In: *Künstliche Intelligenz*, 27(2), 2013, pp. 137–143, DOI 10.1007/s13218-013-0241-8
The original publication is available at link.springer.com

Co-constructing Grounded Symbols – Feedback and Incremental Adaptation in Human–Agent Dialogue

Hendrik Buschmeier · **Stefan Kopp**

the date of receipt and acceptance should be inserted later

Abstract Grounding in dialogue concerns the question of how the gap between the individual symbol systems of interlocutors can be bridged so that mutual understanding is possible. This problem is highly relevant to human–agent interaction where mis- or non-understanding is common. We argue that humans minimise this gap by collaboratively and iteratively creating a shared conceptualisation that serves as a basis for negotiating symbol meaning. We then present a computational model that enables an artificial conversational agent to estimate the user's mental state (in terms of contact, perception, understanding, acceptance, agreement and based upon his or her feedback signals) and use this information to incrementally adapt its ongoing communicative actions to the user's needs. These basic abilities are important to reduce friction in the iterative coordination process of co-constructing grounded symbols in dialogue.

Keywords symbol grounding · dialogue · feedback · adaptation · human–agent interaction

1 Introduction

The classical 'symbol grounding problem' [15] refers to the constitution of meaning for a symbolic token through linkage to experiential knowledge about some external world. An agent links a symbolic token such as, for example, APPLE to its meaning by associating it to the perceptual category of apple-like objects. But what happens when two such agents come to interact through dialogue?

Dialogue is carried out to a large extent by exchanging linguistic symbols using speech, and can be seen as a symbol system in the classical sense. The symbolic tokens of a language (i.e., its words) are arbitrary and fortuitous [8, p. 198] as is common in symbol systems, but at the same time they are also conventionalised within a speech community. Despite being conventionalised, the symbol systems of any two interlocutors—even within the same speech community—differ because of variations in live experience. The same symbolic token can evoke at least slightly different meanings, potentially leading to miscommunication and misunderstanding. In addition, language use cannot draw upon conventions all the time. Often, a conventionalised symbol to denote a certain meaning does not readily exist, making it necessary for communicating agents to create new symbols and establish them as an 'ad hoc convention.' Further, the meaning of words is often too vague or coarse, and the semantics of composite symbols cannot always be derived solely from syntax. Language use thus has a pragmatic dimension that is not part of its lexical and compositional semantics (e.g., reference, deixis, the cooperative principle). Utterances must be 'situated,' i.e., interpreted in their context (previous discourses, the external situation), to determine their intended meaning.

Given this, how can agents participating in dialogue be sure that they share their individual 'meaning'—or at least that it is sufficiently similar—to understand each other? This is the 'grounding' problem in dialogue [11,9], which concerns the question of how interlocutors can actually establish 'common ground' in a conversational interaction. Solving this task requires interlocutors to continuously cooperate and to coordinate with each other. This problem poses key challenges for artificial agents (e.g., embodied conversational agents or robots), most of which remain unsolved.

H. Buschmeier · S. Kopp
Sociable Agents Group, CITEC, Bielefeld University
PO-Box: 10 01 31, 33501 Bielefeld, Germany
E-mail: hbuschme@uni-bielefeld.de
ORCID: http://orcid.org/0000-0002-9613-5713

S. Kopp
E-mail: skopp@uni-bielefeld.de
ORCID: http://orcid.org/0000-0002-4047-9277

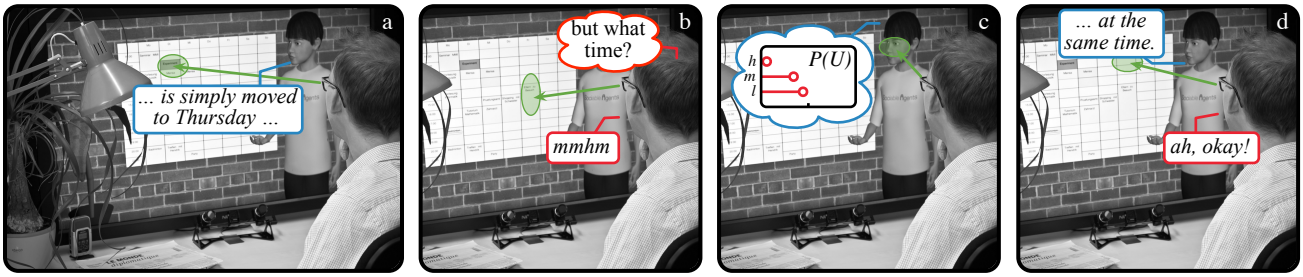

Figure 1 Dialogue coordination in a calendar assistant scenario: The agent 'Billie' informs a user that an appointment was moved from Tuesday 11 AM to Thursday 11 AM. (**a**) Billie tells the user that the appointment was 'simply' moved to Thursday. (**b**) It is not clear to the user that by 'simply,' Billie meant that the time remained unchanged. This confusion and uncertainty are displayed through gaze and verbal-vocal feedback. (**c**) Based on this feedback, Billie attributes a medium to low understanding to the user and (**d**) adapts by elaborating what was implied by 'simply.'

In this article we report work on endowing artificial agents with the basic abilities for this dynamic collaborative process of co-constructing symbol meaning, that makes dialogue so robust, efficient and versatile. Specifically, we focus on the use of communicative feedback and subsequent incremental adaptation, one of the most basic and fast mechanisms involved in dialogue. In our work at the Sociable Agents group of the Center of Excellence 'Cognitive Interaction Technology' (CITEC), we are currently developing a conversational agent that is attentive to the immediate, subtle feedback signals produced by its interlocutor, and that can respond to them by incrementally adapting its communicative behaviour, as illustrated in Figure 1. We begin by discussing how grounded symbols in dialogue are constructed jointly and often without explicit negotiation. We focus on the pivotal role of feedback and adaptation in this process, and present our approach to modelling these abilities such that they support mutual understanding and groundedness of information in human–agent dialogue.

2 Symbol meaning in dialogue is jointly constructed

We begin with the observation that to establish common ground, dialogue partners must ascertain whether and to what extent jointly used symbols relate to the same denotata. Further, if gaps in understanding are found, a method to overcome them must be determined. This task is usually framed as a coordination problem, requiring dialogue partners to behave cooperatively [14] and collaboratively [9]. Interlocutors overcome the differences between their individual symbol systems by using symbols with a sufficiently high certainty of being shared as a starting point for the interaction. By relying on this set of safely understood symbols, they can negotiate what is meant by a symbol in the dialogue context, what the meaning of more difficult symbols is and construct new ones if needed.

This behaviour has been demonstrated in several studies. Clark and Wilkes-Gibbs [12], for example, analysed how participants in an experiment referred to different Tangram figures for which a symbolic description was not readily available. Two participants collaborated in a task where they had to agree on the ordering of twelve Tangram figures. In each of six rounds they started with a random ordering of twelve cards with one participant explaining to the other how his or her cards were ordered. When analysing the dialogues between the participants, Clark and Wilkes-Gibbs found that creating references to the Tangram figures was a collaborative and iterative process. In the first round, the participants needed to establish a reference for the first time. The directing participant always described the Tangram figure, making a proposal of what he or she thought could be an acceptable 'perspective' to see the figure. The second participant then either signalled acceptance of this proposal (when able to understand it and being sufficiently certain that he or she identified the denoted Tangram figure) or signalled difficulty. In the latter case, the directing participant could then adapt his or her proposed reference by repairing part of it, expanding on it or replacing it with a different proposal. Over the course of the following rounds, participants preserved the previously created perspectives and drew upon prior references, refining them slightly but still using them in a definite way.

In later experiments, Brennan and Clark [4] showed that such joint conceptualisations of objects persisted over time. Interlocutors formed 'conceptual pacts' to which they adhered even if the context allowed for a much simpler conceptualisation in later situations (e.g., when the established reference was overly specific). Brennan and Clark also showed that the conceptualisation of an object formed with one interlocutor was usually not directly reused with a different interlocutor. Instead, the process of jointly constructing reference was begun anew.

In sum, when participants begin without symbolic tokens for reliably referring to an entity, they revert to describing the figure using symbols that are likely to be shared. This description opens up a perspective of how the figure/object could be conceptualised. This conceptualisation might then be refined until it is mutually accepted by both participants,

such that it provides a well-grounded basis for creating a novel symbol that can be used as a definite reference for the figure/object. This collaborative effort results in a shared conceptualisation and grounded symbols that have a sufficiently familiar denotation for both interlocutors.

3 Communicative feedback as signals of grounding

A crucial feature of natural dialogue is that the previously described process of jointly constructing a shared conceptualisation as a basis for symbol grounding is not entirely based on explicit negotiation of symbols. Rather, an important part of this coordination is realised through faster and more proactive adaptation: Interlocutors reveal their mental states during this process, indicating understanding, acceptance, and agreement (as well as their opposites non-understanding, rejection, and disagreement) in a variety of ways. At the same time, they are attentive to such signals by the interlocutor, continuously assessing them and responding to them by pro-actively adapting their communicative actions.

Clark and Schaefer [11] characterised dialogue as a sequence of 'contributions,' each consisting of two phases. In the 'presentation phase,' one dialogue participant presents an utterance. This presentation is followed by an 'acceptance phase' (which can also serve as the next presentation phase) in which the other dialogue participant accepts what has just been presented by providing 'evidence of understanding' (or acceptance or agreement).

Such evidence can be given in different ways. As Clark and Schaefer note [11, p. 267]: the interlocutor can show continued attention, initiate the next relevant contribution, demonstrate understanding, display the presentation verbatim, or provide 'communicative feedback' in the form of head gestures (e.g., nodding, shaking), facial expressions (e.g., smiling, raising an eyebrow) or short verbal-vocal expressions, 'backchannels,' such as 'uh-huh,' 'm,' or 'yeah.'

Contrary to common belief, communicative feedback is not merely a way of signalling the interlocutor to continue speaking, but a powerful mechanism that enables listeners to express their mental state towards the speaker's utterance. According to Allwood and colleagues [1,19], communicative feedback signals express the basic communicative functions[1] 'contact' (being "willing and able to continue the interaction"), 'perception' (being "willing and able to perceive the message"), 'understanding' (being "willing and able to understand the message"), and 'attitudinal reactions' (being "willing and able to react and (adequately) respond to the message") such as 'acceptance' or 'agreement' [1, p. 3]. These functions are related to each other hierarchically [1,11] such that higher functions imply lower functions (when signalling feedback of type understanding, for example, successful perception and contact are implied) and lower functions block higher functions (e.g., feedback of failed perception entails a problem in understanding).

The expressivity of feedback, however, is much richer and goes beyond these basic communicative functions. Feedback signals can take a vast number of different forms (if not infinitely many). Verbal-vocal feedback, although usually expressed with a small number of different quasi-lexical item such as 'yeah,' 'okay,' or 'huh?,' can be varied by generating new forms by combining or repeating several of them (e.g., 'hm okay,' 'yeah yeah yeah'). Even more variation can be generated by changing the prosodic overlay [25]. Consisting mostly of sonorants, verbal-vocal feedback signals can easily be lengthened or shortened, altered in their intonation and intensity, and modulated with voice quality.

This richness in form makes it possible for listeners to express evidence of understanding in more subtle ways than suggested by the five basic communicative functions. Listeners can, for example, indicate the strength of their understanding or non-understanding; their confidence in having understood correctly; and whether they are still in the process of understanding. They can also express precise attitudes such as surprise, boredom, or interest.

Providing evidence through feedback is common in dialogue because it has the advantage of being expressive but short and therefore only moderately restricted in its placement. Where the next relevant contribution can only be provided at a 'transition-relevance place' [22] and requires the speaker's willingness to pass on the turn (or involves a fight for the turn), verbal vocal feedback is short and unobtrusive enough [16] to be given at marked points within a speaker's turn [13]. Non-verbal communicative feedback can even be provided at any point in time and concurrently with the speaker's turn.

This flexibility in placement makes communicative feedback an ideal coordination device. Evidence of understanding can be signalled incrementally, as soon as it becomes relevant. When an addressee, for example, feels the need to communicate a problem in understanding to the speaker, he or she can do so immediately. Similarly, speakers can elicit evidence of understanding (with the help of feedback elicitation cues; see for example, [3,13]) from their addressees when it might help them tailor their utterance. In this way, speakers can monitor their addressees for understanding while an utterance is unfolding [10].

This incremental evidence of understanding makes the symbol grounding process in dialogue even more interactive than suggested in Section 2. Not only do both interlocutors in a dialogue contribute to the construction of shared conceptualisations and thus enable the grounding of symbols, but the collaborative activity even takes place at the sub-utterance level. A proposed conceptualisation that is easily understood

[1] Clark and Schaefer [11] describe a similar set of communicative functions for providing evidence of understanding in general.

by an interlocutor can be accepted right away, even before it has been fully explained; a poor conceptualisation proposal, on the other hand, can be altered as soon as difficulties become apparent or can even be rejected before more time is spent on it. This streamlines and speeds up the symbol grounding process [2].

4 Using human feedback in human–agent dialogue

Since it is such a prevalent and elemental mechanism in natural dialogue, it seems natural to make communicative feedback available for spoken language human–machine interaction such as in spoken dialogue systems, embodied conversational agents, or robots. Two aspects of communicative feedback can be modelled for technical systems. Firstly, systems can be endowed with the capacity to provide feedback while the user is speaking. This requires the system to know when feedback should be provided, what kind of feedback should be provided, and how such feedback can be expressed. Secondly, systems can be given the ability to process user feedback even while the system is speaking. For this, a system needs to be able to recognise and interpret a feedback signal in its context, reason about the feedback giver's intention in providing this feedback signal and incrementally adapt its ongoing language generation process.

So far, most research on feedback in dialogue agents has concentrated on the first of these two aspects, with particular emphasis placed on models of appropriate timing of backchannel feedback [26, 20, 18]. Recently, the increase in the capabilities of incremental natural language understanding, has directed attention to the question of what type of feedback should be provided [19, 24, 23].

Our interest lies in using the human interlocutor's feedback signals to make the agent's behaviour more adaptive to the user in order to support and maximise understanding in human–agent interaction and, as a result of this, to facilitate the process of creating shared conceptualisations. In previous work [6], we argued that an 'attentive speaker agent' needs to be able to (1) invite feedback from its users; (2) detect and interpret communicative feedback of its users; and (3) incrementally adapt its ongoing and subsequent utterances to its users' needs. The first point is a basic requirement to obtaining feedback from human interlocutors. The second and third points, however, reflect exactly the mechanisms humans use when jointly creating shared conceptualisations as a foundation for co-constructing grounded symbols, as discussed in Section 2. The remainder of this section will discuss the interpretation aspect and Section 5 will focus on adaptation.

As described above, mapping feedback signals onto meaning is very complex and depends on the discourse context, the dialogue situation, and, being only loosely conventionalised, also on the individual feedback giver (see, for example, [17]). When interpreting listener feedback, one therefore has to deal with a large amount of uncertainty beyond noise in a channel. Accordingly, we model feedback understanding probabilistically by adopting a Bayesian network approach to model the uncertainty about and (in-)dependencies between specific aspects of feedback. This formalism is well suited for our task as (1) its flexibility allows for easy interfacing with separately developed models of context (a more detailed description and examples are given in [7]), (2) it allows us to do causal as well as diagnostic reasoning and thus eventually provides a symmetric model for both feedback processing and production.

Based on ideas from [19], we interpret and describe feedback meaning in terms of an abstract representation of the listener's mental state [6]. Our attentive speaker agent 'Billie' is therefore equipped with a minimal 'Theory of Mind' that allows it to reason about the mental state that the user was most likely in when producing a feedback signal. This model, which we call 'attributed listener state' (ALS), is conceptualised in terms of the same mental categories that Allwood and colleagues [1, 19] assume to underlie the basic communicative functions of feedback.

The Bayesian network thus consists of five discrete random variables/nodes: C (being in contact), P (perceiving the utterance), U (understanding the utterance), AC (accepting the utterance), and AG (agreeing with the utterance). Each of these random variables can take the three states *low, medium,* and *high*, which model the strength of the underlying mental state (e.g., $U = high$ means that the utterance was understood very well, whereas $U = low$ means that it was understood rather poorly). The attributed listener state is the set of probability distributions over these five variables. It is interpreted as degrees of belief in the strength of the underlying mental state of the dialogue partner.

The hierarchical relationship between the underlying mental states (Allwood and colleague's hierarchy of feedback functions/Clark's ladder of actions [1, 11]) are reflected in the way the five ALS-variables influence each other. C influences P, P influences U, and U influences AC and AG (see Figure 2 for a graphical depiction of the model and these influences). This way, for example, a high degree of belief in the listener having perceived the utterance increases the likelihood of the listener also having understood it, and vice versa. On a more abstract level, an estimated grounding status is derived from the ALS variables. This is modelled with one variable GR (also with states *low, medium,* and *high*) that depends on all five ALS-variables, each exerting a different influence on groundedness.

The model's parameters are hand-crafted based on theoretical considerations as well as on intuition gained from annotating feedback use in human–human dialogue. To reduce the number of parameters that need to be specified, the conditional probability tables of the ALS-variables are

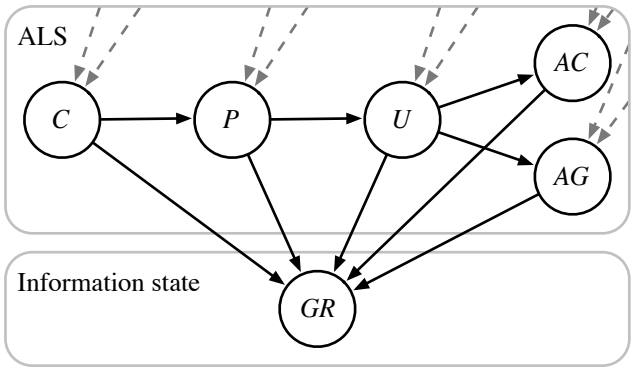

Figure 2 The 'attributed listener state' (ALS) and its effect on the grounding status *GR* of an entity in the information state: five ALS-nodes *C* (contact), *P* (perception), *U* (understanding), *AC* (acceptance), and *AG* (agreement) model the underlying mental states of Allwood and colleagues' basic communicative functions of feedback [1, 19]. The influences between the nodes model the hierarchical relationship between these functions [1, 11]. The grey dashed arrows indicate influences from feedback signal and discourse context (cf. [7] for further details).

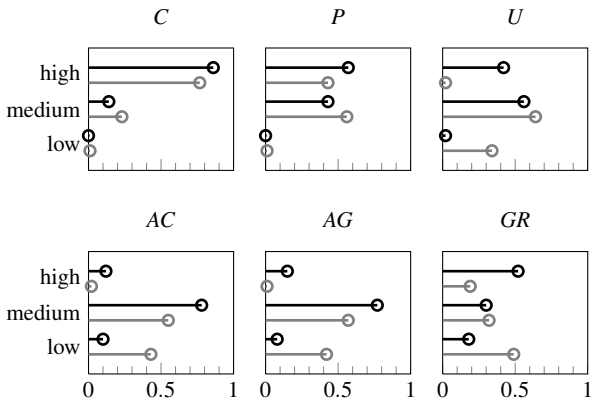

Figure 3 Two sample contrasting belief states: The *x*-axes show the degree of belief in a variable's states. The black bars show the result for the model processing feedback of function understanding, the grey bars for feedback of function non-understanding. Note how the degree of belief in different states shifts for variables *U*, *AC*, *AG* and *GR*—but only very little in variables *C* and *P*.

generated from structured representations [7]. This allowed for a high level definition of the model's behaviour, which made it straightforward to express the relationships between the ALS-variables without making micro decisions for every single state combination. The hand-crafted model can then serve as a starting point to a process that, e.g., by online or active learning, leads to a domain- and/or user-specific model.

Figure 3 illustrates the model's behaviour for two contrasting conditions by showing the belief state of the ALS-variables after Bayesian network inference had been run. In the first condition, drawn in black, feedback of positive understanding is provided. This results in the model having a high degree of belief in the interlocutor being in good contact and perceiving and understanding the utterance moderately or well. Further, the model estimated that acceptance and agreement with the utterance were moderate. Overall the utterance was estimated to be grounded fairly well, but still with a significant chance of being not grounded.

In the second condition, drawn in grey, feedback of non-understanding was received. In this case, the degrees of belief for the variables *U*, *AC, AG* and *GR* were shifted towards medium and low, whereas the belief state of the variables *C* and *P* were almost not affected due to the modelled hierarchical relationship of the feedback functions.

The ALS does not solely depend on the listener's feedback signal. It is also influenced by feedback-external factors such as discourse context or communicative situations, hinted at with the grey dashed arrows in Figure 2. As an example, consider how the meaning of listener feedback interacts with the difficulty of the speaker's corresponding utterance. In [7], we sketched a simple model of utterance difficulty based on the utterance's length, whether the information it conveyed was novel, and how surprising it would be for the listener. This model exerts an influence on the variables *P* and *U* and interacts with listener feedback roughly in the following manner: If the utterance is complex, receiving no feedback from the listener should result in a degree of belief of *P* and *U* being mostly *medium* to *low.* Conversely, for a simplistic utterance, receiving no feedback from the listener, does not indicate a problem and should thus result in a degree of belief of *P* and *U* being mostly *medium* or *high.*

5 Adapting to the interlocutor's needs

As a result of the feedback interpretation process, the attributed listener state captures what went right and/or what went wrong during the coordination and grounding process in dialogue. It provides an abstract—but rich—representation that reflects some of the subtle details (e.g., the degree to which a mental state holds) that lie at the heart of the expressiveness of communicative feedback. This expressiveness is crucial for grounding and coordination in human dialogue, as it allows the interaction partners to identify, with some precision, where problems are located and which aspects of language production and joint construction of shared conceptualisations might need adaptation. Capturing the subtle aspects of a feedback signal in a rich semantic representation of feedback meaning is therefore an ideal foundation for making informed adaptations, which help the interlocutor understand and agree, and will therefore advance the joint project of creating shared conceptualisations and grounded symbols conventionalised in an ad-hoc manner.

Depending on difficulties that a listener encounters, the speaker needs to adapt an utterance such that the specific problem will be resolved. Based on the level where the problem originates in (perception, understanding, etc.), different

adaptation mechanisms and strategies need to be considered. A problem at the level of perception might be resolved by simply repeating the utterance or the problematic phrase or word. If perception was impaired, e.g., by noise in the environment, it might help to adapt the level of realisation by hyper-articulating and raising the speech volume while repeating the misheard fragment. A misunderstanding or non-understanding caused, e.g., by the interlocutor conveying information only implicitly, can be resolved by explicating it (see the example in Figure 1), or making future utterances more redundant. When it becomes apparent that the interlocutor's conceptualisation deviates considerably, the best strategy might be to try a different perspective, perhaps taking the interlocutor's as a starting point. It might also sometimes be necessary to combine adaptation mechanisms that operate on different levels. This way, several problems can be tackled at once, or a set of possible solutions can be attempted when the exact problem is unclear.

So far, we have modelled and implemented two adaptation mechanisms that operate incrementally on the surface form of utterances, and one mechanism that operates on the level of discourse structure [5]. At the microplanning stage of our incremental natural language generation system, we can influence the amount of redundant information that one increment of an utterance (a 'sub-utterance chunk') contains (within the increment or with respect to the previous discourse context). The microplanner is also able to produce more or less verbose versions of a chunk. On the level of micro-content planning, our generator can decide how to structure an utterance, e.g., whether a chunk should be realised as planned, postponed for an adapted repetition of the previous chunk, or skipped completely. Table 1 shows the result of these mechanisms operating on example (sequences of) sub-utterance chunks.

In other work, Reidsma and colleagues [21] have focussed on mechanisms on the level of realisation. Their behaviour realiser 'Elckerlyc' can increase the speech rate and the volume of the speech synthesis flexibly at any point in time. In ongoing work, we are incorporating all of these mechanisms in a larger framework of incremental behaviour generation and adaptation for conversational agents.

Table 1 Examples of adapted natural language output, subject to variation due to different adaptation mechanisms. Redundancy can either be prohibited (**a**) or permitted (**b**). Verbosity can take different strength, from low to high (**c–d**). On a structural level, sub-utterance chunks can be skipped (**f**), produced as planned (**g**), or postponed for an adapted repetition of the previous chunk (**h**). English gloss of German NLG-output is provided in italics; a '◇' marks sub-utterance chunk boundaries.

Mechanism	***x***	**Generated output**
Redundancy	a	'morgen ◇ von 12 bis 14 Uhr' *(tomorrow ◇ from 12 to 14 PM)*
	b	'morgen den 21. Dezember ◇ 12 bis 14 Uhr' *(tomorrow December 21 ◇ from 12 to 14 PM)*
Verbosity	c	'Vorlesung KI' *(Lecture AI)*
	d	'Betreff: Vorlesung KI' *(subject: Lecture AI)*
	e	'mit dem Betreff Vorlesung KI' *(with the subject: Lecture AI)*
Structure	f	'ε ◇ 12 Uhr ◇ Vorlesung KI' *(ε ◇ 12 PM ◇ Lecture AI)*
	g	'morgen ◇ 12 Uhr ◇ Vorlesung KI' *(tomorrow ◇ 12 PM ◇ Lecture AI)*
	h	'morgen ◇ 12 Uhr ◇ ähm ◇ von 12 bis 14 Uhr ◇ Vorlesung KI' *(tomorrow ◇ 12 PM ◇ uhm ◇ 12 to 14 PM ◇ Lecture AI)*

6 Conclusion

In summary, we have discussed how the grounding of symbols extends from the level of the individual agent that associates it with subjective experiential qualities, up to the level of dialogue where two or more agents use symbolic communication in order to establish mutual understanding. This work is part of a larger research programme in the Sociable Agents group of the Center of Excellence 'Cognitive Interaction Technology' (CITEC), which aims to enable robotic or software agents to engage in human-like fluid and adaptive conversational interactions, and in this way establish common ground with their users.

We argued that the process of symbol grounding in dialogue requires a shared conceptualisation and an agreement about what is denoted by a certain symbolic token. Neither of these requirements are given a priori, but must be co-constructed cooperatively and collaboratively. We reported work on endowing artificial agents with basic abilities required in this process. Specifically, instead of targeting heavyweight models of detailed mentalising and explicitly negotiating mental states, our model relies on fast and incremental adaptations based on minimal—but rich—attributed mental states that are updated continuously and probabilistically from the feedback information an interlocutor provides.

Acknowledgements This research is supported by the Deutsche Forschungsgemeinschaft (DFG) in the Center of Excellence EXC 277 in 'Cognitive Interaction Technology' (CITEC).

References

1. Allwood, J., Nivre, J., Ahlsén, E.: On the semantics and pragmatics of linguistic feedback. Journal of Semantics **9**, 1–26 (1992)
2. Bavelas, J.B., Coates, L., Johnson, T.: Listeners as co-narrators. Journal of Personality and Social Psychology **79**, 941–952 (2000)
3. Bavelas, J.B., Coates, L., Johnson, T.: Listener responses as a collaborative process: The role of gaze. Journal of Communication **52**, 566–580 (2002)

4. Brennan, S.E., Clark, H.H.: Conceptual pacts and lexical choice in conversation. Journal of Experimental Psychology: Learning, Memory, and Cognition **22**, 1482–1493 (1996)
5. Buschmeier, H., Baumann, T., Dosch, B., Kopp, S., Schlangen, D.: Combining incremental language generation and incremental speech synthesis for adaptive information presentation. In: Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue, pp. 295–303. Seoul, South Korea (2012)
6. Buschmeier, H., Kopp, S.: Towards conversational agents that attend to and adapt to communicative user feedback. In: Proceedings of the 11th International Conference on Intelligent Virtual Agents, pp. 169–182. Reykjavík, Iceland (2011)
7. Buschmeier, H., Kopp, S.: Using a Bayesian model of the listener to unveil the dialogue information state. In: SemDial 2012: Proceedings of the 16th Workshop on the Semantics and Pragmatics of Dialogue, pp. 12–20. Paris, France (2012)
8. Chao, Y.R.: Language and Symbolic Systems. Cambridge University Press, Cambridge, UK (1968)
9. Clark, H.H.: Using Language. Cambridge University Press, Cambridge, UK (1996)
10. Clark, H.H., Krych, M.A.: Speaking while monitoring addressees for understanding. Journal of Memory and Language **50**, 62–81 (2004)
11. Clark, H.H., Schaefer, E.F.: Contributing to discourse. Cognitive Science **13**, 259–294 (1989)
12. Clark, H.H., Wilkes-Gibbs, D.: Referring as a collaborative process. Cognition **22**, 1–39 (1986)
13. Gravano, A., Hirschberg, J.: Turn-taking cues in task-oriented dialogue. Computer Speech and Language **25**, 601–634 (2011)
14. Grice, H.P.: Logic and conversation. In: P. Cole, J.L. Morgan (eds.) Syntax and Semantics 3: Speech Acts, pp. 41–58. Academic Press, New York, NY (1975)
15. Harnad, S.: The symbol grounding problem. Physica D **42**, 335–346 (1990)
16. Heldner, M., Edlund, J., Hirschberg, J.: Pitch similarity in the vicinity of backchannels. In: Proceedings of INTERSPEECH 2010, pp. 3054–3057. Makuhari, Japan (2010)
17. de Kok, I., Heylen, D.: The MultiLis corpus – Dealing with individual differences in nonverbal listening behavior. In: Proceedings of the 3rd COST 2102 International Training School, pp. 362–375. Caserta, Italy (2011)
18. de Kok, I., Ozkan, D., Heylen, D., Morency, L.P.: Learning and evaluating response prediction models using parallel listener consensus. In: Proceedings of the 12th International Conference on Multimodal Interfaces. Beijing, China (2010)
19. Kopp, S., Allwood, J., Grammar, K., Ahlsén, E., Stocksmeier, T.: Modeling embodied feedback with virtual humans. In: I. Wachsmuth, G. Knoblich (eds.) Modeling Communication with Robots and Virtual Humans, pp. 18–37. Springer, Berlin, Germany (2008)
20. Morency, L.P., de Kok, I., Gratch, J.: A probabilistic multimodal approach for predicting listener backchannels. Autonomous Agents and Multiagent Systems **20**, 70–84 (2010)
21. Reidsma, D., de Kok, I., Neiberg, D., Pammi, S., van Straalen, B., Truong, K., van Welbergen, H.: Continuous interaction with a virtual human. Journal on Multimodal User Interfaces **4**, 97–118 (2011)
22. Sacks, H., Schegloff, E.A., Jefferson, G.: A simplest systematics for the organization of turn-taking for conversation. Language **50**, 696–735 (1974)
23. Schröder, M., Bevacqua, E., Cowie, R., Eyben, F., Gunes, H., Heylen, D., ter Maat, M., McKeown, G., Pammi, S., Pantic, M., Pelachaud, C., Schuller, B., de Sevin, E., Valstar, M., Wollmer, M.: Building autonomous sensitive artificial listeners. IEEE Transactions on Affective Computing **3**, 165–183 (2012)
24. Wang, Z., Lee, J., Marsella, S.: Towards more comprehensive listening behavior: Beyond the bobble head. In: Proceedings of the 11th International Conference on Intelligent Virtual Agents, pp. 216–227. Reykjavík, Iceland (2011)
25. Ward, N.: Non-lexical conversational sounds in American English. Pragmatics & Cognition **14**, 129–182 (2006)
26. Ward, N., Tsukahara, W.: Prosodic features which cue back-channel responses in English and Japanese. Journal of Pragmatics **38**, 1177–1207 (2000)

Hendrik Buschmeier is a PhD-student at CITEC, Bielefeld University. He is interested in dialogue phenomena and the mechanisms underlying dialogue processing. Right now, Hendrik works on a computational model of dialogue coordination based on linguistic feedback and adaptive language production.

Stefan Kopp is research group leader at the Center of Excellence in 'Cognitive Interaction Technology' (CITEC) and principle investigator in the SFB 673 'Alignment in Communication' at Bielefeld University. His research interests center around intelligent systems that can engage in human-like conversational interaction and cooperation, for which he combines empirical studies with cognitive modeling and machine learning techniques. Stefan Kopp is the current president of the German Cognitive Science Society (GK).