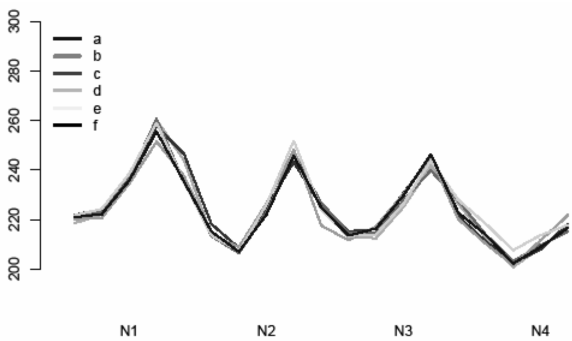

Figure 2: *Time-normalized and interpolated F0 tracks (in Hz) of the six conditions.*

3. Conclusions

The results for German show that prosodic structure reflects syntactic grouping and embedding in a precise way, both for duration and for pitch. Proximity and Similarity account for the prosodic structure emerging from the syntactic structure. The first principle, Proximity, accounts for the lower pitch and shorter duration observed on the left member of groupings. Its corollary, Anti-Proximity, has the opposite effect and strengthens the boundary between two constituents by lengthening the right member of groupings. The second principle, Similarity, accounts for the observation that simplex elements in an expression containing groupings have increased duration and higher pitch to achieve similar prosody to complex elements at the same level of syntactic embedding. As a result, German uses prosody in a sensitive way, interpreting syntactic structure with exactitude. This property of German correlates with its general intonational system. German, as an intonation language, is able to change pitch accents and boundary tones in a variety of ways to express pragmatic meanings. Pitch scaling is a fine-grained device which supports this use of intonation, as shown in [13]. Our experiment demonstrates that prosody as a whole supports the rendition of syntactic structure.

Hindi, by contrast, shows a surprising lack of correlation between syntactic structure and prosody. Neither Proximity nor Similarity were supported by the Hindi data. Also, neither SBR nor LRB are able to make the correct predictions for Hindi. These results can only be understood when Hindi intonation is considered as a whole. Hindi is a phrase language, according to the sentence-based typology of intonational systems. The melody of sentences arise primarily because of the distribution of phrasal tones which are, as the name of these tones indicates, assigned at the level of the phrases, and not because of pitch accents. High tones in a syntactically simple Hindi sentence are always in a downstep relation, and are only marginally sensitive to information structure (see [5]).

The results of this double experiment have several implications for theoretical considerations about the role of prosody in reflecting syntax.[1] Clearly it seems that the prosodic reflection of syntactic and or semantic structure is not a universal property as has been suggested by [14] for the parameter of duration. Instead, whether and to what extent syntactic relations are reflected in prosodic structure crucially depends on the language. The apparent lack of prosodic marking of syntactic grouping in Hindi might be the cost for the clear and consistent marking of prosodic phrases, which might not adjust to pragmatic conditions in the same way as in German. More research is needed to confirm this hypothesis. The difference between the two langauges is important for understanding the role that prosody plays in language comprehension. Languages may differ in this dimension much more than assumed until now. The research on the role of prosody on speech processing has largely concentrated on intonation languages, which do use pitch changes and pitch scaling for the communication of syntax and semantics, and has often ignored other types of languages, such as those which rely more on phrasing for this parameter. We hope to have revealed the need for well-designed experiments for elucidating this issue.

4. Acknowledgements

This paper is part of the project Prosody in Parsing, financed by the DFG. Thanks to Shravan Vasishth and Frank Kügler. Many thanks are due to Umesh Patil who prepared the Hindi material and who run the production experiment in Delhi, as well as to Caroline Magister and Verena Thießen who run the experiment for German. We are also grateful to Daniel Quernheim, for technical assistance.

5. References

[1] Frazier, L., Carlson, K., Clifton, C., “Prosodic phrasing is central to language comprehension”, Trends in Cognitive Sciences 10 (6), 244-249, 2006.

[2] Schafer, A., Speer, S., Warren, P., White SD., “Intonational Disambiguation in Sentence Production and Comprehension”, Journal of Psycholinguistic Research 29 (2), 169-182, 2000.

[3] Steinhauer, K., Alter, K., Friederici, A., “Brain potentials indicate immediate use of prosodic cues in natural speech processing”, Nature Neuroscience 2 (2), 191-196, 1999.

[4] Féry, C., “Indian languages as intonational ‘phrase languages’”, in I. Hasnain and S. Chaudhury [Eds] Festschrift to honour Ramakant Agnihotri, 2008.

[5] Patil, U., Kentner, G., Gollrad, A., Kügler, F., Féry, C. and Vasishth, S., “Focus, word order and intonation in Hindi”, Journal of South Asian Linguistics 1, 55-72, 2008.

[6] Lehiste, I., “The many linguistic functions of duration.” Zürich: Symposium Prosody. 1983.

[7] Ladd, D.R.., “Intonational phrasing: the case for recursive prosodic structure”, Phonology Yearbook 3, 311-140, 1986.

[8] Wagner, M., “Prosody and recursion”, Doctoral Dissertation, Massachusetts Institute of Technology, Cambridge, Mass., 2005.

[9] Hunyadi, L., “Grouping, the cognitive basis for recursion in language”, Argumentum 2, 67-114, 2006.

[10] Watson, D. and Gibson, E., “The relationship between intonational phrasing and syntactic structure in language production”, Language and Cognitive Processes 19/6, 713-755, 2004.

[11] Kentner, G., “Length, ordering preferences and intonational phrasing: Evidence from pauses”, Proceedings of Interspeech, 2337-2340, 2007.

[12] Boersma, P. and Weenink, D., “PRAAT - doing phonetics by computer”, Online resource: http://www.praat.org, 1997-2009.

[13] Féry, C. and Kügler, F., “Pitch accent scaling on given, new, and focused constituents in German”, Journal of Phonetics 36(4), 680-703, 2008.

[14] Yang, Li-chiung. “Duration and pauses as boundary markers in speech: a cross-linguistic study”, Proceedings of Interspeech, 458-461, 2007.

[1] Results of perception experiments will be published independently.