5th International Conference on Spoken Language Processing (ICSLP 98)
Sydney, Australia
November 30 - December 4, 1998

De-accentuation: Linguistic Environments and Prosodic Realizations

Alter, K., Steinhauer, K., Friederici A.D.

Max-Planck-Institute of Cognitive Neuroscience Leipzig, Germany

{alter; steinhau; angelafr}@cns.mpg.de

ABSTRACT

In this paper we present a preliminary speech production study concerning the prosodic realization of the syntactic and information structure in German. Firstly, we made predicitions for the relative prominence and their assignment with tonal patterns. Secondly, exhaustive acoustic analysis were used to test the expectations. The data of a production experiment with seven non-instructed normal subjects were analyzed and then compared with the data of one patient with prosodic disorders.

1. Introduction

The prosodic realization, e.g., accent positions, accent types and prosodic phrasing in an utterance can be modified by several linguistic parameters such as syntactic structure and information structure, e.g., focus-background structure (FBS), and topic-comment structure (TCS). According to Junghanns (1997) we assume focus [F] and topic [TOP] to be syntactic features underlying the informational structuring. Particularly the [F]-feature requires a special realization of the prosodic surface structure. In spoken German, elements assignated with [F] are prosodically more salient than other elements in a sentence. Prosodic salience should be typically realized through assignment of a pitch accent (Féry 1993) to the focused constituent. Modifications of FBS may result in a de-accentuation of the elements normally carrying (neutral 'default') accents. The question of how topics are prosodically realized will not be discussed here.

Furthermore, different speakers seem to realize accents by different prosodic parameters (e.g., pitch, loudness, or duration variations). Below we will show that not only the pitch parameter is used to mark prosodically focused constituents.

One goal of the study was to lay the foundations of a data base for statistical analyses taking into account the variability of prosodic realizations across different subjects. Preliminary data in fact confirm a high variability among subjects for specific prosodic parameters such as pitch. Other prosodic parameters, namely durational measures, however, display relatively stable patterns in normal subjects, but seem to be impaired in the patient.

Taking these inter-individual differences into account, it is possible to examine the linguistic influences on prosody.

2. Syntax-Prosody Mapping

A linguistic theory of syntax-prosody mapping must consider the underlying syntactic structure in termes of its hierarchical organization, especially if the syntax of a given language allows different directions of branching as it is the case in an Object-Verb-language (OV) such as German. In German, the *syntactic* OV-parameter means that in structures with verb-final word order, i.e., in most subordinate clauses, the verb takes its argument from its left.

For reasons of explanatory adequacy, we make use of theories which consider the information structure. In Jacobs (1993), for both the so called 'normally intonated' sentences, e.g., widely focused sentences, and sentences containing narrowly focused constituents, accent positions are predictable by termes of *integration*. Accent positions in termes of their relative prominence (e.g., the weight of accents distributed over a syntactic structure) is thus calculable if the position of [F] is fixed.

In a first step of syntax-prosody mapping, the syntactic structure and the position of [F] have to be determined. Let us assume two basic syntactic structures reflecting the superficially linear VO-order (A1) and OV-order (B1):

(A1) Peter [verspricht]verb_1 Anna zu [ARBEITEN]verb_2

Peter promises Anna to WORK

(B1) Peter [verspricht]verb_1 ANNA zu [entlasten]verb_2

Peter promises to support ANNA

Sentences (A1) and (B1) are structurally locally different as to whether NP2 'Anna' is the object of verb_1 (as in A1) or of verb_2 (as in B1). If A1/B1 are focused widely, i.e., if [F] is located at a very high branching node in the syntactic structure (cf. Figures 1), the 'normal' default accentuation has to be applied: in A1 the second verb 'arbeiten' is accented.

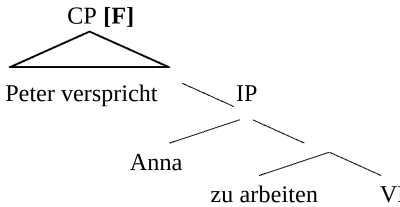

Figure 1: The syntactic structure of the first clause of (A1) with [F] at the highest node, e.g., with wide focus. Note that the NP 'Anna' is the argument of verb_1.

This syntactic structure causes a deeper embedding of the NP2 'Anna' in (B1). In B1, the accented category is the object of verb_2 - namely 'Anna' (marked by small CAPITALS in A1/B1). The corresponding main accent positions can be directly derived from the syntactic structure when applying the algorithm proposed by Jacobs (1993). This can be illustrated via bracketed metrical grids assigning the highest column of beats (‘*’) to the designated constituents (cf. Figure 2). The highest column indicates the position of the main accent.

Peter verspricht....

```
               *              *
               *        )     (*                  )
   *           *        ))    (*        *         ))
..Anna zu   arbeiten]Iph1   [Anna zu   entlasten]Iph2
```

Figure 2: The bracketed metrical grids for the relevant parts of (A1/B1). In (A1), the highest column of beats is assigned to the lexically stressed syllable of verb_2 ‘arbeiten’ whereas in (B1), the highest column is assigned to the lexically stressed syllable of NP2 ‘Anna’.

One advantage of using bracketed metrical grids is the possibility to translate syntactic constituents directly into prosodic domains. Brackets in the metrical grid mark boundaries of Intonational Phrases (IPh). (A1/B1) are thus prosodically restructured in the following way:

(A1) [IPh1 Peter verspricht Anna zu ARBEITEN]

Peter promises Anna to work

(B1) [IPh1 Peter verspricht] [IPh2 ANNA zu entlasten]

Peter promises to support Anna

The relevant part of (A1) consists of only one Iph, (B1) on the other hand is restructured by two Iph.

The location of [F] can be manipulated by a question-answer test. If a question focuses on the whole answer, we have to apply the default ‘normal’ prosody as described in (A1/B1). The question involved in these cases was ‘What happens?’. In the second case (A2/B2) described in this paper, we changed the location of [F] by asking: ‘TO WHOM does Peter promise to do verb_2?’ in (A), and ‘WHO does Peter promise to verb_2?’ in (B). There, the verb ‘promise’ as well as ‘Peter’ and ‘verb_2’ constitue the background information. The *new*, and thus focused constituent consists only of the NP2 ‘Anna’.

(A2) Peter [verspricht]verb_1 ANNA zu [arbeiten]verb_2

Peter promises ANNA to work

(B2) Peter [verspricht]verb_1 ANNA zu [entlasten]verb_2

Peter promises to support ANNA

As the answer to this question-context focuses on the second NP (‘Anna’), a shift of the accent position should only be observed for the default (A1) but not in (B1) as indicated in (A2/B2). When asking for ‘Anna’, [F] has to be assigned to this NP2.

Superficially, (A2), (B1) and (B2) should carry the same prominence pattern. This is due to the assignment of [F] to ‘Anna’, as it is shown in Figure 3 for (A2):

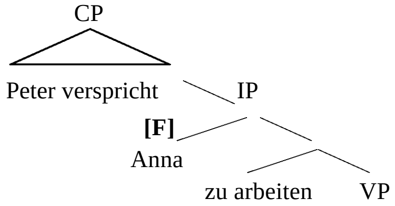

Figure 3: The position of [F] if in (A2) the object NP ‘Anna’ is narrowly focused by applying a question-answer test.

The feature [F] is directly associated to the NP ‘Anna’ in (A2) and the default accentuation is overridden. The verb_2 becomes de-accented and the main accent is assigned to ‘Anna’ as shown in Figure 4 with the appropriate metrical grid:

Peter verspricht....

```
    *
   (*    )              )
   (*    )       *      )
..[Anna]   zu  arbeiten]
```

Figure 4: The bracketed metrical grid for the relevant parts of (A2). The highest column of beats is assigned to the lexically stressed syllable of ‘Anna’.

If the question-context focuses on the second NP ‘Anna’, a shift of the accent position should only be observed in (A) but not in (B) when compared to wide focus. That is, even with wide focus ‘Anna’ already carries the neutral ‘default’ accent in (B), but not in (A).

3. Association with Tonal Sequences

According to recent tonal sequence models (Reyelt, Grice, Benzmüller, Mayer, and Batliner 1996 for German), the main accent positions derived from the syntactic and information structure as described above serve as anchor points for the association of tonal sequences. In German being generally an intonational language, accents can be assumed to be realized preferrably by tonal/pitch variations.

We refer to the German-ToBI system (Reyelt et al 1996) in order to predict the correct tonal sequences: Concerning the four conditions (A1/2 and B1/2), we assume the main accents to be associated with rising tonal sequences of the type L+H*. Notice that the L+H*-sequence is associated via the metrical grid with the lexically stressed syllable of the verb_2 in (A1), and of ‘Anna’ in all other conditions. As is illustrated above, we do not expect differences between the conditions (B1/A2/B2). Only when a different information structure is involved in (A1), the main accent shifts from verb_2 to the NP2 ‘Anna’.

We further expect the boundaries of the Iph to be marked by boundary tones. Sentence internal boundaries are marked by high boundary tones (H%). In the condition (A1), only one prosodic boundary appears after verb_2 whereas in the condition (B1), as well as in both other cases of narrow focus (i.e., (A2) and (B2)) a boundary is expected after the second NP ‘Anna’ as well in the cases of narrow focus on this NP both in condition (A2) and (B2).

According to the findings of Féry (1993) for German focused constituents to be marked obligatorily at their right boundary. We therefore predicted a boundary after NP2 in (A2/B2). In the case of condition (B1) we predicted a boundary at the right of verb_1 as it is the 'normal' default phrasing derived directly from syntactic constituency. Concerning the focused conditions (A2/B2), we leave open the question if the focused NP 'Anna' is also boundary-marked at its left.

In the following presentation, lexically stressed syllables are marked by small CAPITALS, and the boundaries of Iph with the squared brackets:

			L+H*
(A1) [Peter verspricht	Anna	zu	ARbeiten]
	L+H*		
(B1) [Peter verspricht]	[ANna	zu entlasten]	
	L+H*		
(A2) [Peter verspricht	?[ANna]	zu	arbeiten]
	L+H*		
(B2) [Peter verspricht]	?[ANna]	zu	entlasten]

To summarize the predictions of syntax-prosody mapping, we have to differentiate between three predictable prosodic parameters namely (1) accent position in termes of relative metrical prominence, (2) accent type in termes of the association of the latter with tonal sequences and (3) boundary marking in termes of the tonal realization of the edges of Iph.

We expect differences for the conditions (A1) vs. (B1) for the default syntax-prosody mapping, for (A1) vs. (A2) in the case of de-accentuation. We predict only slight differences for the conditions (A2) vs. (B1) vs. (B2).

4. Method and Material

Differences of syntactic structure combined with differences of information structure were tested comparing pairs of sentences such as described above. We only present data concerning the widely focused (A1/B1) and the narrowly focused NP2 (A2/B2) variations.

We analyzed speech production data of 8 native German speakers (7 naive normal subjects, and 1 patient with prosodic disorders). The sentences were locally ambiguous and had to be produced as answers to questions with either wide focus or narrow focus on the object in a randomized order. All subjects were uninstructed concerning the expected prosodic realizations. The speech corpus also contained filler sentences for the normal subjects.

Speech signals were recorded in a sound proof chamber, and digitized (44,1 kHz/16 Bit sampling rate). All speech signals (i.e., 30 to 96 sentences per subject) were extensively analyzed with respect to word and pause durations and F0 contours (pitch tracking).

5. Acoustic analysis

Two important acoustic parameters were extracted from recorded speech in order to test the predicitions on the prosodic behavior made above, namely pitch (fundamental frequency (F0)) and duration.

5.1 The realization of tonal sequences

In order to test the predictions on the tonal behavior of the four conditions (A1-2/B1-2), the pitch values at 16 different positions and their latencies were extracted. The results of the corresponding statistical analysis are reported here verbally: When comparing (A1) with (B1), the seven normal uninstructed subjects realized a L+H*-sequence as predicted. In (A1) on the verb_2, in (B1) on the second NP, a local rise was produced but in most of the cases, the adjacent constituents - NP2 in (A1) and verb_2 in (B1) are also marked by a L+H*-sequence. The difference between these two local rises was found in the global pitch contour over all constituents in the sentence. In fact, we can detect different strategies to mark the underlying prosodic differences between (A1/B1) via pitch variations: (B1) has an additional boundary tone at verb_1 (n=5), in one case the sentence initial onset is remarkably lowered for (B1).

Concerning the de-accentuation from (A1) to (A2), we find the expected pitch pattern in three of the subjects; two subjects use rather rhythmical changes by producing early peaks in (A2) and thus reduce peak-to-peak distances between the sentence initial rise and the focused one. One subject makes immediately focus differences by a higher sentence onset pitch range (cf. Alter & Pirker 1997 for a similar study) in (A1); and the other subject makes the difference by producing a higher pitch range on NP2 in (A2).

Comparing (B1) with (B2), differences appear in the local accent pattern and boundary marking (see table 1).

The patient with prosodic disorders, however, did not use any of the prosodic strategies described above. The pitch patterns do not reflect differences across conditions and each single constituent seems to be accented and marked by a L+H*-sequence.

5.2 The realization of duration

In order to verify the predicitions about the prosodic structure, we performed duration measurements of the constituent length and the pauses.

The statistical analysis reveals significant differences accross all normal subjects for the different conditions. These differences can be summarized as follows: Comparing (A1) vs. (B1), we expected a boundary before verb_2 in (B1). This boundary in (B1) was realized by a significant prefinal lengthening of verb_1 and a subsequent significant pause (for a perceptual study of this effect, cf. Steinhauer, Alter and Friederici in this volume). The comparison of the lengths of NP2 and verb_2 in (A1/B1) confirms our predicitions: In (A1) the verb_2 is longer as in (B1) because it is accented, in (B1) the NP2 is longer than in (A1).

For (B1) vs. (B2), weaker effects were expected but the statistical analysis for all normal subjects show that for (B2) with narrow focus on the NP2, the sentence initial constituents NP1 as well as verb_1 are slightly shortened (cf. Alter 1998 for similar effects). No differences were found in the length of the pause after verb-1, furthermore the NP2 is lengthed in (B2).

Concerning the de-accentuation in (A2) compared with (A1), the following picture appears: Again, both the NP1 and the verb_1 are shortened in (A2) vs. (A1). There was no pause insertion before but after the NP2. Additionally, the NP2 is significantly longer in (A2) than in (A1). All three strategies - the shortening of material preceding the focused target NP2, the lengthening of NP2 when focused and the pause insertion after the NP2 are considered to make this focused constituent more salient.

Cond.	parameter	**normal subjects**	**patient**
A1/B1	pitch	higher pitch range on NP2 in B1 and on verb_2 in A1; boundary tones in B1 before NP2; pitch lowering on NP1 in B1	–
	duration	final lengthening of and pause insertion after verb_1 in B1; lengthening of NP2 in B1	–
A1/A2	pitch	early peaks in A2; pitch lowering on NP1 in A2; higher pitch range on NP2 in A2	–
	duration	shortening of NP1 and verb_1 in A2; pause insertion after NP2 in A2	–
B1/B2	pitch	falling H+L* in B2; higher pitch range on NP2 in B2; pitch lowering on NP1 in B2	–
	duration	shortening of NP1 and verb_1 in B2; lengthening of NP2 in B2	–

Table1: The different prosodic strategies used to make differences between the conditions. Notice that the patient fails to make differences.

To summarize, our predictions concerning the prosodic phrasing in (A1/B1)have also been confirmed for (A1/A2). For the narrowly focused conditions (A2/B2), the normal speakers seem to reach as fast as possible the focused target by shortening the sentence initial material. We therefore assume that the local durational patterns (e.g., the local lengthening of accented constituents) as well as the relational durational patterns (e.g., the shortening of target-preceding constituents) are highly reliable cues for prosodic analyses.

Analysing the patient’s data, he seems to equalize the distances between the word onsets and word offsets.

6. Summary and conclusion

We presented parts of a method to test the reliability of linguistic predicitions on the prosodic behavior by analysing the speech data of 7 normal subjects in a question-answer test. This study was conducted in order to understand more about inter-subject variability in speech production when both the syntactic and information structure are highly controlled. The method presented here can be used to detect prosodic disorders in the acoustic signal if the variability of the strategies of normal subjects is known. We detect in fact that the predictions derived from the syntactic and information structure by the metrical grids and tonal sequences are sufficient. However, caution is required concerning the deviating production strategies across subjects. Only if these strategies are known we can conduct more extensive patient studies.

The production strategies observed in our study are the following: (1) they differ in the tonal realization by the use of local pitch rising of accented material, by the introduction of boundary tones, by a global variation of the sentence initial pitch range and by the use of other local tonal patterns.

(2) they differ less in the use of the durational patterns across subjects. For all conditions, relative stable patterns can be found statistically.

9. Acknowledgements

We thank Annett Schirmer, Franziska Kopp, Nicole Giesa and Janina Gatzky for their assistance in analyzing the speech signals. The study is part of the project ‘Temporal segmentation of acoustic inputs: Prosodic and syntactic processing’ supported by the Deutsche Forschungsgemeinschaft (FR 519/17-1).

8. REFERENCES

1. Alter, K., and Pirker, H., „On the specification of sentence initial f0-patterns in German“, *ESCA workshop Intonation: Theory, models and applications*, Athens, 25-28, 1997.

2. Alter, K., „Discovering time structure in German“, *Proc. of Nordic Prosody VII*, Joensuu, 35-48, 1998.

3. Féry, C.,,German Intonational Patterns,“ Linguistische Arbeiten 285, Tübingen, 1993.

4. Jacobs, J., „Integration,“ *Linguistische Arbeiten 306, Reis, M., (ed.), Wortstellung und Informationsstruktur*, Tübingen, 63-116, 1993.

5. Junghanns, U., „Features and Movement,“ Alexiadou, A. et al. (eds.) *ZAS Papers in Linguistics vol. 9*, Zentrum für Allgemeine Sprachwissenschaft, Sprachtypologie und Universalienforschung Berlin, 74-88, 1997.

6. Reyelt, M., Grice, M., Benzmüller, R., Mayer, J., and Batliner, A., „Prosodische Etikettierung des Deutschen mit ToBI“, *Gibbon, D. (ed.): Natural Language Processing and Speech Technology*, Mouton de Gruyter, Berlin, 144-155, 1996.

7. Steinhauer, K., Alter, K., and Friederici, A.D., „Don’t blame it (all) on the pause: Further ERP evidence for a prosody-induced garden-path in running speech,“ *this volume*. 1998