"Talk to You Later": Doing Social Robotics with Conversation Analysis. Towards the Development of an Automatic System for the Prediction of Disengagement

Nicolas Rollet

Télécom Paris, Institut polytechnique de Paris, I3 CNRS

Chloé Clavel

Télécom Paris, Institut polytechnique de Paris, Université Paris-Saclay

Author Note

Nicolas Rollet, Department of Economics and Social Sciences, Télécom Paris; Chloé Clavel, LTCI, Télécom Paris, Université Paris-Saclay.

The authors thank Softbank robotics for their help in the design of the Pepper scenario, Atef Ben Youssef for his previous work on engagement prediction and data collection, and Christian Licoppe for valuable insights and suggestions. This work has been supported by the Maison des Sciences de l'Homme (MSH), the Centre National de la Recherche Scientifique (CNRS), the European project H2020 ANIMATAS (ITN 7659552) and a grant overseen by the French National Research Agency (ANR-17-MAOI).

Correspondence concerning this article should be addressed to Nicolas Rollet and Chloé Clavel, Télécom Paris, 19, place Marguerite Perey F-91120 Palaiseau, France.

E-mail: nicolas.rollet@telecom-paris.fr; chloe.clavel@telecom-paris.fr

Abstract

This article presents an applied discussion of the possibility of integrating conversation analysis (CA) methodology into that of machine learning. The aim is to improve the detection of that which resembles disengagement in the interaction between a robot and a human. We offer a novel analytical assemblage at the heart of the two disciplines, and namely on the level of the annotation schemes provided by conversation analysis transcription methods. First, we demonstrate that the need for a stable structure in establishing an interaction scenario and in designing robot behaviours does not prevent the emergence of ordinariness or creativity among the participants engaged in this interaction. Secondly, based on an actual case, we emphasize the possibility of systematicness in CA transcription to support the choice (a) of the categories targeted by prediction methods and defined by the annotation scheme, and (b) of the verbal and non-verbal features used to create prediction models.

Keywords: social robotics, engagement, machine learning, multimodal features, annotation schemes, conversation analysis, transcription, closings

"Talk to You Later": Doing Social Robotics with Conversation Analysis. Towards the Development of an Automatic System for the Prediction of Disengagement

In recent human-agent interaction studies, topics such as artificial agent's sociality or engagement in interaction have given rise to a great deal of publications and discussions. (Sidner, Lee, Kidd, & Rich, 2005; Dautenhahn, 2007; Pelachaud & Glas, 2015a; Clavel, Cafaro, Campano, & Pelachaud, 2016; Jones, 2017). With the development of so-called social robotics, one observes a tension between *(i)* building computational models using implementable traits to make an artificial agent sociable or *socially interactive* (Fong, Nourbakhsh, & Dautenhahn, 2003; Breazeal, 2003); and *(ii)* the recognition that this sociality is the product of a local organization, emerging from the interaction (Suchman, 2007; Straub, 2016; Sabanovic, Chang, 2016). Such a recognition suggests an increased focus on the mechanisms that organize an environment in which interactions can take place, and where robot sociality and human engagement can emerge. Such a recognition calls for interdisciplinarity.

As Pélachaud and Glas (2015a: 945) have stated, Sidner and Dzikovska (2002) provide a definition of engagement, that is commonly used in the field of human-agent interaction research, as a collaborative endeavour: "the process by which two (or more) participants establish, maintain and end their perceived connection". For interactionists, engagement does not necessarily involve verbal practices but rather any deployment of orientation from one participant to another (or others) - in particular, the fact that participants take into account the actions of others to produce, adjust their own (Goffman, 1983). Engagement is a form of presence.

From a conversation analysis (CA) perspective, the question of engagement / disengagement is linked to the observable deployment of interactional behaviours at the scale of turns-at-talk and conversational sequences such as closings and pre-closings (Sacks & Schegloff, 1973; Button, 1991), distinct orientations towards turn-taking systems (Jefferson, 1978; Zimmerman, 2006), interruptions (Schegloff, 2002). In addition, verbal behaviours can co-occur with other behaviours: gaze, body orientation, etc. (Goodwin, 1981). Engagement can also refer to the multiple orientations of a person who must respond to external events and reorganize, *in situ*, the practical achievement of multi-activity, for example looking at his smartphone and driving (Licoppe & Figeac, 2014), talking and playing an instrument (Rollet, 2010).

Human-robot (HRI) and Human-agent (HAI) interaction studies using CA as a data exploration and analysis method are recent – about two decades (see for example Cassel et al., 1999)[1]. Nevertheless, the combination of CA and coding practices in affective computing / machine learning is quite uncommon. Although quite new, applications are growing and CA is even recently presented in a new handbook on HRI (Bartnek et al., 2019). Moreover, one can distinguish between studies that use CA more as a thematic basis or conceptual underpinning (Sadazuka, Kuno, Kawashima, & Yamazaki, 2007; Yu et al., 2013; Pelachaud & Glas, 2015b) and those that, in addition to this thematic use, conduct a meticulous analysis of the interactions themselves, for themselves, and therefore often contain detailed transcriptions in their publications (Pitsch et al., 2009; Dickerson, Robins, & Dautenhahn, 2013; Pelikan & Broth, 2016;

[1] There are earlier discussions in the Computer Science community, for example Chapman, D. (1992), that claims that "an interactionist computational interpretation of the conversation analytical rules is possible".

Rollet, Jain, Licoppe, & Devillers, 2017; Porcheron, Fischer, Reeves, & Sharples, 2018).

We pursue the idea that interdisciplinarity in social robotics offers novel and ambitious opportunities for design (Fong, Nourbakhsh, & Dautenhahn, 2003; Bartnek et al., 2019), especially if it focuses on objects at the heart of the respective disciplines. In this sense, we can discuss this interdisciplinarity by considering the nature and purpose of collaboration between a so-called interactionist sociological approach and a computational model of human-robot interaction. Such interdisciplinarity raises multiple questions with regard to both the design and detection of behaviours – especially how to address these in terms of segmentable and annotatable flows of interaction. We address some of these questions through the subject of disengagement in a human-robot interaction. Drawing on a study conducted by (Ben-Youssef et al., 2017), which develops a system to predict engagement breakdown in the context of face-to-face interaction with the robot Pepper (Softbank robotics), we present a conversation analysis viewpoint of the methodology adopted.

CA is a sociological approach that addresses language and especially talk-in-interaction as a social organization. Its general topic lies in the description of the details of this organization through which social interaction is made possible in an orderly and intelligible way (Sacks, Schegloff, & Jefferson, 1974; Levinson, 1983; Sacks, 1992). One recognizes here the affiliation to ethnomethodology's perspective which relies on the intelligibility of the methods (defined as 'accountability') and on the participants' point of view to produce its scientific analyses (Garfinkel, 1967). In addition, CA considers an utterance in conversation in its *sequential conditions of emergence*, i.e., as a contribution retrospectively and prospectively referring to the temporal stream of interaction locally managed by participants. In that sense, social action is context-shaped and context-renewing (Heritage, 1991) – property defined as *reflexivity*. The conversational approach provides a methodological and argumentative framework in which social interaction itself constitutes a powerful resource for analyzing and understanding the meaning of being engaged, adapting, collaborating, disengaging, and sharing an experience between co-participants.

The computational models of human-robot interaction discussed in this article are based on "supervised" machine learning (Mohri, Rostamizadeh, & Talwalkar, 2012). The supervision consists of using audiovisual recordings of human-robot interactions that have been annotated into categories of behaviours to predict (here, engagement-related categories), in order to learn the models associated with each of these categories. The methodology is broken down into different stages that structure this article.

The first stage consists of collecting audiovisual recorded data that will be used for learning behavioural models. This requires an interaction scenario to be defined, that will be followed by the robot during its interaction with the participant. In the first section, we present the chosen scenario and correlate it with the notion of context as understood in conversation analysis, in particular with regard to the notion of *relevance*.

The second stage consists in annotating the data into engagement categories that will be used for learning the supervised model. The second section of this article presents a conversational perspective on the methodology for creating affective computing annotation schemes, and shows how conversation analysis can contribute to identifying features relevant to the development of our human behaviour detection system. Specifically, this approach emphasizes the details constituting social behaviours, on the scale of turns-at-talk, sub-units composing turns (Turn Constructional Units), and sequences. Based on an actual case of human-robot interaction, this section discusses the problems of categories and of segmentation, and proposes leads for an interdisciplinary *assemblage*.

Finally, Section 3 offers a summary and an extension simultaneously aiming at short-term applications in the context of a project underway, and lines of reflection that expand the horizon of possible collaboration between interactionist sociology, machine learning, and Affective Computing.

1. Interaction Scenarios in Social Robotics and the Notion of Context in CA

When seeking to develop methods to predict a participant's behaviour in his or her interaction with a machine, it is essential to examine the situation or setting (*e.g.* museum entrance hall (Campano, Clavel, & Pelachaud, 2015) or negociation game (Langlet & Clavel, 2018) in which we want our prediction system to function— both to define the interaction scenario used for data collection and to better understand the data itself. Indeed, the participant's behaviours faced with the robot could depend heavily on features of the situation of interaction.

The notion of context has been highly discussed and re-specified in conversation analysis and sociolinguistics. We give a clarification below in order to better understand the issues related to defining an interaction scenario and its impact on the type of processing foreseen (Paragraph 1.a). We then give a contrastive view of CA and affective computing approach regarding stability and emergence (Paragraph 1.b).

a. Social robotics scenarios and the CA perspective

The main goal of the interaction scenario that we have used, as defined in (Ben-Youssef et al., 2017), is to collect the data on which the model to predict an engagement breakdown will be built. Specifically, the goal is to collect data on a situation where a participant is liable to exit the interaction before the end of the scenario. In this study, the scenario defines the following aspects:

- The place of the interaction: the robot (in this case, Softbank's Pepper) is placed in a hall where people frequently pass by.
- The mechanisms of entering into interaction: the robot starts speaking when it detects the presence of a person, to invite him or her to interact, and the participant is free to enter into the interaction or not.
- The mechanisms of exiting the interaction: the participant is free to exit the interaction whenever he or she so wishes.
- The participant's engagement area: a space delimited by a semicircle with a radius of 1.5 m.
- The phases of the scenario (note that the scenario was intentionally long, thus including multiple phases, in order to trigger engagement breakdowns prior to completion): the *welcome phase* (the robot introduces itself using very lively animations, and gives instructions; the *dialog phase* (set of open questions that the robot asks the participant about his or her tastes and personality); the *cucumber phase* (with self-mockery and in the form of a game, the robot presents its perceptive capabilities to show that, from its viewpoint, the difference between a cucumber and a human is the human's face); and the *final*

phase (the robot concludes the interaction with questions intended to evaluate the participant's interaction).

Now, one fundamental aspect of the *emic* perspective of conversation analysis (that is, in which the orientation of the analysis is based on the participants' viewpoints emerging in the interaction itself) regarding context, is characterized by *relevance*.

By following what participants make relevant themselves in the ongoing interaction, it is possible to provide a characterization of the participants and of the context. Such characterization provided *in situ* by the participants themselves is called *internal setting relevancies* (Schegloff, 1987). In this sense, the selection of cues relevant to the context from the multitude of available contextual elements, corresponds to what is carried out visibly (accountably) by the interactants in the immediacy of the interaction. These contextual aspects "internal to the interaction" heavily weigh on the interaction, whilst they are not exclusive: background elements can also be important, as well as the structure of the place, time, etc.[2]

The scenario consists of shaping a set of robot behaviours (of a finite number, by definition) for interaction with humans. The robot's verbal and non-verbal behaviours are defined in "interaction phases" (welcome phase, cucumber phase, etc.). From this viewpoint, the scenario is designed asymmetrically: it consists of creating robot's behaviours as ingredients of the different phases that make up a whole—the scenario—but in which other ingredients of these phases, and namely human behaviours, constitute hypothetical, fictional participation's opportunities.

[2] For a multi-dimensional definition of context, inspired by CA and linguistic anthropology, see (Duranti & Goodwin, 1992; Duranti, 1997)

With respect to the actions of the designer, this asymmetry consists in pre-allocating turns-at-talk, which will then be experienced by a human participant.

Here are two examples of the same turn's occurrence:

(Extract 1)[3]

01 R comment tu t'appelles ?
what's your name?
02 P oui, Evelyne
yes, Evelyne

(Extract 2)

01 R comment tu t'app@elles /
what's your name?
02 P @looks towards Robot's face
03 (1s)#(1,7s)
04 P #leaves

In the two examples taken from the data, the robot (R) addresses a question about the name (Line 1) to the participant that stands in front of it. Pepper produces what CA calls a first part of an adjacency pair, namely a question-answer sequence (Sacks & Schegloff, 1973; Schegloff, 2007). This is a very basic (ordinary) sequence in which the first pair calls for possible seconds, that is: for the next speaker, the first pair part is a context in which some relevant actions are expected, namely giving something recognizable as an answer (i.e. a second pair part). That is what happens in the first extract, but not in the second: the participant looks at the robot (L2) and after 1 second pause, just leaves (L4). Contrary to what is expected in an interaction between two

[3] Some conventions of transcription are given below in 2.a.

humans, the strong pressure exerted by the first part of the pair on the possible second part, in this situation, is visibly not addressed by the participant through his unilateral disengagement. This particular observation raises an interesting point regarding the categorization of the robot as a machine rather than as a social partner. For the latter case, mitigation marks would generally be produced before leaving (Goffman, 1973; Sacks & Schegloff, 1973).

Indicating that "the participant is free to exit the interaction" in the protocol illustrates the rather logical asymmetry in the process of its assembling involving the design of robot behaviours: the scenario is designed by projecting a hypothetical participant. Whilst, in ordinary social interaction each apparently identical turn occurs in particular circumstances, often as *another first time*. Hence, the naming of phases such as the "welcome phase" or expectations such as "the participant is free to exit the interaction" doesn't give any details on how this is factually, in a particular moment of the interaction, both relevant and experienced[4]. Reversing the reasoning, we acknowledge that the design of robot behaviours is based on the designer's ordinary interactional knowledge: he/she assumes that these hypothetic interactions should start with a welcome phase, and that saying hello will trigger a hello.

In addition, it is not given in advance that a framework externally considered artificial or experimental implies that the participants treat it as such sometimes, or even never. Context influences practices, but practices actualize and coproduce context as well. In ethnomethodology, this refers to *reflexivity*: social practices pre-suppose (context-shape) and constitute (context-renew) the framework of the interaction

[4] Although the interaction strategies defined within the phases of the scenario take into account the participant's responses/reactions, they are still part of a planning process: they do not predict the particular circumstances under which actions will occur (Suchman, 2007).

embedded within them (Heritage, 1991). There is nothing to prevent a particularly constrained framework, such as a scenario based on question-answer games, from seeing the emergence of unexpected, creative, natural behaviours.

b. Emergence, spontaneity, stability

The emerging dimension (i.e. the situated character of actions) is at the heart of the organization of social interactions (human-human). Conversation analysis' analytical approach puts emphasis both on the *accountability* of actions (that which the participants make visible themselves for themselves to coordinate their actions and structure an activity) and on their normative aspect. Moreover, if the human treats the robot like a conversation partner (and not an answering machine), then despite being "scripted", the interaction will nonetheless adopt an emerging quality (fully on the human side, and in the form of a tree diagram on the robot side). The fact that the context is not only a set of imposed external characteristics but also a set of resources to organize the interaction affords the participants the opportunity to demonstrate creativity and spontaneity based on this framework. This question of spontaneity is addressed from another viewpoint below (Section 3) regarding affective computing annotation schemes in contrast with conversation analysis transcription practices. In both cases, the problem of the reification of emerging behaviours is raised.

Now, the problem is that the contrast between the CA approach and affective computing using supervised machine learning lies at the intersection of a problem of stabilizing cues and the fundamentally emerging nature of social interaction. On the one hand, on the CA side, we address the versatility of the context, or to put it another way, the situated character of actions, as a strong resource to analyze the meaning of these same actions. Actions are to be analyzed in their sequential deployment. (Where versatility does not mean that there is no stability: it is punctual, circumstantial.). Few data are processed generally, CA researchers work on singular cases and collections as well but they are in relatively limited number. On the other hand, on the supervised Machine Learning side, there is a technically justifiable need for stability of features and identification of large classes that must be detectable and in limited numbers to optimize the mass annotation work. Thus, from the viewpoint of a system to automatically analyze participant behaviours, the question of spontaneity is addressed as follows: how can defining constraints limiting the interaction steer the automatic participant behaviour analysis system. Large amount of data is to be processed in this approach.

To summarize, we gather the elements of this contrast in Figure 1.

	Conversation Analysis	Supervised Machine Learning for Affective Computing
Actors	sociologists, anthropologists, linguists	linguists, computer scientists, annotators
Goals	analyze the intimacy of interaction, the practical reasonings, the improvised choreography, account for the intelligibility of actions, the sequential conducts	define classes that can be learned by a system, define cues for these classes, detect human behaviour, present a robust system
Steps in scientific production	audio-video recordings ; transcription ; text (analysis)	audio-video recordings ; annotation ; programming (model training)
Data Scale	small corpus (collections)	Big data

Figure 1. Variety of shared and distinct aspects of conversation analysis and affective computing approaches

The underlying idea is that by defining that which is potentially stable (the context-shaped induced by the interaction scenario) and by integrating it into our behaviour prediction models, we can improve the performance of prediction systems.

Our goal is to explore the possibility of improving the acuity of the detectable cues, while guaranteeing some stability and not excluding the fact that this stability can be temporary (on the scale of a turn or an adjacent pair, for example).

2. Affective Computing Annotation of Recordings in Social Robotics vs. CA Transcription

In this section, we present a comparative study of two productions independently obtained with the same recording data from interactions with the robot Pepper: an affective computing annotation to develop behaviour prediction (Figure 2) and a CA transcription (Transcript 1). The goal of this section is to provide an in-depth comparison of what is produced by both approaches. The first paragraph contrasts the categories produced by the affective computing annotations with the principle of emergence in conversation analysis from a methodological point of view. The second paragraph aims to compare productions of each approach, showing: *i)* similar results that have emerged from both affective computing annotation and transcription processes and *ii)* complementary results that show how both processes could benefit from each other. The third paragraph discusses the affective computing segmentation processes from a conversation analysis perspective. We use as a guideline the framework concerning the development of the system to predict participant engagement breakdowns, and illustrate our study with examples of affective computing annotation and CA transcription.

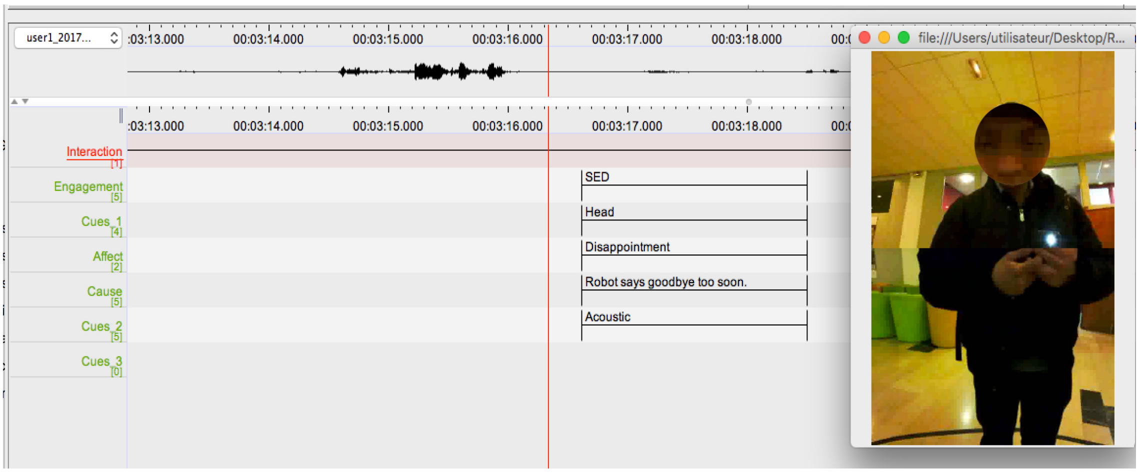

Figure 2. Affective computing annotation of interaction recordings conducted with the Elan software[5] corresponding to lines 1-4 in the transcript 1 below.

Transcript 1 user1_2017-03-03

01 P ablas esp[agnol
hablas español
02 R [une autre fois\\
another time
03 P <oh: ((look at smartphone)) (0,5s) > (0,5s) ok (..)je ne
04 sais pas qué: qu'est-ce qué tou (1,1s) dire\\
oh, ok I don't know what, what do you..say

Transcript 1[6]. In this excerpt from an interaction between the robot Pepper and a human, turns are delimited to the left by a letter for each participant (R for robot and P for human participant), and to the right by the end of the line or several lines below (e.g. the first robot's turn starts in L02, while the first participant's turn starts in L01, then he produces another turn that starts in L03 and ends in L04). Pauses are indicated in seconds. The font used (courier) allows for ideal vertical alignment. Such a layout is

[5] https://tla.mpi.nl/tools/tla-tools/elan
[6] We use an adaptation of ICOR transcription conventions: http://icar.cnrs.fr/projets/corinte/documents/2013_Conv_ICOR_250313.pdf

used to retain the outline of the course of the interaction over time by seeking to reproduce the details of verbal behaviours, overlaps (marked by square brackets [], intonations indicating the end of a turn (going up or down with the signs / or \), but also bodily behaviours (in this case looking at the smartphone, indicated in double parentheses in L03 with the signs "< >" that delineate the co-occurrence of this behaviour with the verbal conduct). Intended to be as neutral as possible, such a transcription is then subject to analysis that can lead to it being refined, for instance by detailing the timing of the participant's orientation towards his smartphone with respect to his turn in L03.

a. Categories vs. emergence principle: methodological comparison

The goal of an affective computing annotation scheme is to define the macro-categories that can be learned by the system. These categories must be sufficiently represented in the data and relatively easy to annotate. The quality of the models learned will depend heavily on the quality and quantity of the annotations obtained. All the difficulty lies in defining an annotation protocol to establish convergence between the annotations of multiple annotators, knowing that socio-emotional behaviours are highly subjective phenomena that are difficult to define in an annotation (Cowie & Cornelius, 2003). The performances of the models learned are also evaluated using these annotations as a reference, acknowledging that it is sometimes difficult to determine what annotation is the most relevant between the automatic annotation of the system and human annotation (Clavel & Callejas, 2016).

In the case of the study of disengagement, we conducted an annotation of the different videos collected, an example of which is given in Figure 2. Four categories were defined upstream to delimit the phenomena to annotate:

- BD (engagement BreakDown): phase when the participant leaves the interaction;
- EBD (Early sign of future engagement BreakDown): the first precursor sign of an engagement breakdown (necessarily results in an engagement breakdown, and is therefore different from a SED and a TD);
- SED (Sign of Engagement Decrease); and
- TD (Temporary disengagement) (TD): phase during which the participant interrupts the interaction before returning (this is a disengagement related to an external interruption, such as a third person).

Modelled on the work in (Clavel, Vasilescu, & Devillers, 2011), a sub-characterization of these macro-categories was also proposed and consisted of the following tasks:

1. Defining the main verbal and non-verbal cues that characterize the annotated phenomena (no sub-segmentation provided): speech, facial expressions, or gestures (see *Cues1 field: Head* in Figure 2).
2. Specifying if emotions were expressed in these annotation segments (no sub-segmentation provided) and identifying them in the following list of negative emotions: frustration, boredom, nervousness, disappointment, anger, submission (see *Affect Field: Disappointment* in Figure 2).
3. Providing an interpretation of the participant's disengagement (see *Cause field: "Robot says goodbye to soon"* in Figure 2).

4. Identifying secondary cues (see *Cues2 field: Acoustic* in Figure 2)

The objective of this subcategorization was to provide cues to understand how the system functions and to interpret the reasons (explanatory cues) for which the system detected the emergence of this category of phenomena. Indeed, when analyzing the performance of the machine learning models of the marco-categories, the subcategories can explain the behaviours of the system. For example, if the errors of the automatic detection of an engagement breakdown are always located in segments where boredom is expressed, it may think that the system missed to model this type of expression of engagement breakdown.

The tool ELAN was used for these affective computing annotations (see Figure 2). It is an annotation tool for multimodal dialogue. This tool allows us to define our own annotation scheme. For example, the segment annotated on Figure 2 is constructed as a so-called "parent" category (SED) followed by associated cues or comments. In this case, there is a bodily cue called Cue 1 ("head"), a non-lexical cue called Cue 2 ("acoustic"), an emotional cue ("disappointment") and, last of all, a Cause comment ("Robot says goodbye too soon"). Note here that the choice of these categories is guided by the task, that is, by what the system must detect. Affective computing annotation can be considered top-down given that the categories are what guide the annotation of explanatory cues—which greatly contrasts with the transcription mentality of conversation analysis.

In conversation analysis transcription is a textual translation of repeated observations from audio-visual data. In this sense, it constitutes a particular configuration of the recorded reality, which is itself simply a particular configuration of overall reality. From a methodological point of view, it is a research support that is not sufficient in itself: the analysis is always carried out, transcription at hand, with repeated visualizations of the corresponding audiovisual data. Transcription is both a necessary and a reifying tool: it leads to decision-making and materializes a graphic layout (Ochs, 1979; Mondada, 2008), as we can see below in Figures 3 and 4, which show such examples of transcription.

To summarize:

- Each type of transcription corresponds to a position and a goal for the researcher or the transcriber.
- Each type of transcription corresponds to a status ascribed to the verbal and non-verbal, and to the relationship maintained between the two.
- Transcription involves theoretical assumptions on the part of the transcriber, which have a configuring effect on this transcription.

Transcription and *a fortiori* affective computing annotation result in de-contextualization, that is, extraction from the singular context of production and transformations (for example, certain phenomena that appear anecdotal in the field become worthy of interest and fixed during transcription / affective computing annotation practices).

Nevertheless, the status of transcriptions in conversation analysis is radically different from that of the computational approach's annotation diagrams—even though both "describe" an undertaking to categorize interactional behaviours, whether by the conversation analysis researcher or multiple affective computing annotators. In affective computing annotation schemes, macro-categories (that can be learned by the system) are associated with cues (which are relatively limited), and the practical reasoning used by annotators is to some extent invisible. By contrast, conversation analysis transcription is an intermediary that attempts to be as neutral as possible between the raw data and the researcher's analysis. This tendency of relative neutrality in the production of transcripts refers to an orientation of the researcher towards the analysis of participants' practical reasoning in the here and now of their social interactions. Note that CA community is familiar with ELAN interface as a mean to visualize and account for multimodal phenomena during collective work processes ('data sessions') and as screenshots for publications (Mondada, 2006, 2008).

Nonetheless, conversation analysis is not fundamentally prevented from seeking out systematicity (Sacks, Schegloff, & Jefferson, 1974; Stivers, 2015). Even though debate exists within the CA community, it is not unreasonable to want to improve or challenge the macro-categories resulting from the description modes of the computational approach. In this approach, using the analysis of micro-phenomena can feed a diversity of cues associated with these macro-categories in annotation schemes.

b. Categories vs. emergence principles: comparison of two types of production

To demonstrate this, let us contrast the affective computing annotation presented in Figure 2 with the transcription of the same excerpt, taken from a common corpus of interactions between the robot Pepper and humans. Note that both processes have been carried out independently. This affective computing annotation is characterized by the annotated segment as a "sign of engagement decrease" (SED). The complete transcription of this segment and of what goes after is presented in Transcript 2.

Transcript 2 user1_2017-03-03

01 P ablas esp[agnol

hablas español

02 R [une autre fois\

another time

03 P <oh: ((look at smartphone)) (0,5s) > (0,5s) ok (..)je ne
04 sais pas qué: qu'est-ce qué tou (1,1s) dire\

oh, ok I don't know what, what do you..say

05 (3,6s)

06 P <((with greeting gesture)) au revoir/>

goodbye

07 (6,2s)

08 P <((with greeting gesture and body torq)) au revoir Pepper\>

goodbye Pepper

Transcript 2. Transcription associated with Figure 3 and 4

In order to visualize the correspondence that could be found between the two productions, we manually integrated the segments of transcriptions corresponding to the annotation environment in Figure 3 and 4. These figures show that in CA, analyzing the excerpt reveals much more detailed information than that annotated, and may provide explanatory indications of signs of disengagement: CA approach could give some hints about how an identified closing (Figure 3) could be analyzed by scrutinizing also what happens right before in the interaction (Figure 4).

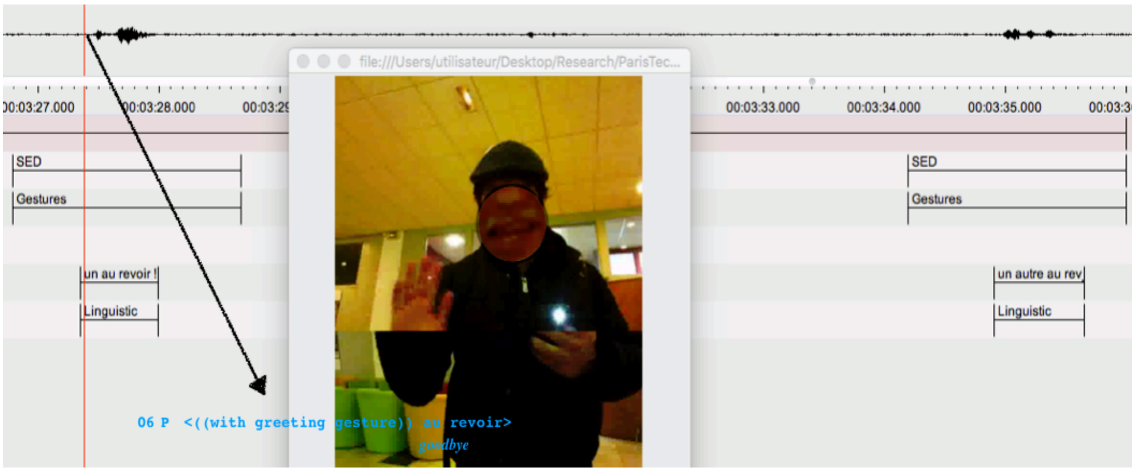

Figure 3. Assemblage of CA transcription – affective computing annotation

Figure 4. Assemblage of CA transcription – affective computing annotation (continued)

A first example of interesting details provided by CA transcription is given in Figure 3. Within the segment annotated signs of engagement decreased (SED), the affective computing annotation indicates globally that it is characterized by linguistic ("un au revoir") and gestural cues, while the CA transcription details the type and the exact timing of gesture that accompanied the "au revoir".

A second example is given in Figure 4. The affective computing annotation indicates a decrease in engagement by using the annotation of the SED category, and specifies that it consists of acoustic cues and cues related to head movements (Head). Here, the CA transcription reveals different cues with their interpretation allowing one to anticipate the engagement breakdown that are not given by the affective computing annotation: *i)* in (L01-02), the CA transcription allows us to identify a potential cause of participant's engagement decrease that occurs before the SED affective computing segment: an overlap between participant turns (the robot (R) produces a turn that causes an overlap (L01-02)) and a violation of adjacency principle by the robot (the participant's turn is a question addressed at R concerning a language skill but R produces a turn that is topically inconsistent with P's question) ; *ii)* the "oh" produced quietly and transcribed in the CA transcription (L03) follows this overlap and marks thus a reaction to this transgression. This cue occurs during SED segment but is just signaled as acoustic cues by affective computing annotation; *iii)* the head cue annotated in the affective computing annotation in Figure 4 is more precise in CA transcript. It is detailed as an accompaniment of the verbal behaviour "oh" with the orientation of P towards his smartphone (L03); *iv)* the CA transcription indicates in L03-L04 linguistic cues denoting the re-engagement of the participant in the interaction. Reengagement is not so far included in the affective computing annotation categories. In CA transcription, we note that the "ok" seems to mark a re-engagement, a *springboard* (Beach, 1993; Rollet, 2013) towards a new orientation: that of moving towards the end of the interaction. This orientation is made visible by the production of a unit ("je ne sais pas qué: qu'est-ce qué tou (1,1s) dire\") that topicalizes an unresolvable non-understanding; *v)* the pre-closing phenomenon, that is, an interactional behaviour that sequentially precedes and serves to project the closing as such (Sacks & Schegloff, 1973). This phenomenon is illustrated here by the re-engagement cues to move towards the interactions described above. In this case, an "interactional blank cheque" is being given by P, which is followed by a 3.6s slot granted to R which is therefore transcribed on a separate line (L05).

Another interesting aspect of comparing the two forms of notation of this extract concerns the relationship between turns L02, L03 and L04. An initial analysis offered by the transcription is that in L03-04, P is marking disengagement according to two mechanisms (gestural—he looks at his cell phone—and verbal). And, still according to this analysis, the robot's turn L02 can be described as a transgression of the "one speaker at a time" rule (Sacks, Schegloff, & Jefferson,1974), made visible by the reaction that this transgression provokes (L03 "Oh"). In other words, in this case, a transgressive value is ascribed to the robot's turn L02, and P is attributed the initiative to move in two steps towards the end of the interaction.

However, an alternative to this analysis is possible, and can be derived from the affective computing annotation itself. Specifically, in the annotation, there is a comment associated with SED under the "cause" section: *Robot says goodbye too soon*. The first comment we can make is that such a "cause" is consistent with categorizing a segment as a "decrease in engagement". However, we can go even further. This comment is an analysis of an entirely different level than, for example, the Cue1 section with "head". This is a categorial and sequential analysis that has a significant influence on the subsequent interpretation of P’s behaviour. The fact that the annotator comments on R’s turn, “another time” as “Robot says goodbye too soon”, and because this is not a “goodbye” in the strict sense, shows that he considers this turn to be a behaviour projecting a closing, such as a “goodbye”: this is precisely the work accomplished by a pre-closing, or a *junction*[7] in general (Button, 1991). If this is the case, P’s “oh” may be an indication of disappointment due not to technical incompetence revealed by the emergence of an overlap (with a thematically incongruent turn), but to the participant analyzing Pepper’s turn as a pre-closing. Hence, P’s turn, and in particular the unit “je ne sais pas qué: qu’est-ce qué tou (1,1s) dire\”, could be analyzed as an alignment (constructed through topicalizing a non-understanding) with the end of the interaction, initiated by the robot: something along the lines of “I guess we don’t understand each other”. Following this analysis, the disengagement is therefore not initiated by P but rather, from P’s viewpoint, by R. Contrary to a disengagement, it is rather an *affiliation* (Stivers, 2008) of the participant towards the disengagement of the robot.

These two comparative analyses show to what extent the interpretation of a behaviour - even that of a robot - is not as univocal as it may seem, as soon as it is examined in its interactional framework. Moreover, regardless of the interpretation prioritized, a central and well-described sequence in conversation analysis literature emerges as an essential phenomenon for analyzing disengagement: pre-closing.

Cues, such as verbal (the exclamation “oh”) and non-verbal cues (posture, gaze) and conversation analysis’ interpretation of them also facilitate our understanding of

[7] The *junction* refers to a conversational pivot action that switches from the current topic to the closing. It is done in the current topic. There are several forms of action that put interaction on a closing track, e.g. projecting future activities, announcement of departure, or formulating summaries.

engagement breakdowns and can subsequently be integrated into the annotation scheme or automatically annotated in order to be quantified. If they are sufficiently frequent and representative, they can be added to the features extracted from videos for the machine learning of the system to detect engagement breakdowns.

c. Segmentation

The allocation of categories by the annotation schemes used in social robotics discussed above is generally based on a prior stage of segmenting audio/video feeds that allows one to delimit annotation segments. In this case, we present a conversation analysis viewpoint of the segments and phenomena thus delimited.

Annotating consists of two main stages. In the first stage the segments associated with the above-mentioned categories are identified according to the following steps:

1. Detecting the occurrences of one of these categories of phenomena (TD, EBD, etc., see 2.a).
2. Segmenting those phenomena over time by defining their time boundaries (segments that appear in the form of rectangles in Figures 2 and 5). Attributing the category identified to these time segments. These annotation time segments are called annotation units.

We believe to be crucial, first, discussing the segmentation of the interaction flow and, second, the categorization of the segments as well as the analysis level that they underpin. If we once again consider the comparative analysis of affective computing annotation vs. transcription presented above, a second conclusion that emerges is that annotation segments are “too macro”. Concretely, P’s turn in L03-04 can be successively analyzed as an indication of disengagement, with the “oh + looks at cell phone”, and a re-engagement initiated by “ok”. Specifically (Figure 5):

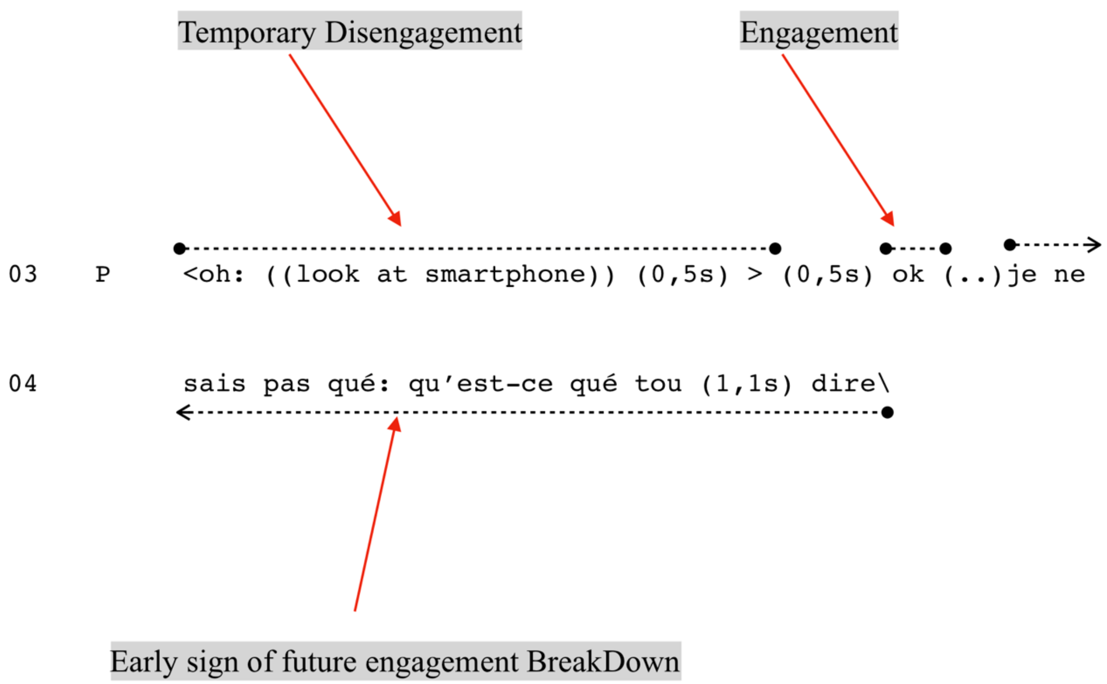

Figure 5. Assemblage diagram with CA transcription (from Transcript 2) and affective annotation (from Figure 4)

Such a breakdown suggests a methodological viewpoint already mentioned above: transcription attempts to provide analysis with the means of rendering (accounting for) participants’ ways of doing. This is particularly true for segmentation. Even though analysis requires one to extract and isolate interaction segments, this task can run the risk of losing its “natural” explanatory basis by becoming de-contextualized—with researchers in that case running behind something that was nonetheless already there. For if we, the second-hand observers, manage to follow a conversation or to understand a fragment of an interaction, it is because an initial analysis undertaking *in praesentia* took place.

By reconsidering the example in Figure 5, we note that the work accomplished by the participant through what he says and does can be described in three stages. Three stages for a turn; three *turn-constructional units* (Sacks, Schegloff, & Jefferson, 1974; Ford & Thompson, 1996) which each accomplish something different, as we describe above (Paragraph 2.a). A turn is a participation unit. It can be composed of multiple sub-units which are often delimited in terms of syntax, intonation, semantics, or pragmatics—and even gesturally or rhythmically. A turn is a contribution to the progressiveness of a situated interaction, and an initial signifying segmentation undertaking is conducted by the interactants themselves in the here and now of their social activities.

From a machine learning viewpoint, the question of the analysis unit is fundamental: what time frames of analysis should be used to extract social signals, and which units should be considered a frame for decision-making for the phenomenon to predict? Conversation analysis contributes to understand how people themselves break down the stream of language, and can help provide indications around the choice of analysis unit or the decision frame.

3. Conclusion and Prospects

In this article we have presented an affective computing annotation protocol dedicated to a project to detect disengagement in human-robot interactions. In the frame of an interdisciplinary collaboration between machine learning and conversation analysis within social robotics, this critical and constructive viewpoint can be applied to the analysis of:

- Context and relevancy: combining a multidimensional viewpoint with that of the interacting participants, affords a perspective that encompasses the situated nature of social behaviours.
- Issues surrounding annotation and the segmenting of interaction flows: as stated in Paragraph 2a., breaking down and categorizing the behaviours appearing in interactions are not meaningless practices: they relate to forms of representation with respect to the relationship between the verbal and the non-verbal, as well as regarding the level of exogenous ('etic') production of meaning.
- The question of sequentiality as a new "explanatory feature": this is a dimension that is all too often neglected but can nonetheless constitute a fundamental resource in the endogenous production of meaning, in the same way as the syntactic, semantic, melodic, and pragmatic levels. The identification of phenomena such as pre-closings and junctions as cues of disengagement not only concerns a set of typical actions ("I gotta go", "talk to you later", etc.), but also a space of interaction that highlights the significant relationship between "what has just happened" and "what could happen next".

An initial line of collaboration between the two disciplines is based on the idea that interaction is more fluid through robotic behaviours that tend towards a form of ordinariness (that is, practices that are recognizable as being able to appear in an ordinary, daily, and routine social interaction (Sacks, 1984)). In other words, collaboration in creating scenarios can consist in providing designers with the viewpoint of a competent participant of everyday life who has developed a reflexive perspective with respect to his or her own (ordinary) practices, which is then refined through detailed observations of diverse social interactions. The following interaction cues serve as examples:

- the acoustic forms of a robot's utterances which project an action in a sequential process;
- the construction of turns as pragmatic units that are not necessarily primarily based on syntactic or semantic considerations; and
- the orientation of actions from a sequential viewpoint, that is, in a logic of turn-taking system and establishing interactional episodes or activities.

As another prospect for collaboration, attention can be drawn to the explanatory cues of an annotation category. In the scheme presented above (Section 1), a number of cues ranging from the human-robot distance to acoustic features and spatial orientation are highlighted. Moreover, analyzing the interaction excerpt between the robot Pepper and a human (cf. Paragraph 2.a) reveals a fundamental feature in explaining disengagement in social interaction in general. Specifically, beyond the pre-closing phenomenon as such, the sequential dimension appears to be central and the machine learning chosen must be able to integrate this sequentiality. To analyze how a robot's turn (such as "another time") is treated by the participant ("je ne sais pas qu'est-ce qué tou dire"), semantic, acoustic and pragmatic contents are not enough: the sequential positioning highly contributes on the grasp of what takes place, what follows, and what the participant makes accountable. It thus becomes necessary to establish a way of categorizing / codifying this sequential dimension each time that the annotator, aware of this dimension, observes its consequential nature. This awareness means considering a certain teaching process intended for annotators, and questioning how far this can be taken. The first step would be to sensitize annotators on pre-closings and conclusive junctions as familiar phenomena they do experience in their ordinary life even if they've never 'conceptualized' it – a step we've been just started to test. Affective computing annotation work is already moving in this direction, consisting of the canonical sequential format in interaction, and namely adjacent pairs (Langlet & Clavel, 2014), such as the question-answer sequence. Rather than leaving this field completely open to the annotator, the idea here is to enrich the explanatory cues of a macro-category (TD, EBD, BD, SED) by implementing the "Cause" section of the best-demarcated sub-cues: first comes the question of sequentiality, but we can also consider the pragmatic or even topical dimension. In this sense, the precision regarding the annotation segment addressed in the Cause section becomes crucial. These sub-cues can either be used as subcategories to be predicted by machine learning, or for the design of the input features of machine learning in order to improve disengagement prediction models.

These prospects raise the question of the extent to which the phenomena observed in the data can be sufficiently formalized to be processed for machine learning in order to improve machines' ability in detecting these phenomena. Moreover, to what extent is the tension between the principle of describing the uniqueness of cases - defining the analytical mentality of Conversation Analysis - and the requirement of generalization for the training of automatic models bearable? That is to say, how closely can we model the uniqueness and emerging nature of social interaction?

References

Bartneck, C., Belpaeme, T., Eyssel, F., Kanda, T., Keijsers, M., & Šabanović, S. (2019). *Human-robot interaction: An introduction*. Cambridge: Cambridge University Press.

Beach, W. A. (1993). Transitional regularities for 'casual' "Okay" usages. *Journal of Pragmatics, 19*, 325-352.

Ben-Youssef, A., Clavel, C., Essid, S., Bilac, M., Chamoux, M., & Lim, A. (2017). *UE-HRI: a new dataset for the study of user engagement in spontaneous human-robot interactions*. Paper presented at the 19th ACM International Conference on Multimodal Interaction, Glasgow, UK.

Button, G. (1991). Conversation-in-a-series. In D. Boden & D. H. Zimmerman (Eds.), *Talk and social structure. Studies in Ethnomethodology and Conversation Analysis* (pp. 251-277). Cambridge: Polity Press.

Campano, S., Clavel, C., & Pélachaud, C. (2015). *"I like this painting too": when an ECA shares appreciations to engage users*. Paper presented at the 14th International Conference on Autonomous Agents and Multiagent Systems (AAMAS'15), Istanbul, Turkey.

Cassell, J., Torres, O., & Prevost, S. (1999). Turn taking vs. discourse structure: How best to model multimodal conversation. In Y. Wilks (Ed.), *Machine Conversations* (pp.143–154). The Hague: Kluwer.

Chapman, D. (1992). Computer rules, conversational rules. *Computational Linguistics, 18*(4), 531-536.

Clavel, C., Vasilescu, I., & Devillers, L. (2011). Fiction support for realistic portrayals of fear-type emotional manifestations. *Computer Speech & Language*, *25(1)*, 63-83.

Clavel, C., Cafaro, A., Campano, S., & Pelachaud, C. (2016). Fostering User Engagement in Face-to-Face Human-Agent Interactions: A Survey. In A. Esposito & L. C. Jain (Eds.), *Toward Robotic Socially Believable Behaving Systems – Volume II: Modeling Social Signals* (pp. 93-120). Cham: Springer International Publishing.

Clavel, C., & Callejas, Z. (2016). Sentiment analysis: from opinion mining to human-agent interaction. *IEEE Transactions on affective computing*, *7(1)*, 74-93.

Cowie, R., & Cornelius, R. R. (2003). Describing the emotional states that are expressed in speech. *Speech communication*, *40(1-2)*, 5-32.

Dautenhahn, K. (2007). Socially intelligent robots: dimensions of human-robot interaction. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 362*(1480), 679-704.

Dickerson, P., Robins, B., & Dautenhahn, K. (2013). Where the action is: A conversation analytic perspective on interaction between a humanoid robot, a co-present adult and a child with an ASD. *Interaction Studies, 14*(2), 297-316.

Duranti, A. (1997). *Linguistic anthropology*. New York: Cambridge University Press.

Duranti, A., & Goodwin, C. (1992). Rethinking context: an introduction. In A. Duranti & C. Goodwin (Eds.), *Rethinking Context, language as an interactive phenomenon* (pp. 1-42). Cambridge University Press.

Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and Autonomous Systems, 42(3-4)*, 143-166.

Garfinkel, H. (1967). *Studies in ethnomethodology*. Engelwood Cliffs: Prentice-Hall.

Goffman, E. (1973). *La mise en scène de la vie quotidienne* (Vol. 2. Les relations en public). Paris: Les Editions de minuit

Goffman, E. (1983). The interaction order. *American Sociological Review, 48(1)*, pp: 1-17.

Goodwin, C. (1981). *Conversational organization: interaction between speakers and hearers*. New York: Academic Press.

Heritage, J. (1991). L'Ethnométhodologie : une approche procédurale de l'action et de la communication. *Réseaux CNET, 50, 89-130.*

Jefferson, G. (1978). Sequential aspects of storytelling in conversation. In J. Schenkein (Ed.), *Studies in the organization of conversational interaction* (pp. 219-248). New York: New York Academic Press.

Jones, R. A. (2017). What makes a robot "social"? *Social Studies of Science, 47*(4), 556-579.

Langlet, C., & Clavel, C. (2014). *Modelling user's attitudinal reactions to the agent utterances: focus on the verbal content*. Paper presented at the 5th International Workshop on Corpora for Research on Emotion, Sentiment & Social Signals (ES3 2014), Reykjavik, Iceland.

Langlet, C., & Clavel, C. (2018). *Detecting User's Likes and Dislikes for a Virtual Negotiating Agent*. Paper presented at the 20th ACM International Conference on Multimodal Interaction, Boulder, USA.

Levinson, S. C. (1983). *Pragmatics*. Cambridge University Press.

Licoppe, C., & Figeac, J. (2014). L'Organisation temporelle des engagements visuels dans des situations de multi-activité équipée en milieu urbain. *Activités, 11*(1).

Mondada, L. (2006). Participants' online analysis and multimodal practices: projecting the end of the turn and the closing of the sequence. *Discourse Studies, 8*(1), pp. 117-129.

Mondada, L. (2008). Documenter l'articulation des ressources multimodales dans le temps : la transcription d'enregistrements vidéos d'interactions. In M. Bilger (Ed.), *Données orales. Les enjeux de la transcription* (Vol. 37), pp. 127-156. Perpignan: Presses Universitaires de Perpignan.

Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2012). *Foundations of machine learning*. MIT press.

Ochs, E. (1979). Transcription as a theory. In B. B. Schieffelin (Ed.), *Developmental Pragmatics* (pp. 42-72). Academic Press.

Pelachaud, C., & Glas, N. (2015a). *Definitions of Engagement in Human-Agent Interaction*. Paper presented at the International Workshop on Engagement in Human Computer Interaction (ENHANCE), Xi'an, China.

Pelachaud, C., & Glas, N. (2015b). *Topic transition strategies for an information-giving agent*. Paper presented at the the 15th European Workshop on natural Language Generation, Brighton, UK.

Pelikan, H. R. M., & Broth, M. (2016). *Why That Nao?: How Humans Adapt to a Conventional Humanoid Robot in Taking Turns-at-Talk*. Paper presented at the 2016 CHI Conference on Human Factors in Computing Systems, San Jose, California, USA.

Pitsch, K., Kuzuoka, H., Suzuki, Y., Süssenbach, L., Luff, P., & Heath, C. (2009). *"The first five seconds": Contingent stepwise entry into an interaction as a means to secure sustained engagement in Human-Robot-Interaction*. Paper presented at the IEEE International Symposium on Robot and Human Interactive Communication ROMAN 2009, Toyama, Japan.

Porcheron, M., Fischer, J. E., Reeves, S., & Sharples, S. (2018). *Voice Interfaces in Everyday Life*. Montreal QC, Canada: Association for Computing Machinery.

Rollet, N. (2010). *All the things you are*. Activité multimodale, frontières et musiques improvisées en répétition In N. Andrieux-Reix (Ed.), *Frontières. Du linguistique au sémiotique* (pp. 279-302). Limoges: Lambert-Lucas.

Rollet, N. (2013). "D'accord". Approche conversationnelle et multimodale d'une forme située dans les appels au Samu-Centre 15. *L'Information grammaticale, 139*.

Rollet, N., Jain, V., Licoppe, C., & Devillers, L. (2017). Towards Interactional Symbiosis: Epistemic Balance and Co-presence in a Quantified Self Experiment. In L. Gamberini, A. Spagnolli, G. Jacucci, B. Blankertz, & J. Freeman (Eds.), *Symbiotic Interaction: 5th International Workshop, Symbiotic 2016, Padua, Italy, September 29–30, 2016, Revised Selected Papers* (pp. 143-154). Cham: Springer International Publishing.

Sacks, H. (1984). On doing 'being ordinary'. In J. M. Atkinson & J. Heritage (Eds.), *Structures of Social Action* (pp. 413-429). Cambridge: CUP.

Sacks, H. (1992). *Lectures on conversation* (Jefferson, G. ed.). Oxford: Blackwell.

Sacks, H., & Schegloff, E. (1973). Opening up closings. *Semiotica, 8*, 289-326.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A symplest systematics for the organization of turn-taking for conversation. *Language, 50*, 696-731.

Sadazuka, K., Kuno, Y., Kawashima, M., & Yamazaki, K. (2007). *Museum Guide Robot with Effective Head Gestures*. Paper presented at the International Conference on Control, Automation and Systems, Seoul, Korea.

Schegloff, E. (1987). Between micro and macro: contexts and other connections. In B. Giesen, J.C. Alexander, R. Münch, & N.J. Smelser (Eds.), *The Micro-macro link* (pp. 207-234). Los Angeles: University of California Press.

Schegloff, E. (2002). Accounts of conduct in interaction. Interruption, overlap, and turn-taking. In T. J.H. (Ed.), *Handbook of sociological theory* (pp. 287-321). New York: Kluwer Academic/Plenum publishers.

Schegloff, E. (2007). *Sequence organization in interaction. A Primer in Conversation Analysis*. (Vol. 1): Cambridge University Press.

Sidner, C. L., & Dzikovska, M. (2002, 16-16 Oct. 2002). *Human-robot interaction: engagement between humans and robots for hosting activities*. Paper presented at the 4th IEEE International Conference on Multimodal Interfaces, Pittsburgh, USA.

Sidner, C. L., Lee, C., Kidd, C. D., & Rich, C. (2005). Explorations in engagement for humans and robots. *Artificial Intelligence, 166*, 140-164.

Stivers, T. (2008). Stance, Alignment, and Affiliation During Storytelling: When Nodding Is a Token of Affiliation. *Research on Language & Social Interaction, 44*(1), 31-57.

Stivers, T. (2015). Coding Social Interaction: A Heretical Approach in Conversation Analysis? *Research on Language and Social Interaction, 48*(1), 1-19.

Šabanović, S., & Chang, W.-L. (2016). Socializing robots: constructing robotic sociality in the design and use of the assistive robot PARO. *AI and Society, 31*(4), 537-551.

Straub, I. (2016). “It looks like a human!” The interrelation of social presence, interaction and agency ascription: a case study about the effects of an android robot on social agency ascription. *AI and Society, 31*(4), 553-571.

Suchman, L. (2007). *Human-Machine Reconfigurations. Plans and situated actions, 2nd edition*. New York: Cambridge University Press.

Yu, Z., Scherer, S., Devault, D., Gratch, J., Stratou, G., Morency, L., & Cassell, J. (2013). *Multimodal Prediction of Psychological Disorder: Learning Verbal and Nonverbal Commonality in Adjacency Pairs*. Paper presented at the 17th Workshop Series on the Semantics and Pragmatics of Dialogue, Amsterdam, Netherland.

Zimmerman, D. (2006). *How closing matters in emergency telephone calls*. Paper presented at the Annual meeting of the American Sociology Association, Montréal, Canada.