

ON THE SPECIFICATION OF SENTENCE INITIAL F0-PATTERNS IN GERMAN

K. Alter[†] and H. Pirker

Austrian Research Institute for Artificial Intelligence
Schottengasse 3, A-1010 Vienna, Austria
e-mail: hannes@ai.univie.ac.at

[†]Max-Planck-Institute of Cognitive Neuroscience
Inselstr. 22-26, D-04103, Leipzig, Germany
e-mail: alter@cns.mpg.de

ABSTRACT

It is widely accepted that linguistic high level information like information structure (focus-background-division, topicalization) influences accent placement and accent type (rising, falling, hat pattern etc.) in German. Previous research was concentrated on tonal patterns associated with focus and topic ([1],[2]). We will demonstrate that 1) sentence initial tonal variation and 2) syllable duration depend on the location and the type of focus in the sentence. We assume that focussed material forms its own prosodic domain. Both prosodic parameters show a significant variation depending on whether they are associated with prosodic categories inside or outside of focussed domains.*

1. INFORMATION STRUCTURE AND PROSODY

Prosody, i.e. the variation of fundamental frequency (F0) and duration is an important means to convey higher level linguistic information such as Information structure (IS - often also called "discourse structure") and grouping of speech items (i.e. phrasing). Prosodic parameters and their acoustic realizations are of course linked to informational related aspects of utterances ([3]).

A common assumption implicit in the phonologically orientated analysis of prosody is that IS is preferably encoded in German by tonal patterns which are acoustically realized as F0-variation. Recently, a second general basic effect of information structure on the processing of prosody was encountered ([4]).

Durational variation (i.e., segmental length in different positions: the length of prosodic heads, distances between heads of prosodic constituents, lengthening in the vicinity of prosodic boundaries) is one of the possible prosodic cues to distinguish different meanings related to information structure. We suppose, following recent analysis in this field (e.g., [5] for American English), that high linguistic discourse entities are part of the factors which are responsible for the variation in syllable duration.

According to these studies, two main factors have a more or less direct association with IS: F0 *and* duration. We found that varying the information structure e.g., changing a widely focussed declarative sentence into a sentence with a narrow focus on a non-sentence-initial constituent, caused the speaker to alter his strategies for encoding this discourse information prosodically. Both the F0- and the duration cue were altered. This alternation of prosodic parameters can be explained in the terms of articulation ([6],[7]). In order to signal important information to the hearer as fast as possible and to convey to

* This work has been supported by the Austrian Fonds zur Förderung der Wissenschaftlichen Forschung (grant no. P10822-MAT). Financial support for the Austrian Research Institute for Artificial Intelligence is provided by the Austrian Federal Ministry of Science and Transport. We thank Harald Trost and Johannes Matiasek for helpful comments.

communicative agreements, the speaker tends to make more salient the focussed parts of a sentence.¹ To achieve this task, general language specific abstract parameters for an economical use of articulation may be activated. The most simple one is: stress the most important information prosodically.

The question is how salience of the most important part of a sentence, e.g., the focussed one, is prosodically realized. In the next section, we present recent methods to analyze IS and the association of elements of IS with tonal patterns.

1.1 Information structure and tonal patterns

According to ([8],[9]), sentence initial constituents may be analyzed as topic. Syntactic material in sentence initial position does not always take part in focus projection and may form its own informational category, i.e., topic. Topic may be understood as 'what the rest of the sentence is talking about' or as the constituent referring to the previous discourse. Topic belongs to the non-focus part of the sentence (cf. [8] p. 48).

Focus on the other hand has to be interpreted as the part of a question which has to be filled by the speaker in the answer. The background corresponds to the part of the answer which is less important from the speaker's view and is neither topicalized nor focussed. Concerning the association of different accent types in German, topics are marked with a special L+H* topic accent ([8],[3]).

In (1-4), we present two examples from a recorded question-answer-experiment. Many observations show that IS in German is obviously realized through tonal movement. Stress positions (i.e. positions of lexical stress) serve as anchor points for topic- and focus accents. Brackets indicate topicalized and focussed constituents. Constituents which are not in brackets form the background of the sentence. Capitalized syllables are accented.

What happened?

(1) [REgen]topic [bringt HOCHnebelwolken im alpenvorland]wide-focus

What does rain bring in the alpine foothills?

(2)[REgen]topic bringt [HOCHnebelwolken] narrow-focus im alpenvorland
rain brings low stratus clouds in the alpine foothills

In (1), the subject noun phrase 'Regen' may be analyzed as the topic of the sentence and the rest is focussed. The topic is located outside of the focus projection. To produce this sentence interpretation, a contextual information (weather forecast) is introduced before starting the question. The topic 'Regen' may be interpreted as discourse topic.

In (2), the subject noun phrase 'Regen' is again the topic, the object constituent 'Hochnebelwolken' is narrowly focussed, the rest of the sentence is in the background. Sentences (3) and (4) have the same distribution of focus but a different word order:

(3) [im ALpenVORland]topic [bringt REgen HOCHnebelwolken]wide-focus

(4) [im ALpenVORland]topic bringt REgen [HOCHnebelwolken]narrow-focus
in the alpine foothills brings rain low stratus clouds

In (3) and (4), the local adjunct is situated in the sentence initial position and forms the topic. The verb and the subject is part of the focus in (3), the direct object in (4) is narrowly focussed.

1.2 Information structure and F0

¹ The important, e.g., focussed parts of an utterance are often located in the right peripherie of a sentence (the so called 'late focus') following the concept of the 'normal' word order, at least in German.

Sentences (1-4) of different focus were recorded and the F0-contour was extracted². The following diagrams show the F0-contour of these four sentences.

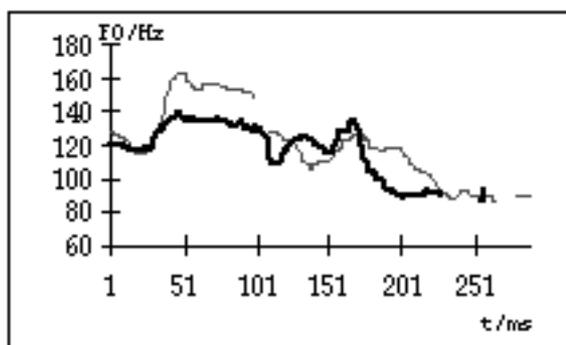


fig. 1: The pitch contours of the sentences (1/2)

The fat line indicates the sentence with the direct object under narrow focus (2/4).

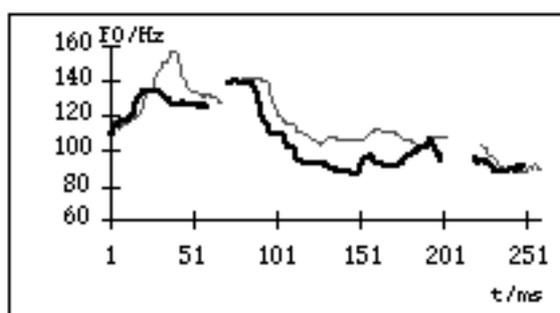


fig. 2: The pitch contours of the sentences (3/4)

1. The F0-behavior of sentence initial parts differs for wide-focus and narrow-focus readings. The initial pitch range in (2) and (4) is smaller than in the widely focussed sentences (1) and (3).

2. Beside this difference in pitch range, both readings have the same initial F0-contour-type, i.e., a rising one.

Our data of the whole corpus confirm that the phonological content of the sentence initial pitch accent is a bitonal sequence of the form L+H*. In the paradigm of tone-sequence-models, the acoustic difference in the pitch range of sentence initial stress cannot be simply reflected in terms of declination. The range between these two readings differs up to 30%. The extended rise in the sentence initial positions seems to be related to the meaning or discursual load of the following constituents. What can be concluded is the fact that in cases of "late" narrow focus, i.e. if there is no narrow focus on the sentence initial constituent, triggers a significant decrease in "early" pitch range.

1.3 . Information structure and duration

Durational phenomena are often related to phenomena of isochrony. The claim of many investigations is that some languages tend to have temporally equivalent distances between lexically stressed syllables. ([6]) discuss durational effects for lexically stressed syllables in nuclear and pre-nuclear accent positions. They argue that stressed syllables in nuclear accent-positions have to be longer because they have to ensure articulatory movements for both F0-variation and prominence relation to stressed syllables in prenuclear positions.

² The corpus is built up on a variety of German spoken in Vienna and contains 54 question-answer-pairs and 10 minutes of spoken monologues from a single speaker. Recording and digitalizations were performed in cooperation with Thomas Keznikl from the Technical University of Vienna.

([10]) demonstrates that in German the differentiation between primary and secondary stress is marked by the duration of syllables.

Our data show that durational effects are related to IS at a very high structural level.

Figure 3 shows the length of all unstressed and stressed syllables of the constituents located before the focussed domain in (1-4), regardless whether they are topics or in the background. We can see that in the widely focussed sentences (1) and (3), both stressed and unstressed syllables of sentence initial non-focussed constituents are remarkably longer than in (2) and (4).

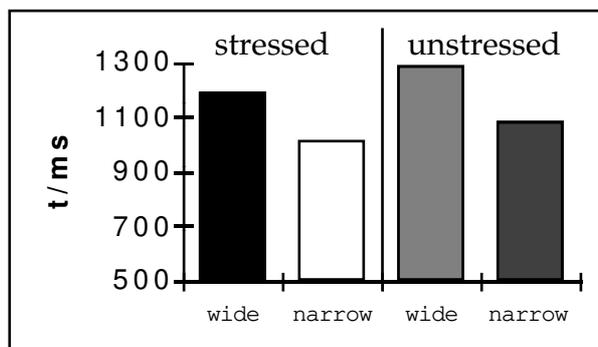


fig. 3: duration of unstressed and stressed syllables of constituents before focus

The durational relation between stressed and unstressed syllables in sentences with wide vs. late narrow focus (1/3 vs. 2/4) seems to be invariant.

To summarize, the duration of both stressed and unstressed syllables is affected by shortening in cases with narrow focus on the object constituents (2/4). We can assume that syllables tend to be shortened in non-focussed positions before focussed constituents.

2. MODELING F0 AND DURATION IN GERMAN

Independent considerations of the human articulatory system determine that the "quality" of articulatory movements is dependent on prosodic information included in phonological structure ([7]). In our interpretation of this framework, targets for articulatory movements have to be determined by high level linguistic information such as focus placement and boundaries of focussed domains. This is what we call major targets. Furthermore, rhythmical relations, (i.e., the durational relation between stressed and unstressed syllables) also have to be fixed. This means that stressed syllables have to be longer than unstressed ones. That requires more articulatory effort for stressed syllables. In figure 4 they are named minor targets. Both types of targets have to be hyperarticulated in the sense of ([7]).

Syllables marked with " ´ " carry lexical stress.

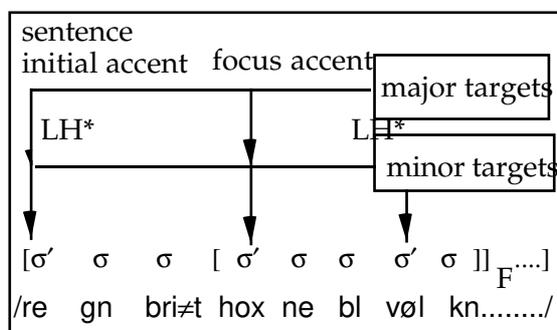


fig. 4: different target positions in a sentence with late narrow focus

The constituents in sentence initial position carries the default LH*-tone. In order to reach the major target, i.e., the narrow-focus accent position as fast as possible, syllables which are not assigned any major information - syllables in pre-focal positions - become shortened. The shortening-rule will then intervene for a discrimination between (1/3) vs. (2/4). In the case of wide focus, no shortening will apply. This rule for durational adaptation is sensitive to two informational entities: the position of the focus feature (late vs. early, i.e., the position of this feature at a lower or higher node in the syntactical structure ([11])) and the type of the focus feature (wide vs. narrow). By this way, it seems possible to reduce the prosodically relevant information: the behavior of prosodic parameters (F0 and duration) is dependent rather on focus than on topic. We could assume that sentence initial constituents is a default LH*-pitch accent assigned to.

The informational categorie 'topic' has thus no prosodic content while 'focus' seems to be more important for the prosodic realization. To 'focus' means in prosodic terms to make a category more salient. This may be performed in different ways. In our corpus we find that hypoarticulation which means less articulatory effort in some prosodic domains, is one of the strategies to make a constituent more salient. In hypoarticulated parts the pitch range is less extended and the the duration of syllables is reduced.

On the prosodic level focussed domains are indicated via feature recognition. The position and the type of the focus feature determine which pitch accent type has to be assigned to the focussed constituent which forms its own prosodic domain. Constituents which are not focussed and are located higher in the syntactical hierarchy are per-se not included in the focussed domain.

On the syllabic level, pitch accents are associated with the most prominent syllables in focussed and pre-focal constituents. Concerning the sentences (2/4), both the lexically stressed syllables of the sentence initial constituents and of the focussed direct objects are associated with a rising LH*-pitch accent. On the phonetic level the durational and F0-parameters are adapted. Pre-focal syllables become reduced, the pre-focal pitch accent decreases.

Hyperarticulation for the salient focussed constituents is then produced through hypoarticulation of prosodic constituents which are specified to support less articulatory effort. By this procedure, an economical relation between unfocussed and focussed parts of a sentence is established. The speaker tends to reach the target - the focussed constituent - as fast as possible.

3. SUMMARY

To conclude, we can establish the following: Using a look-ahead-algorithm which determines the position and the type of focus in a sentence we are able to specify the prosodic behavior in focal and pre-focal constituents.

- (a) A phonological rule assigns an L+H*-pitch accent to the most prominent syllable of the sentence initial constituent.
- (b) For the case of wide focus, the pitch range of the sentence initial L+H*-sequence increases.
- (c) In the case of narrow focus on a consti-tuent in non-sentence-initial position, the pitch range of the sentence initial L+H*-sequence decreases and syllables situated before the focussed domain become shortened.

It is well established in actual phonological and prosodic research that IS has important influences on prosody. Nevertheless, reducing IS relevant information in prosody only to pitch accent variation seems not very satisfying for synthesis purposes. It seems that it is both shortening and variation in pitch range that is responsible for marking focus sensitive salience inside sentences.

References

- [1] Alter, K./Matiasek, J./G. Niklfeld (1996). Modeling Prosody in a German Concept-to-Speech System. In D. Gibbon (Ed.) *Natural Language Processing and Speech Technology. Results of the 3rd KONVENS Conference*, Bielefeld, October 1996. Berlin Mouton de Gruyter.
- [2] Niklfeld, G./K. Alter (1996). Covering prosody in concept-to-speech via an extended two-level phonology component. In *Computational Phonology in Speech technology. Proceedings of SigPhon. Association of Computational Linguistics*. Santa Cruz. 59-66.
- [3] Féry, C. (1993): *German Intonational Patterns*. Tübingen. (Linguistische Arbeiten 285).
- [4] Alter, K. (1996), Discovering time structure in German. t. app. in: *Nordic Prosody VII*. Joensuu.
- [5] Wightman, C. W./Shattuck-Hufnagel, S./Ostendorf, M./P.J. Price, (1992). Segmental Durations in the Vicinity of Prosodic Phrase Boundaries. In *The Journal of the Acoustical Society of America*, March 1992.
- [6] Beckmann, M.E./J. Edwards (1994). Articulatory evidence for differentiating stress categories. In P.A. Keating (Ed.) *Phonological structure and Phonetic Form. Papers in Laboratory Phonology III*. Cambridge. 7-33.
- [7] Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In Hardcastle, W.J., A. Marchal (Eds.). *Speech production and speech modelling*. (pp. 403-439). Dordrecht.
- [8] Büring, D. (1995). *The 59th Street Bridge Accent: On the meaning of Topic and Focus*. Diss. Universität Tübingen.
- [9] Vallduvi, E. (1993), *Information packaging: A survey*. Human Community Research Center. U. of Edinburgh. Research Paper HCRC/RP-44.
- [10] Dogil, G.: Phonetic correlates of word stress. in: *AIMS (Arbeiten des IMS)*, Vol.2, NO2, pp1-60. Stuttgart, 1995.
- [11] Jacobs, J. (1994). Informational Autonomy. In P. Bosch; R. van der Sandt (Eds.). *Focus & Natural Language Processing*. Vol. 1. Intonation and Syntax. Working Paper 6. Dec. 1994 (pp. 113-132). Heidelberg.