Speech Prosody 2016
31 May - 3 Jun 2106, Boston, USA

Contrastive topic constituents in German

Sabine Zerbian[1], Giuseppina Turco[2], Nadja Schauffler[1],
Margaret Zellers[1], Arndt Riester[3]

[1] Department of Linguistics: English, University of Stuttgart, Germany
[2] LPP (CNRS/Sorbonne Nouvelle), France
[3] Institute for Natural Language Processing, University of Stuttgart, Germany
{sabine.zerbian, giuseppina.turco, nadja.schauffler, margaret.zellers}@ifla.uni-stuttgart.de, arndt.riester@ims.uni-stuttgart.de

Abstract

This article reports on a study investigating the intonational realization of context-changing and context-preserving contrastive topics in German. Results of the pilot study show that both kinds of topics can be marked by different pitch accents on the topic constituent, although speaker-specific differences emerge. When speakers do use different pitch accents, low rises (L*H) exclusively occur on context-changing contrastive topics whereas simple rises ((L)H*) are frequently used for both types. The qualitative results are complemented by a fine-grained phonetic analysis using Functional Data Analysis. Together, the results provide empirical evidence from a carefully controlled study for the impressionistic descriptions found in the literature.

Index Terms: contrastive topic, German, pitch accent, Functional Data Analysis

1. Introduction

Contrastive topics (henceforth CT) are a subset of the information-structural category of topics, specifically, in the current study, a subset that involve an element of sub-informativity, according to [1]. Going beyond this simple characterization, however, [1] differentiates between two kinds of CTs, namely context-changing (1a) and context-preserving CTs (1b, 1c).

(1) Examples of CTs (following [1])
a. *context-preserving*
Haben sich deine Nachhilfeschüler verbessert?
(Have your tutoring students improved?)
Daniela kommt gut voran, Janina muss noch üben.
(Daniela makes good progress, Janina still needs to practice.)
b. *context-changing: move insertion*
Hat sich dein Nachhilfeschüler verbessert?
(Has your tutoring student improved?)
Daniela kommt gut voran, Janina muss noch üben.
(Daniela makes good progress, Janina still needs to practice.)
c. *context-changing: strategy shift*
Gefallen dir deine neuen Klamotten?
(Do you like your new clothes?)
Daniela findet sie toll.
(Daniela thinks they are awesome!)

These different kinds of CTs are all instances of topic-related sub-informativity in that more than one topic-comment relation is established. However, they differ semantically, as can be shown by the questions under discussion that they answer (QUDs; [2]). This is shown in (2).

(2) QUDs of contrastive topics
a. *context-preserving*
QUDs: {What about x's performance? What about y's performance?}
b. *context-changing: move insertion*
original QUD: {What about x's performance? [x=male]}
new QUDs: {What about x's performance [x=female]? What about y's performance?}
c. *context-changing: strategy shift*
original QUD: {What do YOU think about your clothes?}
new QUD: {What does D. think about your clothes?}
+ conversational implicature

[1] claims that the two kinds of CTs are differentiated by their intonation in German. Only context-changing CTs can be realized with the so-called root contour ([3]). The root contour consists of a low (L*) or scooped accent (L*+H), followed by a H phrase accent, yielding the contour L*+H H- (also referred to as a low rise). This accent might be simplified to a simple peak accent, H*, in connected speech. Context-preserving CTs do not occur with the root contour but rather with a hat contour. The hat contour consists of a scooped accent, (L+)H*, followed by a high pitch plateau and a falling focus accent.

The intonation of German contrastive topics as opposed to non-contrastive topics has been widely discussed in the literature, and empirical evidence on the phonetic realization has recently been presented (cf. [4] for overview and data). However, little is known about the phonetic realization of the hat and root pattern (cf. discussion in [5: 29]). Therefore, a production study was designed to gather empirical data on whether and how German speakers produce an intonational difference between context-preserving and context-changing CTs on the topic constituent. The results of the pilot study are presented in this paper. The production study is presented in section 2, the results in sections 3 and 4, and a discussion of the

 doi: 10.21437/SpeechProsody.2016-71

results together with an outlook for future research is presented in section 5.

2. Read production study

2.1. Participants

Three German speakers (1 male, 2 female) participated in the read production study. They were between 23 and 38 years old and spoke a Northern variety of German. At the time of the recording they were graduate students or staff members of the University of Stuttgart. None of them had any known speech problems. All were unaware of the purpose of the experiment.

2.2. Materials

Twenty-four experimental sentences following the schema in (1) were constructed. Sentences were matched for lexical stress position of the topic constituent (i.e. the stress fell always on the second syllable) and segmental make-up (i.e. the stressed syllable was composed of sonorant material). Of these 24 experimental sentences, 8 were preceded by a question triggering a context-preserving reading (1a), another set of 8 by a question triggering a move-insertion (context-changing) reading (1b) and another set of 8 a strategy-shift (context-changing) reading (1c). To facilitate pitch comparison across the three contexts, answer-sentences contained the same contrastive topic constituent (e.g. *Daniela*).

The three sets of experimental sentences were interspersed with 48 filler sentences. Twenty-four of the fillers contained topical referents elicited in a non-contrastive context. The other fillers were comprised of yes-no question-answer pairs. Thus the whole experiment consisted of 72 trials. The question sentences for all experimental and filler trials were pre-recorded by a German native speaker. The pre-recorded sentences served to facilitate coherent interpretation of the context across participants and the elicitation of the relative answer.

2.3. Procedure

Participants were given written instructions to read the sentences at normal rate. Question-answer pairs were presented one-by-one on a computer screen and initiated with a button click by the participant. In cases of hesitations, participants were asked to read the sentence again at the end of the whole session. All recordings were made in a sound-proof booth in the Linguistics Laboratory at the University of Stuttgart.

In each trial participants first heard the pre-recorded question and then had to read the answer aloud. Question-answer pairs were presented in a block-wise randomized fashion. The recording session lasted approximately 15 minutes.

3. Qualitative Analysis

3.1. Procedure

The pitch accent on the topic constituent was the focus of interest in the current study. Therefore, an analysis of the intonation contours on the contrastive topic constituents was carried out by the first three authors, relying on auditory impression and visual evidence from the pitch track and following the DIMA annotation guidelines for German intonation ([6]).

Three phonetically-based labels emerged from this first round of annotations (LH*, L*H, and >L*H), which were then grouped into one of two intonational categories by the first author. For this, the pitch accent on the topic constituent was considered and categorized according to the primary perceptual impression on the accented syllable as high or low, resulting in[1]:

- (L)H*: A perceptually high pitch accent which might either be H* or L+H*. The peak is reached in the accented syllable or early in the posttonic syllable. The peak might be reached from relatively low (L+H*) or not (H*). In most cases this accent will be perceived as a rise. Where the perceptual impression relates to a fall, this has been noted separately through the transcription of H*+L, a label for a bitonal pitch accent which is not available in GToBI ([7]).
- L*(H): A perceptually low accent, in which the low target clearly lies in the accented syllable. It may be preceded by a shallow fall (root contour, low rise). The H follows either in the immediately following syllable or later. In GToBI notation, these would correspond to two different accent categories, namely L*+H and L*, since the second tone of a bitonal pitch accent must immediately follow the starred tone (see [8: 35]).

3.2. Results

The results of the qualitative analysis are presented in tables 1 and 2. The data obtained from the three speakers support the view that context-preserving CTs are distinguished intonationally from context-changing CTs in German. Production data, annotated for type of pitch accent used, show that context-changing CTs are produced with a late rise (L*H) more often than context-preserving CTs.

Table 1. *Counts of pitch accents ((L)H*, L*(H), H*L; invalid) produced on topic constituents across three CT contexts (preserving, move insertion, strategy shift). Invalid cases are those in which a speech error or other problem made it impossible to apply an accent label.*

	preserving	move	strategy
(L)H*	16	10	4
L*(H)	6	12	16
H*L	1	1	2
invalid	1	1	2
total	24	24	24

1 This categorization basically merged two of the three DIMA-labels into one category. Note that the resulting category labels, (L)H* and L*H, are not necessarily identical to the binary opposition used in [4], which distinguishes non-contrastive topics from contrastive topics. From first inspection, it seems that the L* in [4]'s context-preserving contexts is earlier than in our context-changing contexts. But this requires further examination.

	speaker M			speaker K			speaker F		
	P	M	S	P	M	S	P	M	S
(L)H*	6	4	1	8	6	3	2	-	-
L*(H)	-	4	5	-	1	4	6	7	7
H*L	1	-	2	-	1	-	-	-	-
invalid	1	-	-	-	-	1	-	1	1

Table 2. *Counts of pitch accents by speaker*

The speaker-specific results in table 2 show clear speaker-dependent differences in the realization of CTs in German. Speaker F uses low rises (L*H) almost exclusively.

4. Quantitative Analysis

4.1. Functional Data Analysis

One difficulty with carrying out intonational analysis is that annotators, even within the same system, may not always agree. Since we are interested in category differences in the intonational production of different kinds of contrastive topics, it is essential to validate the intonational labels. To this end, we have adopted techniques from Functional Data Analysis (FDA). FDA allows for the extension of classic statistical tools to the domain of functions ([9]). Entire contours, represented by functions, are the input to the analyses, rather than data points. In the current work we will specifically make use of Functional Principal Component Analysis (PCA). Functional PCA results in a compact description of the main shape variations (or Principal Components, PCs) that are present within a dataset of curves; in this case, the set of F0 contours from the contrastive topics. Once the PCs are identified, each original curve is associated with a PC score, which quantifies where in the continuum of each shape variation described by a PC a specific curve is located. Since PC scores are numbers, they can be easily used in further analyses, including being correlated with manual labels (as already shown by [10], [11]). Since they are directly related to the shape of the contours, they provide an objective method of identifying consistency and variation across a set of contours.

4.2. Methodology

The input to the Functional PCA was the set of pitch contours occurring on the contrastive topic word in the above dataset (i.e. *Daniela* as in (1)). Using Praat ([12]), the contours were extracted as sets of pitch samples at 10ms intervals. The rest of the analysis was carried out in R, using the 'fda' package ([13]) and a freely available analysis procedure ([14]). In order to normalize the data for use in FDA, the pitch values were converted from Hertz into semitones, and the average value of each contour was subtracted from each point. This helps reduce the influence of global variation such as that based on speaker gender. Then, all functions are interpolated using a B-splines basis function.

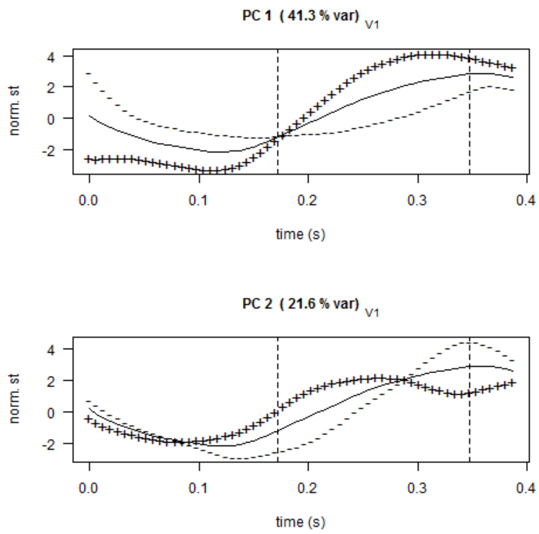
The use of the common function basis means that all the contours must be defined on a common time interval. Thus the contours must be re-registered in time such that they all fill this same time interval. In order to reduce the effects of time-warping, the contours were registered not only to the word boundaries (initial and final), but also to the boundaries of the stressed syllable (initial and final). Since the pitch features on or in the vicinity of the stressed syllable are most important for accentuation, this landmark registration ensures that parts of the contour that are similarly meaningful are time-aligned in the analysis.

After exclusion of the four invalid contours, Functional PCA was applied to the remaining 68 registered functions. It is important to note that since the current study is a pilot, this dataset is smaller than what would normally be considered appropriate for an FDA analysis.

4.3. Results

The functional PCA applied to our data results in 3 significant PCs, of which the first two will be discussed here. The first principal component (PC1) explains 41.3% of the variation in the data, while the second principal component (PC2) explains an additional 21.6% of the variation. (The third PC, not reported here, explains a further 15.5%.) The first two PCs are shown in Figure 1.

Figure 1. *Principal components 1 (top) and 2 (bottom). In both cases, the solid line represents the mean of all contours. The line made up of + signs represents the contour with the greatest positive value for the PC, while the line made up of – signs represents the contour with the greatest negative value for the PC. The x-axis is normalized time, and the y-axis normalized semitones. The vertical dashed lines indicate the normalized position of the stressed syllable.*

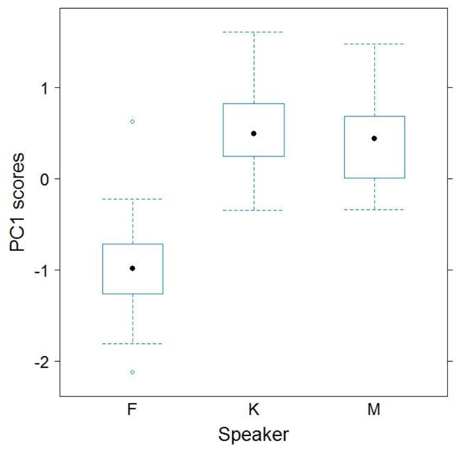


A visual inspection of the two principal components gives us an idea of the aspects of the contours that they represent. Contours with a positive value for PC1 have a steeper rise from the first through the middle (stressed) syllable, while contours with a negative value for PC1 have a flatter shape. Looking at PC2, we see that contours with a positive value have a peak in the early-to-middle part of the stressed syllable, while contours with a negative PC2 value have their peak later.

Since each contour in the dataset is assigned one value for each PC, we can test these PC values to see how contour shape is affected by different features of the data. For example, the qualitative analysis found that speaker F used a different strategy than the other two speakers. Statistical tests show that this different behavior is also supported by the quantitative data in PC1; speaker F's PC1 values are significantly different from those of speakers M and K (ANOVA; $F(2, 61)=53.36$, $p<0.001$; cf. Figure 2). Note that since variation due to speaker gender, etc. was eliminated during the normalization stage (cf. section 4.2), we can take this variation to represent different speaker strategy with confidence.

PC1 has been shown to be able to distinguish between different speakers' strategies in a general way. However, the question still remains to what extent the PCA may be able to distinguish between strategies for the different contrastive topic categories in question. None of the PCs in question correlated consistently with any of the three types of contrastive topic. However, PC2 was strongly correlated with the annotator's labels of the intonational contours (ANOVA; $F(2, 61)=13.68$, $p<0.001$; cf. Figure 3). Since the PCA was blind to the category labels, this constitutes independent verification of the accuracy of the category labels.

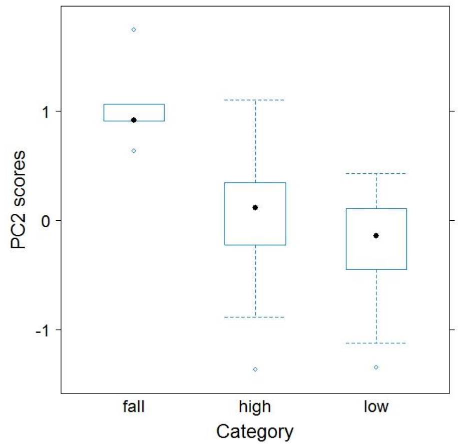
Figure 2. *PC1 scores by speaker.*



5. Discussion

Although no clear difference for the topic categories emerges in the quantitative study (possibly due to the small size of the data set), we have evidence that the intonational labels used are capturing real variation in the data. Together with the qualitative results, the pilot study can thus be argued to confirm the intonational difference on the topic constituent of different kinds of contrastive topics in German. This has been suggested in the literature by attributing different ToBI labels (L)+H* and L*+H to context-preserving CTs and context-changing CTs respectively. To our knowledge, this pilot is the first to provide empirical data on this distinction with controlled and comparable stimuli and several speakers.

Figure 3. *PC2 scores by annotator's category. Fall corresponds to H*L, high to (L)H*, and low to L*(H) in Tables 1 and 2 above.*



Within the context-changing topics, the L*+H accent is used most reliably in the case of strategy shift, where the common ground modification crucially implies a conversational implicature.

Although the data were elicited from prosody-aware speakers, there emerge speaker-dependent production strategies involving the alignment and possibly also the scaling of tones, which deserve further attention (see also [15]).

6. Conclusions

Against the background of this pilot study, a follow-up study is currently being conducted to corroborate these initial findings with a larger speaker group of naïve speakers. Special attention will be paid to inter-speaker variation and to the intonational cues used for the semantic distinction of contrastive topics in German.

The intonational distinction on the topic constituent is also predicted to be relevant in perception. Recent psycholinguistic research by [16] on contrastive vs. non-contrastive topic accents has shown that listeners begin disambiguating at the time of the topic constituent even though they need time to establish the contrastive interference of prenuclear CT accents.

7. Acknowledgements

This study forms part of project A7 of the SFB 732 (University of Stuttgart), funded by the German Research Foundation (DFG). Thanks to Sabrina Beer, Yulia Pavlyuchenko, and Nicole Stirm for assistance in data collection and preparation for analysis, and to Michele Gubian for assisting with the FDA implementation.

8. References

[1] Gast, V. 2010. “Contrastive topics and distributed foci as instances of sub-informativity: A comparison of English and German”. In Breul, C. & E. Göbbel (eds.): Comparative and Contrastive Studies of Information Structure, 15-50. Amsterdam: John Benjamins.

[2] Büring, D. 2003. “On D-Trees, Beans, and B-Accents”. Linguistics & Philosophy 26:5. 511-545

[3] Jacobs, J. 1997. “I-Topikalisierung”, Linguistische Berichte 168:91-133.

[4] Braun, B. 2006. “Phonetics and phonology of thematic contrast in German”, Language and Speech 49(4): 451-493.

[5] Grabe, E. 1998. Comparative Intonational Phonology: English and German. MPI Series in Psycholinguistics 7, Wageningen, Ponsen en Looien.

[6] Kügler, F., Smolibocki, B., Arnold, D., Baumann, S., Braun, B., Grice, M., Jannedy, S., Michalsky, J., Niebuhr, O., Peters, J., Ritter, S., Röhr, C.T., Schweitzer, A., Schweitzer, K., Wagner, P. 2015. “DIMA – Annotation guidelines for German intonation.” Proceedings of ICPhS 2015 Glasgow.

[7] Grice, M., Baumann, S. & R. Benzmüller. 2005. “German Intonation in Autosegmental-Metrical Phonology”, in S.-A. Jun (ed.) Prosodic Typology. Oxford: Oxford University Press. 55-83.

[8] Peters, J. 2014. Intonation. Heidelberg: Winter.

[9] Ramsay, J.O. & Silverman, B.W. 2005. Functional Data Analysis (2nd Ed.) New York: Springer.

[10] Gubian, M., Cangemi, F. & Boves, L. 2010. “Automatic and data driven pitch contour manipulation with Functional Data Analysis.” In Proceedings of Fifth International Conference on Speech Prosody, Chicago, USA.

[11] Zellers, M., Gubian, M. & Post, B. 2010. “Redescribing intonational categories with Functional Data Analysis.” In Proceedings of 11th Interspeech, Makuhari, Japan.

[12] Boersma, P. & Weenink, D. 2015. Praat: doing phonetics by computer [Computer program]. Version 6.0.05, available http://www.praat.org/

[13] Ramsay, J.O., Hookers, G. & Graves, S. 2009. Functional Data Analysis with R and MATLAB. New York: Springer.

[14] Gubian, M. 2010. Functional Data Analysis for speech research. [online] Available http://lands.let.ru.nl/FDA/.

[15] Niebuhr, O., D’Imperio, M., Gili Fivela, B. & Cangemi, F. 2011. “Are there “shapers” and “aligners”? Individual differences in signaling pitch accent category”, Proceedings of the 17th ICPhS, Hong Kong, 120-123.

[16] Braun, B. & Y. Asano. 2013. “Double contrast is signalled by prenuclear and nuclear accent types alone, not by f0-plateaux”, Proceedings of the 14th Annual Conference of the International Speech Communication Association. Lyon, France. 263-265.