

# Classification of Functional-Meanings of Non-isolated Discourse Particles in Human-Human-Interaction

Alicia Flores Lotz<sup>(✉)</sup>, Ingo Siegert, and Andreas Wendemuth

Institute of Information and Communication Engineering, Cognitive Systems Group,  
Otto von Guericke University, 39016 Magdeburg, Germany  
[alicia.lotz@ovgu.de](mailto:alicia.lotz@ovgu.de)

**Abstract.** To enable a natural interaction with future technical systems, not only the meaning of the pure spoken text, but also meta-information such as attention or turn-taking has to be perceived and processed. This further information is effectively transmitted by semantic and prosodic cues, without interrupting the speaker. For the German language we rely on previous empirically discovered seven types of form-function-concurrences on the isolated discourse particle (DP) “hm”.

In this paper we present an improved automatic classification-method towards non-isolated DPs in human-human interaction (HHI). We show that classifiers trained on (HCI)-data can be used to robustly evaluate the contours of DPs in both HCI and HHI by performing a classifier adaptation to HHI data. We also discuss the problem of the pitch-contour extraction due to the unvoiced “hm”-utterances, leading to gaps and/or jumps in the signal and hence to confusions in form-type classifications. This can be alleviated by our investigation of contours with high extraction completion grade. We also show that for the acoustical evaluation of the functional-meaning, the idealized form-function prototypes by Schmidt are not suitable in case of naturalistic HHI. However, the precision of acoustical-meaning prediction with our classifier remains high.








**Keywords:** Human-human interaction · Human-computer interaction · discourse particles · Automatic form-function classification

## 1 Introduction

In HHI the behavior of the speaker is mainly characterized by semantic and prosodic cues such as short feedback signals. These signals further the progress and coordination of interaction. These cues transmit so called meta-information about certain dialogue functions such as attention, understanding, confirmation or other attitudinal reactions [1]. DPs such as “ja”, “so”, “wie” or “hm” belong to these feedback signals and serve as independent small utterance units, occurring at communicative decisive points of the conversation, without interrupting the

speaker [16]. As these signals only have a very small semantic content, special attention needs to be paid to their prosody. By changing the intonation, a DP can have several different meanings. Therefore, “hm” is one of the most diverse DP. For the German language seven form-function-relations for the isolated speech signal of the DP “hm” have been revealed by Schmidt [12]. They describe the relation between the functional-meaning of the DPs and specific pitch-contours, see Table 1.

**Table 1.** Form-function relation of DP “hm” according to [12], the terms are translated into appropriate English ones.

Name	Form-Prototype	Description
DP-1		attention
DP-2		thinking
DP-3		finalization signal
DP-4		confirmation
DP-5		decline
DP-6		positive assessment
DP-7		request to respond

Furthermore, DPs are verifiably used in both HHI and HCI [5, 12, 15] and specific form-function-relations could be confirmed [8, 9]. Also influences of specific subject characteristics, as age and biological gender, as well as certain personality traits on the use of DPs were uncovered [6, 14].

To obtain a more human-like interaction in communicating with a technical system, the system needs to be capable of understanding these feedback signals and react appropriately. This enables a more natural interaction and the system thus becomes the users’ attendant and ultimately his *companion* [2].

To achieve this aim, we developed a classifier to distinguish between the different functional-meanings of the DP “hm” [8]. The classifier is trained on speech material from HCI. In this paper we describe the adaptation of this classifier to a HHI corpus and show that it is possible to assign non-isolated DPs from HHI. Furthermore, we show the possibility to assign one of the seven prototypes obtained by Schmidt by only applying small changes to the existing algorithm. Compared to the results of HCI we assume an increase of partner-oriented feedback signals in case of HHI [5].

The remainder of the paper is structured as follows: Sect. 2 shortly describes the utilized dataset providing HHI as well as HCI. Then, Sect. 3 describes the DP-classifier developed on the HCI data in detail. Afterwards, Sect. 4 describes the conducted manual labeling process as well as the adaptation of the classifier

based on the new obtained HHI data. Section 5 presents and discusses the results of the improved DP classification algorithm and analyses the manual labeling. Finally, Sect. 6 concludes the paper and provides an outlook for further research directions.

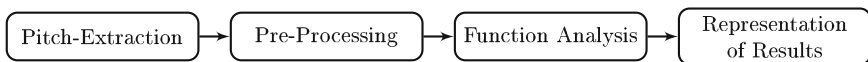
## 2 Datasets

The *LAST MINUTE Corpus (LMC)* contains 130 high-quality multi-modal recordings of German speaking subjects during Wizard-of-Oz experiments [10, 11]. The setup of the HCI revolves around an imaginary journey to the unknown place Waiuku. With the help of an adaptable technical system, the subjects have to prepare the journey, by packing the suitcase, and select clothing and other equipment by using voice commands. Each experiment takes about 30 min. For a sub-set of the corpus a total number of 259 DPs are extracted from 25 h of speech material received from 56 subjects. They serve as training data for the basis classification algorithm presented in [8].

The *ALICO Corpus* is recorded to investigate feedback behavior changes in HHI concerning distraction of the listener [4]. The corpus consists of  $2 \times 25$  dialogues in German language, in which one person is telling the other participant two stories. In the first talk the listener gets the instruction to pay attention, make remarks and ask questions. In the second one an additional distraction task is given to the listener. He gets the instruction to press a button on a hidden counter every time the story teller utters a word starting with the letter ‘s’. A sub-set of 40 dialogues is annotated, resulting in 1505 feedback signals from 3 h of speech material. Out of these signals 537 are marked as “hm” and used for the investigation presented in this paper.



## 3 The Discourse Particle Classifier

The DP-classifier uses a rule-based approach. Before the classification a pre-processing step is conducted, see Fig. 1. The following function analysis is based on regression calculation on the pitch-contour. Depending on certain threshold values the classifier decides to which form-prototype the given pitch-contour belongs. The thresholds were set by preliminary investigations. During these investigations, additionally to the seven form-function-prototypes of Schmidt, two more frequently occurring formtypes were identified, see Table 2. No information about the functional use of these formtypes is given, as the work [8] only deals with the course of the pitch-contour and not the identification of its meaning.



**Fig. 1.** Structural flowchart of the classification-algorithm

**Table 2.** Additional form types as stated in [8].

Name	Form-Prototype
DP-8	
DP-9	

Depending on the coefficient of determination ( $R^2$ ) either a first, second or third order regression is performed. If the coefficient exceeds the threshold value:

$$TH_{determination} = 0.9 \quad (1)$$

It is assumed that the regression function describes the original pitch-contour sufficiently. Only regarding the slope of the first order regression line, the tendency of the contour is determined (horizontal, rising or falling). This allows the classifier to neglect some prototypes for further assignment. For example, if the first order regression line has a rising tendency, all prototypes with a horizontal (DP-2) or falling tendency (DP-1, DP-3, DP-5 or DP-8) can be neglected. This leaves only the prototypes DP-4, DP-6, DP-7 or DP-9 for the later assignment.

Considering the order of regression, the prototype-courses can be distinguished into 3 types: linear (DP-2, DP-3, DP-7), quadratic (DP-4, DP-5, DP-8, DP-9) and cubic ones (DP-1, DP-6). If a pitch-contour is sufficiently described by a linear regression function, it can be unambiguously assigned to a linear prototype by looking at its course tendency. In case of contours described by higher order regression, the course is divided into sections limited by their turning points. Based on the timeratio and slope of these sections the prototypes are assigned. The thresholds for these parameters are:

$$TH_{timeratio} = 1/3 \quad (2)$$

$$TH_{slope} = 40 \text{ Hz/s} \quad (3)$$

If two consecutive sections exceed  $TH_{timeratio}$ , both sections will be taken into account for further assignment. Otherwise, only the dominant part (longer section) will be considered. For  $TH_{slope}$ , the courses (linear regression) and sections (higher order regression) are assumed to be horizontal if the absolute slope is below the threshold and rising or falling for slopes exceeding the threshold, respectively.

As the investigation is implemented on a HCI corpus, most DP samples express talk-organizing or expressive functions. These are mainly DPs of type DP-2. Therefore, an exception in the algorithm is made for regression lines with horizontal tendencies. Disregarding  $TH_{determination}$ , if the standard deviation of the original pitch-contour is less than 7 Hz the type DP-2 is assigned. Variations of this size can be neglected, as the human ear is not able to perceive them [17]. More information on the classification algorithm especially the pre-processing of the original pitch-contours is given in [8].

To evaluate the implemented classifier a manual annotation of the pitch-contours is performed. A consistency of the results is obtained in approx. 79% of the given DPs contained in the LMC HCI corpus.

## 4 Adaptation of the Classifier

As the existing classifier is only trained on the LMC HCI corpus, the amount of training material for partner-oriented prototypes is recognizably low, see row HCI in Table 4. As most of the prototypes detected in HCI are of type DP-2, DP-3 and DP-7, having a linear course of the pitch-contour, a verification of contours described by higher dimensional functions is not given. Another problem is the low number of DPs used to evaluate the classifier, as not all pitch-contours occurring in HHI are considered in the original implementation, leading to impossible assignment of prototypes. Therefore, an adaptation of the classifier is needed.

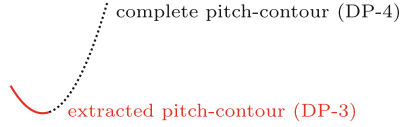
### 4.1 Manual Labeling

To verify the results of the algorithm [8] and adapt the classifier to non-isolated DPs from HHI we first utilized a manual labeling. The labeling was conducted in two stages: First, a visual assignment of the intonation curve, based on extracted pitch-values, to one of the seven idealized pitch-contours was performed. DP-8 and DP-9 were left unconsidered on purpose, as no functional-meaning of these types is known. We assumed that these formtypes are variants of the verified form-function-prototypes by Schmidt. Second, an acoustic annotation of the functional-meaning without knowledge of the pitch-contour was conducted (see Table 1). For the acoustic annotation, a modified version of *ikannotate* (cf. [3]) was used. To clarify the results stated in Sect. 5.2: For the acoustic annotation the labelers were not instructed to annotate the perceived intonation-curve but the functional-meaning according to the audible context the DPs occurred in.

### 4.2 Adjustment of the Threshold Parameter

As already stated in our earlier investigation on DPs in HCI [7], one major problem of a classifier using the intonational-curve of the signals is the subject of pitch-extraction. In our case the considered DP “hm” is a unvoiced utterance. This makes it hard to get a robust pitch-estimation, as there is no stimulation of the vocal cords at the glottis. The pitch-extractor will not be able to continuously estimate a pitch-contour. This leads to gaps and/or jumps in the signal. To ensure the reliability of the classifier, we assumed that at least 50% of the contour needs to be available to the function analysis algorithm. For all contours with a lower percentage of existence a correct assignment of the prototypes’ functional-meaning by the classifier is not implied. Figure 2 depicts a frequently occurring mismatch of prototypes: While the extracted pitch-contour will lead to a DP-3 (finalization signal) assignment, the complete contour will get assigned to DP-4

(confirmation). The made assumption has no influence on the actual assignment of the prototypes, but there is no point in correctly assigning a formtype, if it is not possible to assign the functional-meaning correctly.



**Fig. 2.** Frequently occurring mismatch in prototype.

To optimize the performance of the classifier, first, the function analysis algorithm was extended. Therefore, a new threshold was introduced  $TH_{sloperatio}$ . In cases of the description of the pitch-contours using higher order regression functions, not only the timeratio but also the sloperatio of two consecutive sections needs to be considered for the assignment of the prototypes. This will lead to a better differentiation of linear and higher polynomial (quadratic and cubic) prototypes (cf. Sect. 3), which were rarely considered in the earlier investigation on DPs in HCI [8]. Furthermore, the thresholds were adjusted:

$$TH_{timeratio} = TH_{sloperatio} = 1/6 \quad (4)$$

This is necessary to make the classifier more sensible for higher polynomial prototypes.

## 5 Results

In this section the results of the original and optimized classifier are compared. The performance of the classifier is determined depending on the visual and acoustic annotation of the labelers. If not indicated differently, all results refer to the 537 extracted DPs obtained from the ALICO HHI Corpus.

### 5.1 Classification-Algorithm

In the pre-processing of the classifier, all pitch-contours that are too short (114) or failing the requirement made in Sect. 4.2, saying that at least 50 % of the pitch-contour needs to be available to the classifier, (177) were not considered for the classification. In total 537 DPs were given as input to the function analysis loop, leaving 423 for further investigation to the original and 246 to the optimized classifier. The results can be found in Table 3.

In both cases (original and optimized classifier) a high number of DP-2 was recognized. In case of the original classifier there was an assignment of linear courses of type DP-3 and DP-7. This is explainable by the development of the rule-based classifier on a corpus for HCI, including mostly linear prototypes.

**Table 3.** Frequency distribution of the results of the classification-algorithm. (a) original classifier (423 DPs) b) adjusted classifier (246 DPs); additionally used labels: NSP “no specification possible”

Label		DP-1	DP-2	DP-3	DP-4	DP-5	DP-6	DP-7	DP-8	DP-9	NSP
Item [%]	(a)	0	30.5	15.4	18.0	0.7	2.1	24.6	5.7	1.7	1.4
	(b)	0	31.7	11.8	30.5	0.8	4.9	6.5	11.4	2.4	0

The optimization of the classifier led to a decrease of these linear types and an increase of quadratic and cubic prototypes of type DP-4, DP-6 and DP-8, as assumed and supported by the results of the following Sect. 5.3. In comparison to HCI (cf. Table 4), we obtained a remarkable number of detected DP-4 which corresponds to the assumption made in the introduction, that the number of partner-oriented feedback signals will increase in HHI. This is expectable, as the style of the dialogue is more interview-like leading the listener to use feedback signals such as the DPs to express confirmation.

**Table 4.** Frequency distribution of the results of the original classifier comparing HCI (LMC corpus) with HHI (ALICO corpus)

Label		DP-1	DP-2	DP-3	DP-4	DP-5	DP-6	DP-7	DP-8	DP-9
Item [%]	HCI	0.5	49.3	30.6	3.7	2.7	1.8	7.8	2.7	0.9
	HHI	0	30.5	15.4	18.0	0.7	2.1	24.6	5.7	1.7

## 5.2 Manual Labeling

The manual labeling was assessed using majority voting. For both assignments (visual and acoustic) 10 labelers were consulted. Then a majority voting of the results of the 5 labelers with the highest inter-rater reliability (Krippendorff’s  $\alpha_{visual} = 0.64$  and  $\alpha_{acoustic} = 0.16$ ) was carried out. The low reliability for the acoustic labeling is due to the many degrees of freedom the labelers had, as no further restrictions was given except for the functional-meanings. These meanings, defined by Schmidt, describe sometimes quite similar functional-meanings, which are hard to distinguish acoustically. The low values of  $\alpha_{acoustic}$  are rather typical across several databases, as it was shown in [13] for various emotional assessments.

The results of the majority voting are stated in Table 5. Considering the inter-rater reliability and the results of the majority voting we can already see, that it is hard for all labelers to agree on one functional-meaning. The low inter-rater reliability of the annotation of the functional-meaning can also be justified by the calculation algorithm of Krippendorff’s  $\alpha$ . As the majority of the data is rated as DP-4, all single mismatches are rated as unlikely and therefore are less reliable.

**Table 5.** ALICO data: Frequency distribution of the majority voting. (a) visual assignment of the idealized pitch-contours (b) acoustic annotation of the functional-meaning; additionally used labels: NSP “no specification possible” and NM “no majority”

Label		DP-1	DP-2	DP-3	DP-4	DP-5	DP-6	DP-7	NSP	NM
Item [%]	(a)	0.6	35.6	7.8	16.8	0.7	1.9	15.3	13.6	7.8
	(b)	0	0.2	0	89.0	0.4	2.6	0	0.7	7.1

### 5.3 Pitch-Contour Assessment

For the manual labeling process DP-8 and DP-9 were left unconsidered, as no statement about the functional-meaning of these prototypes is given. The pitch-contours assigned to DP-8 or 9 by the classifier will, accordingly, never match with the manually annotated labels and therefore, not be taken into account to calculate the performance of the classifier concerning the visual annotation results. Without any adaptation of the function analysis algorithm or threshold parameters, the original classifier correctly identified 70.5 % of all DPs in non-isolated HHI. By optimizing the classifier this value was raised up to 74.5 %.

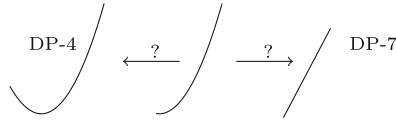
**Table 6.** Confusion matrix of the prototype assignment of the original classifier compared to the manual labeling; additionally used labels: NSP “no specification possible”

		Manual Labeling									
		DP-1	DP-2	DP-3	DP-4	DP-5	DP-6	DP-7	DP-8	DP-9	NSP
Classifier	DP-1	0	0	0	0	0	0	0	0	0	0
	DP-2	1	113	0	7	0	1	0	0	0	7
	DP-3	2	28	29	0	0	1	0	0	0	5
	DP-4	0	6	0	61	0	2	0	0	0	7
	DP-5	0	0	0	0	2	0	0	0	0	1
	DP-6	0	1	0	2	0	4	0	0	0	2
	DP-7	0	16	0	14	0	1	61	0	0	12
	DP-8	0	6	10	2	0	0	0	0	0	6
	DP-9	0	1	0	0	2	0	3	0	0	1
	NSP	0	2	0	2	0	1	0	0	0	1

Table 6 shows the confusion matrix of the manual labeling of the formtypes and the results of the original classification algorithm. Most of the mismatches are noticed between DP-2 and DP-3, DP-7. In these cases only a disagreement in the slope of the course was found. It can be concluded, that the labelers were more likely to identify the perceived signals as sloping than horizontal.

Moreover, we recognized a high confusion between the formtypes DP-7 and DP-4, illustrated in Fig. 3. This also explains the unexpected high number of DP-7 in the visual annotation of the pitch-contours. In these cases the labelers





**Fig. 3.** Illustration of the confusion of real signals between the prototypes of DP-4 and DP-7.

were not sure on which prototype to decide on. Both cases of mismatch were minimized by the adjustment of the classification algorithm described in Sect. 4.2. For the optimized classifier no clear majority in the mismatches of prototype assignment was recognized.

#### 5.4 Functional-Meaning Assessment

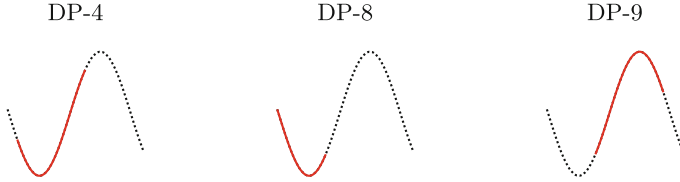
We now considered the results of the acoustic functional-meaning assessment. Regarding the results stated in Table 5 (b) a clear majority of DP-4 (confirmation) was recognized. In comparison the results of the pitch-contour assignment showed a distribution over all considered prototypes. The high deviation of both assignments will clearly lead to a low performance of the classifier concerning the functional-meaning of the DPs. Only considering the annotation results of the labelers, will lead to a consistency in 28.45 % of all cases. The classifier obtained a slightly higher performance of 33.49 %. This makes it impossible for the user to get a reliable statement on the functional-meaning of the pitch-contours. A reason for this low agreement is the number of overlaps contained in the DP audio-samples. To still be able to rate the performance of the classifier for the ALICO Corpus a new approach is introduced. As the annotation of the functional-meaning has a strong majority of DP-4 we reduced the classification problem from a seven class problem to a two class problem containing the classes “DP-4” and “other DP”. The confusion table of this new classification problem is depicted in Table 7, resulting in a (foreseeable) recall of 31.05 % and a (remarkable) precision of 90.67 %.

This means, almost all prototypes assigned to type DP-4 are true confirmation signals. Furthermore, looking at the idealized pitch-contours of DP-6, DP-4, DP-8 and DP-9 we noticed, that all of these prototypes can be represented within DP-6. This phenomenon is depicted in Fig. 4. As the functional-meaning of DP-6 and DP-4 state similar descriptions (confirmation and positive assessment →

**Table 7.** Confusion table of the reduced 2-class classification problem. Recall = 31.05 %, precision = 90.67 %

	classified DP-4	classified other DP
Acoustically labeled DP-4	68	7
Acoustically labeled other DP	151	20

positive feedback) we generalized these functional-meanings into one class, also containing DP-8 and DP-9. The acoustic evaluation of these formtypes also confirms this assumption. This led to an even higher recall of 51.10 % and precision of 95.08 %.



**Fig. 4.** Comparison of pitch-contours with similar functional-meanings

In case of the LMC HCI dataset we got a higher match (cf. Table. 8) of the acoustic labeling. Only the ratios of the assignments were available and not the direct assignments of the data-samples. Nevertheless, a high agreement in the frequency occurring of DP-2 and DP-4 is recognized. We assume that a direct assignment of the data-samples will also lead to a better match of pitch-contour and functional-meaning compared to the results of the seven class problem presented earlier in this section. This is explainable by the style of interaction: In almost no cases the computer interrupts the speaker. This leads to a low number of overlaps in the DP-samples and an almost “isolated” occurrence of the pitch-contours.

**Table 8.** LMC data: Frequency distribution of the a) classifier b) acoustic annotation of the functional-meaning; additionally used labels: NSP “no specification possible”

Label	DP-1	DP-2	DP-3	DP-4	DP-5	DP-6	DP-7	NSP
Item [%] a)	0	66.4	10.9	12.5	3.9	0.8	5.5	0
b)	2.9	77.0	2.2	14.2	0	1.1	0	2.6

## 6 Conclusion and Discussion

As a conclusion, we state that a high congruence of the visual manual labeling and the DP-classifier for pitch-contours could be shown (74.53 %). Thus, a cross-usage of HCI-trained classifiers applied on HHI data is viable, and our classifier can be used, to robustly evaluate the contours of DPs in both HCI and HHI. For the acoustical evaluation of the functional-meaning, the idealized form-function prototypes by Schmidt are not suitable in case of naturalistic HHI. For the given

ALICO dataset it was possible to reduce the seven class classification problem to a two class problem obtaining a foreseeable recall of 31.05 % and a remarkable precision of 90.67 %.

Furthermore, we can state that the given audio material of the ALICO corpus is not the most suitable dataset for this investigation. As the annotated feedback signals were not assigned to the different speakers, an overlap of the speakers is possible. This can lead to failures in the pitch-estimation, resulting in a high number of correct assignments of pitch-contours to the prototypes but no correlation of their functional-meaning. A mapping of the DPs to the speakers could lead to an increase of matches in the functional-meaning, as for both, speaker and listener separate headset microphone recordings are available. Additionally, most of the DPs contained in the corpora were of type DP-4. To get a significant statement about the classification of the functional-meaning the dataset should include a balanced distribution of all form-function-prototypes identified by Schmidt in an appropriate audio quality. As the investigation so far only deals with the classification of “isolated” DPs, in terms of no surrounding content of the considered DPs, the desired dataset can be merged from different corpora of the same conversational style (HCI/HHI) and level of naturalness (naturalistic/acted).

**Acknowledgment.** The work presented was done within the Transregional Collaborative Research Centre SFB/TRR 62 “Companion-Technology for Cognitive Technical Systems” ([www.sfb-trr-62.de](http://www.sfb-trr-62.de)) funded by the German Research Foundation (DFG).

## References

1. Allwood, J., Nivre, J., Ahlsén, E.: On the semantics and pragmatics of linguistic feedback. *J. Semant.* **9**(1), 1–26 (1992)
2. Biundo, S., Wendemuth, A.: Companion-technology for cognitive technical systems. *Künstliche Intelligenz* **30**(1), 71–75 (2016)
3. Böck, R., Siegert, I., Haase, M., Lange, J., Wendemuth, A.: ikannotate – a tool for labelling, transcription, and annotation of emotionally coloured speech. In: D’Mello, S., Graesser, A., Schuller, B., Martin, J.-C. (eds.) *ACII 2011, Part I*. LNCS, vol. 6974, pp. 25–34. Springer, Heidelberg (2011)
4. Buschmeier, H., Malisz, Z., Skubisz, J., Włodarczak, M., Wachsmuth, I., Kopp, S., Wagner, P.: Alico: a multimodal corpus for the study of active listening. In: *Proceeding of the 9th LREC 2014*, Reykjavik, Iceland (2014)
5. Fischer, K., Wrede, B., Brindöpke, C., Johantokrax, M.: Quantitative und funktionale Analysen von Diskurspartikeln im computer Talk. *Int. J. Lang. Data Process.* **20**(1–2), 85–100 (1996)
6. Giles, H., Coupland, N.: *Language: Contexts and Consequences*. Cengage Learning, Boston (1991)
7. Lotz, A.F.: *Differentiation von Form-Funktions-Verläufen des Diskurs Partikels “hm” über unterschiedliche mathematische Herangehensweisen*. Master’s thesis, Otto von Guericke University Magdeburg (2014)

8. Lotz, A.F., Siegert, I., Wendemuth, A.: Automatic differentiation of form-function-relations of the discourse particle “hm” in a naturalistic human-computer interaction. In: Günther, W. (ed.) *Elektronische Sprachsignalverarbeitung 2015, Studientexte zur Sprachkommunikation*, vol. 78, pp. 172–179. TUDpress (2015)
9. Paschen, H.: *Die Funktion der Diskurspartikel HM*. Master’s thesis, University Mainz (1995)
10. Rösner, D., Haase, M., Bauer, T., Günther, S., Krüger, J., Frommer, J.: Desiderata for the design of companion systems - insights from a large scale wizard of oz experiment. *Künstliche Intelligenz* **30**(1), 53–61 (2016)
11. Rösner, D., Frommer, J., Friesen, R., Haase, M., Lange, J., Otto, M.: Last minute: a multimodal corpus of speech-based user-companion interactions. In: *Proceeding of the 8th LREC 2012*, pp. 2559–2566, Istanbul, Turkey (2012)
12. Schmidt, J.E.: Bausteine der intonation. In: *Neue Wege der Intonationsforschung, Germanistische Linguistik*, vol. 157–158, pp. 9–32, Georg Olms Verlag (2001)
13. Siegert, I., Böck, R., Wendemuth, A.: Inter-rater reliability for emotion annotation in human-computer interaction: comparison and methodological improvements. *J. Multimodal User Interfaces* **8**(1), 17–28 (2014)
14. Siegert, I., Haase, M., Prylipko, D., Wendemuth, A.: Discourse particles and user characteristics in naturalistic human-computer interaction. In: Kurosu, M. (ed.) *HCI 2014, Part II. LNCS*, vol. 8511, pp. 492–501. Springer, Heidelberg (2014)
15. Siegert, I., Prylipko, D., Hartmann, K., Bück, R., Wendemuth, A.: Investigating the form-function-relation of the discourse particle “hm” in a naturalistic human-computer interaction. In: Bassis, S., Esposito, A., Morabito, F.C. (eds.) *Recent Advances of Neural Networks Models and Applications. SIST*, vol. 26, pp. 387–394. Springer, Heidelberg (2014)
16. Sievers, E.: *Grundzüge der Phonetik*, 5th edn. Breitkopf Härtel, Leipzig (1901)
17. Wendemuth, A.: *Grundlagen der stochastischen Sprachverarbeitung*. Oldenbourg, Munich (2004)