Variability in Perception of Prosody: Focus Marking Mismatches in German

Farhat Jabeen, Jutta M. Hartmann, Petra Wagner

University of Bielefeld

Abstract

The interplay between accentuation patterns and information structure has received considerable attention in the field. The current study aims to investigate the extent to which German listeners are tolerant of (mis)matching accentuation presented in wide and contrastive narrow focus contexts. In a perception experiment, we presented our participants with sentences produced with congruent accentuation in terms of the elicited focus context as well as with mismatching accentuation. The mismatches comprised accent type only or both accent type and position. The participants rated the appropriateness of each accentuation pattern presented in narrow or wide focus contexts. The analysis of their ratings showed that German listeners are sensitive to (mis)matches in accentuation type and position in different focus contexts. The listeners dispreferred mismatching accent type in wide focus when L* nuclear accent was replaced with L+H*. However, replacing L+H* in narrow focus with L* received high appropriateness ratings. Participants also preferred the earlier placement of nuclear pitch accent in wide focus, while dispreferring the mismatching accent position in narrow focus context. Furthermore, they varied in their tolerance for different types of mismatching input. This indicates that German listeners are sensitive to variation in terms of accent type and position as well as focus marking.

Index Terms: contrastive narrow focus, wide focus, accentuation, Information Structure, mismatch, German

1. Introduction

Prosody and information structure marking as well as their interaction have been major issues in the last decades in linguistic research. Different studies on German intonation have investigated accent type and placement to convey focus as well as deaccentuation to mark givenness [1, 2, 3]. In a sentence produced in wide focus, the position of the nuclear accent in German has variously been described as the right-most accent in an Intonational Phrase (IP) [4], realized on the final argument in an IP [5], or the final pitch accent in an intonation unit [6]. From an information structural perspective, new and/or contrastive information usually carries the nuclear accent in German [7, 8, 9]. This results in similar nuclear accent placement for wide and a subset of narrow focus contexts.

Furthermore, utterances allow for a wide variety of (nested) information structural conditions [10] and listeners are confronted with varied accentuation patterns in less controlled environments [11], [12] have shown that listeners pay attention to variation in accentuation and in information structure. They played audio files without a focus-triggering context to their participants and, using Rapid Prosody Transcription, asked the respondents to identify the prominent words. Their analysis showed that discrete prosodic factors i.e. accent type, position, and the presence/absence of accent were highly associated with prominence marking. (1) shows the relevant hierarchy for perceived prominence in German. [12] point out that this ranking alludes to the interplay between accent type (high vs. low) and position (e.g. prenuclear) in the perception of prominence.

(1) High + rising nuclear accents > Low + falling nuclear accents > Prenuclear accents > Deaccentuation

Regarding the processing of accentuation, ERP studies have reported listeners’ neuro-cognitive reactions to unexpected accentuation patterns [13, 14]. [6] offers a detailed analysis of different degrees of perceived givenness and how it affects the perception of prosodic prominence. Using ERP patterns, they studied the processing of first occurrence focus (+new, +focus), second occurrence focus (-new, +focus), quasi second occurrence focus (+new, ±focus), and background (-new, -focus) in German. They found that pitch accents presented in first occurrence focus were processed differently than lack of accentuation in second occurrence focus (phrase accents marked by longer duration only) and deaccentuation of background information. They reported increased processing effort for lack of accents and deaccentuation as compared with nuclear pitch accents (2).

The authors explained that prominence marking indicated by pitch accents leads to less processing cost for listeners as the relevant information is available at hand.

(2) Processing cost for prominence:
Lack of accentuation/deaccentuation > Nuclear accents

These studies show that German listeners are sensitive to the type and position of pitch accents as well as to different degrees of prominence. Moreover, this difference in perceived prominence results in varied processing costs. While listeners are used to variable accentuation in different contexts, there is less research on how they deal with different types of mismatches of accentuation and information structure. In this study, we investigate whether listeners are more likely to accept mismatching accentuation and information structural input, when the prosodic information is partially congruent (e.g. congruent position of the nuclear accent vs. congruent accent type).

We report appropriateness ratings for wide and two positions of contrastive narrow focus accentuation presented in wide vs. narrow focus question contexts. For each target context, we used a matching and two mismatching accentuation patterns. A mismatch refers to an incongruity in the context of original recording and the context used to present it to our participants.

1.1. Hypotheses

Based on the studies discussed above, we anticipate that our participants will vary their appropriateness ratings on the basis of (mis)matching accent type, position, and information structural input (cf. 2.1). Therefore, we hypothesize the following:

1. In the wide focus context, sentences with L+H* on N2
(mismatching accent type only) will receive higher ratings than the ones with L+H* nuclear pitch accent on N1 (mismatching accent type and position).

2. In the contrastive narrow focus context, we expect that sentences with L* pitch accent on N2 (mismatching accent type) will be rated higher than their variants with L+H* nuclear pitch accent on N1 (mismatching position and accent type as N2 is deaccented).

2. Methodology

2.1. Stimuli

For the experiment, we used six transitive sentences containing a complex direct object which embedded a possessive Prepositional Phrase (PP). Each target sentence was produced in wide focus as well as in contrastive narrow focus. The former was elicited by using a wh-question inquiring the general state of affairs (3-a). Contrastive narrow focus was induced by a wh-question about the direct object (3-b). The wh-question was followed by an alternative question offering two proper nouns as possible answers, targeting the possessors in the embedded PP. A congruent answer to such a question specifies a noun only from the available set.

(3) a. Was ist bei Euch im Dorf los?

   what is with you in village up
   ‘What’s happening in your village?’

   b. Wessen Bienen hat der Imker
do bees have the.Nom beekeeper
   letztes umgesiedelt? Die von Lena oder
   last time relocate.Part the of Lena or
die von Nora?
   the of Nora
   ‘Whose bees did the beekeeper relocate? Those
   from Lena or those from Nora?’

A female native speaker of German produced the target sentences with three accentuation patterns: Wide focus (4)a, contrastive narrow focus on the object (N1) shown in (4)b, or the embedded object in the PP (N2) given in (4)c. The F0 contour of sentences in (4) is given in Figure 1. It shows that this target sentence, when produced in wide focus, carries an H* pitch accent on N1 and an L* nuclear pitch accent on N2. On the other hand, the N1 or N2 produced in contrastive narrow focus carries an L+H* pitch accent.

(4) a. Der Imker hat letztens die Bienen von Lena umgesiedelt.

   Wide focus
   ‘The beekeeper recently relocated Lena’s bees.’

   b. Der Imker hat letztens die Bienen von Lena umgesiedelt.

   Prominent N1
   ‘The beekeeper recently relocated Lena’s bees.’

   c. Der Imker hat letztens die Bienen von Lena umgesiedelt.

   Prominent N2
   ‘The beekeeper recently relocated Lena’s bees.’

To record the stimuli, the speaker was presented with the target sentences on MS Ppt slides and directed to produce those sentences as naturally as possible. In order to make sure that the speaker knew where to produce contrastive narrow focus, the target noun was presented in bold face. The sentences with wide focus were displayed entirely in bold face. Along with the target sentences, thirty-two distractors were also recorded by the same speaker.

Figure 1: F0 contour of a segment identical sentence produced in wide focus, contrastive narrow focus on the object (N1), and on the embedded object (N2).

2.2. Apparatus

The online perception experiment was set up using PsyToolkit [15, 16]. The participants were presented with a wh-question in written form eliciting either contrastive narrow focus on N2 or wide focus. Each question was followed by an audio stimuli. The participants were instructed to read the question displayed on the screen, play the answer, and rate the appropriateness of the audio as an answer to the given question. The target sentences were to be rated on a five-point Likert scale (1 = Not appropriate; 5 = Appropriate).

Every target context had one matching and two mismatching accentuation patterns e.g. wh-question inducing wide focus paired with (a) declarative sentence with matching prosody, (b) declarative with contrastive narrow focus on N1 (mismatching L+H* instead of H*), and (c) declarative with narrow focus on N2 (mismatching L+H* instead of L*). Similarly, the wh-question inducing contrastive narrow focus was presented either with a declarative with matching or mismatching prosody (wide focus with L* on congruent noun, narrow focus on N1 with L+H* & deaccented congruent noun). Noticeably, narrow focus realization on N1 was prosodically mismatching in both
wide and contrastive narrow focus contexts. The target sentences were interspersed with nine distractors.

The online survey was divided into three parts: Training phase consisting of five sentences, the test phase, and finally the questionnaire for demographic information. To avoid an overly long test phase, target sentences were divided into two lists. Each list comprised eighteen target sentences (2 focus contexts x 3 accentuation patterns x 3 items) and nine distractors. The lists were devised so that each participant was presented with all the items but in different focus contexts and accentuation patterns. In each list, the target sentences and distractors were presented in a pseudo-random fashion. Participants were assigned randomly to either list. The average duration of the experiment was nine minutes.

2.3. Participants
Twenty-eight German speakers (25 females) participated in the online survey. The participants were aged between twenty-one to fifty years. They were given a lottery-based prize for their participation in the experiment.

2.4. Data analysis
The statistical analysis of participants’ ratings of target sentences was carried out using Linear Mixed Effects Regression (LMER) models [17] in R [18]. The context (matching, mismatching) was added as a fixed factor and an interaction was run with accentuation patterns (wide, L+H* on N1 or N2). Items and participants were used as random effects in the models. Random slopes were kept in a model when found to be significant at the α-level of 0.05. The comparison between multiple levels of accentuation patterns was carried out using the ‘emmeans’ package which used the Tukey method of adjustment.

For the descriptive analysis of listener based variation in their ratings of stimuli, the first author went through individual speakers’ responses and categorised them as expected vs. unexpected as per our hypotheses given in subsection 1.1. Higher ratings for sentences presented in matching contexts were analysed as expected, whereas better ratings for sentences presented in one mismatching context vs. another were categorised as unexpected. The complete explanation for each of these categories in wide and contrastive narrow focus contexts is presented later in subsection 3.2.

3. Results

3.1. Regression analysis
The regression analysis showed a significant interaction between the question context and (mis)matching accentuation pattern (p = 0.001). The average ratings of the target sentences presented in wide and narrow focus context are illustrated in Figure 2. The figure shows that when the stimuli were presented in the matching wide focus context, they were rated significantly better (µ = 3.73, SE = 0.1) as compared with the sentences produced with contrastive focus accentuation on N1 (µ = 3.15, SE = 0.2, t.ratio = 2.6, p_{Adjusted} = 0.03) and on N2 (µ = 2.73, SE = 0.2, t.ratio = 4.0, p_{Adjusted} = 0.001). Moreover, the sentences with contrastive narrow focus accentuation on N1 were rated significantly better than the sentences with similar accentuation on N2 (β = -1.79, SE = 0.2, t.ratio = -8.0, p_{Adjusted} < 0.0001). The low number of participants in this group shows that this lack of distinction between contexts and accentuation is not an artifact of our experimental set up. The second unexpected finding in this context

Figure 2: Average ratings of sentences presented in wide and contrastive narrow focus (NaF) contexts. 1 = Inappropriate; 5 = Appropriate. The whiskers indicate 95% Confidence Interval.

3.2. Descriptive analysis
The descriptive analysis of listener based variation showed that only 43% (n = 12) of the participants were consistent in their responses in both wide and narrow focus contexts. There is overlap among participants in the categories given in the tables below. A given participant might have behaved as expected in one context for a few items but unexpectedly in another context or different items in the same context. Therefore, the numbers in each table do not add up to 28.

Table 1 presents the response categories when stimuli were presented in the wide focus context. It shows that as expected, the target sentences with matching wide focus accentuation were rated better as compared with their counterparts containing contrastive narrow focus accentuation on N1 or N2. On the other hand, two participants showed no variation in the rating of target sentences regardless of their accentuation pattern. This is categorised as an unexpected finding as we had presumed that mismatching accentuation pattern will lead to lower ratings in comparison with matching accentuation.

The low number of participants in this group shows that this lack of distinction between contexts and accentuation is not an artifact of our experimental set up. The second unexpected finding in this context
is that the target sentences with contrastive narrow focus on N1 (mismatching position and type of nuclear pitch accent) were rated higher than their variants with narrow focus on N2 (mismatching nuclear accent type only). The higher rating of N1 prominence may not be explained with reference to nuclear accentuation as this would predict better ratings for sentences with narrow focus on N2 i.e. the final argument in the IP [5]. The most unexpected finding in this context is when six participants rated target sentences with matching accentuation as less appropriate than the sentences with mismatching accentuation (L+H* on N1 or N2).

Table 1: Speaker variation in ratings when stimuli were presented in wide focus context.

<table>
<thead>
<tr>
<th></th>
<th>Expected</th>
<th>Unexpected</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Matching &gt; N1 and/or N2 with L+H* (n = 18)</td>
<td>No variation (n = 2)</td>
</tr>
<tr>
<td></td>
<td>N1 &gt; N2 with L+H* (n = 11)</td>
<td>Matching &lt; N1 and/or N2 with L+H* (n = 6)</td>
</tr>
</tbody>
</table>

Table 2 shows the analysis of speaker related variation in ratings when the target sentences were presented in the contrastive narrow focus context. It shows that a large number of participants behaved as expected and rated the target sentences with matching prosody as more appropriate than the sentences with narrow focus accentuation on N1 as well as the sentences originally produced in wide focus. However, there is one participant who, unexpectedly, did not respond to experimental manipulation at all. Furthermore, three listeners rated the sentences with matching accentuation as less appropriate than their variants carrying narrow focus on N1 or wide focus accentuation.

Table 2: Speaker variation in ratings when stimuli were presented in contrastive narrow focus context.

<table>
<thead>
<tr>
<th></th>
<th>Expected</th>
<th>Unexpected</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Matching &gt; L+H* on N1 (n = 24)</td>
<td>No variation (n = 1)</td>
</tr>
<tr>
<td></td>
<td>Matching &gt; wide focus (n = 16)</td>
<td>Matching &lt; wide focus (n = 3)</td>
</tr>
</tbody>
</table>

Interestingly, two participants consistently preferred prosodically and contextually mismatching input in wide and narrow focus contexts. Our participants also showed idiosyncrasies in their preference for particular accentuation. For example, one participant consistently preferred wide focus accentuation, whereas two of them dispreferred it in both wide and narrow focus question contexts. Furthermore, two participants consistently down rated narrow focus accentuation on NP1. For two participants, wide focus accentuation was always acceptable regardless of its context. But their rating of L+H* pitch accent on NP2 varied as they preferred it only in the matching context.

4. Discussion

The results presented above show an interesting interplay between wide and contrastive narrow focus contexts and different patterns of accentuation in German. We found that our participants predictably preferred the input that matched in terms of context and accentuation. Furthermore, they distinguished between different types of accentuation and contextual mismatches. In the contrastive narrow focus context, mismatches of nuclear accent type (L* on N2 in wide focus accentuation) were deemed as more appropriate compared with wrongly placed L+H* on N1 (mismatching position & type of nuclear pitch accent), thereby confirming our second hypothesis. This is interesting in view of [6]'s finding that higher prosodic prominence leads to reduced processing cost for the listeners. As contrastive narrow focus (-New information) is less prominent than the pitch accents (+New information) found in wide focus accentuation, the latter is less costly for listeners than the former. Moreover, participants preferred sentences with an L* nuclear pitch accent on N2 produced in wide focus than the sentences where N2 is deaccented after narrowly focused N1. This is also congruent with [6]'s findings who reported that German listeners showed lower processing cost for low nuclear pitch accents as compared with the lack of accentuation.

Our first hypothesis regarding wide focus context was not supported by our data as the sentences carrying L+H* nuclear pitch accent on N1 (mismatching accent type & position) were given higher ratings than their counterparts with L+H* nuclear pitch accent on N2 (mismatching accent type only). Thus, mismatching accent type is penalized by the listeners in this context, whereas they accept mismatching position of nuclear pitch accent on N1. This contrasts with the ratings of sentences presented in narrow focus context where the mismatching nuclear accent type was less problematic for our listeners.

This contrast in the ratings of mismatching accent type and position in wide and narrow focus contexts may be explained with reference to the prominence hierarchy given in (1) and processing cost (2). [6] have shown that the rising nuclear pitch accents are more prominent and thus less costly for listeners as compared with the low and falling pitch accents. Furthermore, prenuclear accents have a higher processing cost as compared with the nuclear pitch accents. Hence, when the prenuclear H* pitch accent on N1 produced in wide focus is replaced with the L+H* nuclear pitch accent (position based difference), this results in increased perceived prominence and lower processing cost for the listeners. This is also the case when L* nuclear pitch accent on N2 in wide focus is replaced with a rising nuclear pitch accent (accent type based difference). Although both these mismatches cost less processing effort, the reduction in this cost is greater in the former than the latter mismatch. Moreover, listeners showed variation in their (dis)preference and accepted prosodically “challenging” patterns at least in some contexts. This is reflected in their preference for placement of L+H* on N1 when presented in wide focus, whereas the sentences with L+H* on N2 were dispreferred in this context. However, this does not explain why the rising pitch accent on N2 is dispreferred when presented in the matching narrow focus context.

In the contrastive narrow focus context, the deaccentuation of N2 following the production of L+H* on N1 leads to higher processing cost for the deaccented noun. The very low appropriateness rating for this accentuation pattern could indicate that the loss of relevant information combined with high processing cost was penalized more severely than the replacement of accent type (L* on N2) in sentences produced in wide focus context.

5. Conclusions

This study illustrates that German listeners are sensitive to different types of incongruent contextual and prosodic information. Our results underline listeners’ competence to deal with the complexity of information structural contexts and accentuation variation they face in spontaneous everyday interactions.
6. Acknowledgements

We would like to thank the participants of the workshop on Creativity and Variability organized in conjunction with the Phonetics and Phonology in Europe 2021 (PaPE) conference as well as the TAI 2021 conference for valuable feedback on our work.

7. References


