Multimodal Aspects of Affectivity in Interaction and Language Acquisition – An ICCA 2010 report

Maxi Kupetz

1. Introduction

This ICCA’s conference theme ‘multimodal interaction' stressed the relevance of understanding social interaction in its holistic gestalt, taking into account all possible resources participants may use to achieve meaningful interaction. Even though the term ‘multimodality’ is largely debated, here, it is deployed to describe the simultaneity of different modalities of interaction. This means that all resources are considered equally: linguistic and para-linguistic resources such as lexico-semantics, morphosyntax, phonetics, prosody, as well as visible resources such as gesture, gaze, facial expression, body orientation, manipulation of objects etc. (for further discussion see Deppermann/Schmitt 2007).

In this paper, I will provide a summary of the plenaries held by Charles Goodwin and Lorenza Mondada, as well as an overview of two panels, i) Affect and Emotion in Interaction (Part 1: Everyday Interaction), and ii) Multimodality in Language Acquisition. In addition to a short summary of these papers, I will address the issues which were subject to discussion in order to provide insights into the current discourse of the Conversation Analysis (CA) scholars.

2. Plenaries

2.1. Goodwin: Building Action by Combining Unlike Resources

In his plenary, Charles Goodwin showed how, in social interaction, humans have the capacity to construct meaning by tying language to visible, embodied displays in a relevant action. Goodwin pointed out the collaborative aspect of social interaction by stressing the role of alignment and cooperation between participants. Ongoing actions are constructed through embodied co-participation. The ‘division of labor' - defined here as the simultaneous organization of diverse semiotic materials used by different actors within a sequential framework – is decisive, as humans "inhabit each other’s actions".

Actors may use a variety of resources in order to adapt to different contingencies and locally changing circumstances. This variety includes linguistic and para-linguistic features, bodily behavior, as well as objects and tools present in the interaction. These unlike phenomena are joined together within a public framework to create a new whole that is greater than the sum of its constituent parts. The notion of epistemic ecology describes how actions are built through the sequential intertwining of i) language, ii) body/gesture, iii) objects/structure in environment.

1 I would like to thank Margret Selting, Elizabeth Couper-Kuhlen, Friederike Kern, and Dagmar Barth-Weingarten for their enriching comments on this report and Andrea Donohue for proof-reading.
Another aspect which is central to human interaction compared to the communication of primates is the process of decomposition of actions. A video extract in which a mother shows her daughter how to make pancakes was used to illustrate i) how the manipulation of a tool, e.g. a bowl, can link two or more actors into a coherent course of joint action, and how tools become integrated components of interaction, and ii) how collaborative multiparty action is achieved through instruction and decomposition of action, the latter being a pedagogic practice unique to humans. These collaborative processes are essential for meaning making; the joint involvement in cooperation makes it possible to understand each other and to understand each other’s actions.

In the discussion, questions were raised with regard to the relationship between the resources deployed by actors and to the primacy of a modality. In response, Goodwin pointed out that actors project what is to come by recognizing the constant transformation of the local environment; therefore, they constantly reallocate and shift their resources. He furthermore argued that there is no such thing as a primacy of language since humans are very flexible and action in interaction can take many forms. Hence, it is crucial to look at the embodiment of actions and how humans inhabit the environment to form their actions.2

2.2. Mondada: Interactional Space: The Embodied Organization of Participation and Turn-Taking

In her plenary, Lorenza Mondada focused on the visual and public accountability of embodied and spatial resources used for the organization of turn-taking in institutional interaction. She explored in detail practices for the achievement of speakership in multiparty settings. The video-data were collected from participatory democracy meetings in Lyon, where a chairman organizes the discussion by moving between tables and pointing at persons who would like to take the floor.

Within this dynamic participation framework, the chairman very early on identifies the next speaker and pre-publicizes them as a next speaker through pointing. This pre-selection can be accomplished for one or more next speakers. While securing the public visibility of all current pre-selector, the chairman defends the current speaker and his right to be the one speaker at a time and to be listened to. This double orientation is achieved through resources such as gesture, gaze, and circulating in the room. It is striking how these resources are used independently from each other to constantly transform the participation framework as e.g. head and hands may show orientations to different candidate speakers.

Mondada stressed that the achievement of speakership in a public debate is not only a matter of visibility but also of publicity. Next speakers do not only need to be identified, they also need to be promoted publicly and put into the focus of everyone’s attention by the chairman. The latter works out the visible sequential structure of the debate, thus, turn-taking procedures become transparent and distinctively shape the parties’ opportunities for action. The achievement and public promotion of multiple speakerships are distinctive ways of enabling participation.
and of constituting public figures. Interestingly, public visibility makes transparent the rules, norms and values of the debate. Here, overlaps are not treated as organizational problems but as moral problems.

In the discussion, it was pointed out that further aspects need to be considered, e.g. i) the role of the pieces of furniture and their arrangement in the room since there seems to be a conflict between the resources 'body' and 'objects' at times when the chairman moves around, and ii) the relevance of types of gestures and hand shapes used by the chairman since through their economic transformation various states of the debate seem to be managed.

3. Panels


In recent years, studies of everyday interaction (Couper-Kuhlen 2009; Reber 2008; Selting 1994, 1996, 2010; Wilkinson/Kitzinger 2006) as well as institutional interaction (Heath 1988; Sandlund 2004; Ruusuvuori 2005, 2007) showed the relevance of affect displays in social interaction. At the international conference 'Prosody and Interaction' in Potsdam in 2008, scholars agreed that the study of affectivity in interaction requires the analysis of phonetic-prosodic as well as visual resources (Barth-Weingarten/Reber 2009). In order to describe the nuances and variants of affect-laden talk, researchers need to identify the clusters of resources that participants use to make them recognizable.

The overall aim of the ICCA panel organized by Margret Selting and Elizabeth Couper-Kuhlen was to assemble scholars involved in the analysis of emotion and affect displays in everyday spoken interaction. The scholars invited were to focus either on practices used by participants to enact affect displays or on activities in which affect displays are made relevant and locally managed by the speakers. First, I will concentrate on papers that focused on practices such as prosodic matching, prosodic intensification, laughter, and facial expression used to display affective stances in interaction, and possibly involved in emotion regulation. Second, I will summarize papers which focused on specific activities such as storytelling, teasing, or compliments, and the practices used to display affectivity within these activities. After a short introduction of each paper, a summary of the discussion and remaining open questions will be given.

Against the background of previous research on matching and non-matching prosody in interaction (e.g. Szczepek 2001), Mikko Kahri analyzed video-recordings of 4 to 6-month-old infants and their caretakers in order to reveal the relationship between prosodic matching and/or non-matching and emotion regulation. He argued that emotion regulation can be achieved in interaction through prosodic matching and 'adding' new features (non-matching). E.g. if harshness in the infant’s vocalization makes a potentially negative affect recognizable, the caretaker will move away from that affect and add 'playfulness' by matching her phonetic sound production but deploying markedly different pitch register and smiley voice.

Richard Ogden showed that 'intensifying emphasis' is used as a resource for making something recognizable as extraordinary. It is related to the practice of phonetic upgrading (e.g. Ogden 2006) and may display and/or project a certain
stance. Ogden focused on the prosodic features of assessments in story beginnings and found that intensifying emphasis can be prosodically achieved through pitch prominence, tense articulation, gaps, and glottal stops as well as faster speech before the intensified item. He also drew on other sequential and linguistic features of talk. Possible functions of intensifying emphasis are, among others, to upgrade an assessment, to provide a non-neutral interpretation, to signal contrast, to handle epistemic issues, or to project the tellable as extraordinary.

Defining laughter as a socially organized action, Markku Haakana focused on stylized laughter which participants orient to as 'fake'. Through sample analyses of Finnish telephone conversations it became obvious that participants themselves make various sorts of laughter relevant, e.g. in the course of joking activities. However, Haakana made it clear that various issues still need to be addressed with regard to i) the terminology of the phenomenon, ii) its specific characteristics (when and how is laughter to be recognized as 'fake'?), iii) the division of labour between the tokens 'hah' and 'heh', and iv) the role of the actual affect involved. In the discussion, it was mentioned that Bachtin’s concept of 'layering voices', where "two voices [and intentions, MK] are dialogically interrelated" (Günthner 1997:259), might be helpful to describe the discrepancy between the distancing from laughter and the invitation to laugh at the same time.

Drawing on previous studies on verbal pursuits of response (e.g. Pomerantz 1984), Johanna Ruusuvuori explored the pursuit of response by emotional facial expression. An analysis of video-recordings of Finnish university students at lunch revealed how stance can be reciprocated through facial expression before any further vocalization occurs. Hence, an ongoing activity can be maintained, e.g. by contextualizing it as a 'joke'. A story may be assessed through facial expression in lieu of verbal assessment. The pursuit of response by facial expression maintains the ongoing activity and preserves the intersubjectivity of the participants. In the discussion, Ruusuvuori herself raised the question of how the results of her analysis are relevant for the study of emotion regulation. Furthermore, it was mentioned that an integration of facial and prosodic pursuit resources might complete the analysis in the future.

Starting from a social constructionist conception of affectivity, Margret Selting analyzed the display of affect in talk-in-interaction in German. The subject of her presentation was a case study of a complaint story and subsequent complaint stories in which affective stances such as 'anger' and/or 'indignation' are displayed. Affect displays were realized at specific sequential positions in the stories by a combination of lexicosemantic and rhetoric devices such as exaggerations and swear words, marked prosodic resources such as high pitch register, conspicuous pitch jumps or laugh particles, and visual devices such as raised shoulders or feigned 'cheeky' face. These resources are used to make a story recognizable and interpretable as a complaint story displaying anger. It is treated as such by the co-participant when she uses similar resources to display a similar kind of affect in her subsequent story. The case study made obvious that affectivity is a locally constructed public display which needs to be perceived and responded to in social interaction (see also Selting 2010).

Susanne Günthner showed how participants suggest affective meaning by deploying syntactic constructions in storytellings which Günthner calls 'dense constructions'. These marked syntactic structures are constructions which traditio-
nal German grammar would classify as 'elliptical'. They consist of a pronoun and a prepositional or participle phrase, but lack the typical German sentence brace. However, the actor or agent of an event is introduced and the event is narrated (e.g. *ICH (.) in die bahnhofshalle*, / 'i (.) in the station concourse*'). Günthner argued that these constructions are performative devices which minimize the distance between the participants and the events. The syntactic and prosodic form of these 'dense constructions' illustrate past events as immediate and dramatic and make them interpretable for recipients by indexing the teller's specific affective stance.

Elizabeth Couper-Kuhlen explored the notion of affiliation (Stivers 2008) in storytelling in English telephone conversations. She showed that story recipients can display emotional reciprocity through response cries with prosodic matching. However, there seems to be a relationship between the way in which participants’ further substantive responses are realized and their 'entitlement to experience' (Sacks 1992) of this type of affect. If there is an asymmetric distribution of 'entitlement to affectivity' between the storyteller and the recipient, the latter must work to manage this 'feeling differential' (cf. 'knowledge differential' in Drew/Heritage 1992). This can be done by avoiding substantive responses to affect displays or by realizing these without prosodic matching. Both response cries and substantive responses are only realized with prosodic matching if the recipient can claim to have the same right to experience. In light of these subtle differences, Couper-Kuhlen pleaded for the use of the term 'emotional reciprocity' rather than 'mirroring' as is used in psychology.

Niklas Norén explored 'embarrassment' in teasing activities in a corpus of everyday Swedish conversations between elderly female friends. He drew on a previous study on embarrassment in academic settings (Sandlund 2004) and focused on the sequential organization of embarrassment. This involves 1) a teasing action (eliciting action) which makes 2) an immediate response (potential spot for the display of embarrassment) relevant, followed by 3) the management of the teasing (e.g. through mutual laughter), and 4) the closing of the sequence (e.g. through the introduction of a new topic).

With reference to previous work on the preference structure of compliment responses in English (Pomerantz 1978) and in German (Golato 2005), Mia Halonen showed how in Finnish, compliment responses are finely tuned to manage the balance between self-praise and shallowness. Participants seem to achieve this balance by agreeing to and upgrading the compliment, and then providing an (eventually negative) account, a reason, or an explanation for the accessibility to the praised object. This is interpreted as showing the participants’ orientation to the truth value of the compliment (epistemic stance). Halonen concluded that specific cultural codes and functions of compliments need to be studied in the future.

Charles and Marjorie Harness Goodwin were the panels’ discussants. They provided an overview of the topics discussed by all papers and of further challenges in the study of affectivity in interaction. It was pointed out that there was a resonance throughout the papers that emotion or affect displays are invitations to participate in various ways in the interaction. On the one hand speakers make use of various resources to pursue responses by the recipient, on the other hand prosodic and nonverbal resources are important features to align or even affiliate with the speaker.
In order to grasp the relation between several types of resources, the notion of 'lamination' was brought up. Speakers do not just produce speech on various linguistic levels; they perform dense human activities and invest themselves in taking stances, pursuing responses, and aligning their perspectives with each other. Therefore, it is important to identify what each type of resource does in interaction, what features serve particular functions or carry particular meanings, and how meaning is achieved through the combination of these features.

In the future, it should be discussed in more detail what kind of phenomenon is meant by the notion of 'emotion regulation' and to what extent this might reach into the field of psychology. The strength of micro interaction analyses of emotion regulation lies in the opportunity to describe how (affective) responsiveness is pursued by actors, and to explore the role of various resources such as mutual gaze or matching prosody. Another aspect to be studied is the relevance of the participation framework, e.g. the relationship between persons present in an interaction and third persons talked about should be investigated. This also leads to the issue of 'entitlement to experience' (Sacks 1992) as discussed in Couper-Kuhlen’s paper. A further topic to be taken up in the future is 'irony' and how it is realized and interpreted in the midst of social activities.

In a concluding summary, Margret Selting outlined two perspectives on how speakers put unlike resources together to make affects recognizable, i) a psychological perspective where notions such as 'mirroring', 'internal processes', and 'emotion regulation' come into play, and ii) a CA or interactionist point of view where notions such as 'reciprocity', 'public processes', 'negotiation of affect', and 'affect and stance as interactional phenomena' are considered. The main questions to be addressed are, on the one hand, how specific affects can be recognized in interaction, and, on the other hand, how they are handled by participants. The assumption is that there are ordered sequences which may be affect specific, i) so called 'unmarked trajectories' where the recipient goes along with the display of affect suggested by the speaker, ii) 'marked trajectories' where the affect displayed is resisted to, and iii) cases of 'fake' affect which need to be situated on a continuum between 'fake' and 'real' affect and whose communicative function still needs to be explored in detail. Interestingly, in cases of 'fake' affectivity, participants themselves seem to treat the negotiation of meaning as a problematic interactive issue.

Despite the fruitful discussion at the ICCA, there remain a number of unresolved questions which need to be addressed in future conversation analytic research. For instance: notions that are inconsistently used at the moment need to be clarified and differentiated: What do we mean by 'stance', 'affective stance' or 'affective meaning'? Which role does the term 'affiliation' play in the study of affectivity? What is the relationship between 'affiliation' and 'emotional reciprocity'? How can we distinguish between 'emotional reciprocity' and 'empathy'? Further issues arise with regard to the classification of affect displays. How do we know how a certain kind of affect is displayed, e.g. 'embarrassment'? How can we recognize it? And, if we, as conversation analysts, aim at revealing practices used to constitute social actions – then what is an affect display? A social action? A practice?
3.2. Multimodality in Language Acquisition

The panel 'Multimodality in Language Acquisition' was organized by Petra Strähle and Friederike Kern. Its main aim was to bring together researchers on first and second language acquisition with CA background to address the question of how nonverbal and para-verbal means are used by children as discourse-related resources in the acquisition of communicative competences.

In order to reveal how children coordinate tone with lexical and non-vocal acts, Bill Wells analyzed video-recordings of an about 20-month-old boy and his caretaker. Focusing on cases where the child repeats the last word the mother has said, Wells described the tone characteristics of these lexical repeats, compared them to the mother’s utterance and described accompanying non-vocal acts. It was shown how the child varies these resources with regard to the ongoing action (e.g. repair or topic shift). Lexis, tone, and non-vocal acts are thus independent and variably combinable practices. Moreover, in the interaction with the caretaker, there seem to be occasions for the child to learn about tonal contrast and its deployment for specific communicative actions.

Addressing the question of when a child becomes a 'competent' member in social interaction, Michael Forrester analyzed video-recordings of an about 24-month-old girl sitting at the dinner table with her parents. It was shown how the child is constantly interactively engaged by the parents through each other’s mimicking of sounds, the production of collaboratively achieved intonational 'choreographies', as well as parental productions of second pair parts. Opportunities for 'doing talk' seem to be made available through the methodic production of engagement sequences when parents act as if the child wanted something they know of. There seem to be indications of a pre-linguistic 'doing formulating' and related reflexively accountable social actions.

Regarding gesture as an opening wedge into communication which predates language development and touches off opportunities for language learning, Anna Filipi explored the role of pointing in language acquisition. Sample video analyses revealed how 18-month-old children can initiate, e.g., request sequences through pointing and vocalization, and how 24-month-old children can have control over interational organization by making certain 'answers' by parents relevant. In 'working on language' sequences, not only do the parents 'model' the ongoing activity, but does also the child work on his/her own communicative skills. Filipi concludes by stating that by focusing on the sequence structures and the resources used to create sequences, the actions of both parent and child working together become much more explicit. Acquisition becomes visible as parent and child move between interacting to socialize and interacting to work on language skills.

In her longitudinal study, Petra Strähle aimed at exploring children’s discourse skills in greeting routines and at showing, through a comparison of greeting sequences at different ages, how acquisitional processes become evident in multimodal interaction. The analyses of video-recordings of front-door encounter openings in the children’s homes showed how children develop from 12 to 48 months of age by running through various phases. First, they seem to develop a sense of basic structures of a greeting sequence, e.g., making each other recognizable, identifying each other, paying attention to each other. Later on, children make use of greeting gestures such as hugging, kissing, handshakes, well before
acquiring the adequate greeting formulae which consists of a cluster of linguistic, and non-linguistic resources. Strähle argues strongly in favor of a multimodal approach, since nonverbal behavior is shown to be an important resource in social interaction before language acquisition itself.

Amy Sheldon addressed the question of how multiple modes of communication are choreographed in children’s social activities. Her corpus consisted of various video-recordings of three four-year-old children playing at a table at the daycare center. A micro-analysis of an extract in which a girl mis-constructs her turn by saying 'I'll cut my soup' revealed how children jointly enact 'pretend play' activities and how these play sequences are co-constructed. Sheldon claims that spoken discourse is not reducible to a linguistic description and that children’s discourse competences should include a description of their synchronous and co-expressive, co-produced body movements when they interact with each other or others.

Carmen Taleghani-Nikazm was invited as discussant. She provided a short summary of all papers and addressed advantages as well as challenges which arise when approaching language acquisition from a CA perspective. In general, it was shown in the papers how local and sequential issues seem to be closely linked to a child’s emerging membership within society. However, the notion of 'membership' needs to be discussed more in detail: what is constitutive for someone to become a member of a certain culture and then to be treated as such? How is membership conceptualized in terms of skills and competences? What is the meaning of the term 'competence'? The latter question was largely debated. Some CA researchers prefer the term 'communicative repertoire' which underlines that research focuses on communicative practices which can be observed, rather than on internal processes of speakers which are inferred from observations. Other researchers argue that it is the conversation analyst’s task to incorporate the notions of 'competence' and 'development', i) because it has been shown that it is what children sometimes do orient to in interaction, and ii) because CA work should make itself compatible with work from other disciplines rather than isolating itself from them.

It is the strength of a micro interaction analysis that it provides the opportunity to explore the immediate context of the child’s acquisition of communicative competence, e.g., studies on repair revealed that children very early show the ability to identify misunderstandings (e.g. Salonen/Laakso 2009). In the future, CA could provide a description both of the interface of linguistic, social, physical, and pragmatic skills and of how a child develops an understanding of sequential aspects of social activities. However, the crucial problem of integrating theoretical accounts from other research fields such as developmental psychology, cognitive sciences, and psycholinguistics still remains unsolved. As the common assumption is that language is learned in interaction, interactionist approaches from cognitive sciences may be compatible with CA. Yet, the combination of these quantitative approaches with micro interaction analyses remains a challenge.

The panel reflected critically upon the explanatory power of CA as a framework for describing language acquisition regarding the problems of longitudinal corpora and the amount of work which needs to be invested in analyzing them. Furthermore, it remains difficult to compare the results of various studies in terms of stages of acquisition when communicative situations vary and linguistic resources are not comparable. However, there was a plea for changing the perspec-
tive: research should focus on the acquisition of social practices/activities and their complexity rather than ‘only’ on language, as this is what makes someone a member of a society – knowing how to deal with expectations by other members and complying with them.

In sum, all participants stressed the need for an informal network of researchers working on language acquisition and multimodality from a CA perspective. This is necessary in order to tackle the complex issues addressed and make themselves heard and understood inside as well as outside the CA community.

4. Concluding remarks

In general, ICCA 2010 provided an excellent opportunity to immerse in the state of the art in CA research and to get in touch with CA scholars from all over the world.

A general observation: authors often presented video-data, but provided transcripts reflecting only linguistic and para-linguistic resources used by the participants. Hence, there seems to be a strong need for a transcription system which allows us to incorporate visual practices in an efficient and practicable manner.

Furthermore, in many papers the analysis was restricted to just one type of resource and usually in the discussions, these limitations were addressed. However, in his plenary, Goodwin pointed out that a multimodal approach is needed to understand human interaction, multimodality referring to the combination and coordination of all possible resources and their evolvement in time and space. We as analysts should take this plea for a multimodal approach seriously and analyze everything participants provide us with to explore what is made relevant and interpreted by participants in social interaction.

5. References


Maxi Kupetz, M.A.
Universität Potsdam
Institut für Germanistik
Am Neuen Palais 10
14469 Potsdam
kupetz@uni-potsdam.de

Veröffentlicht am 17.3.2011
© Copyright by GESPRÄCHSFORSCHUNG. Alle Rechte vorbehalten.