# Pitch accent, eyebrows and context – effects on the exhaustivity of answers?

*Charlotte Wollermann*[1,2]*, Bernhard Schröder*[2]*, and Ulrich Schade*[1,3]

[1]Speech and Communication, University of Bonn, Germany
[2]German Linguistics, University of Duisburg-Essen, Germany
[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 describe an interpretation study on the effect of accentuation, eyebrow movement and context on the exhaustive interpretation of answers. Our material consists of video recordings of nine German speakers who read dialogues with embedded question-answer pairs in which the answer constituted the pragmatic focus of the utterance. For the dialogues we varied the context regarding (un)certainty and (non-)exhaustivity of the answer. The material was presented to 45 subjects, the focus interpretation was tested by presenting pictures showing the (non-)exhaustive reading. In the current analysis we concentrate on a coordination of two noun phrases functioning as focus constituents. When using the pictures showing the non-exhaustive reading, our data show that the influence of the audiovisual prosody is less evident than the contextual influence.

**Index Terms**: audiovisual prosody, exhaustivity, focus

## 1. Introduction

In this paper we investigate to what extent accent type, eyebrow movement and context influence whether listeners prefer an exhaustive or a non-exhaustive interpretation of a focus utterance. Our study is motivated by methodological approaches from research on focus, experimental pragmatics and audiovisual prosody.[1]

### 1.1. Focus and exhaustivity of answers

Various focus phenomena can be found in natural language and also different terminology is used in the literature [e.g. 1, 2]. In languages like English, German or Dutch it can be observed that pitch accent correlates with new information in utterances, whereas old information is often deaccented. The relation between different focus types and different accent types has been discussed among others by [3]. The *pragmatic focus* usually applies to the constituent in the answer which corresponds to the interrogative pronoun in the question. In semantic-pragmatic focus theories [4, 5] it is assumed that in the context of such a question, pitch accent is correlated with focus. If the background question is interpreted as a *mention-all* question, the precondition for an *exhaustive* interpretation is given. Consider example 1 from [6: 526]. If the hearer of (1b) concludes that *only* John kissed Mary, she interprets the answer *exhaustively*.

(1a) Who kissed Mary?

(1b) $[\text{John}]_F$ kissed Mary.

[1]Many thanks to Till Krempel for helping us with the experimental set-up and to Kelly Rein for helpful comments on this paper.

Exhaustivity depends on the epistemic knowledge which is ascribed to the speaker by the hearer [7]. The findings of [2] suggest that not only accentuation affects the pragmatic focus interpretation, but also the context of the focus utterance. Clear effects of the syntactic structure of the focus constituent on exhaustivity have not been found in [2]. All in all, the role of different accent types for the interpretation of answers needs more investigation, and we hope to contribute to this by our paper.

### 1.2. Eyebrow movement and focus/prominence perception

With respect to the production side in [8] it is reported that contrastive focus in French can be marked by a raising of eyebrows. In accordance to that, other studies suggest that eyebrow movements can be used to perceive focus and prominence. According to [9] eyebrow movement can affect the perception of prominence. The results of [10] suggest that both pitch accent and eyebrow movement affect the perception of focus of attention, even though the impact of accent is stronger. In the current paper we are interested in the interpretation effects of focus indicators, not just in their perception.

### 1.3. Expression of uncertainty

Speakers use different cues for signalling and detecting uncertainty in communication. Rising intonation, delays, fillers and lexical cues were found as relevant cues [11, 12] as well as smiles and funny faces [13]. According to [14], with respect to English, a fall-rise intonation contributes to the assumption that the speaker is uncertain. For German it has been discussed whether a falling-slightly-rising contour can signal uncertainty [15].

### 1.4. Assumption

We assume that if the speaker – due to her epistemic knowledge – uses audiovisual cues of (un)certainty for the encoding of the focus utterance, the hearer uses this information for decoding the focus utterance. The hearer will assume that the speaker is (un)certain, and, thus, the interpretation should be biased towards (non-)exhaustivity.

## 2. Related work

In [16, 17] we tested the influence of intonation and context on the exhaustivity of answers. In general results show a preference for the exhaustive interpretation, but the context can bias the interpretation towards non-exhaustivity. The prosodic influence is weaker than expected. In our production study [18] we tested the audiovisual marking of pragmatic focus utterances. In the following we describe the material from that study since

we use it in the study presented here.

## 3. Material from the production study

The material consists of six question-answer pairs embedded into different dialogues. The scenario is a fictitious party where different groups of students act differently. For every action, there is a question-answer pair. The focus exponent in the answer is either a noun phrase (NP) referring to one group of students (dialogue 1, 4 and 6) or a coordinated NP referring to two groups of students (dialogue 2, 3 and 5). We refer to one group by *focus sentence with one NP* and to two groups by *focus sentence with a coordination of two NPs*. In addition, two variants of context were generated. i) Variant I is intended to have a bias towards certainty and exhaustivity. After some conversational turns, a question follows which is congruent to the focus utterance (see 2a, b). No alternatives are given in the context. Further, a sentence indicating certainty about the answer follows (2c). ii) Variant II is characterized to have a bias towards uncertainty and non-exhaustivity. A "competing" discourse entity is introduced at the beginning of the dialogue (3a). A general question follows (3b) being attached to broad focus (3c). A sentence indicating uncertainty about the answer follows (3d). In this case further elements of the set of the alternatives are explicitly given.

(2a) Wer hat die Nachbarn durch lautes Lachen gestört? *Who disturbed the neighbours by laughing loudly?*

(2b) [Die Mathematiker und Designerinnen]$_F$ haben die Nachbarn durch lautes Lachen gestört.
[*The mathematicians and designers*]$_F$ *disturbed the neighbours by laughing loudly.*

(2c) Das waren die Einzigen, da bin ich mir sicher.
*I am certain that they were the only ones.*

(3a) ...Die Linguisten lachen überall... *...The linguists usually laugh...*

(3b) Was ist passiert? *What happened?*

(3c) [Die Mathematiker und Designerinnen haben die Nachbarn durch lautes Lachen gestört.]$_F$
[*The mathematicians and designers disturbed the neighbours by laughing loudly.*]$_F$

(3d) Könnte aber sein, dass die Geografinnen auch mitgelacht haben. *It is possible that the geographers also laughed.*

Nine German speakers were instructed to read the six dialogues. Each speaker read for three dialogues the contextual variant I and for the other three dialogues the contextual variant II. In total, there were 54 dialogues and 18 filler-dialogues; the two filler-dialogues per speaker were always the same.

We annotated the data with respect to type of pitch accent, occurrence of eyebrow and head movement (for a detailed description see [18]). L+H* was produced most frequently for the marking of the focus constituent, followed by H*, no accentuation and a few realizations of other accent types. Our data show a tendency that H* accompanied by a raising of eyebrows or head appeared more often for the contextual variant intended to be biased towards uncertainty and non-exhaustivity. We interpret this as a possible manifestation of the biological codes [19]: high pitch expresses uncertainty and continuation on the pragmatical level, whereas low pitch expresses certainty and finality. In a next step, the audio material from [18] was presented to subjects for testing the influence of different accent types and context variation on focus interpretation [20]. Our data suggest a significant influence of both context and prosody in general, but the contextual influence is more evident. In the current study we present the *audiovisual* material to subjects.

## 4. Interpretation study

### 4.1. Goal

The goal of this study is to test to what extent pitch accent, eyebrow movement and context affect the exhaustivity of answers.

### 4.2. Material for the current study

We used the material from [18], but removed the sentence indicating (un)certainty about the answer for each dialogue to avoid lexical effects of uncertainty on the recipients' judgements.

### 4.3. Testing focus interpretation

For measuring focus interpretation we used pictures. From the subjects' choice of a picture we inferred the preference of interpretation. This way, we are trying to avoid that the subjects' linguistic awareness is directed to the phenomenona tested here. For half of the dialogues (dialogue 2, 3 and 6) we used pictures showing the exhaustive reading, i.e., there was only the discourse entity performing the question under discussion illustrated in the picture. For the other half of the dialogues (dialogue 1, 4 and 5) pictures illustrating the non-exhaustive interpretation were presented, which means that other elements of the set of the alternatives were also illustrated here.

### 4.4. Hypothesis

Based on our previous studies [16, 17] we assume that the contextual variant II combined with H* and/or raised eyebrows (as prosodic indicators of uncertainty) for producing the focus constituent is biased towards non-exhaustivity. In contrast, we assume that the contextual variant I combined with L+H* for producing the focus utterance is biased towards exhaustivity.

### 4.5. 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 + 2 filler-dialogues) produced by the same speaker. Thus, 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). The results were statistically analysed using the Wilcoxon Rank Sum Test; our level of significance was 5%.[2]

## 5. Results

### 5.1. Results for focus sentences with one noun phrase

The results for dialogues with *focus sentences with one NP* (dialogue 1, 4 and 6) have already been described in [21]. We observed a contextual effect, but weaker than theoretically expected. Furthermore, there is a tendency that L+H* affects the focus interpretation in favour of exhaustivity. Influences of eyebrow movement barely occur. In the following we do a more fine-grained analysis of dialogues with *focus sentences with a coordination of two noun phrases* since we expect that a higher degree of freedom in the choice of prosodic realizations could bear more information as input to the interpretation.

[2]We also performed a Kruskal Wallis Test on our data ($p<0.01$), but with this test it is not possible to test the influence of the different factors, i.e., by context and audiovisual prosody.

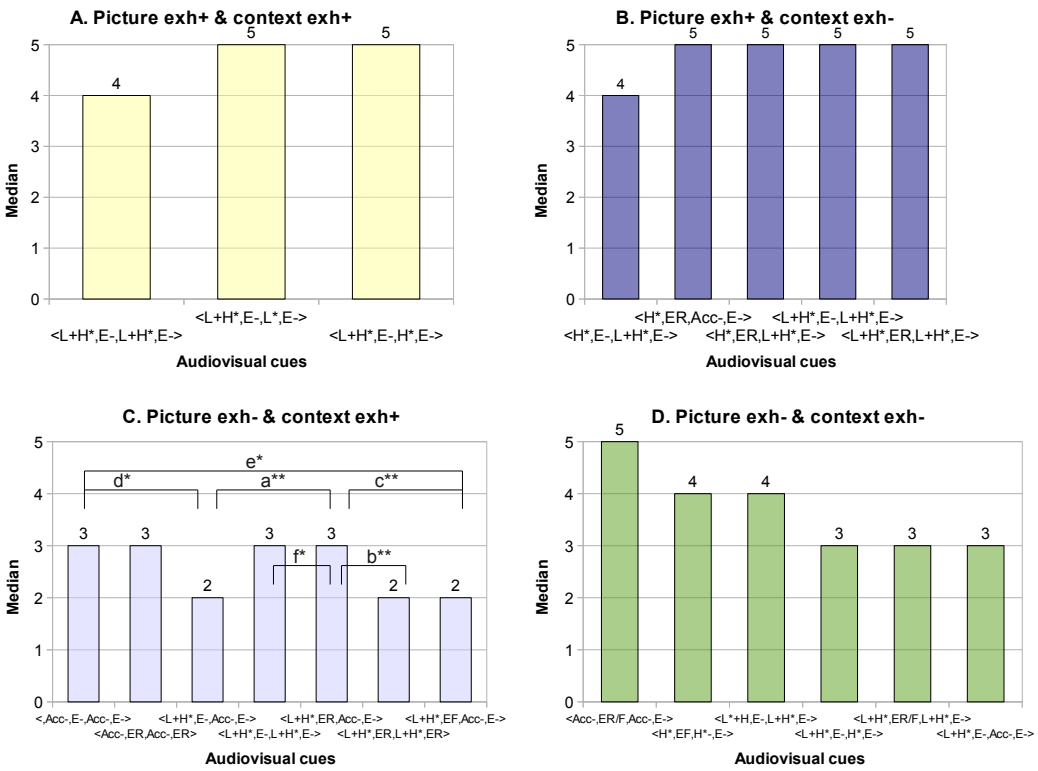


Figure 1: Recipients judgements for dialogues with focus sentences with two NPs (medians); **: significant differences between judgements ($p<0.05$), *: marginally significant differences between judgements ($p<0.1$)

### 5.2. Results for focus sentences with a coordination of two noun phrases

For our current analysis we concentrate on the cases in which we have dialogues with a coordination of two NPs functioning as focus constituents (dialogue 2, 3 and 5). There are 27 dialogues (9 speakers x 3 dialogues). In diagrams A-D of figure 1 the medians are shown for different conditions. *Picture exh+* refers to the picture showing the exhaustive reading, whereas *picture exh-* refers to the picture showing the non-exhaustive reading. Furthermore, *context exh+* refers to contextual variant I, whereas *context exh-* refers to contextual variant II. In the cutlines' 4-tuples <w,x,y,z>, *w* refers to the accent type of the first NP (L+H*, H*, L*+H or Acc- (no accentuation)), *x* refers to the eyebrow movement for the first NP (EF (eyebrows frown), ER (eyebrows raised), ER/F (eyebrows raised and then frown), E- (no eyebrow movement)), *y* refers to the accent type of the second NP (same values as *w*), and *z* refers to the eyebrow movement for the second NP (same values as *x*). Each bar represents 5 judgements with the following exceptions: i) diagram B, <H*,ER,Acc-,E->: 10 judgements, ii) diagram C: <L+H*,E-,L+H*,E->: 15 judgements; <L+H*,ER,Acc-,E->: 10 judgements, and iii) diagram D, <L+H*,E-,H*,E->: 10 judgements; <L+H*,E-,L+H*,E->: 10 judgements.

Diagram A shows the medians when we present the *picture exh+* combined with the *context exh+*. The recipients' judgements do not differ significantly. When we use the *picture exh+* combined with the *context exh-* (see diagram B) the statistical analysis again does not show significant differences between the judgements. Comparing the judgements for <L+H*,E-,H*,E-> combined with the *picture exh+* & *context exh+* with those for <H*,E-,L+H*,E-> combined with the *picture exh+* & *context exh-*, we observe a marginal significant difference between the judgements ($p<0.1$). In this case, the context is varied and the accent types are also inverted on the two NPs. We infer that the context influences the exhaustivity as expected and possibly overrules the influence of the accent type.

In diagram C, the results are shown for the *picture exh-* combined with the *context exh+*. The statistical analysis shows significant differences ($p<0.05$) between the judgements for the following comparisons (the 4-tuple with the higher rating is listed first): a) <L+H*,ER,Acc-,E-> vs. <L+H*,E-,Acc-,E->, b) <L+H*,ER,Acc-,E-> vs. <L+H*,ER,L+H*,ER>, and c) <L+H*,ER,Acc-,E-> vs. <L+H*,EF,Acc-,E->. We infer that in the case of a) and c), L+H* combined with the raising of eyebrows for producing the first NP affects the interpretation in favour of non-exhaustivity. On the other hand, we conclude for b) that the absence of both accentuation and eyebrow movement for the second NP has a greater impact on non-exhaustivity than L+H* combined with a raising of eyebrows. For the following comparisons our data show a difference between the judgements in a marginally significant way ($p<0.1$): d) <Acc-,E-,Acc-,E-> vs. <L+H*,E-,Acc-,E->, e) <Acc-,E-,Acc-,E-> vs. <L+H*,EF,Acc-,E->, and f) <L+H*,ER,Acc-,E-> vs. <L+H*,E-,L+H*,E->. We conclude for d) and e) that the audiovisual prosody is judged as less adequate with respect to non-exhaustivity than the total absence of the audiovisual information. For f) we infer that L+H* combined with a raising of eyebrows for realizing the first NP and the absence of prosodic marking for the second NP is judged as more adequate than L+H* without eyebrow movement for

both NPs. When we present the *picture exh-* combined with the *context exh-* (see diagram D) our data do not show significant differences between the judgements.

Finally we compared the judgements presenting the *picture exh-* while varying the context. In the following cases we observe significant differences between the judgements ($p<0.05$): <Acc-,ER/F,Acc-,E-> combined with the *picture exh-* & *context exh-* is judged as more adequate than the *picture exh-* & *context exh+* combined with g) <L+H*,E-,Acc-,E->, h) <L+H*,E-,L+H*,E->, i) <L+H*,ER,L+H*,ER>, and j) <L+H*,EF,Acc-,E->. Here we infer that the context affects the interpretation in favour of non-exhaustivity, possibly combined with a raising and then frowning of eyebrows for producing the first NP. Furthermore significant effects ($p<0.05$) also occur for for the following cases: <L+H*,E-,Acc-,E-> combined with the *picture exh-* & *context exh-* is judged as more adequate than *picture exh-* & *context exh+* combined with k) <L+H*,E-,Acc-,E->, l) <L+H*,ER,L+H*,ER>, and m) <L+H*,EF,Acc,E->. For k) we observe that the context influences the focus interpretation as expected. For l) and m) we infer that the context determines the non-exhaustive interpretation, the effect of the prosodic information is not clear. The following comparisons show a difference between the interpretations in a marginally significant way ($p<0.1$): <H*,EF,H*,E-> combined with the *picture exh-* & *context exh-* is judged more adequate than the *picture exh-* & *context exh+* combined with n) <L+H*,E-,Acc-,E->, o) <L+H*,E-,L+H*,E->, p) <L+H*,ER,L+H*,ER>, and q) <L+H*,EF,Acc-,E->. Here we infer that the context – possibly combined with H* for producing both NPs and accompanied by a frowning of eyebrows for producing the first NP – affects the non-exhaustive interpretation. Furthermore, <L+H*,E-,H*,E-> combined with the *picture exh-* & *context exh-* is judged as more adequate than the *picture exh-* & *context exh+* combined with r) <L+H*,E-,Acc-,E->, s) <L+H*,ER,L+H*,ER> and t) <L+H*,EF,Acc-,E->. Here we can not observe clear effects of audiovisual prosody, the context possibly overrules the prosodic information. The recipients judgements are also higher for <L+H*,E-,L+H*,E-> combined with the *picture exh-* & *context exh-* in comparison to the *picture exh-* & *context exh+* combined with u) <L+H*,E-,Acc-,E-> and v) <L+H*,EF,Acc-,E->. For u) and v) we infer that the context is the relevant factor for the interpretation whereas the audiovisual effect is not clear. Finally, <L+H*,E-,Acc-,E-> combined with the *picture exh-* & *context exh-* is judged as more adequate than <L+H*,E-,L+H*,E-> combined with the *picture exh-* & *context exh+*. Again, the context seems to affect the focus interpretation.

## 6. Conclusion

We presented an empirical study on the effect of accent type, eyebrow movement and context on the exhaustive interpretation of answers. Our focus utterances consisted of a coordination of two noun phrases and were embedded into short dialogues. For testing the focus interpretation, we used pictures showing the (non-)exhaustive reading. When we use the pictures showing the non-exhaustive reading our data suggest a significant influence of the context on exhaustivity. The effects of audiovisual prosody can be observed when the information illustrated in the picture and the information given by the context contradict to each other, but those effects are generally weak. Since we have several independent variables in our experiment the comparisons in our analysis are based on a limited number of recipients' judgements. This could explain why clear effects of audiovisual prosody can not be found. However, our data suggest that the utterance context plays a more important role than proposed in semantic-pragmatic theories [4, 5]. This observation is in line with the findings of [2].

In our future work, we would like to gain a higher amount of recipients' judgements. Furthermore, we would like to find a scenario where the production of the focus utterance can be controlled and where the speakers signal uncertainty as an expression of their own epistemic state at the same time. In a further step, this material could be given to listeners to test the effects on the focus interpretation.

## 7. References


[1] Krifka, M. (2007). Basic Notions of Information Structure, in: Fery, C., Fanselow, G., Krifka, M. (eds.), *Working Papers of the SFB632, Interdisciplinary Studies on Information Structure (ISIS)* 6, Potsdam: Universittsverlag Potsdam, 13-56.

[2] Fisseni, B. (2011). *Focus: Interpretation? Empirical Investigations on Focus Interpretation*. Duisburg: Universitätsverlag Rhein-Ruhr.

[3] Baumann, S., Grice, M., Steindamm, S. (2006). Prosodic Marking of Focus Domains - Categorical or Gradient?, in: *Proceedings of Speech Prosody*, 301-304, Dresden, Germany.

[4] Groenendijk, J., Stokhof, M. (1984). *Studies on the Semantics of Questions and the Pragmatics of Answers*. Dissertation, University of Amsterdam.

[5] Rooth, M. (1992). A theory of focus interpretation, in: *Natural Language Semantics*, 1, 75-116.

[6] Szabolcsi, A. (1980). The semantics of topic-focus articulation, in: Groenendijk, J., Janssen, T., Stockhof, M. (eds.), *Formal methods in the study of language*, 2, Mathematical Centre Tracts, 513-540.

[7] van Rooij, R., Schulz, K. (2006). Pragmatic Meaning and Non-monotonic Reasoning: The Case of Exhaustive Interpretation, in: *Linguistics and Philosophy*, 29, 205-250.

[8] Dohen, M., Loevenbruck, H., Hill, H. (2006). Visual Correlates of Prosodic Contrastive Focus in French: Description and Inter-Speaker Variability, in: *Proceedings of Speech Prosody*, Dresden, Germany, 221-224.

[9] Granström, B., House, D., Lundeberg, M. (1999). Prosodic Cues in Multimodal Speech Perception, in: *Proceedings of the International Congress of Phonetic Sciences*, San Francisco, CA, 655-658.

[10] Krahmer, E., Ruttkay, Zs., Swerts, M., Wesselink, W. (2002). Pitch, Eyebrows and the Perception of Focus, in: *Proceedings of Speech Prosody 2002*, Aix en Provence, France, 443-446.

[11] Smith, V., Clark, H. (1993). On the course of answering questions, in: *Journal of Memory and Language*, 32, 25-38.

[12] Brennan, S.E., Williams, M. (1995). The feeling of another's knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers, in: *Journal of Memory and Language*, 34, 383-398.

[13] Swerts, M., Krahmer, E. (2005). Audiovisual prosody and feeling of knowing, in: *Journal of Memory and Language*, 53(1), 81-94.

[14] Ward, G., Hirschberg, J. (1985). Implicating Uncertainty: The Pragmatics of Fall-Rise Intonation, in: *Language*, 61(4), 747-776.

[15] Ambrazaitis, G. (2005). Between Fall and Fall-Rise: Substance-Function Relations in German Phrase-Final Intonation Contours, in: *Phonetica*, 62(2-4), 196-214.

[16] Wollermann, C., Schröder, B. (2008). Does Uncertainty Effect the Case of Exhaustive Interpretation?, in: *Proceedings of the ISCA Tutorial and Research Workshop on Experimental Linguistics*, Athens, Greece, 233-236.

[17] Wollermann, C., Schröder, B. (2008). Certainty, Context and Exhaustivity of Answers. *Paper presented at the workshop on Speech and Face to Face communication*, Grenoble, France.

[18] Wollermann, C., Schröder, B. (2009). Effects of Exhaustivity and Uncertainty on Audiovisual Focus Production, in: *Proceedings of the International Conference on Auditory-Visual Speech Processing*, Norwich, UK, 145-150.

[19] Gussenhoven, C. (2002). Intonation and interpretation: Phonetics and Phonology, in: *Proceedings of Speech Prosody*, Aix-en-Provence, France, 47-57.

[20] Wollermann, C., Schade, U., Schrder, B. (2011). Variation of Accent Type and of Context - Influences on Pragmatic Focus Interpretation, in: *Proceedings of Interspeech*, Florence, Italy, 1077-1080.

[21] Wollermann, C., Schröder, B., Schade, U. (2011). Pragmatic focus interpretation: interplay between context and audiovisual prosody? in: *Proceedings of Gesture and Speech in Interaction*, Bielefeld, Germany.