PROSODY OF PRE-FOCAL BACKGROUND DEPENDS ON FOLLOWING FOCUS

Simon Roessig

Department of Linguistics, Cornell University
mail@simonroessig.de

ABSTRACT

This study is concerned with the realization of focus structure by prosodic means in German. Results of a production study are presented that support previous findings about differences between focus types. Crucially, however, the realization of given (background) words in the pre-focal region also depends on the focus type that follows. The data are analyzed with classic scalar measures (F0 maximum and syllable duration) as well as by assessing the evolution of the F0 trajectories over time in the intervals of interest. The two analyses show converging results: Before corrective focus, pre-nuclear words are realized with lower F0 peaks, larger F0 excursions, and shorter duration than before non-corrective narrow focus. In the focal domain, the picture is almost reversed with higher F0 peaks and larger F0 excursions. These results suggest that nuclear *and* pre-nuclear accents contribute to the interpretation of the focus structure of the sentence.

Keywords: prosody, information structure, pre-nuclear, F0, duration

1. INTRODUCTION

In West-Germanic languages, focus is expressed by a prominence on the focused word (or focus exponent) [1], [2] – more specifically by the placement of the nuclear pitch accent, and post-focal deaccentuation or compression [3], [4]. In addition, the realization pattern of the accented syllable is correlated with the *type* of focus. Accented syllables in narrow and corrective focus are realized with higher F0 targets [5], longer durations [6] and more extended articulatory movements [7] compared to their counterparts in broad focus. More recently, it has been shown that there are even differences between narrow and corrective focus [6], [8], [9], although these differences appear to be subtler.

While post-focal deaccentuation and the modulation of nuclear accents in the focal region of a sentence as a function of focus types are well-documented, fewer studies have looked at the pre-focal region. The predominant view is that, as [10] suggested, pre-nuclear accents are optional ("ornamental"), especially when pre-focal. Similarly, [11] posited that pre-nuclear accents normally do not mark information structure but rather are correlates of the metrical structure of the sentence. This view is supported by results from a perceptual learning experiment presented by [12] with English speaking individuals: While younger children pay equal attention to the beginning *and* end of intonation contours, older children and adults seem to have internalized that the end contains the most important information and hence pay more attention to the later part of the contour.

Results from other studies are at odds with the idea that pre-nuclear accents are unaffected by information structure. For German, [5] showed that pre-nuclear accents exhibit lower F0 when they appear pre-focally before a narrow focus compared to pitch accents in the same position in a broad focus sentence (in which case they are part of the focus). Similar results were obtained for Bulgarian by [13]. These findings from production are in line with perception results of [14] for Dutch in which participants judged the excursion of two pitch peaks (pre-nuclear and nuclear) in relation to a focus structure. Compared to broad focus, a smaller pre-nuclear peak was judged as optimal for words preceding corrective focus. The results for German presented in [6] indicate that words in pre-focal position preceding corrective focus are realized with longer durations compared to the same words in broad focus sentences. These studies demonstrate that the pre-nuclear domain can indeed be affected by information structure.

[9], also for German, used speech material that includes narrow and corrective focus. Interestingly, there seemed to be a lower probability for the placement of a pre-nuclear accent before corrective focus than before narrow focus – although in both cases the pre-nuclear domain is given, i.e., pre-focal. This finding suggests that the pre-focal region may actually contain information about the focal region. It raises the question as to how distributed information structure is across the phrase and whether pre-focal elements can contribute to the marking of information structure of the following focal element.

The present study concentrates on this question. In doing so, it takes both the pre-focal and the focal region into account and compares the prosodic patterns of the two regions in sentences with narrow

and corrective focus. A corpus of approximately 1000 productions from 27 speakers is analyzed regarding the realization of F0 and syllable duration. The prosodic patterns are investigated in terms of simple scalar measurements, namely F0 maximum in the word and stressed syllable duration, as well as in terms of the evolution of F0 trajectories over time using generalized additive mixed models (GAMM) – with converging results. The results of the paper contribute to our understanding of the marking of focus structure in German and the dispersion of information in the acoustic signal.

2. METHODS

2.1. Speech material

The speech material analyzed in this study consists of sentences produced with two different focus structures. In order to elicit these focus conditions, question-answer pairs were used. The answers are the analyzed target sentences and were always of the form *Er hat den/die <A> auf die <B> gelegt* ('He put the <A> on the <B>') with two nouns A and B. The questions served as triggers for the focus structure of the answer. In both focus conditions, word A is given and occurs pre-focally, i.e., in the background. The difference lies in the focus type of the focal word B: In the first focus condition, *background-narrow*, word B is in narrow focus. In the second focus condition, *background-corrective,* word B is in corrective focus. The question to elicit *background-narrow* followed the scheme *Wo hat er den/die <A> hingelegt?* ('Where did he put the <A>?'); the question to elicit *background-corrective* followed the scheme *Hat er den/die <A> auf die <C> gelegt?* ('Did he put the <A> on the <C>?') where C is a contrasting alternative referent. Table 1 illustrates the focus conditions with examples. Square brackets and subscript F indicate the focused elements. The full data set comprises two additional focus conditions, *corrective-background* and *broad-broad*, that are not reported in the current paper.

background-narrow	
Question:	*Wo hat er den Hammer hingelegt?* 'Where did he put the hammer?'
Answer:	*Er hat den Hammer [auf die Wohse]$_F$ gelegt.* 'He put the hammer on the Wohse.'
background-corrective	
Question:	*Hat er den Hammer auf die Mahse gelegt?* 'Did he put the hammer on the Mahse?'
Answer:	*Er hat den Hammer auf die [Wohse]$_F$ gelegt.* 'He put the hammer on the Wohse.'

Table 1: Focus conditions with question-answer pairs

As targets for the pre-focal word (A), ten German disyllabic nouns denoting common tools with stress on the first syllable were used: *Amboss* ('anvil'), *Besen* ('broom'), *Bohrer* ('drill'), *Bürste* ('scrub brush'), *Hammer* ('hammer'), *Pinsel* ('paint brush'), *Rolle* ('paint roller'), *Säge* ('saw'), *Schere* ('scissors'), and *Zange* ('pliers').

As targets for the focal word (B), twenty German sounding disyllabic nonce words with a $C_1V_1C_2V_2$ structure were created. All nonce words had stress on the first syllable. C_1 was chosen from the set of {/n/, /m/, /b/, /l/, /v/}, V_1 from {/a/, /o/}, and C_2 from {/n/, /m/, /z/, /l/, /v/}. V_2 was always Schwa. Examples for target word B are *Nahne*, *Mohme*, and *Bahle*.

2.2. Speakers and recordings

27 monolingual native speakers of German (19-35 yrs.; 17 female) were recorded. The subjects were prompted to produce the target utterances by involving them in an interactive game on a computer screen. In the game, their task was to help an animated robot retrieve tools. The robot's questions served as triggers for the focus structure of the answer. A training session with different target words preceded the actual recording session. The recordings were carried out at the University of Cologne using a head-mounted condenser microphone. In addition to the acoustic signal, the articulators' movements were recorded (EMA). This paper only deals with the acoustic data.

2.3. Annotations and measurements

The boundaries of the two target words (words A and B) and their stressed syllables were annotated. Additional segmental annotations were obtained from forced alignment using Kaldi [15] through the Montreal Forced Aligner [16]. Furthermore, the low boundary tone at the end of each sentence was labelled as the last reliable F0 point. Using these annotations, the following measurements were performed. First, the stressed syllable durations of both target words (pre-focal/A and focal/B) were measured. Second, F0 was calculated over the whole sentence using Praat [17] through the Python interface parselmouth [18]. For each speaker, the floor for the F0 calculation was set separately as the F0 value of the lowest L-% boundary tone of that speaker minus 10 Hz. From the F0 track, the maximum in each target word was obtained. Furthermore, time-normalized F0 was extracted in 49 equal time steps over the words in positions A and B, and in 149 equal time steps over the whole sentence. All F0 values in this analysis are expressed in semitones relative to the 5th percentile of the distribution of all L-% boundary tones of the speaker

that produced the utterance. Productions that had a clear phrase boundary between word A and word B were excluded to ensure that the pre-focal word was always pre-nuclear. After this exclusion, the data set comprised 991 recordings. The statistical analyses were carried out in R [19] using the libraries brms [20] for Bayesian regression, mgcv [21], itsadug [22] and tidymv [23] for fitting and visualizing GAMMs, as well as tidyverse [24] and zoo [25] for data processing and plotting. Data and code are available publicly on OSF: https://osf.io/2an3v/.

3. RESULTS

3.1. Analysis of F0 contours

The top panel of Figure 1 presents scatterplots of all F0 points with average contours superimposed. The bottom panel "zooms" into the F0 trajectories by giving the average contours for the pre-focal (left) and focal words (right). The average contours were obtained by averaging over all F0 measures of a point in normalized time. Only those time points for which more than 25% of measures existed entered the calculation of the average contour. The average contours are smoothed by applying a rolling mean with a window size of 3. Comparing the pre-focal and the focal regions, the peak relations are reversed: In the pre-focal region, *background-narrow* (blue) exhibits a higher peak than *background-corrective* (red), while the opposite is true in the focal region.

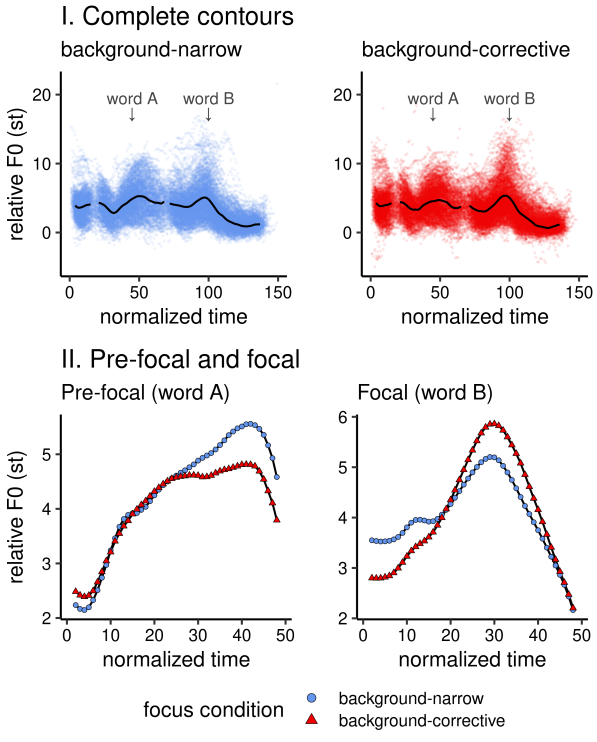

Figure 1: Average contours. Top: Complete sentence. Bottom: pre-focal (word A) and focal (word B) words.

The differences in the contours are assessed more formally by fitting a GAMM to each region – pre-focal and focal. For this analysis, the F0 trajectories are interpolated linearly. As fixed effects, the models include FOCUS CONDITION as a parametric term and smooths over TIME for FOCUS CONDITION. The models were fit such that the smooth over TIME for the condition *background-narrow* represents the reference smooth and the model contains a difference smooth for the condition *background-corrective* in relation to the reference smooth. In addition, random factor smooths per focus condition were included for the individual levels of SPEAKER and TARGET WORD. To account for autocorrelation of the residuals, an AR1 error model was incorporated. The smooths obtained from the GAMMs for the two focus conditions are visualized in the top panel of Figure 2.

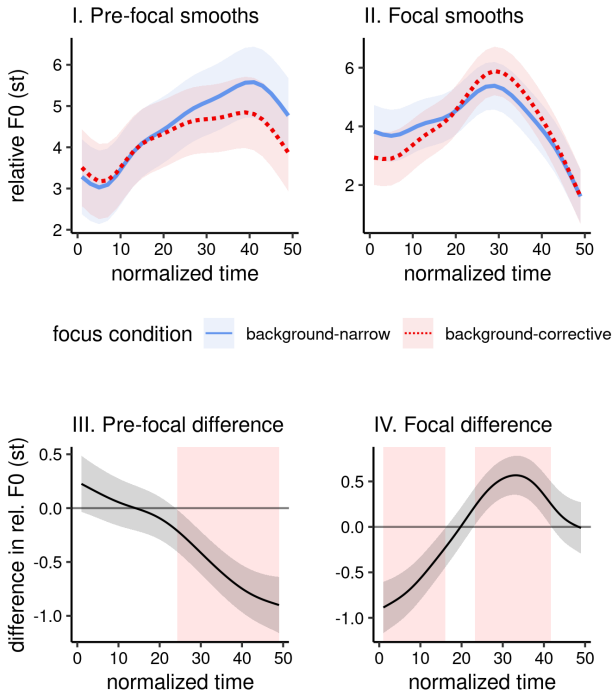

Figure 2: GAMM results for pre-focal (left) and focal regions (right). Top: smooths. Bottom: difference smooths between the two conditions.

Both models have significant difference smooths for *background-corrective*, indicating that the course of F0 over time in this condition is different from *background-narrow*. It should be emphasized that this is not only true for the focal region but also for the pre-focal region – although the information structure remains constant in this region (i.e., background). The difference plots in the bottom panel of Figure 2 visualize where and how *background-corrective* differs from *background-narrow*. In these plots, the red shaded areas indicate the regions of a significant difference. In the pre-focal region (Figure 2 bottom left), the second half of the contour over the

word in *background-corrective* takes a lower course (the difference is negative). In the focal region (Figure 2 bottom right), the start of the contour and the region of the peak are different: the contour of *background-corrective* starts lower (negative difference at the beginning) and reaches a higher peak (positive difference around the peak).

3.2. Analysis of F0 maximum and syllable duration

Figure 3 shows the means and standard errors for F0 maximum and stressed syllable duration. The pre-focal region (purple circles) exhibits a lower F0 maximum and shorter syllable duration in *background-corrective* than in *background-narrow.* The picture is again reversed in the focal region (green triangles).

Bayesian regression models were fit with either F0 MAXIMUM or SYLLABLE DURATION as dependent variable and FOCUS CONDITION as fixed effect. Random intercepts for SPEAKER and TARGET WORD were included as well as by-SPEAKER and by-TARGET WORD random slopes for FOCUS CONDITION. *background-narrow* is the reference level in the models (intercept). The regression coefficient β for *background-corrective* thus indicates the difference between the two focus conditions.

Table 2 lists the estimates β with their 90% credible interval (CI). For the pre-focal region, the table additionally gives Pr(β<0), the probability that β is negative. For the focal region, the table gives Pr(β>0), the probability that β is positive. In the pre-focal region, the models provide strong evidence for lower F0 maxima and shorter syllable durations in *background-corrective* than in *background-narrow* (the estimated β are negative). In the focal region, the models provide strong evidence for higher F0 maxima but not for longer syllable durations in *background-corrective*: here, Pr(β>0) is only 0.88.

Pre-focal (word A)			
Parameter	*β*	*90%-CI*	*Pr(β<0)*
F0 maximum	-0.71	[-0.93 -0.50]	1.00
Syllable duration	-6.96	[-11.01 -2.80]	1.00
Focal (word B)			
Parameter	*β*	*90%-CI*	*Pr(β>0)*
F0 maximum	0.28	[0.05 0.51]	0.98
Syllable duration	2.11	[-0.87 5.13]	0.88

Table 2: Estimates from the Bayesian regression models

4. DISCUSSION

The presented results reveal that not only the focused element is realized differently depending on the focus type; the realization of pre-focal elements depends on the following focus type as well. In the case of syllable duration, the effect is even stronger in the pre-focal region. One interpretation is that flatter, lower F0 and shorter durations in the pre-focal region help to boost the prominence perception of the following focus. In this case, the effect on pre-focal elements would be indirect. A more direct interpretation would be that speakers intend to differentiate background before narrow from background before corrective focus. In both perspectives, the prosodic marking of information structure is distributed across the phrase and not localized in the focus. Hence, this study contributes evidence for the significance of pre-nuclear words. The effects presented here are certainly relatively small. Future research will have to test the perceptual relevance of pre-nuclear accents in information structure marking, a research question to which the findings of [14] and [26] give first positive hints. It is also interesting to investigate the relative scaling of pre-nuclear and nuclear pitch accents in relation to the phenomenon known as the *Gussenhoven-Rietveld effect* [27], [28] – the unexpected finding that raising the pre-nuclear peak boosts the perceived prominence of the nuclear peak. While this effect generally predicts the opposite of the results found in this study, it underlines the importance of taking prominence relations into account.

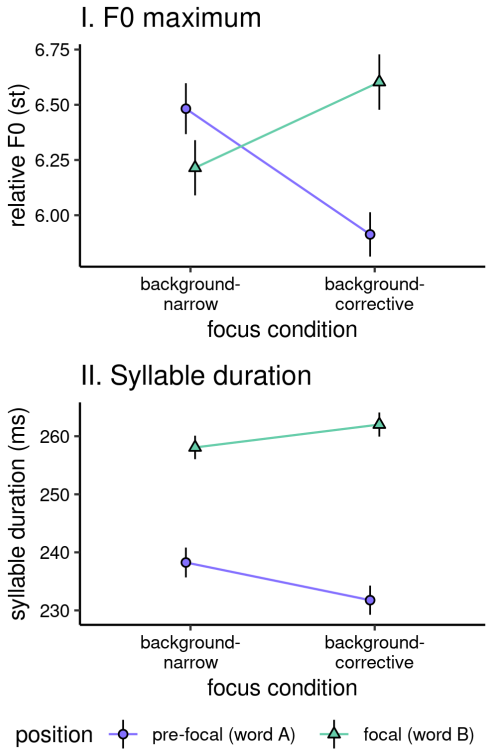

Figure 3: Means (with standard errors). Top: F0 maximum in the word. Bottom: syllable duration.

5. ACKNOWLEDGEMENTS

This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – RO 6767/1-1 (Walter Benjamin program).

6. REFERENCES

[1] M. Rooth, "Association with focus," Dissertation, University of Massachusetts, Amherst, MA, 1985.

[2] D. R. Ladd, *The structure of intonational meaning: Evidence from English*. Indiana University Press, 1980.

[3] D. R. Ladd, *Intonational Phonology*. Cambridge University Press, 2008.

[4] M. Wagner, "Asymmetries in Prosodic Domain Formation," in *Perspectives on Phases*, Cambridge, MA: MITWPL 49, pp. 329–367.

[5] C. Féry and F. Kügler, "Pitch accent scaling on given, new and focused constituents in German," *J. Phon.*, vol. 36, no. 4, pp. 680–703, Oct. 2008.

[6] F. Kügler, "The role of duration as a phonetic correlate of focus," in *Proc. Speech Prosody 2008*, 2008, pp. 591–594.

[7] D. Mücke and M. Grice, "The effect of focus marking on supralaryngeal articulation – Is it mediated by accentuation?," *J. Phon.*, vol. 44, pp. 47–61, 2014.

[8] S. Roessig, D. Mücke, and L. Pagel, "Dimensions of Prosodic Prominence in an Attractor Model," in *Proceedings of Interspeech 2019*, 2019, pp. 2533–2537.

[9] S. Baumann, J. Becker, M. Grice, and D. Mücke, "Tonal and Articulatory Marking of Focus in German," in *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken, Germany, 2007, pp. 1029–1032.

[10] D. Büring, "Intonation, Semantics and Information Structure," in *The Oxford Handbook of Linguistic Interfaces*, G. Ramchand and C. Reiss, Eds. Oxford University Press, 2007.

[11] S. Calhoun, "The centrality of metrical structure in signaling information structure: A probabilistic perspective," *Language*, vol. 86, no. 1, pp. 1–42, 2010.

[12] V. Kapatsinski, P. Olejarczuk, and M. A. Redford, "Perceptual Learning of Intonation Contour Categories in Adults and 9- to 11-Year-Old Children: Adults Are More Narrow-Minded," *Cogn. Sci.*, vol. 41, no. 2, pp. 383–415, Mar. 2017.

[13] B. Andreeva, W. J. Barry, and J. Koreman, "Local and Global Cues in the Prosodic Realization of Broad and Narrow Focus in Bulgarian:," *Phonetica*, vol. 73, no. 3–4, pp. 256–278, Feb. 2017.

[14] H. H. Rump and R. Collier, "Focus Conditions and the Prominence of Pitch-Accented Syllables," *Lang. Speech*, vol. 39, no. 1, pp. 1–17, Jan. 1996.

[15] D. Povey *et al.*, "The Kaldi Speech Recognition Toolkit," in *IEEE 2011 Workshop on Automatic Speech Recognition and Understanding*, Hilton Waikoloa Village, Big Island, Hawaii, US, Dec. 2011.

[16] M. McAuliffe, M. Socolof, S. Mihuc, and M. Wagner, "Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi," in *Proceedings of INTERSPEECH, 20-24 August, Stockholm, Sweden*, 2017, pp. 498–502.

[17] P. Boersma and D. Weenink, "PRAAT, a system for doing phonetics by computer," *Glot Int.*, vol. 5, no. 9/10, pp. 341–345, 2001.

[18] Y. Jadoul, B. Thompson, and B. de Boer, "Introducing Parselmouth: A Python interface to Praat," *J. Phon.*, vol. 71, pp. 1–15, Nov. 2018.

[19] R Core Team, *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2021. [Online]. Available: https://www.R-project.org/

[20] P.-C. Bürkner, "Advanced Bayesian Multilevel Modeling with the R Package brms," *R J.*, vol. 10, no. 1, pp. 395–411, 2018.

[21] S. N. Wood, "Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models," *J. R. Stat. Soc. B*, vol. 73, no. 1, pp. 3–36, 2011.

[22] J. van Rij, M. Wieling, R. H. Baayen, and H. van Rijn, "itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs." 2022.

[23] S. Coretta, *tidymv: Tidy Model Visualisation for Generalised Additive Models*. 2022. [Online]. Available: https://CRAN.R-project.org/package=tidymv

[24] H. Wickham *et al.*, "Welcome to the tidyverse," *J. Open Source Softw.*, vol. 4, no. 43, p. 1686, 2019.

[25] A. Zeileis and G. Grothendieck, "zoo: S3 Infrastructure for Regular and Irregular Time Series," *J. Stat. Softw.*, vol. 14, no. 6, pp. 1–27, 2005.

[26] J. Bishop, "Focus projection and prenuclear accents: evidence from lexical processing," *Lang. Cogn. Neurosci.*, vol. 32, no. 2, pp. 236–253, Feb. 2017.

[27] C. Gussenhoven and A. C. M. Rietveld, "Fundamental frequency declination in Dutch: testing three hypotheses," *J. Phon.*, vol. 16, pp. 355–369, 1988.

[28] D. R. Ladd, J. Verhoeven, and K. Jacobs, "Influence of adjacent pitch accents on each other's perceived prominence: two contradictory effects," *J. Phon.*, vol. 22, pp. 87–99, 1994.