Coping with turn-taking: investigating breakdowns in human-robot interaction from a Conversation Analysis (CA) perspective

Béatrice Arend
Joint Activity Team, University of Luxembourg
Esch-Belval, Luxembourg

and

Patrick Sunnen
Joint Activity Team, University of Luxembourg
Esch-Belval, Luxembourg

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, 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.

Keywords: Human-robot interaction, Social Sciences, conversation analysis, museum, breakdown.

1. INTRODUCTION

This paper seeks to make a contribution to the field of human-robot interaction (HRI) from the perspective of Social Sciences. Our study provides a video based analysis on how human participants engage in social interaction with NAO robots in a modern art museum. The video footage used for this paper was gathered at the Museum of Modern Art in Luxembourg (MUDAM) in the context of developing interdisciplinary work and sharing joint reflections with Robotics scientists. We recorded visitors attending the workshop 'CoRobots' [1] while they were interacting with NAO robots. When attending to the extracts where visitors and a NAO robot are playing a sports guessing game, we noticed multiple instances of breakdown in human-robot communication. Participants visibly encounter difficulties with turn-taking. Thus, in order to shed light on trouble impeding the conversational flow, we did a fine grained analysis of relevant moments in the unfolding HRI.

In our case study, we rely on a conversation analytic (CA) approach [2], [3], [4], [5] to investigate through our case study how breakdowns may occur (and be overcome) in human-robot communication. It is only recently that Conversation Analysis is applied as a method for investigating human robot communication, mainly in the context of joint research work of multidisciplinary teams in Robotics and Social Sciences [6], [7], [8].

In a not too distant future, human-shaped social robots are supposed to collaborate with humans in diverse contexts. There are indeed manifold scenarios envisioning robots as learning assistants or caregivers [9]. When offering functional and personalized assistance in dynamic environments, robots will have to engage in dialogic real-time communication with humans and thus need to be capable of 'equally' participating in smooth turn-taking interactions [10]. Indeed, "if we want robots to engage effectively with humans on a daily basis in service applications or in collaborative work scenarios, then it will become increasingly important for them to achieve the type of interaction fluency that comes naturally between humans" [10].

Robots are expected to address (and to 'listen' to) their communication 'partners' while building on 'internal' representations required for engaging in a situated and contextualized process of responsive understanding. Many robotics scientists agree that HRI "share one critical component: the need for effective communication" [9]. But 'humans and robots achieving reciprocally other-oriented smooth interactions' seems to be an exciting vision "far ahead of what has been realized" [11]. The success appears to be 'moderate' in the sense that achieving fluid interactions is sought from the human 'user' who is required to understand the actions of the robot and to adapt to its turn-taking system.

To summarize thus far, many researchers in Robotics do point out that social and interactive 'capacities' have become increasingly important for robots which interact and collaborate with humans. Especially humanoid robots are expected to act as recipients to co-participants. Thus, turn-taking is a relevant matter and poses big challenges regarding running human-robot communication.

2. THEORETICAL AND METHODOLOGICAL ISSUES

Turn-taking and 'other-orientation'

In the following, we will raise the issue of turn-taking from the perspective of Social Sciences and emphasize key features of human communication that are relevant for the purpose of our study. Adopting a CA approach, we assert that *human* dialogue builds on the coordinated temporal and sequential unfolding of participants' turn-taking [2]. Hence, CA is an appropriate approach to investigate turn-taking, i.e. the (above mentioned) complex interactional phenomenon of 'mutual understanding', in *human-robot* communication. We shall demonstrate in our analysis that CA can provide a substantial contribution to investigating breakdowns in HRI.

Conversation analysis studies the methods human participants orient to when they organize social action through talk. In other words, CA is concerned with how people achieve courses of action in and through talk and how they make their respective understanding of the actions accountable to each other. Thus, conversation analytic research states that humans always adjust their actions to a specific recipient. Sacks et al. (1974) 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". That means that by building on assumptions about the interactional partner's knowledge and expectancies, participants adjust their turns to the recipient, thus constituting a continuously modified 'partner model' [12] . In human communication, a display of 'recipiency' elicits an action, a turn at talk; participants show to each other that they are 'ready to listen' or that they will go to talk. Heath [13] emphasizes that participants by displaying 'availability' construct a "pre-initiating activity providing an environment for the occurrence of a range of actions".

When we refer to 'other-orientation' in communication while focusing on fluent HRI, certain questions arise. (To what extent) are *robots* able to make continuously adapted assumptions about their partners' expectancies and provide partner oriented responsive accounts of availability and recipiency? Thus, in the light of CA giving access to the complex reciprocally shaped process of mutual orientation in talk-in-interaction, we can get an idea of the challenges (to face) regarding effective human-robot communication.

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 in socially organized joint activities [14], [15]. The participants' gaze, gesture, body movement and other modes of embodied conduct co-occurring with speech, jointly contribute to set up the participation framework and by that way the sense-making process.

"There is order at all points" (Sacks, 1984, 22)

Furthermore, building on its basic position that *there is order at all points* [2] in human communication, CA studies the participants' sense-making devices in their produced orderliness [16], [17]. CA offers insights into the 'interactional machinery' of the turn-taking system and shows when and how participants take the floor. In the following, we will point out some key features (basic rules) of turn taking which are relevant for the purpose of our study.

Taking the orderliness and sequential nature of human communication for granted, a conversation analytic research approach is suggested to be especially appropriate for investigating HRI in terms of turn-taking. According to CA, communication is sequentially organized. Sequences are ordered series of turns through which participants accomplish and coordinate an interactional activity. 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) [3]. 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. (1974) 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.

We further emphasize the following 'basic rules' of turn taking: Only one person talks at 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.

Occurring breakdowns in triadic IRF sequences

The below analyzed HRI involves human participants and a NAO robot playing a sports guessing game which is designed according to the triadic IRF sequence pattern. This structure of talk was first described by Sinclair and Coulthard (1975) and is strongly associated with instructional contexts. It is considered as the most common feature of teacher-student interaction found in the classroom. IRF sequences are generally known in educational contexts as (teacher) initiation, (learner) response, and (teacher) follow-up or feedback (IRF) [18]; as initiation, response and evaluation (IRE) [19]; and as question, answer, comment Q-A-C [20]. The questions asked are mainly 'known information questions'. That means that "the questioner already has the answer, or at least has the parameters in which a reply can properly fall. The questioner is testing the knowledge of the respondent. The respondent to a known information question is placed in the position of trying to match the questioner's predetermined knowledge, or at least fall within the previously established parameters" [19].

In our analysis, we will point out how during the sports guessing game the human participant, beyond trying to provide the 'right' predetermined answer, has to 'match' the questioner's turn-taking system, or at least fall within its previously established rules.

Data construction

To study the human robot communication in its sequential organization as an emergent and interactional phenomenon, we rely on video data which give access to a situated view of social conduct [21]. 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 roving cameras operated by researchers. The six resulting video data streams were connected within one space to generate an 'expanded-around view' of the ongoing event; elsewhere, we termed this apparatus 'joint screen' [22] (Figure 1). For reasons of convenience and to ensure that the chosen frame grabs are not too small to recognize relevant details, we also chose here to rely on images from single camera perspectives to support our analysis.

Figure 1

In order to address our fine grained CA driven analysis we represent the videotaped interactions in a multimodal transcription format. Regarding the transcription, we consider the human participants' verbal conduct and their gestures as well as the verbal utterances of the robot and its 'dadup' sound. The transcription tool 'TranScripter' [23] 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). Furthermore, some relevant screenshots displaying the emerging visible conduct of the participants are connected by a line to the related bars of the partition. The scale of the time line is adapted to the page set up.

3. CONTEXT

The COROBOTS-installation was a part of the visual arts and technology exhibition "Eppur si muove" (9 July 2015– 17 January 2016) at the Museum of Modern Art in Luxembourg" (MUDAM, 2015). Researchers[1] 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[2].

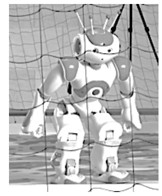
4. ANALYSIS

In the subsequent analysis we will shed light on a 'troublesome' communication episode (of 16 seconds) between a visitor to the museum and a NAO robot. Besides the NAO robot and the visitor, it involves the researcher Patrice, who is part of the setting, she manages the COROBOTS installation and ensures a well-run event.

In terms of turn-taking, we will study how an IRF sequence, which could be considered as 'simply' structured, turns into a recurring question-answer 'game' with pending feedback and multiple repetitions. Despite the instructional work of the researcher Patrice (P) (aiming at a smooth dialogue), we will witness a rather confused visitor appreciating the researcher's support. Furthermore, we will point out how trouble relevant perturbations occur and imped fluid HRI and how breakdowns are overcome.

Analysis (Part I): Bridging 'discrepancy' in 'partner model' knowledge

When we join the scene, the robot has 'performed' a sport ('ski') (Figure 2) and is in 'stand still position' (for 3 seconds) (Figure 3).



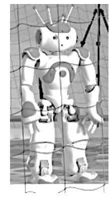


Figure 2 Figure 3 Figure 4

Meanwhile, a visitor of the museum (V) has volunteered to play the sports guessing game and checks with the researcher (P) that she has correctly guessed the sport 'ski' (Figure 4). P validates the suggested answer and after this 'rehearsal', the visitor is reassured to share 'predetermined knowledge' with the questioner(s) (the robot, respectively the Robotics scientists). Thus, when the robot proceeds to provide the verbal utterance 'what was that sport' (i.e., the first turn of the 'inbuilt' IRF pattern), we could expect a smooth and fluent IRF sequence realization. But, matters will be rather different.

Usually, in three-part IRF sequences, only one questioner and one respondent are involved in talk-in-interaction at any one time. In the analyzed excerpt however (see transcription below), a third participant's (the researcher's) utterances as informed by prior expert knowledge about and experience with the turn-taking system of the robot contribute to an expanded multiparty IRF sequence.

Multimodal transcription (Part I)

1	NAO	what [was that sport]
2	P	[gesture *wait*
3	NAO	[*dadup sound*
4	V	it [is ski
5	P	go now
6	V	it is ski
7	P	just one word
8	V	ski
9	P	louder
10	V	SKI ((moving toward the robot))

[1] The installation was designed and implemented by a team of Holger Voos' Automation Research Group (SnT-Interdisciplinary Centre for Security, Reliability and Trust) lead by Patrice Caire.

[2] We analyse further sequences of the sports guessing game from a CA perspective by focussing on the bakhtinian concept of dialogism in an upcoming paper.

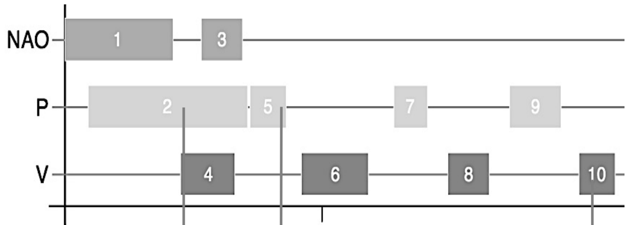


Figure 5a

Figure 5b

Figure 6

Figure 7

The NAO utters ‘what was that sport’ (1), by that way addressing a potential recipient (in this case V) who is supposed to provide an answer. Then, after one second, the robot emits a *dadup* sound (3). We note that the researcher (P) gazes at the visitor and signals her by a gesture of the hand to wait before responding (2), (Figures 5a, 5b). The length of Patrice’s turn, (i.e., her gestural ‘wait’ instruction) extends beyond the verbal utterance of the NAO (1), it includes also the robot’s *dadup* sound (3) following the question. Immediately after the *dadup* sound, by verbal means, P then invites the visitor to perform (5).

In the meantime, the visitor, while gazing at the robot (Figure 5a), selects *when* she answers and proceeds without any delay (4), applying the rule of minimizing gaps. She addresses the robot recipient and does not direct her attention to the researcher’s gesture. Thus, she has already replied to the robot’s input when P gives the instruction ‘go now’ (5). V’s orientation toward Patrice is simultaneous with the researcher’s verbal advice (Figure 6).

However, it is the *dadup* following the question which displays the robot’s recipiency, thereby eliciting the respondent’s answer. From a programming perspective, the NAO is only ‘listening’ for the speaker’s turn after the *dadup* sound. 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. This limited ability to ‘listen’ is a crucial difference to human interaction. Human participants can do both at any point in the interaction, monitoring their conversational partners’ utterances *and* respond to all actions.

Furthermore, the acoustic characteristics of the place 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 answers. We should also mention, by the way, that the *dadup* sound may be compared to a feature of two-way radio communication in which the actual speaker has to release the push-to-talk button (commonly after having uttered a procedure word such as ‘over’) before being able to listen to the other radio operator.

With regard to the above reported multimodally occurring multiparty communication (turns 1 to 5) considered in its temporal (synchronous and continuous) unfolding, we can point out that Patrice makes accountable through her gesture (2) and her verbal conduct (5) that she is orienting to the *dadup* sound as relevant for the NAO-human interaction. Relying on (prior) expert knowledge, P treats the robot’s sound signal as a relevant part of its turn-taking system. We can consider her utterances (2, 5) as instantiating her ‘double’-orientation to two different partner models. In terms of transition relevance place, P shows that she anticipates that the visitor might not be familiar with the robot’s turn-taking system and would probably apply the rule of proceeding to the next turn with very little gap. At the same time, the researcher displays her assumption about a probable ‘discrepancy’ in knowledge regarding the NAO’s interactional features between herself as an experienced Robotics scientist and the visitor to the museum. There is shared agreement (knowledge) about the ‘correct’ word to utter, but the researcher’s surplus of locally mobilized procedural knowledge allows her to mediate the flow of the interactional sequence (which will not be led only by the NAO and the visitor).

Indeed, V appears to be not familiar with the robot’s turn-taking system. Her answer (4) is overlapping with the *dadup* sound (3). The robot’s feedback is pending at that moment as, due to the overlap, it did not ‘hear’ the visitor’s early answer. From a CA perspective however (investigating *human* communication), the natural next action in terms of IRF sequence would be the feedback of the robot; and the robot’s feedback is an expected part of the sports guessing game.

Henceforth, V will carefully follow P’ instructions, by that way acknowledging the scientist’s expertise. V’s further replies will be double-oriented next actions: each new trial will be both, a response to the NAO’s input and simultaneously the re-voicing of Patrice’s actual instruction. Thus, V repeats the ‘correct’ answer ‘it is ski’ (6) immediately after P’s verbal advice (5).

Since the robot’s feedback remains overdue although the preferred and correct response (6) has been uttered *after* the *dadup* sound, the researcher gives another verbal instruction to V ‘just one word’ (7). Here, by making accountable that the turn construction can be relevant, P displays again ‘insider’ knowledge. The NAO is indeed programmed to recognize ‘single’ words. The visitor performs the repair ‘ski’ (8) according to P’s previous instruction. However, there seems to be some acoustic problem: the researcher invites V to speak louder (9). V provides her answer with increased loudness without any delay. (10).

We should note that the researcher, through giving instructions, and the visitor, through instantiating these instructions in ‘new’ repaired responses, are jointly bridging the gap caused by the NAO’s pending feedback. Both human participants display that they aim at accomplishing the IRF sequence. Patrice relies on expert knowledge as a resource for re-adjusting the visitor’s ‘partner model’ (of the robot recipient). The visitor’s re-formulations are responsive instances of acknowledging and understanding the expert’s advices as relevant and competent ‘in the matter’ of HRI. To some extent, by addressing the robot through jointly constructed/instructed re-formulation, P and V are also bridging the knowledge gap concerning the NAO’s turn-taking system. They jointly proceed to match the questioner’s predetermined knowledge, i.e. to fall within the previously established turn-taking parameters [19].

Nevertheless, at that point, the lack of the robot’s feedback becomes disturbing all the more as V has provided the ‘correct’ word three times at an appropriate transition relevance place (6, 8, 10) by closely following P’ instructions for repair.

Analysis (Part II): Last trial and match

After V has responded 'SKI' with increased loudness (10) while leaning forward to the robot, the NAO emits a *dadup* sound (11). Since from the robot's programming (technical) perspective, it has not 'yet' received any answer (overlap, noise), it performs a new input, 'what was that sport' (12). Visibly confused, V turns toward the researcher gazing at her, seeking advice (Figure 8). P puts her left hand onto the visitor's shoulder making accountable that V has to wait before replying once again. Simultaneously with the robot's 'next' *dadup* sound (13) displaying its recipiency, P removes her left hand from V's shoulder, anticipatorily suggesting the visitor's immediate next action, but still giving gestural advice to wait by raising her right hand (Figure, 9). V is already gazing at the robot recipient. Then, subsequently to the *dadup*, through synchronously provided verbal and gestural utterances, the researcher gives the instruction to answer 'at that time' (14) (Figure 10).

Multimodal transcription (Part II)

11	NAO	*dadup sound*
12	NAO	what was that sport
13	NAO	*dadup sound*
14	P	now ((gesture *go*))
15	V	ski
16	NAO	*dadup sound* that is correct

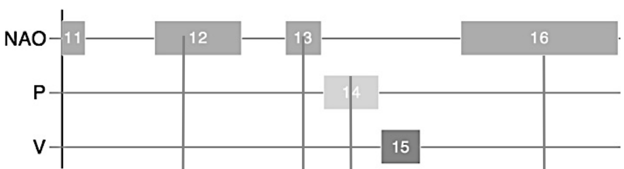


Figure 8

Figure 9

Figure 10

Figure 11a

Figure 11b

V utters 'ski', thus instantiating P's instruction without any delay (15). The robot immediately signals going to a next action by emitting a *dadup* sound (16). Then it utters 'that is correct' with a falling intonation: it is closing the IRF sequence by giving a positive feedback (16). After the verbal feedback, the robot produces an *applauding* sound.

Finally, the visitor and the researcher make mutually accountable to each other (Figures 11a, 11b) that they are pleased to have successfully achieved the interactional issue.

5. CONCLUSION

Our single case study demonstrated how a CA based approach can contribute to generate insights on how to investigate human-robot interaction in terms of turn-taking.

As we already pointed out, especially humanoid robots are expected to act as recipients to co-participants, that means to listen to them "while building the internal representations required for engaging in an effective dialogue within the context of a given interaction" [11]. However, this exciting vision seems far ahead of what has been realized [11].

Indeed, in our analysis, we detected communication breakdowns in the sense that the robot's feedback was pending. The focused participant encounters difficulties with the robot's turn-taking system. Besides knowing the correct answer, the human participant has to cope with the NAO's programming 'technical' characteristics. Above all, the robot's recipiency leads to confusion. A *dadup* sound is used to announce the robot's turn and as 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.

It is actually through to the researcher's locally managed mediation as well as through the visitor's responsive understanding that the interactional flow can go on. We can show that the Robotics scientist's instructions are "constructed or designed in ways which display an orientation and sensitivity to the particular other(s) who are the co-participants" [2]. While mobilizing expert knowledge and simultaneously orienting to the sequential nature of everyday talk-in-interaction, Patrice displays sensitivity to both, the human participant and the robot.

Achieving fluid interactions is sought from the human participants who are required to understand the actions of the robot and to adapt to its turn-taking system. In our study, we show 'a best case' where experts in Robotics and novices jointly achieve the challenging HRI.

6. NOTE ON TRANSCRIPTION CONVENTIONS

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

SKI salient talk

((gesture *go*)) non verbal utterance

[overlapping

7. REFERENCES

[1] P. Caire et al., COROBOTS, exhibition brochure, 9. June 2015 – 16. January 2016. MUDAM Luxembourg.

[2] H. Sacks, E.A. Schegloff, & G. Jefferson, "A simplest systematics for the organisation of turn taking in conversation", **Language**, Vol. 50, No 4, 1974, pp. 596-735.

[3] E. Schegloff, **Sequence Organization in Interaction. A primer in Conversation Analysis**, Cambridge: Cambridge University Press, 2007.

[4] G. Psathas, **Conversation Analysis: The Study of Talk-in-Interaction**, Thousand Oaks, London, New Dehli: Sage Publications, 1995.

[5] I. Hutchby & R. Wooffitt, **Conversation Analysis**, Cambridge: Polity Press, 2008.

[6] H. R. M. Pelikan & M. Broth, "Why that Nao? How Humans Adapt to a Conventional Humanoid Robot in Taking Turns-at-Talk", **Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems**, 2016, pp. 4921-4932.

[7] K. Pitsch, R. Gehle & S. Wrede, "Addressing Multiple Participants: A Museum Robot's Gaze Shapes Visitor Participation", **Proceedings of ICSR 2013**, 2013, pp. 587-588.

[8] A. Nishizaka, "Exploring Interaction between Humans and Robots with Conversation Analysis", **National Institute of Informatics News**, No, 48, (ISSN 1883-1974), 2014.

[9] N. Mirnig & M. Tscheligi, “Comprehension, coherence and consistency: Essentials of Robot Feedback”, J. Markowitz (ed.), **Robots that Talk and Listen**, Berlin: De Gruyter, 2015, pp. 81-124.
[10] Ch. Chao, “Timing multimodal turn-taking for human-robot cooperation”, **Proceedings of the 14th ACM international conference on Multimodal interaction**, 2012, pp. 309-312.
[11] B. Mutlu, S. Andrist & A. Sauppé, “Enabling human-robot dialogue”, J. Markowitz (ed.), **Robots that Talk and Listen**. Berlin: De Gruyter, 2015, pp. 81-124.
[12] A. Deppermann, H. Blühdorn, “Negation als Verfahren des Adressatenzuschnitts: Verstehenssteuerung durch Interpretationsrestriktionen”, **Deutsche Sprache: Zeitschrift für Theorie, Praxis, Dokumentation, Themenheft: Interaktionale Linguistik des Verstehens**, Vol. 13, No 1, 2013, pp. 6-30.
[13] 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.
[14] J. Streeck, C. Goodwin, C. LeBaron, C. (eds.), **Embodied Interaction, Language and Body in the Material World**, Cambridge, Cambridge University Press, 2011.
[15] P. Linell, **Rethinking Language, Mind, and World Dialogically**. Charlotte, NC: Information Age Publishing, 2009.
[16] J. Heritage, **Conversation Analysis**, Cambridge: Polity Press, pp. 233-292.
[17] P. ten Have, **Doing Conversation Analysis**, Los Angeles: Sage Publications, 2007.
[18] J. M. Sinclair & M. Coulthard, **Towards an analysis of discourse: The English used by teachers and pupils**, Oxford: University Press, 1975.
[19] H. Mehan, **Learning Lessons**, Cambridge: Harvard University Press, 1979.
[20] Nicholls, J., “Exchange Structure in the ESL Classroom: Q-A-C and Q-CQ-A-C Sequences in Small Group Interaction”, **Pragmatics and Language Learning**, Monograph Series No, 4, 1993, pp. 183-191
[21] L. Mondada, “Video Recording as the Reflexive Preservation and Configuration of Phenomenal Features for Analysis”, H. Knoblauch, B. Schnettler, J. Raab, H. G. Soeffner, **Video Analysis: Methodology and Methods**, Peter Lang, Frankfurt/Main, 2009 pp. 51-67.
[22] B. Arend, P. Sunnen, P. Fixmer, M. Sujbert, “Perspectives do matter – Expanding Multimodal Interaction Analysis with Joint Screen”, **Classroom Discourse**, Special Issue Multimodality, Vol. 5, no. 1, 2014, pp. 38-50.
[23] B. Arend, “Contribution d’un outil de transcription dans la mise en mots de l’analyse d’un processus de conception collaborative” **Recherches Qualitatives**. Hors série, No 9, 2010, pp. 95-108.