Prosodic Cues and Varying Dialogue Context
Impact on Exhaustivity of Answers
Wollermann, Charlotte; Niebuhr, Oliver; Görs, Karin; Schröder, Bernhard

Published in:
Proceedings of the 10th International Seminar on Speech Production

Publication date:
2014

Document version
Early version, also known as pre-print

Citation for published version (APA):
Prosodic Cues and Varying Dialogue Context -
Impact on exhaustivity of answers
Charlotte Wollermann¹, Oliver Niebuhr², Karin Görs², and Bernhard Schöder¹
¹German Linguistics, University of Duisburg-Essen, Germany
²Department of General and Comparative Linguistics, ISFAS, Kiel University, Germany
charlotte.wollermann@uni-due.de, niebuhr@isfas.uni-kiel.de,
kgoers@arcor.de, bernhard.schroeder@uni-due.de

Abstract
The presented study investigates the contribution of dialogue context and prosody on pragmatic focus interpretation. More specifically, we test to what extent specific contextual variation combined with different pitch accent patterns (hat pattern versus dip pattern) biases the interpretation of answers towards exhaustivity or towards non-exhaustivity. For measuring focus interpretation we asked listeners to judge on scales, how well presented pictures match with the content of the dialogue. The results of our interpretation study suggest that contextual variation in combination with prosodic variation has a significant effect on pragmatic focus interpretation.

Keywords: pragmatic focus, prosody, dip pattern, hat pattern, exhaustivity of answers

1. Introduction
The presented study investigates the contribution of context combined with pitch accent patterns on the exhaustive interpretation of answers to preceding questions in a virtual dialogue scenario. In Section 1 we will present the theoretical background of our work, starting with definitions of focus and of exhaustive interpretation of answers. In this context, we will also discuss the role of prosody for pragmatic focus interpretation. Afterwards, we will introduce a particular type of intonational contrast, i.e. the hat pattern versus dip pattern distinction. In a next step, we relate the concept of focus to the hat pattern vs. dip pattern contrast. Finally, our assumption will be presented.

In Section 2 the material will be described that we use in our study. In Section 3 we present our experimental study, and finally, we summarize our results and discuss them in Section 4.

1.1. Focus interpretation and exhaustivity of answers
Defining the term focus is problematic since the term is used differently in the literature (see Krifka, 2007 and Fisseni, 2011 for an overview). In our approach, we investigate question-answering focus which constitutes a pragmatic focus. The latter is usually applied to that constituent in the answer which corresponds to the interrogative pronoun in the question. According to semantic-pragmatic theories (Rooth, 1992; Groenendijk, Stokhof, 1984), in the context of such a question, pitch accent correlates with focus. If the hearer interprets the (potentially implicit) background question as a mention-all question, an exhaustive interpretation should be triggered. Consider example (1), taken from van Rooij and Schulz (2004: 498).

(1) a. Who called yesterday?
   b. [Peter] called yesterday.

If the hearer takes (1b) to mean that Peter was the only person who called yesterday, the interpretation is exhaustive. But, if s/he infers that Peter was only one of a number of persons who called yesterday, then the interpretation is non-exhaustive.

The results of Fisseni (2011) suggest that focus accentuation does not suffice to trigger pragmatic focus interpretation. Rather, the expectations of the hearer, the language awareness of the hearer and, in particular, the context also play an important role.

In Wollermann (2012) the following was assumed: if the speaker signals uncertainty (e.g. by a rising intonation) when realizing the focus constituent, the hearer uses the prosodic cue and assumes that the speaker is uncertain with respect to his/her epistemic knowledge. Thus, the interpretation should be biased towards non-exhaustivity. Wollermann's findings show that there was an influence of prosody, but weaker than suggested theoretically. In contrast, the contextual effect was stronger. Based on these findings, a model of focus interpretation was developed (Wollermann, 2012; see also Wollermann et al., 2010). This model accounts on the relative contribution of context and of prosody for pragmatic focus interpretation. Information is processed top-down as well as bottom-up. The expectations of the listener are, for instance, affected by contextual information and processed top-down. The prosodic cues influence focus interpretation in the bottom-up direction. According to this model, the following prediction can be made: the influence of prosodic cues increases when the listeners expectations are less clear.

With respect to written language, the study of Skopeteas and Fanselow (2011) suggests that both question and answer type affect the exhaustive interpretation of answers.

In the current study, we use the same method as Wollermann (2012) for investigating the effect of contextual variation on focus interpretation in German, but this time we combine this variation with the intonational hat vs. dip pattern contrast. Both the context and the intonational pattern together constitute the independent variable.
1.2. Pitch accent patterns: Hat pattern versus dip pattern

According to the Kiel Intonation Model (KIM; Kohler, 1991) it is assumed that concatenation patterns in between two pitch accents are phonologically distinctive. Two basic types of concatenations are distinguished: dipped or non-dipped. A succession of two pitch accents with a high-level concatenation in between constitutes a hat pattern. All other patterns are dip patterns. Ambrazaitis and Niebuhr (2008) provided empirical evidence for this difference by showing that a change from hat to dip pattern is reflected in the perceptual and functional interpretations of utterances. Niebuhr and Zellers (2012) corroborate this evidence with different stimulus material, but suggest additionally that the pitch accent paradigms in hat patterns may not be the same as those in dip patterns. Production data of Grice et al. (2009) basically also support the relevance of a dip vs. no dip distinction before a non-initial pitch accent. However, the hat-dip contrast is not a separate phonological feature in GToBI (Grice, Baumann, 2005; Grice et al., 2005). Instead, this contrast is to a certain extent integrated into the pitch-accent paradigm.

1.3. Assumption

It was argued in Ambrazaitis and Niebuhr (2008) that a hat pattern has a bracketing function, whereas a dip pattern indicates a semantic-pragmatic detachment of two focused arguments of the coordination structure. For example, Niebuhr and Ambrazaitis conducted a perception experiment based on the target phrase ‘Samstag oder Sonntag’ (Saturday or Sunday) in an appointment-making scenario. If the phrase was spanned by a hat pattern, then listeners interpreted it as referring to the entire weekend, i.e. both days would suit, and even a two-days meeting would be possible. However, if there was a dip in between the two pitch accents on Saturday and Sunday, then listeners felt that they had to choose between one of them.

Therefore, when it comes to listing persons or entities in conversation, hat patterns are typically used in combination with complete lists, whereas dip patterns are preferably used for incomplete or basically open lists.

We assume on this basis that a hat pattern across two focus arguments in combination with a context conveying certainty about the exclusion of alternatives with respect to the focused item would bias the interpretation towards exhaustivity. In contrast, we expect that a dip pattern combined with a context conveying uncertainty about the exclusion of alternatives with respect to the focused item would trigger a non-exhaustive interpretation.

2. Material

The goal of this study was to investigate the effect of the intonational hat pattern versus dip pattern contrast and also the effect of context on focus interpretation in German. Therefore, question-answer pairs were generated. The answer constituted the utterance with the pragmatic focus and was characterized by a hat pattern or a dip pattern. In a next step, the question answer pairs were embedded into short dialogues, which were also varied.

The dialogues were all about a party at which groups of students (biologists, physicians, designers etc.) performed different actions. Every action described in the dialogue led to a target question in which one dialogue partner asked for the corresponding agents of that action. The target answer with the relevant focused constituents, i.e. a coordination of two student groups (noun phrases), was then provided by the other dialogue partner. An example is given in (2a) and (2b). The pitch accented syllables of the two focused arguments are underlined.

(2) a. Wer hat denn Karikaturen gemalt?
Who has drawn caricatures?

b. [Die Chemikerinnen und Ingenieure] F haben Karikaturen gemalt.
[The chemists and the engineers] F have drawn caricatures.

The dialogue frames provided two kinds of contexts.

(1) Certainty context: one student group is salient during the dialogue. In addition the target answer ended in a sentence in which the speaker signalled certainty with respect to the performing agents (e.g. I am sure that X were the only ones who did/performing as...).

(2) Uncertainty context: the group introduced at the beginning of the dialogue was not identical with the group mentioned in the target answer. Moreover, the target answer was followed by a sentence in which the speakers signalled uncertainty with respect to the performing agents (e.g. I am not really sure that X were the only ones who did/performing as...).

As for the intonational cues, the two coordinated noun phrases that functioned as focus constituents in the target answers were either spanned by intonational hat patterns or dip patterns, each which comparably prominent H* pitch accents on either focus argument (see Figure 1).

It should be mentioned here that we used scripted lab speech in our study, even though this kind of speech lacks the spontaneity of everyday speech. However, it is particularly important for us to vary contextual and prosodic cues in a systematic way, and to minimize the occurrence of confounding factors which may come along with spontaneous speech.
3. Experimental study

In this section we present our empirical study. Firstly, we describe our methodological approach. Afterwards, we outline our experimental procedure. Finally, our results are presented.

3.1 Method

Overall, we generated six different dialogues. There were two variants which differed regarding contextual and prosodic cues. For every dialogue, either variant \( exh^+ \) (exhaustive+) or variant \( exh^- \) (exhaustive-) was used (see Table 1). To test whether there is any evidence for our assumption at all, we use the two 'extreme' combinations of context and prosody in the current study.

That is, for the variant \( exh^+ \), there was only one student group salient during the dialogue. By ‘salient’ we mean here that this student group was frequently and consistently referred to so that other groups become less present in the listeners’ minds. Furthermore, the coordinated noun phrases in the answer with the pragmatic focus were spanned by a hat pattern. Also, a sentence indicating certainty about the exclusion of alternatives followed after the answer.

In contrast, in the case of variant \( exh^- \), one group of students was introduced at the beginning of the dialogue as usually performing the action under discussion. This group was not identical with the group mentioned in the focus utterance. For the realization of the coordinated noun phrases in the answer with the focus, a dip pattern was used. Furthermore, a sentence indicating uncertainty about the exclusion of the alternatives followed after the answer.

Table 1: Two variants of the stimuli with different contextual and prosodic cues.

<table>
<thead>
<tr>
<th>Variant</th>
<th>Context</th>
<th>Prosody</th>
</tr>
</thead>
<tbody>
<tr>
<td>variant ( exh^+ )</td>
<td>only one salient student group, lexical expression of certainty about the exclusion of alternatives</td>
<td>hat pattern</td>
</tr>
<tr>
<td>variant ( exh^- )</td>
<td>another student group mentioned as possible alternatives, lexical expression of certainty about the exclusion of alternatives</td>
<td>dip pattern</td>
</tr>
</tbody>
</table>

Furthermore, we generated three filler dialogues in order to distract subjects from our actual research question. An example dialogue was generated to make subjects familiar with the task.

For measuring focus interpretation we used the method of Wollermann (2012). Drawings were created that depicted either the exhaustive or the non-exhaustive reading. Subjects referred to these drawings when judging the dialogues. In this way, we wanted to avoid any complex and potentially biasing metalinguistic instructions of our subjects. Moreover, we wanted to avoid that the subjects’ linguistic awareness is directed to the goal of our experiment.

As already mentioned in Section 1.3, our hypothesis is the following: we assume that variant \( exh^+ \) biases the interpretation towards exhaustivity, whereas variant \( exh^- \) biases the interpretation towards non-exhaustivity.
3.2 Procedure

Subjects were 18 native speakers of German, 14 females, 4 males, who were on average 23 years old. They were undergraduate students at the University of Duisburg-Essen and participated in the experiment voluntarily at the beginning of a seminar. Each dialogue was played, and afterwards the subjects had to judge each time on a 5-point scale (with 1 = very bad and 5 = very good), how well the subsequently presented drawing matched with the dialogue. Distractor and example dialogues were also included in the experiment. All in all 9 dialogues were played. Dialogues 1, 2, and 5 were judged against drawings signalling exhaustivity. Dialogues 3, 4, and 6 were judged against drawings illustrating the non-exhaustive drawings.

A Wilcoxon Signed Rank Test was used for statistical analysis. Since we perform several comparisons on our data, we use a Bonferroni correction. Starting from a level of significance of \( p \leq 0.05 \), we applied a more conservative significance level, which reflected the number of tested comparisons: 0.05/4, i.e. \( p \leq 0.0125 \).

3.3 Results

The diagrams in Figure 2 show the results. The upper diagram of Figure 2 shows that our subjects judged dialogues 2 and 5 with a median of 5 each time. In contrast, the median for dialogue 1 was about 4 and hence lower. The Wilcoxon Signed Rank Test showed that this judgement difference is statistically significant. For dialogue 5 vs. dialogue 1, the p-value was <0.0025; for dialogue 2 vs. dialogue 1, it was <0.00025. So, the subjects' rankings were significantly higher when the drawing representing the exhaustive reading coincided with the dialogue variant \( \text{exh}^+ \) (dialogues 2 and 5). This variant is marked by contextual and intonational cues intended to bias the interpretation towards exhaustivity. The ranking was lower when there was a mismatch between the information given by the drawing and the information given by the dialogue (dialogue 1).

Figure 2: x-axis shows dialogues and the respective variant (\( \text{exh}^+ \) or \( \text{exh}^- \)), y-axis shows the median for recipients' judgements. Level of significance: \( p = 0.05/4 = 0.0125 \), *: \( p \leq 0.0125 \), **: \( p \leq 0.0025 \), ***: \( p \leq 0.00025 \)

As regards the bottom panel of Figure 2, we can see the following: dialogues 4 and 6 were ranked with medians of 3.5 and 4 respectively. Dialogue 3 yielded the lowest median of 3. The corresponding Wilcoxon Signed Rank Tests resulted in significant differences between the judgements of these two variants. For dialogue 6 vs. dialogue 3, the p-value was <0.0025; for dialogue 4 vs. dialogue 3, it was <0.0125. We can conclude from this outcome that judgements were significantly higher when the picture depicting the non-exhaustive reading was presented in combination with the contextual and intonational dialogue variant which was intended to bias the interpretation towards non-exhaustivity (dialogue 4 and 6). A lower ranking occurred in the case of a mismatch between the information illustrated in the drawing and that conveyed by the contextual and intonational cues (dialogue 3).

4. Discussion

We presented a study dealing with the contribution of varied dialogue contexts combined with different pitch accent patterns (\( \text{hat pattern} \) versus \( \text{dip pattern} \)) on the exhaustive interpretation of answers.

Our results suggest that the contextual variation in combination with intonational variation has a significant effect on the exhaustive interpretation of answers.

In the current study, we have not yet tested the two factors context and intonation independently from each other. The information transmitted by the speaker is regarded to be redundant in our approach. As Fisseni et al. (2013) suggest, the transmission of 'optimal' redundant information and its testing by means of empirical data can help to improve communication (for some earlier remarks on the role of redundancy for focus see also Fisseni, 2011: 227).

The question remains open what the relative contribution of context on the one hand, and \( \text{hat} \) vs. \( \text{dip} \) pattern on the other hand actually is. However, the intonational effect is clearly consistent with communicative function ascribed to the \( \text{hat} \) vs. \( \text{dip} \) contrast by Ambrazaitis and Niebuhr (2008). This also means that our results do not contradict the claim of the Kiel Intonation Model (KIM, Kohler 1991) to represent the intonational difference between \( \text{hat} \) and \( \text{dip} \) as a separate feature in intonational phonology, if a correlation between this prosodic distinction and its pragmatic interpretation can be established.

In our future work, we aim at testing all four combinations of the two factors context and prosody. We would like to collect more data by testing these four combinations for each of the six dialogues. Furthermore, for measuring focus interpretation, it would be interesting to use for each stimulus the drawing showing the exhaustive interpretation as well as the drawing showing the non-exhaustive interpretation.
Acknowledgements
We would like to thank Bernhard Fisseni and Nina Jeanette Hofferberth for helpful comments.

References


