10th International Conference on Speech Prosody 2020
25-28 May 2020, Tokyo, Japan

The appropriateness of prenuclear accent types – Evidence for information structural effects

Stefan Baumann, Janina Kalbertodt, Jane Mertens

I*f*L Phonetik, University of Cologne, Germany
{stefan.baumann, janina.kalbertodt, jane.mertens}@uni-koeln.de

Abstract

In this study, two perception experiments on German were carried out to investigate whether informativeness (comprising information status and focus) affects listeners' appropriateness judgments of accent types in prenuclear position and whether these effects depend on the mode of context presentation (listening vs. reading). Five different prenuclear f0 contours were tested for each target sentence by combining them with four contexts varying the informativeness of the sentence-initial target word. As expected, the results mirror the findings of an earlier production study on prenuclear accents: In general, there is a preference for rising accent types in prenuclear position but we also find subtle effects of informativeness. Listeners clearly prefer deaccentuation and low accents on given referents, whereas rising accents are favored for referents in a contrastive context. Contrary to our expectations, high accents were judged as least appropriate, even in contexts where the referent was newly introduced. Importantly, the effects of informativeness are independent of the mode of context presentation, i.e. silent reading did not significantly alter the appropriateness judgments of accent types. These findings once more challenge the view of prenuclear accents as being merely 'ornamental' and rather suggest that they are sensitive to changes in information structure.

Index Terms: perception, prenuclear accents, information structure, prosody, appropriateness, German

1. Introduction

Compared to nuclear accents, commonly defined as the last pitch accent in an intonation unit making up its structural head, prenuclear accents, i.e. pitch accents preceding the nucleus within the same unit, have received much less attention. Previous investigations on prenuclear accents draw an inconsistent picture as to their status in the phonological system of West Germanic languages. While [1] claimed that prenuclear accents are placed accounting for the rhythmic organization of an utterance, not contributing to its meaning, [2] claimed that prenuclear accents are optional, or 'ornamental', in many cases (especially on prefocal elements). However, recent studies reveal that prenuclear accents are placed consistently and that they do, in fact, vary in their form depending on information structural differences. For German, it was shown that prenuclear accents on given referents were produced with slightly lower f0 peaks compared to those placed on contextually new information [3]. Furthermore, and relatedly, [4] found that accents on contrastive topics had later and higher f0 peaks than accents on non-contrastive topics.

The most recent production study on German [5] has shown that while prenuclear accents are indeed placed for rhythmic reasons, they are nevertheless sensitive to changes in informativeness, which were controlled for by rendering the sentence-initial referents as either *given*, *accessible*, *new* or *contrastive*. In fact, nearly all target referents, even given ones, were accented, indicating a structural need for accents in this position. Moreover, there was a strong preference for rising prenuclear accents, although newer items were produced with a wider range and a steeper rise. Contrastively focused, i.e. allegedly most informative, referents were longer than target words in the other conditions. Surprisingly, however, the other continuous prominence-cueing phonetic parameters such as tonal range and slope, intensity and tonal center of gravity [6], contribute to a less prominent realization of contrastive topics (which can be explained by the fact that the contrast was already expressed by a parallel syntactic and semantic-pragmatic structure of the stimuli).

In the current perception study we examine whether some accent types are more appropriate in certain contexts than other ones, i.e. whether participants' judgments of accent types change depending on the information structural context in which the accent was produced. Further, we assess if appropriateness ratings differ with regard to the mode of stimulus presentation, namely by asking participants either to judge accents following visually presented contexts, or to judge accents following visually *and* auditorily presented material. For this aim, we produced five stylized f0 contours on the prenuclear referent (deaccentuation, L*, H*, L*+H, L+H*; accent types following GToBI [7]) combined with an invariant contour on the nuclear referent (H*). These contours were then presented in the four information structural contexts already used in [5].

Based on previous findings on the relation between accent types and informativeness (not only [5] but also e.g. [8] on German or [9] on American English), we expect to find a general preference for rising accents as well as effects of informativeness: Deaccentuation should be preferred for *given* referents, low accents (L*) for *accessible* ones, high accents (H*) should be favored for *new* items, while rising accents (L*+H and L+H*) are expected to be preferred for *contrastive* items. As to the mode of context presentation we hypothesize to get clearer results in the auditory condition due to insights of prior psycholinguistic priming experiments (e.g. [10]). These suggest that auditorily presented items are stronger primes than visually presented ones due to an *echoic memory* effect [11].

2. Method

2.1. Speech material

2.1.1. Stimuli

The two experiments that were carried out separately operated on the same stimuli in order to tease apart possible effects of the

10.21437/SpeechProsody.2020-33

task (listening to audio recordings vs. silent reading). We chose 16 target words with their appertaining stories from a previous production experiment [5] as well as one training story. All target words had a disyllabic, trochaic structure and consisted of voiced segments only (including voiced obstruents). Each story was made up of three sentences and was modified with regard to the second sentence to obtain information structural differences. Context sentence one (C1) and the target sentence (T) were maintained in all conditions, while context sentence two (C2) varied in order to render the target word either *given*, *accessible*, *new* or *contrastive* (see Table 1).

Table 1: *Example story set. Target and trigger words are marked by italics.*

Context 1 (C1)	Nach dem langen Winter freuten sich alle auf ein paar sonnige Stunden im Freien. 'After the long winter everybody was looking forward to a couple of sunny hours in the open.'
Context 2a (C2a) given	Die *Nonne* kümmerte sich um den Klostergarten. 'The *nun* was looking after the cloister garden.'
Context 2b (C2b) accessible	Im *Klostergarten* blühten die ersten Pflanzen. 'In the *cloister garden* bloomed the first plants.'
Context 2c (C2c) new	Die Sonne schien schon den ganzen Tag und der Schnee war endlich geschmolzen. 'The sun had been shining all day and the snow had finally melted.'
Context 2d (C2d) contrastive	Der *Mönch* hat einen Brombeerstrauch gegossen. 'The *monk* watered a blackberry bush.'
Target (T)	**Die *Nonne* hat einen Mandelbaum gegossen.** 'The *nun* watered an almond tree.'

2.1.2. Recordings and manipulation

All 17 stories were produced by a trained native speaker of Standard German and recorded in a soundproof booth with 44,100 Hz sampling rate and 16 bit resolution. The speaker read the context paragraphs (C1 plus one of the four C2) as naturally as possible. This implied slightly rising accents on the initial referents. In the contrastive condition (which actually triggers a double contrast of subject and object in the target sentence), the context sentence (C2d) was produced with a more steeply rising L*+H pitch accent on the subject (in the example in Table 1 on *Mönch* 'monk'). The target sentences were read with five different tonal contours on the prenuclear referent. The nuclear part of the utterance was kept constant, since the perception experiment was meant to investigate the relation between prenuclear accents and informativeness. The five contours are exemplified in Figure 1 for the target sentence *The nun watered an almond tree*.

In order to obtain only minimally varying target items (i.e. excluding speaker-specific variability in the production) the recorded material was manipulated in *Praat* [12] with respect to the duration of silence between C1 and C2 (300 to 350 ms), the overall loudness of both contexts and target sentences, as well as the loudness, duration and especially the shape of the tonal contour of prenuclear target words.

When manipulating the tonal contours, target sentences were divided into a prenuclear and nuclear part, the prenuclear part ending with the last periodic waveform of the vowel /a/ in the auxiliary *hat* (negative zero crossing). The subsequent nuclear contour was manipulated once and then concatenated in *Praat* with each of the five (manipulated) prenuclear contours.

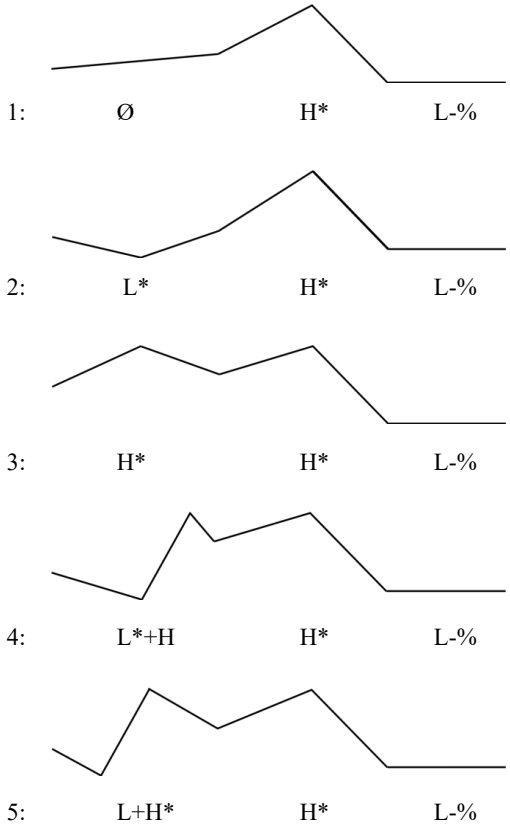

Figure 1: *Stylized contours of the target sentences. Accent types follow GToBI [7]. ∅ stands for deaccentuation. Symbols are associated with underlined syllables; the low boundary tone is associated with the final syllable of the sentence.*

The nuclear contour started with 125 Hz on the indefinite article (*eine* or *einen* 'a' or 'an') and reached the tonal peak at 165 Hz in the middle of the stressed vowel of the nuclear target word (in Fig.1 the /a/ in *Mandelbaum* 'almond tree'). The contour subsequently fell to around 100 Hz towards the third syllable of the target word. The rest of the sentence (i.e. the tail) remained unmodified.

Manipulation of the prenuclear contours began with the definite article (*die* 'the' in Fig.1) and used the following values: In deaccented target words, the prenuclear contour started at 110 Hz and ended at 120 Hz. It hence included only a shallow rise and almost no movement on the prenuclear target word. For L*, the prenuclear contour started at 110 Hz and the tonal valley was reached in the middle of the target word's stressed vowel at 95 Hz. The contour ended at 115 Hz. For H*, the contour started at 130 Hz and the tonal peak was reached in the middle of the stressed vowel at 165 Hz. The contour ended at 140 Hz and both rise and fall of the contour were strictly

linear. For L*+H-accents, the contour began at 115 Hz and the tonal valley was aligned in the middle of the stressed vowel at 95 Hz. The tonal peak was reached in the middle of the target word's unstressed vowel at 165 Hz, and the contour ended at 145 Hz. The distance between valley and peak ranged around 150 ms. For L+H*-accents, the contour started at 115 Hz and the tonal peak was reached at the end of the stressed syllable at 165 Hz. The tonal valley was aligned 130 ms before the peak at 95 Hz. The contour ended at approximately 135 Hz.

2.2. Procedure

In each experiment, 52 subjects participated (experiment 1: 22 f, 30 m; experiment 2: 30 f, 22 m), with an average age of 31 years (range: 18-66 years). The task in both experiments was to judge as to how well the target sentences match the context in which they were presented. In order to reduce the amount of stimuli for each subject, they were only presented with two (pseudo-randomized) conditions each, adding up to 160 test sentences per subject (16 target words * 2 conditions * 5 f0 contours). Following a short training section, the stimuli were displayed in a randomized order. However, the mode of stimuli presentation (of the context sentences) differed: In the first experiment, participants had to *listen* to the recorded context sentences (and could do so maximally twice), whereas in the second experiment they had to silently *read* these contexts before judging the (audio) target sentence. Appropriateness ratings were assigned via a slider bar, ranging from 'very bad match' to 'very good match'. The slider bar returned a continuous measure of the appropriateness rating.

Both experiments were conducted via the online platform *SoSci Survey* [13] and the participants were recruited using *Prolific* [14].

2.3. Analysis

In total, we received appropriateness ratings for 16,640 tokens (160 items * 104 participants). However, two participants of experiment 1 had to be excluded from analysis due to hearing issues which we assessed via a questionnaire. An additional of eleven participants (nine in experiment 1, two in experiment 2) had to be excluded since they completed the task too quickly (less than 15 minutes for rating 160 items) and did hence not produce reliable outcomes. Consequently, 14,560 ratings entered the analysis.

For the statistical analysis of the data, we used linear mixed effects modelling with the *lme4* package [15] in *R* [16]. In a first step, we investigated whether the type of context stimulus presentation had an impact on the outcome. Accordingly, STIMULUS TYPE (auditorily vs. visually presented context sentences) was used as the fixed factor. Since we expect specific combinations of accent types and levels of informativeness to affect the appropriateness ratings, we further tested for an interaction of ACCENT TYPE and CONDITION. In addition to these fixed effects, random intercepts were included for participant and target word.

3. Results

Comparing the results of experiment 1 and experiment 2 shows that the mode of context presentation (auditive vs. visual) does not affect the outcome ($\chi^2(1) = 1.7113$, $p = .19$), i.e. silent reading did not lead to different appropriateness ratings than listening to an audio file (see Fig.2). There is only a tendency for all accent types in the *contrastive* condition to be relatively less appropriate when the context was presented auditorily, especially with a deaccented target word.

When investigating the effects of condition and accent type, we observe two general patterns: First, the rising accent L*+H is preferred for all conditions, as it reaches the overall highest appropriateness ratings; second, the accent type H* is dispreferred for almost all conditions, as it was judged as least appropriate for *given*, *accessible* and *contrastive* referents (see Table 2).

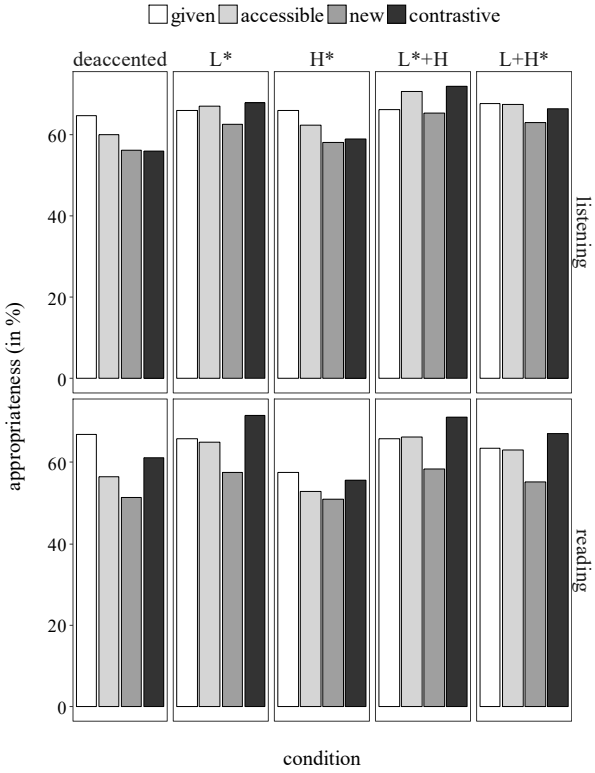

Figure 2: *Averaged appropriateness ratings (in %) of conditions as a function of accent type, separated by mode of context presentation.*

Table 2: *Averaged appropriateness ratings as a function of condition and accent type.*

	Given	Accessible	New	Contrastive
Deacc.	65.8 %	58.1 %	53.4 %	58.6 %
L*	65.8 %	65.9 %	59.7 %	69.7 %
H*	61.5 %	57.3 %	54.1 %	57.2 %
L*+H	65.9 %	68.2 %	61.5 %	71.4 %
L+H*	65.5 %	65.0 %	58.6 %	66.7 %

Besides these general (dis-)preferences, we observe several subtle differences in the appropriateness judgments depending on both the CONDITION and ACCENT TYPES. To begin with CONDITION, Figure 2 and Table 2 show that participants slightly dispref high tonal targets (H*) but prefer deaccentuation and low tonal targets (L*, L*+H) when the referent is *given* (more clearly in the silent reading mode). For referents in the three other conditions, participants prefer rising contours (L*+H, L+H*) as well as monotonal low accents (L*). However, the

appropriateness values for *new*ly introduced referents are in general lower than in the other conditions.

Turning towards the appropriateness ratings as a function of ACCENT TYPE, we observe that deaccentuation is best rated when the referent is *given* (65.8 % opposed to 53.4 % for *new* referents). L* is rated best for referents in a *contrastive* context (69.7 %), whereas H*, which receives the overall lowest appropriateness ratings, achieves its best rating when it is produced in a context where the referent is *given* (61.5 %). The rising accent L*+H is rated as most appropriate when the referent is either *accessible* (68.2 %) or *contrastive* (71.4 %). Finally, the rising accent L+H* is perceived as equally appropriate in all contexts in which the referent is either *given*, *accessible* or *contrastive*, i.e. it is only judged as somewhat less appropriate when the referent is *new*ly introduced. These described patterns strongly suggest an interaction of CONDITION and ACCENT TYPE, and the linear mixed effects analysis confirms this interaction with $\chi^2(19) = 765.52$ and $p < .0001$.

4. Discussion

In this study we investigated the interrelation of prosody and meaning in prenuclear accents. For this purpose, five different accent types were presented in four different information structural contexts and were then rated regarding their appropriateness on a continuous scale. In accordance with previous results of a production study [5] we hypothesized that we would find a general preference for rising accent types as well as a positive correlation between a referent's informativeness and its prosodic marking. To be more precise, we expected the least prominent realization (i.e. deaccentuation) to be judged as most appropriate when the referent is *given*, whereas the most prominent accent types (like L*+H and L+H*) should be judged as more appropriate in more restrictive contexts, e.g. when the referent is produced in contrastive focus. Furthermore, we investigated possible effects of stimulus presentation in that we conducted one experiment in which participants listened to pre-recorded context sentences and one experiment in which the subjects read the context sentences for themselves before listening to the target sentence.

The results of this study showed that the type of context stimulus presentation (auditory vs. visual) did not significantly impact the appropriateness ratings. Nevertheless, the tendency to judge accents – but especially the *lack* of an accent – as less appropriate for *contrastive* items after an auditorily presented context supports the hypothesis that *echoic memory* plays a role here: In the contrastive context sentence, the sentence-initial referent carried a prominent L*+H accent, supposedly triggering the expectation of the same prenuclear accent type on the target word. Thus, deaccentuation on this item, i.e. the least prominent realization, should be particularly inappropriate.

As to the actual ratings, our first hypothesis that subjects would prefer rising accents in prenuclear position was partly confirmed by the data. The rising accent L*+H generally received the highest scores. The other type of rising accent, L+H*, however, received somewhat lower values. This is interesting since another perception study [17] revealed that L+H* was judged as the most *prominent* accent type in German, slightly more prominent than L*+H. It has to be taken into account, though, that the above-mentioned study investigated the degree of perceived prominence in the *nuclear* region, and not in *pre*nuclear position. Possibly, L+H* evokes an impression of prominence exceeding what is expected in the prenuclear region and is therefore dispreferred. Furthermore, the combination of L+H* with the less prominent H* in the nuclear region could sound inappropriate as it might highlight the prenuclear accent to a disproportionate extent. In contrast, L*+H is not only *less prominent* than L+H* but also *more frequent* in prenuclear position (as a closer analysis of the production data in [5] reveal), making this accent type sound particularly appropriate for the subjects in the present study.

Besides this general preference for L*+H, we indeed found an interaction of informativeness and accent type that affected the appropriateness ratings. Yet, our expectations were not fully met by the data. As expected, deaccentuation was judged as most appropriate for *given* referents while rising accents were most appropriate for *contrastive* referents but also for *accessible* and *given* referents (particularly L*+H).

However, especially the monotonal accents L* and H* did not meet our assumptions. In general, H* was the most dispreferred accent type throughout the experiment. Most interestingly and quite unexpectedly, H* received its highest scores when it marked a *given* referent, and its lowest scores when marking a *new* referent. The relative inappropriateness of H* could be due to the manipulation of the audio material. Perhaps the contour comprising two monotonal high accents, with nearly identical f0 movements which were held strictly linear in both the prenuclear and nuclear region, could have sounded less natural to the listener compared to the other contour types which display more variation.

The low accent (L*) was dispreferred by participants for marking *new* referents and preferred for marking *accessible* as well as *given* referents, which is to a large extent in line with our assumptions. Yet, it was also judged as appropriate for marking *contrastive* referents, thus forming a functional group with the rising accent types L*+H and L+H*. In fact, our subjects may have re-interpreted the contour following the low monotonal accent: Since the nuclear accent in each target sentence was a monotonal high accent, the sequence of L* plus H* was probably perceived by participants as a rising contour as well, although with a shallower rise.

One last notable finding of this study was that *new* referents received the lowest appropriateness ratings for each accent type. A possible explanation for this pattern is that listeners generally disprefer an out-of-the-blue introduction of *new* referents, especially in comparison with the other, more contextualized, versions.

5. Conclusion

This investigation confirmed our hypothesis that listeners perceive accents in the prenuclear region dependent on the information structural context. As expected, deaccentuation was favored when referents were *given* while more prominent accent types, e.g. L*+H, were preferred in target sentences where referents are contextually required to get emphasized.

The findings of this study strengthen the assumption formulated in [5] that prenuclear accents are placed mainly for rhythmic reasons but that they are still sensitive to information structural changes.

6. Acknowledgements

This research has been funded by the German Research Foundation (DFG), Project BA 4734/1.

7. References

[1] S. Calhoun, "The Centrality of Metrical Structure in Signaling Information Structure: A Probabilistic Perspective," *Language*, vol. 86, no. 1, pp. 1-42, 2010.

[2] D. Büring, "Intonation, Semantics and Information Structure," in G. Ramchand and C. Reiss (eds.), *The Oxford Handbook of Linguistic Interfaces,* Oxford: Oxford University Press, pp. 445-474, 2007.

[3] C. Féry and F. Kügler, "Pitch accent scaling on given, new and focused constituents in German," *Journal of Phonetics*, vol. 36, no. 4, pp. 680-703, 2008.

[4] B. Braun, "Phonetics and phonology of thematic contrast in German," *Language and Speech*, vol. 49, no. 4, pp. 451-493, 2006.

[5] S. Baumann, J. Mertens, and J. Kalbertodt, "Informativeness and speaking style affect the realisation of nuclear and prenuclear accents in German," *Proceedings of ICPhS 19, Melbourne, Australia*, 2019, pp. 1580-1584.

[6] J. Barnes, N. Veilleux, A. Brugos, and S. Shattuck-Hufnagel, "Tonal Center of Gravity: A global approach to tonal implementation in a level-based intonational phonology," *Laboratory Phonology*, vol. 3, no. 2, pp. 337-383, 2012.

[7] M. Grice, S. Baumann, and R. Benzmüller, "German intonation in autosegmental-metrical phonology," in S.-A. Jun (ed.), *Prosodic typology: The phonology of intonation and phrasing,* Oxford: Oxford University Press, pp. 55-83, 2005.

[8] S. Baumann, and A. Riester, "Coreference, Lexical Givenness and Prosody in German," *Lingua*, vol. 136, pp. 16-37, 2013.

[9] J.B. Pierrehumbert, and J. Hirschberg, "The Meaning of Intonational Contours in the Interpretation of Discourse," in P.R. Cohen, J. Morgan, and M.E. Pollack, (eds.), *Intentions in Communication,* Cambridge: MIT Press, pp. 271-311, 1990.

[10] S. Baumann, and K. Hadelich, "Accent Type and Givenness: An Experiment with Auditory and Visual Priming," *Proceedings of ICPhS 15, Barcelona, Spain*, 2003, pp. 1811-1814.

[11] W. Chafe, *Discourse, Consciousness, and Time*. Chicago/London: University of Chicago Press, 1994.

[12] P. Boersma, and D. Weenink, *Praat. Doing phonetics by computer*, Version 5.4.06, 2015.

[13] D.J. Leiner, *SoSci Survey*, Available from https://www.soscisurvey.de, Version 2.4.00-I, 2014.

[14] Prolific, Available at: https://www.prolific.co, 2019.

[15] D. Bates, M. Maechler, B. Bolker and S. Walker, *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.1-8, 2015.

[16] R Core Team, *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria, 2017.

[17] S. Baumann, and C. Röhr, "The perceptual prominence of pitch accent types in German," *Proceedings of ICPhS 18, Glasgow, UK*, 2015, paper number 298, pp. 1-5.