International Journal of Mechanical, Industrial and Aerospace Sciences
ISSN: 2517-9950
Vol:11, No:5, 2017

Investigating Breakdowns in Human Robot Interaction: A Conversation Analysis Guided Single Case Study of a Human-Robot Communication in a Museum Environment

B. Arend, P. Sunnen, P. Caire

***Abstract*—**In a single case study, we show how a conversation analysis (CA) approach can shed light onto the sequential unfolding of human-robot interaction. Relying on video data, we are able to show that CA allows us to investigate the respective turn-taking systems of humans and a NAO robot in their dialogical dynamics, thus pointing out relevant differences. Our fine grained video analysis points out occurring breakdowns and their overcoming, when humans and a NAO-robot engage in a multimodally uttered multi-party communication during a sports guessing game. Our findings suggest that interdisciplinary work opens up the opportunity to gain new insights into the challenging issues of human robot communication in order to provide resources for developing mechanisms that enable complex human-robot interaction (HRI).

***Keywords*—**Human-robot interaction, conversation analysis, dialogism, museum, breakdown.

I. Introduction

OUR article seeks to make a contribution to the field of human-robot communication. There have been some studies in diverse contexts to understand how humans and robots 'communicate' [1]-[3]. In the present paper, we approach the 'HRI story' from the perspective of Social Sciences. More precisely we draw on CA methods to study HRI by exploring in a single case study how humans conduct multimodally embodied communication with a NAO robot in a modern art museum.

At first, we will outline the theoretical framework and the methodological apparatus we rely on to investigate how humans situatedly proceed when interacting with a robot. After describing the context of the analyzed event, we present a fine grained CA informed analysis on a contextualized troublesome 'episode' of HRI. By relying on video data, we shall show how participants engaging in a sports guessing game, i.e. an IRF sequence, with a NAO robot, mobilize multiple multimodal resources to make interaction work and overcome breakdowns. IRF sequences are generally known in educational contexts as (teacher) initiation, (learner) response, and (teacher) follow-up or feedback.

In terms of dialogic dynamics and recipient design, we will discuss raising questions regarding the 'responsivity' of the NAO robot and its 'recognition ability' to capture verbal utterances. This research effort is focused on exploring how features of human 'dialogue' can shed light on HRI.

Béatrice Arend and Patrick Sunnen are with the Institute of Education and Society, University of Luxembourg, L-4366 Esch Belval, Luxembourg (phone: 352 466644 9214, 352 466644 9211, e-mail: beatrice.arend@uni, patrick.sunnen@uni).

Patrice Caire is with the Interdisciplinary Centre for Security, Reliability and Trust, University of Luxembourg (phone: 352 466644 5413, e-mail: patrice.caire@uni.lu).

II. Theoretical and Methodological Issues

A. Dialogism

To investigate how humans (adults and/or children) engage in joint activities by communicating with each other [4], [5], we rely on an 'extended' concept of Bakhtinian 'dialogism' which is centered on the reciprocal dynamics of real-time interactions accomplished in embodied practices [6]. Thus, we approach human communication as dialogically co-constructed in situated social interactions and consider that 'dialogue' consists in a multi-layered communication process instantiating in verbal and non-verbal utterances (i.e. in language and other semiotic resources), unfolding in time [7], [8].

In line with 'dialogic' theory, we argue that communication arises in a space of difference, i.e. 'dialogue' is oriented to otherness. According to Bakhtin, every utterance is other-oriented and can be considered as the participants' shared territory [9]. We should note here that in 'dialogism', 'addressivity' and 'responsivity' are inherent features of utterances. Bakhtin claims that "an essential (constitutive) marker of the utterance is its quality of being directed to someone, its *addressivity*" [7]. When addressing the other, the speaker (the participant) takes into account the way the other may perceive 'the word' as well as the other's situation. That means that "from the very beginning", the utterance "is constructed while taking into account possible responsive reactions" [7]. By the same means, in terms of responsivity, every utterance is also "filled with various kinds of responsive reactions to other utterances of the given sphere of communication" [7]. The utterance is in fact doubly oriented: it has a responsive backward relation to prior utterances and an anticipatory projective relation to potential next utterances. With regard to Bakhtinian dialogism, we assert that every utterance is dialogic, responding to previous ones and anticipating 'the other's' reply; it is impossible to determine the position of an utterance without correlating it with other

positions [7]. Furthermore, an utterance, "as we know, is constructed between two socially organised persons […] it is precisely *the product of the reciprocal relationship between speaker and listener, addresser and addressee*" [9]. Participants build on the active understanding and response of 'the other', so that it is in communication with 'the other' that 'true dialogic understanding' [9] can be mutually achieved.

From that point of view, other-oriented verbal and non-verbal utterances can be understood as internally social and the dialogic dynamics of human communication can be seen as intrinsically intertwined with social organization. By extension, we consider 'dialogism' as an epistemology for social sciences to study human communication as well as human-robot communication.

B. Conversation Analysis

For a deeper understanding of the pragmatics of 'dialogue', we have to grasp and to visualise human communication in the process of its social occurrence, i.e. in its sequentially unfolding instantiation. Here, CA gives access to the methods participants use to build locally relevant actions and thus sheds light on the above mentioned dialogic dynamics. In a CA informed approach, interactions are considered as 'orderly', and orderliness is seen as a product of the participants' systematic deployment of interactional 'methods' or 'devices'. Thus, in our research, we rely on CA as "a disciplined way of studying the local organisation of interactional episodes" [10] to deal with embodied social practices. Focussing on 'dialogic' communication as an embodied process, we analyse the practices by which co-participants sequentially coordinate and organize their courses of communication. A CA perspective on human interaction as "*organizational* and *procedural*" gives access to people's talk with each other as "an emergent collectively organized event" insofar as "the analytic purpose is not to explain *why* people act as they do, but rather to explicate *how* they do it" [11]. The following provides a brief overview of the CA key terms we will rely on in our subsequent analysis.

CA shares with dialogist traditions the assumption of 'other orientation' in human sense-making [6]. Thus, the concepts of 'addressivity' and 'responsivity', related to both the backward and the projective aspects of utterances, are known in CA as 'recipient design' of turns and sequences [8]. If 'dialogism' points out that "both the composition and, particularly, the style of the utterance depend on those to whom the utterance is addressed, on how the speaker senses or imagines his addressees" [7], CA research states, in a similar but rather (more) practical orientation, that humans always adjust their actions to a specific recipient. In their study on the organization of turn-taking, Sacks et al. [12] refer to "recipient design" as "a multitude of respects in which the talk by a party in a conversation is constructed or designed in ways which display an orientation and sensitivity to the particular other(s) who are the co-participants". We note here that recipient design can also operate in terms of how participants rely on non-verbal resources (gaze, gesture, body movement) as stances of their orientations toward the recipient. Recent CA informed research in human communication indeed underlines the complex coordinating dynamics of verbal and non-verbal utterances (speech, gaze, gesture, body posture) in socially organized joint activities. That means that by building on assumptions about the interactional partner's knowledge and expectancies, participants adjust their (verbal and non-verbal) utterances to the recipient, thus constituting a continuously modified 'partner model' [13]. Heath [14] emphasizes that participants, by displaying 'availability' construct a "pre-initiating activity providing an environment for the occurrence of a range of actions". A display of 'recipiency' elicits an action, a turn at talk. Participants show that they are 'ready to listen' or that they will go on to talk.

As pointed out above, communication is sequentially organized [12]. Sequences are ordered series of turns through which participants accomplish and coordinate an interactional activity. The sequential organization is found in the structure of turn taking; the relevance of any turn is to be understood from its occurrence in a series of turns. Turns are unfolding in time referring to what has been said (done) before. They simultaneously initiate expectations about relevant next turns. The most common type of sequences are dyadic adjacency pairs uttered by two different speakers who produce one turn each. In the analyzed data we will deal with a three-turn IRF sequence. More specifically, turn taking is to be considered in terms of TCU (turn constructional units) and turn allocation at TRP (transition relevance places) [15]. In most instances, turn transition (speaker change) is accomplished smoothly at TRP, and such places are accountably projected. At TRP, the different parties negotiate who is taking the next turn. Sacks et al. [12] propose three options. First, the current speaker can select the next; another option is self-selection; third, if the current speaker does not select the next participant and there is no self-selection from another party at TRP, the current speaker can decide to continue.

For the purpose of our analysis, we further emphasize the following 'basic rules' of turn taking: Only one person talks at a time. Overlap of speech is common but brief. Participants proceed to the next turn with very little gap. Longer gaps and silence should be avoided; when they occur, they are meaningful and are most of the time perceived as trouble.

C. HRI and CA

Multimodal CA is a relatively 'new' method for the investigation of human robot communication [3], [17]. Few studies have focused on how humans talk to a robot with regard to their changing expectations over the course of the interaction. We will point out through our analysis that CA reveals to be a suited approach to investigate the complex sequential organization of HRI. As CA allows to show how humans adjust their actions to a specific recipient with regard to their assumptions about the interactional partner, we can show in this single case study how participants deal with a NAO recipient during a sports guessing game. In the analyzed multi-party communication, the participants display differing partner models in action to conduct the IRF designed game activity, by that way, overcoming breakdowns in the unfolding

event.

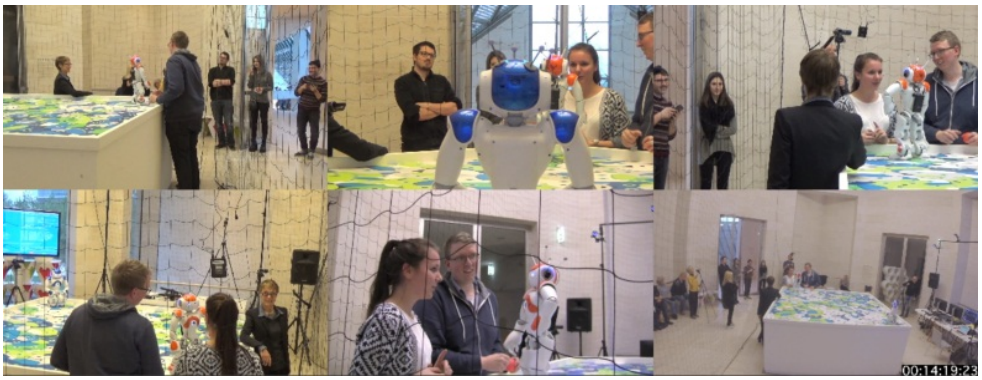

Fig. 1 Joint screen (all six perspectives)

Fig. 2 Joint Screen (perspectives of the two moving cameras)

Our research seeks to demonstrate how a systematic fine-grained understanding of human 'dialogue' can serve as a resource for designing mechanisms that enable complex HRI. To capture the encounter of humans and robots (here: playing a sports guessing game) with great accuracy and detail, we recorded the event from different shots and perspectives. The recording equipment was composed of four fixed cameras, mounted on tripods, and two moving cameras operated by the researchers. The six resulting video data streams were connected within one space (Fig. 1) to generate an 'expanded-around view' of the ongoing event; elsewhere, we termed this apparatus 'joint screen' [4]. For reasons of convenience and to ensure that the chosen frame grabs are not too small to recognize relevant details, we chose here to rely on images from the perspectives of the two moving cameras to support our analysis (Fig. 2).

Regarding the transcription, we consider the human participants' verbal conduct and their gestures as well as the verbal utterances of the robot, its arm-movements, its 'shiny' eyes and its 'dadup' sound. The transcription tool 'TranScripter' [17] allows to generate both a list format transcription, more adequate for the representation of sequentiality, and a graphical transcription in partition format, practical for a multimodal analysis of various simultaneous lines of action. Thus, in order to enhance the readability of the transcript, we will provide a combined version of both: an 'overviewing' representation of the participants' multimodal utterances displayed in their temporal unfolding and their mutually occurring synchronicity (see transcriptions 1-3). Furthermore, some relevant screenshots displaying the emerging visible conduct of the participants are connected by a line to the related bars of the partition.

III. Context

The COROBOTS-installation [18] was a part of the visual arts and technology exhibition "Eppur si muove" at the Museum of Modern Art in Luxembourg. Researchers from the Automation Research Group had reconstructed a small lab allowing visitors to play, among others, a sports guessing game with NAO robots. After introducing itself, the robot explains the game and with the agreement of the human, it starts imitating a sport (e.g. tennis, bodybuilding, skiing) by using gestures and body movements, and by playing sounds (e.g. tennis ball against a racket, ski sliding on the snow). It then asks the participant to guess the sport it is 'performing'. Upon the visitor's proposed guess, the robot replies whether it is the right or wrong answer and proceeds with the next sport.

IV. Analysis

A. Setting the Scene

In the subsequent analysis, the research will shed light on a 'troublesome' communication episode of 39 seconds between Julie, a visitor to the museum, and a NAO robot. Besides the NAO robot and the visitor, it involves the two researchers Patrice and Alex, who are part of the setting, i.e. they both manage the COROBOTS installation and ensure a well-run event.

In terms of recipient design and responsivity, we will study how an IRF sequence, which could be considered as 'simply'

structured, turns into a recurring ('even bothering') question-answer 'game' with pending feedback(s) and multiple repairs. Despite the instructional work of the two researchers, Patrice and Alex (aiming at a smooth dialogue), we will witness a rather confused visitor (Julie) giving accounts of understanding the robot's silence and its replies as unexpected.

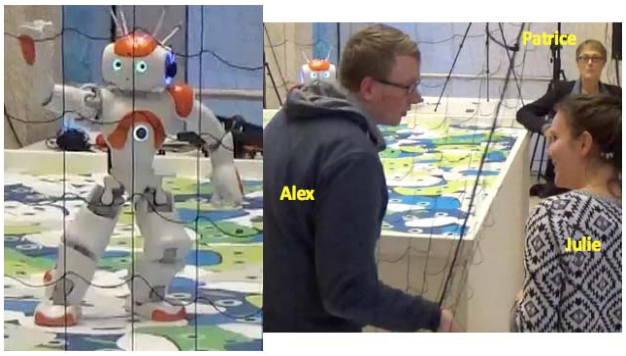

Fig. 3 Joining the scene

When we join the scene, the robot has performed a sport ('bodybuilding') and is pausing. The robot is pausing before taking another initiative and asking again 'what was that sport'. Julie volunteers to play the game and moves into position to address the robot (Fig. 3). While the robot is still in its 'pausing position', Julie checks with the two researchers that she has correctly guessed the sport 'bodybuilding'. Both Patrice and Alex confirm and validate 'bodybuilding' as the 'preferred' response. Thus, after this 'rehearsal', Julie can proceed to 'perform' the right answer.

We could now expect a fluent realization of the 'inbuilt' IRF pattern: the robot is supposed to put a question (**I**nput), Julie to give the right answer (**R**esponse) and then the NAO to provide a positive feedback (**F**). However, matters will be somewhat different.

V. Analysis

A. Analysis 1: "Bodybuilding"

The NAO takes the initiative by moving its arms (1), then it emits a 'dadup' sound (2) and, after one second, utters 'what was that sport' (5), by that way addressing a potential recipient who is supposed to provide an answer. Then, there is a second 'dadup' sound (6) (announcing the robot's recipiency) and then the NAO moves its arms announcing a 'rest' position (9).

From a certain point onward, the NAO has 'shiny eyes' (3). Throughout our data we can observe that co-participants orient to the robot's eyes to determine whether and how it is attending to what is being said. With regard to recipient design, we could consider the robot's 'shiny eyes' as an account of availability "which can serve to mark the completion of establishing co-presence and readiness for talk on topic" [14]. However, it is the *dadup* (6) following the question which displays the robot's recipiency, thereby eliciting the co-participant's answer. From a programming perspective, the robot is only 'listening' for the speaker's turn after the second *dadup* sound (6). We should note here that a continuous hearing is technically not possible for the robot since it would recognize its own speech as an answer and run the risk of an infinite loop. Furthermore, the acoustic characteristics of the location of the installation (a semi-open room with a high ceiling) contributed to the development of noise and echoes, which impacted on the robot's ability to capture targeted verbal utterances.

In the following, we will point out several issues related to the participants' 'addressivity' and 'responsivity'.

Transcription 1

1	NAO	((arm movement))
2	NAO	*dadup* sound
3	Nao	((shiny eyes))
4	Alex	((gesture 'wait'))
5	NAO	what was that sport?
6	NAO	*dadup* sound
7	Alex	((gesture 'go'))
8	Julie	bodybuilding
9	NAO	((arm movement))

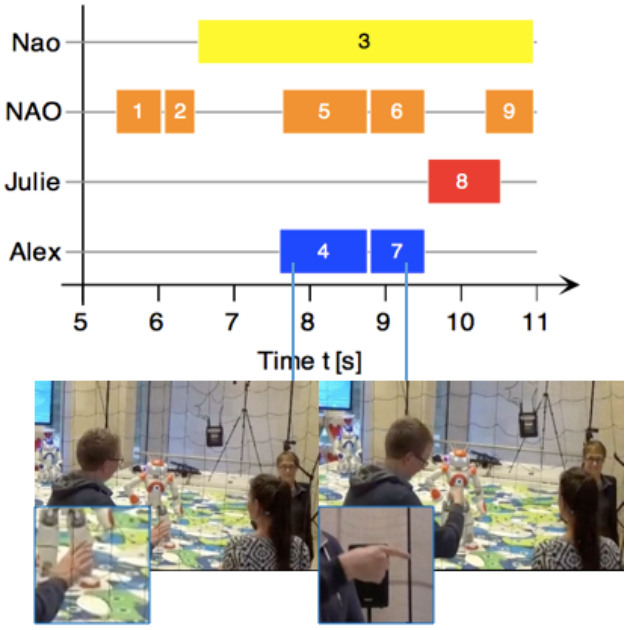

Fig. 4 Multimodal transcription 1

While gazing at the robot, concurrently with the NAO's verbal utterance (5), Alex addresses Julie with a hand gesture and signals her to wait before responding (4) (Fig. 4). Then, synchronically with the robot's *dadup* sound, by means of a pointing gesture, Alex invites Julie to perform (7), (Fig. 4). Julie immediately follows Alex's instruction and takes the next turn while facing the robot addressee (8). If it is true that she self-selects as the next speaker, it is however Alex who instructs her *when* she has to answer. The NAO is indeed hearing, thus recognizing verbal utterances as 'right' or 'wrong', when Alex gives the 'go' signal (see the above

mentioned programming details).

We can point out here that Alex makes accountable through his gestures that he is orienting to the *dadup* sound as relevant for the HRI. At the same time, in terms of the transition relevance place, Alex shows that he anticipates that Julie is about to apply the rule of proceeding to the next turn with very little gap [12]. Thus, he instructs her to wait before answering. Doing so, he displays a simultaneous orientation to the turn-taking system of the robot (as different from the one of humans) and to the sequential nature of everyday talk-in-interaction. Furthermore, through his gestures, he addresses the assumption that the visitor may not be familiar with the NAO's interactional features. We can consider Alex' gestures as instantiating his simultaneous 'double' orientation to two different partner models. Julie seeks to 'get it right' and makes Alex' gestures relevant as 'instructional' when she provides her verbal utterance at the indicated moment.

From a CA perspective (investigating *human* communication), the natural next action in terms of IRF sequence would be the feedback of the robot, but the feedback is overdue, i.e. it is absent. The lack of feedback, respectively the 'pausing position' the robot holds, are disturbing, all the more as Julie has answered at a timely transition relevant place projected by the *dadup* sound; and the robot's feedback is an expected part of the sports guessing game. Instead, feedback will be provided by the researchers eliciting Julie to speak louder and more slowly. The robot was in fact in its listening position but could not capture Julie's verbal utterance (due to the presumed low volume and the high speed of her speech flow).

B. Analysis 2: Bodybuilding: Second and Third Trial/Repairs and Confusion

Since the robot's feedback is pending (it is in its 'pausing' position, which means it is technically not the recipient), the researchers instruct Julie to increase the volume level of her voice and to decrease her speaking rate (10, 11), suggesting that there was an acoustic problem. The instructions are also accounts of Alex and Patrice's assumptions that Julie will reformulate the utterance or elaborate further on it. Julie acknowledges with "ok" while facing the robot and proceeds immediately to answer with partly increased pitch (12). With the first phrase of her turn ("ok") oriented to Patrice and Alex, Julie is approving the researchers' preceding instructions. The second part of her turn "bodybuilding" can be regarded as "a link in the chain of speech communication" in a Bakhtinian sense [7]. Julie's response "is filled with various kinds of responsive reactions to other utterances of the given sphere" (ibid.). The repaired "bodybuilding" (12) addressed to the NAO can be considered as the dialogic display of Julie's responsive understanding of what the immediate previous turn was about (next-turn proof procedure), instantiated in a 'new' response to the robot's prior input (5). The whole utterance "ok bodybuilding" has a double orientation: it has a responsive backward relation to prior utterances (the researchers' instructions and the NAO's question) and an anticipatory projective relation to potential next utterances (the NAO's feedback) [7].

Transcription 2

10	Alex	louder
11	Pat	louder more slowly and louder
12	Julie	ok BODYbuilding ((facing the robot))
13	Alex	louder
14	Pat	louder
15	NAO	((arm movement))
16	Julie	((laughing))
17	NAO	*dadup* sound
18	Nao	((shiny eyes))
19	NAO	what was that sport?
20	Alex	((moves the robot closer))
21	Julie	BODYBUILDING ((gazing at Pat))
22	NAO	*dadup* sound
23	NAO	((arm movement))

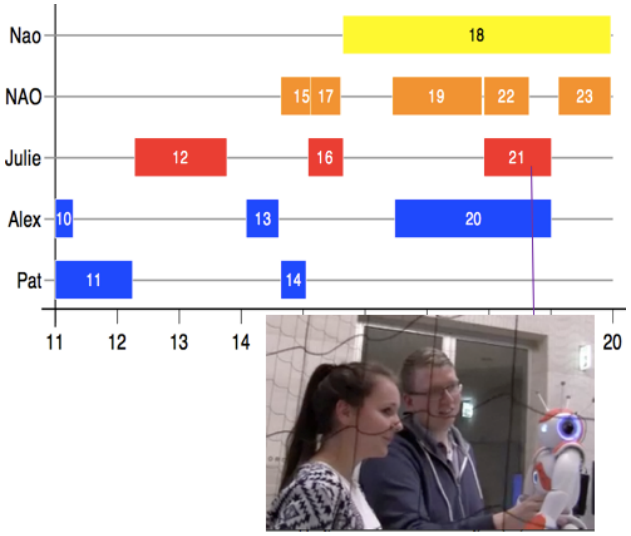

Fig. 5 Multimodal transcription 2

We should point out here that Julie selects (12) *when* she speaks and proceeds without any delay, applying to the rule of minimizing gaps [12]. However, avoiding or minimizing gaps is a pattern commonly mobilized in *human* talk-in-interaction. Here, the robot is not the recipient (since it is still in pausing position) when Julie addresses it (12). This shows that Julie considers the robot's turn-taking system as being similar to that of humans. She does not treat the robot's sound signal as a relevant part of its turn-taking system.

Subsequently, Patrice and Alex repeat their advice for repair (13, 14). Julie replies by laughing (16) which occurs concurrently with the robot's 'waking up' (15, 17). Julie selects laughter as an appropriate response to the researchers' recurrent instructions. According to Holt [19], recipients may treat overdone phrases as laughable. Here, we can assume that Julie treats overdone *instructions* as laughable, thus displaying a certain embarrassment. By now she has performed the correct answer twice without any success.

Meanwhile the robot displays that it leaves its 'pausing

position'. We observe the same 'self'-selecting procedure of the NAO as described above: arm-movement (15), *dadup* sound (17), verbal utterance (question) (19), *dadup* sound announcing imminent recipiency (22), arm-movement (23), and shiny eyes (18). Concurrently to the robot's question and to its *dadup* sound, Alex moves the NAO closer to Julie to support its hearing ability (20) (Fig. 5). By his gesture, the researcher makes accountable that he is understanding the lack of feedback from the robot as related to its (non)-ability of capturing verbal utterances when positioned at a remote distance from the speaker.

The NAO gives a new input "what was that sport" (19) followed by *dadup* (22). Julie provides her answer (with increased loudness) to the robot's question without any delay (21). Her talking is overlapping with the robot's sound signal (22). Due to the overlap, the robot does not capture Julie's turn and cannot treat it as the correct answer. Furthermore, Alex's manipulation of the NAO also affords an immediate response from Julie. Multimodally uttered synchronicity is of major importance here. Alex brings the robot closer to Julie (Fig. 5) holding it like a 'doll' in his right hand (20). Simultaneously, the NAO poses its question (19). Julie's answer is thus a doubly oriented next action responding to the robot's input as well as to the researcher's mobilization of the robot. We note also that at this time Alex does not give the 'usual' gestural advice (to wait or to go). We will see in the following that Julie has to repeat her answer three more times to be finally successful.

C.Analysis 3: "Bodybuilding": Success after Two Other Failures

In the preceding parts of our analysis, we showed how, in the studied multi-party co-constructed IRF sequences, the NAO's feedback remains overdue although the preferred and correct response is uttered three times by a co-participant (Julie). Let us recall here that, after the robot has put forward the question for the first time, Julie gives the right answer at the right moment (8), i.e. at an appropriate transition relevance place that follows the researcher's gestural instructions. However, there is an acoustic problem. Then, Julie provides another utterance (12) (during the NAO's 'rest' position) which the robot cannot 'perceive'. Since from the robot's programming (technical) perspective, it has not 'yet' received any answer, it performs the input (the question) a second time. Here Julie provides her 'third' response (21) in an overlap with the robot's turn, which results in another failure.

Julie will provide three more answers, two of which are given when the robot is in 'rest' position. Finally, Julie's last trial (see transcription 3) will be successful.

Transcription 3

33	Nao	((shiny eyes))
34	NAO	did you guess?
35	Julie	((laughing))
36	Alex	((gesture 'wait'))
37	NAO	*dadup* sound
38	Julie	((clearing the throat))
39	Nao	((shiny eyes))
40	NAO	*dadup* sound
41	NAO	bodybuilding
42	NAO	((raising its arms, *applauding* sound))

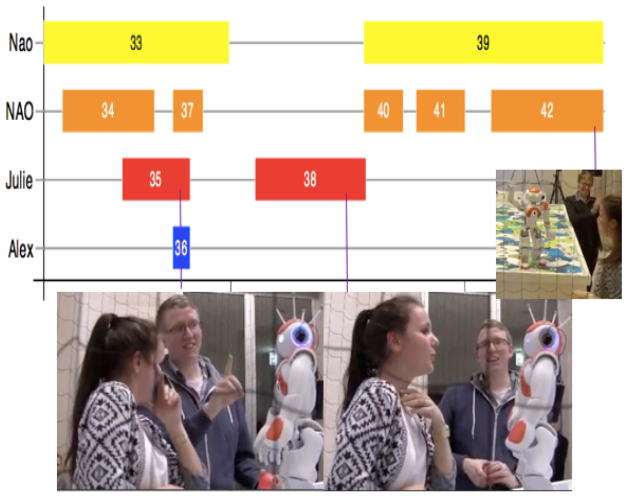

Fig. 6 Multimodal transcription 3

The NAO re-starts (33). Since the robot has not captured any of the five previously uttered responses, it provides another input "did you guess?" (34) followed by a *dadup* sound (37). Julie, at first, replies by laughing (35), and lowering her head and hiding her mouth with her left index finger (Fig. 6). Thus, she displays understanding the robot's re-formulation of the question as an account of impatience or even of rebuke. The change of word selection in the input signals that the speaker asks for a prompt response, it is the robot's 'final call' for Julie. Note here that the robot's question, a polar interrogative, would elicit either a "yes" or a "no" in human communication.

Alex addresses a gestural advice (36) to Julie to 'go' for an answer. Julie follows the instruction: she raises her head, clears her throat after the *dadup* sound while gazing at the robot (Fig. 6) and clearly articulates "bodybuilding" (38). The robot immediately signals going to a next action by emitting a *dadup* sound (40) and displaying shining eyes (39). Then, the NAO utters "bodybuilding" with a falling intonation: it is closing the IRF sequence by repeating the right response (41). After its verbal feedback, the robot raises its right arm and produces an *applauding* sound (Fig. 6) (42).

VI. Conclusion

Through a single case study, it was demonstrated how a CA based approach can contribute to generate insights on how to investigate HRI. The researchers suggest that CA is well suited for studying human robot communication in its sequential unfolding [16]. Relying on CA, it was possible to point out, in a fine grained analysis, how sequentiality and orderliness are instantiated in 'real time' multimodally embodied HRI.

Especially humanoid robots are expected to act as recipients

to co-participants, which means to listen to them "while building the internal representations required for engaging in an effective dialogue within the context of a given interaction" [20]. However, "this exciting vision (…) is far ahead of what has been realized" [20]. Indeed, in the analyzed episode we detected communication breakdowns in the sense that a participant had to provide the right (known) answer six times before finally receiving positive feedback from the NAO robot. Besides some technical issues, the focused participant (Julie) encounters difficulties with the robot's turn-taking system. The NAO's programmed turn-taking reveals to be quite different from the one of the human participants. We pointed out that especially the robot's *dadup* sound leads to confusion. The sound is used both to announce the NAO's imminent verbal turn and as a turn completion signal. For technical reasons, the robot is only able to capture the participant's response after its turn completion, i.e. after the *dadup* sound.

With regard to the Bakhtinian concept of dialogue, we can assert that the NAO robot and the human participants here engage in a 'dis-balanced' communication. Currently, the NAO is able to capture verbal utterances (with certain above mentioned restrictions); however, the robot is neither able to anticipate the 'other's' next action nor to cope with the 'other's' anticipatory responsive understanding. "Bodybuilding" could be considered, at least on the surface, as the right shared answer, but not as the robot's and the human participants' 'shared territory' [9]. The word may well be part of the NAO's linguistic repertoire and can be invoked according to a programmed algorithm. However, the robot can neither seize nor contribute to the inherent social nature of the word dialogically realized in its phenomenological depth.

Note on Transcription Conventions

Talk was described according to conventions commonly used in Conversation Analysis.

BODYBUILDING salient talk
((arm movement)) non verbal utterance

Acknowledgment

We would like to acknowledge with appreciation the support provided by Alexander Eyjolfsson, Pierre Fixmer, Germain Kerschen and Holger Voos.

References

[1] M. Goodrich & A. Schultz, "Human-Robot Interaction: A Survey", in *Foundation and Trends in Human-Computer Interaction*, vol. 1, no 3, 2007, pp. 203-275.
[2] J. Markowitz (ed.), *Robots that Talk and Listen*. Berlin: De Gruyter, 2015.
[3] H. R. M. Pelikan & M. Broth, "Why that Nao? How Humans Adapt to a Conventional Humanoid Robot in Taking Turns-at-Talk", in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2016, pp. 4921-4932.
[4] B. Arend, P. Sunnen, P. Fixmer & M Sujbert, "Perspectives do matter – Expanding Multimodal Interaction Analysis with Joint Screen", in *Classroom Discourse, Special Issue Multimodality*, vol. 5, no 1, 2014, pp. 38-50.
[5] B. Arend, P. Sunnen, "Dialogic Classroom Talk - Rethinking 'Messy' Classroom Interaction", in *Proceedings of the 10th EAPRIL (European Association for Practitioner Research on Improving Learning in Education and Professional Practice) conference*, 2016, pp. 424-434.
[6] P. Linell, "Dialogism and the Distributed Language Approach: a rejoinder to Steffensen", in *Language Sciences*, vol. 50, 2015, pp. 120-126.
[7] M.M. Bakhtin, *Speech Genres and Other Late Essays*. Austin: University of Texas Press, 1986.
[8] P. Linell, *Rethinking Language, Mind, and World Dialogically*. Charlotte, NC: Information Age Publishing, 2009.
[9] V.N. Volosinov, V.N., *Marxism and the Philosophy of Language*. Cambridge: Harvard University Press, 1973, p. 86.
[10] P. ten Have, "Methodological Issues in Conversation Analysis", 1990. http://www.paultenhave.nl/mica.htm Accessed on 19/04/2016.
[11] P. ten Have, *Doing Conversation Analysis*. Los Angeles: Sage Publications, 2007.
[12] H. Sacks, E.A. Schegloff, & G. Jefferson, "A simplest systematics for the organisation of turn taking in conversation", in *Language*, vol. 50, no 4, 1974, pp. 596-735.
[13] A. Deppermann, H. Blühdorn, "Negation als Verfahren des Adressatenzuschnitts: Verstehenssteuerung durch Interpretationsrestriktionen", in *Deutsche Sprache: Zeitschrift für Theorie, Praxis, Dokumentation, Themenheft: Interaktionale Linguistik des Verstehens*, vol. 13, no 1, 2013, pp. 6-30.
[14] C. Heath, "Talk and recipiency: sequential organization in speech and body movement", in J. Maxwell Atkinson & J. Heritage (eds.), *Structures of Social Action. Studies in Conversation Analysis,* Cambridge: Cambridge University Press, 1984, p. 250.
[15] E. Schegloff, *Sequence Organization in Interaction. A primer in Conversation Analysis.* Cambridge: Cambridge University Press, 2007.
[16] K. Pitsch, R. Gehle & S. Wrede, "Addressing Multiple Participants: A Museum Robot's Gaze Shapes Visitor Participation", in *Proceedings of ICSR 2013*, 2013, pp. 587-588.
[17] B. Arend, "Contribution d'un outil de transcription dans la mise en mots de l'analyse d'un processus de conception collaborative observé dans sa profondeur phénoménale", in*Recherches Qualitatives*, *Logiciels Pour l'analyse Qualitative: Innovations Techniques et Sociales*, vol. 9, 2010, pp.95-108.
[18] P. Caire et al., COROBOTS, exhibition brochure, 9. June 2015 – 16. January 2016. MUDAM Luxembourg.
[19] E. Holt, "On the nature of 'laughables': laughter as a response to overdone figurative phrases", in *Pragmatics*, vol. 21, no 3, 2011, pp. 393-410.
[20] B. Mutlu, S. Andrist & A. Sauppé, "Enabling human-robot dialogue", in J. Markowitz (ed.), *Robots that Talk and Listen.* Berlin: De Gruyter, 2015, pp. 81-124.