Pragmatic Focus Interpretation:
Interplay between Context and Audiovisual Prosody?

Charlotte Wollermann\textsuperscript{1,2}, Bernhard Schröder\textsuperscript{2} and Ulrich Schade\textsuperscript{1,3}

\textsuperscript{1}Speech and Communication, University of Bonn, Germany
\textsuperscript{2}German Linguistics, University of Duisburg-Essen, Germany
\textsuperscript{3}Fraunhofer Institute for Communication, Information Processing and Ergonomics FKIE, Germany
cwo@sk.uni-bonn.de, bernhard.schroeder@uni-due.de, ulrich.schade@fkie.fraunhofer.de

Abstract

We present an experimental study on the effect of context and of audiovisual prosody on pragmatic focus interpretation. The audiovisual material was elicited from our previous production study on the exhaustivity of answers. Here we recorded nine German speakers who were instructed to read dialogues with embedded question-answer pairs in which the answer constituted the pragmatic focus of the utterance. We varied the context regarding (un)certainty and (non-)exhaustivity. It was characterized by the occurrence of different accent types accompanied by different visual cues. In our current study we presented the audiovisual material to 45 subjects to test to what extent the variation of audiovisual prosody and of context effects exhaustivity. Focus interpretation was tested by using pictures showing the (non-)exhaustive reading. Our results suggest an influence of the context and accentuation on the focus interpretation, but weaker than expected. Expected effects of eyebrow movement barely occur.

Index Terms: audiovisual prosody, exhaustivity, focus interpretation, contextual variation

1. Introduction

In this paper we empirically investigate the role of audiovisual prosody and of context for pragmatic focus interpretation. We address the following questions: Does audiovisual prosody affect the exhaustive interpretation of answers? Is there an interplay between accentuation, eyebrow movement and contextual information?

1.1. The concept of pragmatic focus

The term \textit{focus} often refers to the intuition that pitch accent correlates with new information in utterances, whereas old information is deaccented. This phenomenon can be observed in West Germanic languages like English, German or Dutch [e.g. 1]. Even though focus is a complex phenomenon and labelled differently [2], most theories agree that focus can be defined as the answer to an – explicitly or implicitly given – question [3]. This phenomenon is also referred to as \textit{pragmatic focus}.

In the framework of semantic-pragmatic focus theories [4, 5] it is often suggested that focus licenses an exhaustive interpretation of the focussed constituent(s). If the background question is interpreted as a \textit{mention-all} question, the precondition for an \textit{exhaustive} interpretation is given. In the case of \textit{exhaustive} interpretation the hearer of (1b) concludes that only John kissed Mary (and no one else). In the case of \textit{non-exhaustive} interpretation there might be also other persons from a set of alternatives who kissed Mary. The crucial factor for exhaustivity is the knowledge about the situation under discussion, which is ascribed to the speaker by the hearer [6].

(1a) Who kissed Mary?
(1b) John \textit{F} kissed Mary.

We assume that when the speaker signals uncertainty it is due to a lack of knowledge, e.g. she is not completely informed about the answer. In [4, 5] it is argued that, in the context of an adequate question, pitch accent is highly correlated with focus; then the interpretation should be biased towards exhaustivity. In these theories, contextual influence is generally not considered, or the immediate context, i.e. the type of question, is discussed exclusively.

In [2] a series of interpretation studies was conducted to test the influence of accentuation on pragmatic focus interpretation. Results suggest that the influence of accentuation is weaker than expected, but also the hearer’s expectations, the sensitization for focus phenomena and in particular contextual factors play a crucial role. For written speech, it was also found that the question and also the answer type influence the exhaustivity of answers [7].

However, it is less clear which role visual prosody plays for pragmatic focus interpretation.

1.2. Audiovisual prosody

The term \textit{visual prosody} has been used to broaden the definition of the traditional notation of prosody to the visual domain. It refers to facial expressions, arm and body gestures that communication partners use in face-to-face conversation [8]. In [9] the following definition is given: "Linguistic information in the text is used to generate visual cues for relevant prosodic categories such as prominence, phrasing and emphasis. These cues generally take the form of eyebrow and head movements [...]"

Different studies suggested evidence for the co-expressivity of speech and gesture or facial expression. According to the \textit{metaphor of up and down} [10: 202 ff] a rising or falling of pitch is accompanied by rising or falling of gesture and facial expression, e.g. eyebrow movement. The study of [11] suggests empirical evidence for this assumption, an interplay between rising of eyebrows and rising of pitch which is linguistically and communicationally motivated can be observed. Further, the data of [12] suggest a correlation between head movement and pitch.

However, accentuation and visual cues do not necessarily co-occur as prosodic cues. In [9] it was reported that eyebrow movement generated by means of a Talking Head can effect
the perception of prominence. In a follow-up study [13] it was found that that both eyebrow and head movement expressed by a Talking Head contribute to the perception of prominence. Also in the context of human-machine interaction, [14] investigated whether pitch accents and eyebrow movements expressed by a Talking Head effect the perception of focus of attention. Results show evidence for an influence of both cues, but the impact of accent is stronger.

1.3. Expression of uncertainty

The question of how speakers and listeners produce and perceive uncertainty in face to face communication has been investigated in several studies. [15, 16] reported that uncertainty is expressed and also perceived by different prosodic cues: rising intonation, delays, fillers and lexical cues. According to [17] fall-rise intonation contributes to a context-independent meaning of utterance interpretation conveying the speaker’s uncertainty in English. Regarding the visual channel smiles and funny faces, i.e. marked facial expressions, were found as indicators of uncertainty which are also perceptually relevant [18].

1.4. Assumption

We assume that if the speaker signals uncertainty in a question-answering situation by audio and/or visual cues, the hearer uses the prosodic information for decoding the utterance. The hearer will assume that the speaker is uncertain regarding the answer, then the interpretation should be biased towards non-exhaustively. In contrast it is assumed when the speaker produces audio and/or visual cues of certainty, the hearer will assume that the speaker is certain regarding the answer in this case the interpretation should be biased towards exhaustivity.

2. Related work

In [19] and [20] we tested the influence of intonation and context on the exhaustivity of answers. Our data provide evidence that the exhaustive reading is generally preferred. However, the interpretation is biased towards the non-exhaustive interpretation when the micro and macro context are biased towards non-exhaustivity and uncertainty. Compared to the contextual influence the prosodic influence on exhaustivity is relatively weak. From these findings we derived a model of pragmatic focus interpretation [21]: micro and macro context is relevant for raising the hearer’s expectations; these expectations are top-down influences and the prosodic information influences the focus interpretation bottom-up.

Since our data suggest a weaker influence of prosody on focus interpretation than expected, we conducted a production study [22, 23] in order to see what prosodic cues speakers realize when producing pragmatic focus utterances. The following accent types for realizing the focus constituent were found: L*+H was realized most frequently for the marking of the focus constituent, followed by H* or no accentuation. A few realizations of L* also occurred. The detailed analysis of the analysis of the frequencies of accent types can be found in [22, 23]. Regarding the visual modality we found a tendency that H* accompanied by a raising of eyebrows or head appears more often for the contextual variant intended to be biased towards uncertainty and non-exhaustivity than for the contextual variant intended to be biased towards certainty and exhaustivity. We interpreted this as a possible manifestation of the biological codes [24]: high pitch expresses uncertainty and continuation on the pragmatical level, whereas low pitch expresses certainty and finality. We use the material of that study for our current study.

In a next step, we presented the audio material to listeners to test to what extent the variation of accent type and of context affects the exhaustivity of answers [25]. Interpretation was tested by using pictures intended to illustrate the (non-)exhaustive reading. When presenting the picture illustrating the non-exhaustive reading, results show in general a significant influence of the context and accentuation, but the prosodic influence is weaker.

In our current study we present the audiovisual material from our production study [22, 23] to subjects for testing the impact of audiovisual prosody and of context on pragmatic focus interpretation.

3. Interpretation study

3.1. Goal

The aim of the present study is to test empirically to what extent audiovisual prosody and context effects the exhaustive interpretation of answers.

3.2. Material

As already mentioned, the audiovisual material was elicited in our production study [22, 23]. We video-recorded nine German speakers uttering focus utterances which were embedded in short dialogues. The dialogue partner was always the same female speaker. In total, there were six different dialogues with question-answer pairs.

The scenario was a fictitious party where different groups of students acted differently. For every action, there was a question asking about the agent and an answer providing the information. The focus exponent in the answer was either one NP referring to group of students (dialogue 1, 4 and 6) or coordinated NP referring to two groups of students (dialogue 2, 3 and 5). We refer to one group by focus sentence with one NP (noun phrase) and to two groups by focus sentence with (a coordination of) two NPs. The ratio was 50:50. However, for our current study we focus on the focus utterances with one noun phrase, i.e. three dialogues in total, for a first analysis of the role of audiovisual prosody and context on focus interpretation.

Two variants of context were generated. i) The variant I was intended to have a bias towards certainty and exhaustivity. It was characterized by contextual congruity. One student group was salient during the whole dialogue, a question followed which was congruent to the focus utterance (see 2a + 2b). No alternatives were given in the context. Further, a sentence indicating certainty about the answer followed (see 2c).

(2a) ... Wer ist zu spät gekommen? Who was late?
(2b) [Die Mathematiker]F sind zu spät gekommen. The mathematiciansF were late.
(2c) Das waren die Einzigen, die nicht pünktlich waren... They were the only ones not being in time...

ii) The variant II was characterized to have a bias towards uncertainty and continuation. It was marked by contextual incongruity: One “competing” discourse entity was introduced at the beginning of the dialogue (usually the linguists are late for parties) (see 3a). Thus the alternatives were explicitly given. A general question eliciting broad focus followed (see 3b).

The respective students group as focus of the answer (see 3c) was different from the salient discourse entity. A sentence indicating uncertainty about the answer followed (see 3d).

(3a) Wer war zu spät gekommen? Who was late?
(3b) [Die Mathematiker]F sind zu spät gekommen. The mathematiciansF were late.
(3c) Die Mathematiker F waren zu spät gekommen, die Linguisten F nicht. The mathematiciansF were late, but the linguistsF not.
(3d) Die Mathematiker F waren zu spät gekommen, die Linguisten F nicht. The mathematiciansF were late, but the linguistsF not.
Task 1
How well does the picture suit the dialogue?

very bad 1 2 3 4 5 very good

Task 2
As a host, which of the following behavior of your guests would you find acceptable?
A: showing up too late
B: not showing up at all (without an excuse)
C: bringing other guests without asking

Figure 1: Testing the focus interpretation of the stimulus Die Mathematiker sind zu spät gekommen/The mathematicans were late using the picture illustrating the exhaustive reading

Each student group was characterized by an unique accessory, e.g. mathematicians were wearing green glasses, linguists blue hats, geographers red wigs etc.

(3a) ...Die Linguisten waren immer zu spät...
...The linguists were always late...
(3b) Was ist passiert? What happened?
(3c) [Die Mathematiker sind zu spät gekommen] F.
[The mathematicans were late] F
(3d) Wenn ich mich nicht täusche, waren die Linguisten auch nicht pünktlich... If I am not wrong, the linguists were also not in time...

In contrast to our production study [22, 23], in our current study we removed the sentence indicating (un)certainty about the answer for each dialogue from our data for testing the influence of uncertainty as a paralinguistic expression on focus interpretation. Since we annotated the audiovisual material in [22, 23] with respect to accent type, eyebrow movement and head movement we now have for each utterance a prosodic annotation. However, in our following analysis we focus on the role of accentuation and eyebrow movement on focus interpretation.

In the production study [22, 23] there were 54 dialogues with embedded question answer pairs (9 speakers x 6 dialogues) and 18 filler-dialogues (9 speakers x 2 dialogues); the two filler-dialogues were always the same. As already mentioned for the purposes of the current paper we concentrate on the dialogues with one noun phrase (9 speakers x 3 dialogues).

3.3. Testing focus interpretation

We tested the focus interpretation by using pictures. There were two different types of pictures. For half of the stimuli (dialogue 1, 4 and 5) the pictures illustrating the non-exhaustive interpretation were presented. Here the student group performing usually the question under discussion was also illustrated in the picture.

From the subjects’ choice of the picture we inferred the preference of interpretation. By using this method we tried to avoid that the subjects’ linguistic awareness was focussed on the tested question. Thus influences of expectations of the subjects with respect to the goal of the experiment should be minimized.

3.4. Hypothesis

Based on our previous studies [19, 20, 25] we assume that the exhaustive interpretation is influenced by the context and the prosody. We expect that the degree of congruency between the information conveyed by the context and the audiovisual prosody effects the exhaustivity. It is assumed that a context with incongruent information (variant II) combined with H* or raised eyebrows (as prosodic indicators of uncertainty) for realizing the focus constituent is biased towards non-exhaustivity. In contrast we expect a context with congruent information (variant I) combined with L+H* for realizing the focus utterance to be biased towards exhaustivity.

3.5. Procedure

The audiovisual material was presented as an online experiment to 45 subjects; all of them students and German native speakers. Each subject watched the eight dialogues (6 dialogues with the focus utterances and 2 filler-dialogues) produced by the same speaker. This way, we have five subjects per speaker (5 x 9). After each dialogue was played, subjects had to judge on a 5-point Likert-Scale how well the picture suits the dialogue (1=very bad, 5=very good). For every picture there was a caption with the different students groups characterized by the appropriate accessory. Furthermore for each dialogue there was a

1None of the subjects participated in our previous experiment [25], where we presented the material under audio-only condition and tested the focus interpretation.
multiple-choice-question asking for the subjects’ personal opinion about an aspect of the dialogue functioning as a distractor (see figure 1).

As already mentioned, for the purposes of the current paper we analyzed exclusively the judgments for dialogues with focus sentences with one noun phrase. The results were statistically analyzed using the Wilcoxon Rank Sum Test since our data are ordinal-scaled and since we have different group sizes. Our level of significance was 5%.  

4. Results

The results for dialogues with focus sentences with one noun phrase as focus constituent are shown in figure 2. The relevant abbreviations are listed in table 1. For the description of the stimuli we use in the following the quadruple <a,b,c,d>; a refers to the picture and has either the value 1 (the picture illustrating exhaustivity) or 0 (the picture illustrating non-exhaustivity), b refers to the context and has also either the value 1 (variant I) or 0 (variant II), c refers to the accent type (see table 1) and d to the eyebrow movement (see also table 1).

When we present the picture illustrating the exhaustive reading the following can be observed: First, we concentrate on the recipients’ judgments for those cases when we present the picture illustrating exhaustivity combined with the context being biased towards exhaustivity and varied audiovisual prosody. <1,1,L+H*,E-> and <1,1,H*,ER-> are both ranked with a median of 4. Further, the median for <1,1,L+H*,ER-> and <1,1,H*,ER/F-> is each time 3; for <1,1,H*,E-> it is 3.5. None of the comparisons between the rankings yields to a statistical difference, p is each time > 0.05. Here we can neither observe an influence of accentuation nor of eyebrow movement on the focus interpretation.

Secondly, we compare the judgments for those cases when we use the picture illustrating the exhaustive reading combined with a context being biased towards non-exhaustivity and varied audiovisual prosody. <1,0,acc-,E-> and <1,0,H*,E-> obtain a median of 4; <1,0,acc-> and <1,0,L+H*,ER-> are judged with a median of 3. The statistical analysis shows that the judgments for <1,0,acc-,E-> vs. <1,0,L+H*,ER-> differ from each other in a marginally significant way with p<0.1. We infer that the absent audiovisual prosodic information, i.e. no accentuation and no eyebrow movement, might be stronger biased towards exhaustivity than the prosodic information which is ambiguous with respect to the context.

Further, it can be observed that <1,1,L+H*,E-> is judged as more adequate than <1,0,L+H*,ER->. This difference is again marginally significant with p<0.1. Here we find evidence for our hypothesis even though it is weak: the context influences the focus interpretation in favour of the exhaustive interpretation and the absent eyebrow movement as well.

When presenting the picture illustrating the non-exhaustive reading the following occurs: First, we compare the judgments for the stimuli which we present using the picture showing the non-exhaustive reading combined with the context being biased towards exhaustivity. The median of <0,1,L+H*,E-> is 3.5, whereas <0,1,L+H*,ER-> and <0,1,H*,E-> have a median of 2. The comparison <0,1,L+H*,E-> vs. <0,1,L+H*,ER-> shows a marginally significant difference with p<0.01. Against our expectation a raising of eyebrows does not favour the non-exhaustive reading.

Secondly, we concentrate on the recipients’ judgments for the stimuli when we also present the picture illustrating the non-exhaustive presentation in combination with the context which is also biased towards non-exhaustivity. <0,0,H*,EF->, <0,0,L+H*,E-> and <0,0,L+H*,ER-> are judged each time with a median of 2. There is no significant difference between the judgments, p is each time > 0.05.

With respect to the contextual influence the following can be observed: <0,1,L+H*,E-> is judged as more adequate than <0,0,L+H*,E-> with p<0.05 in a significant way. We interpret this result as evidence for a contextual influence on focus interpretation in favour of exhaustivity even though the picture illustrating the non-exhaustive reading is presented. In a similar way, the comparison <0,1,L+H*,E-> vs. <0,0,H*,EF-> shows a significant difference with p<0.05. Here we also conclude that the context and the accent effect the interpretation in favour of the exhaustive interpretation even though the picture illustrating the non-exhaustive reading is presented.

5. Conclusion

We presented an empirical study on the role of the variation of audiovisual prosody and of context on the exhaustive interpretation of answers. For these purposes we used pictures intended to show either the exhaustive or non-exhaustive reading. It was assumed that the degree of the congruity between the contextual information and the prosodic information effects the pragmatic focus interpretation. Our data show a tendency that the context and the prosodic information effect the exhaustivity of answers, but the effects are weaker than theoretically expected. What our data do not suggest is a specific prosodic pattern of accent and eyebrows in favour of (non-)exhaustivity. However, there is a tendency that L+H* effects the focus interpretation in favour of exhaustivity.

For a further investigation of this question we regard it as important to analyze our data with two noun phrases as focus constituent. It would be interesting to analyze the more complex prosodic patterns and to test to what extent an occurrence of the same prosodic patterns for realizing both noun phrases effects the focus interpretation.

With respect to our method we have to mention that we have several independent variables in our experiment (context, accent type, eyebrow movement), such that the comparisons of the current analysis are based on a limited number of recipients’ judgments. However, for our future work we regard it as a challenge to find a method to gain data which are as natural as possible. Here it would be useful to use spontaneous speech and to find a scenario in which exhaustive utterances are elicited. At the same time speakers should use uncertainty as a paralinguistic expression of their own epistemic state. On the other hand these data should be presented to listeners to test the influence on pragmatic focus interpretation.

6. Acknowledgements

We would like to thank Till Krempel for helping us with the experimental set-up. Many thanks to Bernhard Fisseni for helpful comments.

---

2It should be noted that we also performed a Kruskal Wallis Test on our data (p=0.05), but with this test it is not possible to test the influence of the single factors, i.e. context and audiovisual prosody.
Table 1: Accent types and eyebrow movements

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>H*</td>
<td>peak accent</td>
</tr>
<tr>
<td>L*</td>
<td>low accent</td>
</tr>
<tr>
<td>L+H*</td>
<td>rising peak accent</td>
</tr>
<tr>
<td>acc-</td>
<td>no accent</td>
</tr>
<tr>
<td>EF</td>
<td>eyebrows frown</td>
</tr>
<tr>
<td>ER</td>
<td>eyebrows raised</td>
</tr>
<tr>
<td>EF/R</td>
<td>eyebrows frown and then raised</td>
</tr>
<tr>
<td>E-</td>
<td>no eyebrow movement</td>
</tr>
</tbody>
</table>

Figure 2: Medians for the different stimuli; p-values for (marginally) significant differences between recipients’ judgments
7. References


