

# Accentuation, Uncertainty and Exhaustivity – Towards a Model of Pragmatic Focus Interpretation

Charlotte Wollermann<sup>1</sup>, Ulrich Schade<sup>1,2</sup>, Bernhard Fisseni<sup>3</sup> and Bernhard Schröder<sup>3</sup>

<sup>1</sup>Institute of Communication Sciences, University of Bonn, Germany

<sup>2</sup>Fraunhofer Institute for Communication, Information Processing and Ergonomics FKIE, Germany

<sup>3</sup>German Linguistics, University of Duisburg-Essen, Germany

cwo@ifk.uni-bonn.de, ulrich.schade@fkie.fraunhofer.de, bernhard.{fisseni,schroeder}@uni-due.de

## Abstract

This paper presents a model of pragmatic focus interpretation that is assumed to be part of a complete language comprehension model and that is inspired by Levelt's language processing model. The model is derived from our empirical data on the role of accentuation, prosodic indicators of uncertainty and context for pragmatic focus interpretation. In its present state, the model is restricted to these data, but nevertheless generates predictions.

**Index Terms:** prosody, context, pragmatic focus, utterance interpretation, uncertainty

## 1. Introduction

This paper presents a model of pragmatic focus interpretation. It is derived from our empirical data on the role of accentuation, prosodic indicators of uncertainty and context for pragmatic focus interpretation. We will give a brief overview of the theoretical background on focus, uncertainty and audiovisual prosody and finally establish a connection (section 1). Afterwards we explain our general methodological approach and summarize our empirical studies (section 2). From our findings we deduce a model for pragmatic focus interpretation that is assumed to be part of a complete language comprehension model and that is inspired by Levelt's language processing model [1] (section 3). Finally we provide a conclusion (section 4).

### 1.1. Theoretical background of focus

Various phenomena have been labelled with the term *focus* in linguistics and different terminology is used in the literature [2, 3]. The term *focus* often refers to the intuition that pitch accent correlates with new information in utterances, whereas old information is deaccented (e.g. [4, 5]) (at least) in West Germanic languages like English, German or Dutch. It is widely accepted that focus can be characterized as "... the answer to the question being addressed..." [6: 261]; this concept of focus, the *pragmatic focus*, is what we use here. Semantic-pragmatic focus theories [7, 8] assume that focus is associated with a background question. If the latter is interpreted as a *mention-all* question, the precondition for an *exhaustive* interpretation is given.<sup>1</sup> Exhaustivity is one of the possible effects of focus on interpreta-

<sup>1</sup>Consider "Where can I buy a newspaper?" from [7: 278]. Mentioning the closest shop would constitute a non-exhaustive listing, but for all common purposes a 'complete' answer. Similarly, one can answer partially, but with full certainty: "You can buy an Italian newspaper around the corner." We do not consider mention-some questions or explicitly partial answers here.

tion; others might be correction or salience [9]. Consider example (1), from [10: 519], where (1b) serves as answer to (1a).

(1a) Who kissed Mary?

(1b) [John]<sub>F</sub> kissed Mary.

If the hearer of (1b) concludes that *only* John kissed Mary, the answer is interpreted *exhaustively*. Exhaustivity depends on the knowledge about the situation in question which is ascribed to the speaker by the hearer [11]. In our scenarios, the assumption is plausible that the speaker – unless signalling uncertainty – is completely informed. Experiments are therefore designed to test only for exhaustive interpretation which is *epistemically strong* in the terms of [11]. If there are (or may be) also other persons who kissed Mary, the interpretation is – for the purpose of this paper – *non-exhaustive*.

Semantic-pragmatic focus theories [7, 8] expect that in the context of a question, pitch accent is highly correlated with focus. If focus is detected, an exhaustive interpretation should be facilitated. In contrast, contextual influence is only briefly sketched by these theories, i.e. the role of the question is discussed.

In [2] some predictions of semantic-pragmatic focus theories were tested empirically. It was found that not only accent but also the hearer's expectations, the sensitization for focus phenomena and contextual factors are relevant for focus interpretation. Similarly for written speech, the results of [9] show that the type of question and answer affects exhaustivity of answers.

### 1.2. Theoretical background of uncertainty

Speakers and listeners use different cues to signal and detect *uncertainty* in communication. The work of [12–14] shows that uncertainty is expressed and perceived by prosodic cues like *rising intonation* and *delays* and also by *fillers* like "um" and *lexical cues* like "I guess". From the perspective of psycholinguistics *pauses* and *fillers* in speech play an important role in self-repair, also referred to as *c-repairs* [15]. These repairs occur if the speaker recognizes and corrects a slip of the tongue even before the speech signal is produced. A connectionist model of such a kind of repairs can be found in [16]. Furthermore, the analysis of [17] suggests for English that *fall-rise* intonation contributes to a context-independent meaning of utterance interpretation conveying the speaker's uncertainty.

### 1.3. Theoretical background of audiovisual prosody

[18] shows that visual information plays an important role for speech perception. According to the *principle of superadditive*

combination, information conveyed by the audiovisual channel affects speech perception stronger than information transported by unimodal channels [19]. In multimodal speech synthesis systems, so-called *Talking Heads* [20], additional visual information is provided for communication. [21] reports that eyebrow movement expressed by a Talking Head affects the perception of *focus of attention*, but the effect of accent is stronger.

#### 1.4. Assumption

We assume that if the speaker marks the pragmatic focus constituent by audiovisual prosody, the hearer will perceive the accentuation and detect the focus. As already mentioned in the case of exhaustivity, the hearer has reason to assume that the speaker is certain about the answer and fully competent. We assume that if the speaker conveys uncertainty in the audio signal, the hearer doubts the speaker's competence and will be less attracted by an (epistemically strong) exhaustive interpretation.

## 2. Empirical evidence for pragmatic focus interpretation

For testing pragmatic focus interpretation directly we follow the methodological approach developed by [2]. As stimuli we use natural audio recordings which consist of question-answer pairs. The focus of the answer is a noun phrase. The stimuli are embedded in short stories for controlling the influence of macro context, which is not explicitly considered in semantic-pragmatic focus theories. We define *macro context* as the story of the dialogue preceding the question-answer pair; whereas *micro context* refers to the question preceding the answer with the focus. For testing focus interpretation we use pictures intended to illustrate the different readings. From the subjects' choice of the pictures, we infer the preference for interpretation. When in the following we refer to 'the focus' in a stimulus, we mean the constituent that is phonologically or visually marked and for which we expect exhaustive interpretation.

### 2.1. Eyebrows, accent and focus

In [22] we investigate whether audiovisual prosody affects pragmatic focus detection. The stimuli consist of German question-answer pairs, which are structurally ambiguous. Although the 'canonical' constituent order in German main clauses is *subject verb object* (SVO), it is also possible to have sentences with an *object verb subject* (OSV) constituent order. We hypothesize that a question asking for the object combined with audiovisual accentuation of the focus constituent leads to an OSV reading, which is dispreferred in out-of-the-blue contexts. The stimuli are generated by means of a Talking Head [23]; its lip movements are synchronized with recorded natural speech. In (2) an example of a stimulus is given.

(2a) Wer erhielt einen Anruf? *Who received a phone call?*

(2b) [Tim]<sub>F</sub> rief Wim an. [Tim]<sub>F</sub> called Wim.

We vary whether the focus of the answer is marked by *accent* and/or *eyebrow movement*. According to the superadditive principle we expect the strongest impact on OVS for the audiovisual modality, whereas weaker effects should occur for unimodal modalities. 31 subjects are asked to annotate pictures presenting either the OVS or the SVO interpretation. Results suggest that the SVO reading is strongly preferred. But our data also show evidence that accent per se leads more often to an OVS interpretation than eyebrow movement per se. Against our expectation, for the audiovisual channel no stronger effect can be

observed than for the unimodal channels. Since the visual signal is marked by a higher degree of artificiality than the audio signal we assume that listeners might be more sensitive to the prosodic cues in the natural speech.

### 2.2. Uncertainty, exhaustivity and micro context

Our aim in [24] is to investigate the influence of *intonation* and *micro context* on exhaustivity of answers. Our audio stimuli consist of question-answer pairs (see example 3). For the realization of the noun phrase in the answer, intonation is varied by either *rising* intonation for expressing *uncertainty* and *continuation* or by *falling* intonation for conveying *certainly* and *finality* (3c).<sup>2</sup> The preceding question is either *congruent* (3a) with the supposed focus structure of the answer, and then expected to favour *exhaustivity*, or it constitutes a *general* question (3b) expected to favour *non-exhaustivity*.

(3a) Wer ist zu spät gekommen? *Who was late?*

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

(3c) [Die Mathematiker]<sub>F</sub> sind zu spät gekommen.  
[The mathematicians]<sub>F</sub> were late.

We assume that intended *certainly* combined with a *congruent* question favours *exhaustive* interpretation, whereas intended *uncertainty* in combination with a *general* question favours *non-exhaustivity*. For each dialogue there are four combinations of intonation and context for investigating the relative impact of the two variables. For testing interpretation we use three pictures: one picture presenting exhaustive reading (showing only the mathematicians), another one illustrating non-exhaustive reading (showing the mathematicians and linguists) and a third picture functioning as a distractor (showing the geographers). For each dialogue 71 students are asked which picture suits the best. Results suggest that the exhaustive interpretation is generally preferred in our scenario, but we also observe a weak effect of both intonation and micro context on non-exhaustivity.

### 2.3. Uncertainty, exhaustivity and macro context

In [26] we test if *several cues of uncertainty* and variation of *macro context* can favour non-exhaustivity. This time *intonation* is not only varied with respect to the supposed focus constituent, but also for the sentence-final verb. We also vary *pauses* affecting the whole sentence. We generate two kinds of macro context for each dialogue: For the context intended to favour exhaustivity the referent of the focus of the answer is salient in the discourse (*exclusion of alternatives*). Furthermore, question and answer are *congruent*. In contrast, the context intended to favour non-exhaustivity introduces a discourse entity usually carrying out the action under discussion which is different from the focus of the answer, i.e. linguists are always late for parties (*inclusion of alternatives*). This time the question is *general*. Interpretation is tested by presenting one picture, either showing exhaustive or non-exhaustive reading. For each dialogue 160 students are asked to mark on a Likert scale how appropriate the picture is. Results suggest that the exhaustive reading is preferred. But comparing all judgements for pictures illustrating non-exhaustivity, we observe strong effects of the macro context, whereas the impact of prosody is weaker.

<sup>2</sup>Our intonational variation is motivated by Gussenhoven's biological codes, i.e. the frequency and the production code [25].

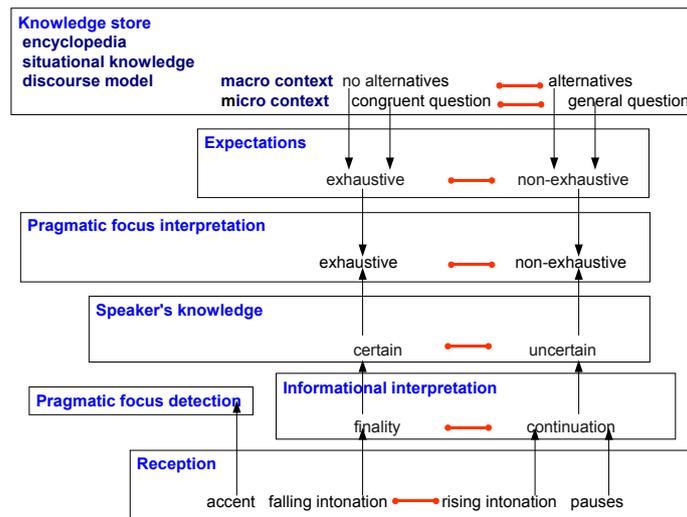


Figure 1: Model of pragmatic focus interpretation. Boxes represent different modules; arrows represent influences; red connections represent dichotomies.

### 3. Model of pragmatic focus interpretation

To guide further research in the area of focus interpretation, a model is needed. In subsection 3.1 we briefly present the main characteristics of Levelt's language processing model [1] since it serves as a source of inspiration for our comprehension model on focus interpretation. In subsection 3.2 we present our model.

#### 3.1. Levelt's blueprint of the speaker

Levelt's model for language production [1] consists of three modules, the *conceptualizer*, the *formulator* and the *articulator*. The *conceptualizer* takes the intent, selects the information to be expressed to realize it and prepares this information in the form of the preverbal message. The *formulator* does grammatical and phonological encoding to generate a phonetic plan which is then executed by the *articulator*. Our model refers to the *conceptualizer*.

The *conceptualizer* exploits declarative knowledge to generate the preverbal message, *encyclopedic knowledge*, *situational knowledge* as well as the *discourse model*, which records what has already been said and what is currently prominent. We call the available knowledge *knowledge store*. With respect to processing, the conceptualizer consists of two modules, one for *macroplanning* and one for *microplanning*. *Macroplanning* is responsible for the content as such and answers the question what is to be expressed next. *Microplanning* adds its perspective. It marks which parts of the message are new information and which parts are already prominent in the discourse. Thus, it is responsible to mark topic and focus. Language production is seen as an incremental process such that uncertainty results in production delays (e.g. pauses and hesitations).

#### 3.2. Model

Levelt's model is about speech production. However, we want to model a comprehension process. It is expected that the knowledge store is shared by production and comprehension. Besides, we assume that the means to mark linguistic aspects as topic and focus in production are used in comprehension to detect exactly these aspects. With this assumption in mind, we

can now present our model of pragmatic focus interpretation (see fig. 1).

As has already been explained, the *knowledge store* contains the *encyclopedic knowledge*, the *situational knowledge*, and the *discourse model*. The discourse model contains the hearer's beliefs about what is shared with the speaker about the content of the discourse so far. We propose that contextual factors play an important role for the hearer's assumption what the speaker's current statement is about. In particular, we assume that a *macro context* which *excludes alternatives* raises the hearer's expectations of *exhaustivity*, whereas a macro context which *includes alternatives* contributes to expectations of *non-exhaustivity*. The same principle holds for the *micro context*: A question which is *congruent* with the focus-intonational structure of the utterance raises expectations of *exhaustivity*. In contrast, a *general* question makes a contribution to expectations of *non-exhaustive* interpretation. Overall, the knowledge store constitutes the input for the next module in which the hearer's expectations are calculated and stored.

The *hearer's expectations* are influenced by the context. They also comprise syntactic commonality. Since the canonical sequence SVO is statistically very dominant, OVS readings are nearly always suppressed if both readings are plausible. As the expectations are calculated from these two factors at least, we regard them as multidimensional. The expectations calculated from context and syntactic commonality influence the pragmatic focus interpretation. If there is no prosodic cue, they determine the interpretation.

On the *reception* side prosodic information contributes to both the detection and the interpretation of pragmatic focus. If the hearer perceives a *pitch accent*, which is used in production to mark *focus*, she interprets that the respective constituent is focused. This contributes to an *exhaustive* interpretation of the statement as such. *Falling* intonation used to express *finality* on the informational level also effects an *exhaustive* interpretation. This interpretation results from the fact that the hearer regards the speaker's knowledge as *certain*. In contrast, the perception of *rising* intonation used to express *continuation* on the informational level and/or *pauses*, as prosodic cues of uncertainty, contributes to a *non-exhaustive* interpretation. This interpreta-

tion results from the fact that the hearer regards the speaker's knowledge as *uncertain*. Thus, prosodic indicators of uncertainty may overrule the exhaustive reading as default interpretation in favour of a non-exhaustive interpretation.

With respect to the relative contribution of the different modules to pragmatic focus interpretation the following is postulated: Firstly, the *hearer's expectations* which result from different sources of knowledge and syntactic commonality do have the strongest impact on focus interpretation. Secondly, *prosody* does also contribute to both detection of focus and focus interpretation, but the effect is relatively weak. Finally we propose that assumptions of semantic-pragmatic focus theories need to be revised: Prosody per se is not sufficient for focus interpretation. It is the interplay between the expectations of the hearer which results from both contextual conditions or other sources of knowledge and from accentuation and prosodic indicators of uncertainty.

#### 4. Conclusion

We presented a model for pragmatic focus interpretation which was derived from our empirical data on the role of accentuation, prosodic indicators of uncertainty and context for pragmatic focus interpretation. According to our model different factors are relevant for pragmatic focus interpretation. We argue that prosody per se is not sufficient to trigger focus interpretation, contrary to what is assumed by semantic-pragmatic focus theories [7, 8]. It is proposed that the macro and micro context are crucial factors for raising the hearer's expectations and that both these expectations and prosodic information on the reception side influence focus interpretation, but the expectations have the upper-hand. Our model is in line with [2], who found that besides accent, the expectations of the hearer, the sensitivity to prosody, structural and in particular contextual factors are relevant for focus interpretation. In its present state, the model is restricted to our data on focus detection and exhaustive interpretation. However, it already generates an interesting prediction: If the expectations are weak, e.g. if syntactic commonality and context contradict each other, the prosodic influence on focus interpretation should increase.

In our future work, we will thus examine the predictions of the model. In addition, we will extend the model to integrate other functions of focus, e.g. correction. It would also be interesting to make use of other constituents than noun phrases as focus exponents. Furthermore, we would like to take the visual information into account for a model of focus interpretation. Here it would be important to use natural speech for consistency. Our study on pragmatic focus production [27] suggests that exhaustive answers correlate with a lowering of the head, whereas non-exhaustive answers are more often accompanied with a rising of the head or with eyebrow movement. It is less clear which role these cues play for pragmatic focus interpretation.

**Acknowledgements** We would like to thank Bernd Möbius for helpful comments on this paper.

#### 5. References

- [1] Levelt, W. J. M., 1989. *Speaking: From intention to articulation*. MIT Press, Cambridge, MA.
- [2] Fisseni, B., to appear. *Focus: Interpretation*. Dissertation, University of Duisburg-Essen.
- [3] Krifka, M., 2007. Basic Notions of Information Structure, in: Féry, C., Fanselow, G., Krifka, M. (Eds.), *Working Papers of the*
- [4] Bolinger, D., 1958, A theory of pitch accent in English, in: *Word*, 14, 109-149.
- [5] Büring, D., 1997, *The meaning of topic and focus: The 59th street bridge accent*, London: Routledge.
- [6] Kadmon, N., 2001. *Formal Pragmatics. Semantics, Pragmatics, Presupposition, and Focus*. Blackwell, Oxford.
- [7] Groenendijk, J., Stokhof, M., 1984. *Studies on the Semantics of Questions and the Pragmatics of Answers*. Dissertation, University of Amsterdam.
- [8] Rooth, M., 1992. A theory of focus interpretation, in: *Natural Language Semantics*, 75-116.
- [9] Skopeteas, S., Fanselow, G., 2009. Exhaustive identification of focused referents and contextual conditions. *Paper presented at DGfS 2009*, Osnabrück, Germany.
- [10] Szabolcsi, A., 1980. The semantics of topic-focus articulation, in: Groenendijk, J., Janssen, T., Stokhof, M. (Eds.), *Formal methods in the study of language*, 2, Mathematical Centre Tracts, 513-540.
- [11] van Rooij, R., Schulz, K., 2006. Pragmatic Meaning and Non-monotonic Reasoning: The Case of Exhaustive Interpretation, in: *Linguistics and Philosophy*, 29, 205-250.
- [12] Smith, V., Clark, H., 1993. On the course of answering questions, in: *Journal of Memory and Language*, 32, 25-38.
- [13] 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.
- [14] Swerts, M., Kraemer, E., 2005. Audiovisual prosody and feeling of knowing, in: *Journal of Memory and Language*, 53(1), 81-94.
- [15] Levelt, W.J.M., 1983. Monitoring and self-repair in speech, in: *Cognition*, 14, 41-104.
- [16] Schade, U., Eikmeyer, H.-J., 1991. "wahrscheinlich sind meine Beispiele soo sprunghaft und und und eh ehm zu zu telegraph" - Konnektionistische Modellierung von "covert repairs", in: Christaller, Th. (Ed.), *GWAI-91 1. Fachtagung für Künstliche Intelligenz*, Springer Verlag, Berlin, 264-272.
- [17] Ward, G., Hirschberg, J., 1985. Implicating Uncertainty: The Pragmatics of Fall-Rise Intonation, in: *Language*, 61(4), 747-776.
- [18] McGurk, H., McDonald, J., 1976. Hearing lips and seeing voices, in: *Nature*, 264, 746-749.
- [19] Massaro, D.W. 1998. *Perceiving talking faces: From Speech Perception to a Behavioral Principle*. MIT Press, Cambridge, MA.
- [20] Beskow, J., 2003. *Talking heads - Models and applications for multimodal speech synthesis*. Doctoral dissertation, KTH, Stockholm, Sweden.
- [21] Kraemer, 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.
- [22] Fisseni, B., Hülsken, F., Schröder, B., Wollermann, C., 2006. Eyebrows, accent and focus detection. *Paper presented at the Szklarska Poreba Workshop 7 on the Roots of Pragmaseantics*, Szklarska Poreba, Poland.
- [23] Virtual Human Project Home Page, 2008. Retrieved 24th of January 2010 from <http://www.virtual-human.org/>.
- [24] 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.
- [25] Gussenhoven, C., 2004. *The Phonology of Tone and Intonation*. Cambridge University Press, Cambridge.
- [26] Wollermann, C., Schröder, B., 2008. Certainty, Context and Exhaustivity of Answers. *Paper presented at Speech and Face to Face communication*, Grenoble, France.
- [27] 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.