Incremental Collection of Activity-based Multimodal Corpora and their Use in Activity-based studies

Jens Allwood & Elisabeth Ahlsén
SSKKII Interdisciplinary Center/Communication and Cognition
University of Gothenburg
jens@ling.gu.se, eliza@ling.gu.se

Abstract

Activity-based communication Analysis is a framework, which puts social activity in focus and analyzes communication in relation to the determining and determined factors of the activity. Given an activity-based approach, it is essential to collect multimodal corpora with a variation of social activities, in order to study similarities, as well as differences between activities and possible influencing factors. The Gothenburg Spoken Language Corpus was collected as a corpus representing communication in a wide range of social activities. The paper describes and briefly discusses the purpose and some of the features of the corpus. The usefulness of activity-based multimodal corpora is exemplified by the analysis of spoken feedback in a specific activity (the physical examination in doctor-patient interaction).

The framework of Activity-based Communication Analysis (ACA) for Studying Multimodal Communication

A communicative situation can to a large extent be identified with the social activity the communicators are engaged in. Activity-based communication Analysis is a framework for the influence of situational factors on communication developed by Allwood (1976, 2000, 2001, 2007), which puts social activity in focus and analyzes communication in relation to influencing and influenced factors of an activity. The framework is on inspired by work in philosophy, linguistics, anthropology, psychology and sociology, most importantly Peirce, Wittgenstein, Austin, Grice, Bühler, Malinowski, Firth, Vygotsky, Rommetveit, Mead, Goffman, Garfinkel and Sacks, Schegloff and Jefferson, and sees communication as action , involving degrees of coordination, collaboration and cooperation, in particular social activities.

Some influencing factors in an activity are global, i.e. influence the activity as a whole, while others are local, i.e. influence specific parts of an activity. Some of the influencing activity factors are collective, which means that they influence all participants in an activity, while others are individual, which means that they influence only individual participants.

Besides influencing (mainly background to communication) factors, there are influenced communication parameters in the interaction, which can also be global or local and collective or individual. Among the collective influenced parameters we find for example, interaction patterns that are produced collaboratively, while examples of individual influenced parameters are particular traits of communicative behavior or particular traits of perceiving/understanding speech and gestures for each of the participants. A summary of the framework is found below.

Influencing factors of an activity

- Collective: purposes and function of the activity, participant roles of the activity, the sub-activity structure of the activity, artifacts and other instruments used in the activity as well as social and physical environment of the activity
- Individual: goals of the individual participants, individual role interpretations, individual artifacts as well as individual interpretations of the environment

Besides the influencing factors associated with the activity type, participant roles, activity instruments and environment, the communicative activity itself also has a "reflexive" influence on the development of the activity, through a continuous influence of a contribution on the contribution that follows it.

Influenced factors in an activity

- Collective: interaction patterns, such as those to be found in interactive communication management (turn management, feedback patterns and sequences)
- Individual: communicative behavior and perception of communication particular to individual participants (e.g. production and perception of vocabulary, grammar, pronunciation, gestures)

The ACA theory and framework rest on a belief that activity factors are important and lead to important differences between social activities, so that normally

only some of the findings based on a study of communication in a particular social activity can be generalized to other social activities. Understanding how activity variation affects features of communication is therefore an important goal of using this framework. Since both the physical conditions (non-communicative, but nonetheless informative) actions as well as the use of gestures, tools etc. can vary between social activities, an analysis of multimodal communication is always relevant.

The Need for Multimodal Corpora

Given an activity-based approach, it is essential to collect multimodal corpora with a variation of social activities. This makes possible a study of similarities, as well as of differences between activities and possible influencing factors. Depending on available resources, an activity-based corpus can be collected during a limited period of time as a project in itself or incrementally, by accumulating multimodal recordings from different projects, involving different activity types, as in the corpora described in this paper. Multimodal corpora are necessary for capturing important aspects of the interactive and individual communicative parameters of ACA, as well as of the setting. The purpose of the corpus is an important initial consideration. The corpora presented in this paper have all been collected with the purpose of studying multimodal interaction in different, mostly naturalistic, settings, with a priority on ecological validity. This emphasis on field recordings means that the best possible quality, given this condition, has been achieved, but that naturalness has been more important than studio quality. Since face-to-face interaction has been prioritized, most of the field recordings have been made using one camera, with all participants visible in the same picture. In studio made recordings, three cameras have sometimes been used, together with separate microphones, in order to make the material useful for in-depth analysis of, for example, facial expressions and speech characteristics.

Incremental Data Collection and Structure of the Corpus

Since the early 1980's, the Gothenburg Spoken Language Corpus (GSLC) has been incrementally collected, i.e. new social activities have gradually been added from different projects and other sources. The corpus consists of mostly videorecorded interactions in Swedish from 25 general activity types. The size of the corpus is around 1 400 000 transcribed words. The included activities are: Arranged discussions, Auction, Bus driver/passenger, Church, Consultation, Court, Dinner, Discussion, Factory conversation, Formal meeting, Games & play, Hearing, Committee on the constitution, Hotel, Informal conversation, Interview, Lecture, Market, Meeting, Phone, Political debate, Retelling of article, Role play, Shop, Task-oriented dialogue, Therapy, Trade fair, and Travel agency, TV.
Since the corpus is dynamic and grows mainly by the inclusion of new activities from new projects, there is more material from some activities and less from others, something which has to be taken into account when activities are compared. One important feature of an activity-based corpus is to have metadata organized, so that different features can be extracted and compared. The GSLC videorecordings, transcriptions and codings have headers with some of the metadata easily available and retreivable. A corpus browser allows different search procedures based on the transcriptions and headers. Below is an example header with metadata. All names are pseudonyms.

@ Activity type, level 1: Consultation
@ Activity type, level 2: -
@ Activity type, level 3: P-D: Radiation
@ Recorded activity title: Patient-Doctor Conversation: Radiation Control
@ Recorded activity date: 890914
@ Recorded activity ID: A500302
@ Transcription name: A5003021
@ Transcription System: MSO6
@ Duration: 00:07:53
@ Short name: Radiation
@ Participant: D = (Dr. Bengtsson)
@ Participant: P = (Patient)
@ Anonymized: yes
@ For external use: no
@ Kernel: yes
@ Transcriber: Unknown
@ Transcription date: 950815
@ Checker: Elisabeth Kovacs
@ Checking date: 950828
@ Project: doctor-patient conversations
@ Comment:
@ Time coding: yes
@ Transcribed segments: all
@ Tape: a5003, ka5003
@ Section: 1: Start
@ Section: 2: Main reason
@ Section: 3: Physical
@ Section: 4: Diagnosis
@ Section: 5: History
@ Section: 6: Ordination
@ Section: 7: Diagnosis end
@ Section: 8: Frame
@ Section: 9: End
@ Stats: Audible tokens: 911
@ Stats: Contributions: 118
@ Stats: Overlapped tokens: 32
@ Stats: Overlaps: 22
@ Stats: Participants: 2
@ Stats: Pauses: 73
@ Stats: Sections: 38
@ Stats: Stressed tokens: 4
@ Stats: Tokens: 924
@ Stats: Turns: 98

As we can see in the example, the activity is further divided into sub-activities or sections, which often have specific characteristics. A more advanced relational database could also be very useful, but requires more administrative effort and is not as easily available to users of the corpus. The videorecorded and/or audiorecorded activities have all been transcribed, using a standardized format the Modified Standard Orthography (MSO6) (Nivre, 1999) and the Gothenburg Transcription Standard (GTS 6.4) (Nivre, 2004). The transcriptions have been checked by a second transcriber and by a transcription checking tool, in order to ensure that they can be merged and that a number of tools for calculating types of behavior, making concordances of words, counting and sorting various features can be used. The transcriptions can be used in different formats: e.g. the transcribed spoken language variant and the written language equivalent variant. This enables a transcription close to speech for spoken language analysis and a written language version for comparisons between spoken and written language. Some of the activities have also been annotated for multimodal communication, either using the comment function of GTS or using multimodal transcription tools, such as Praat and ANVIL. Other annotations have also been made for specific purposes.

In addition to the GSLC, activity based multimodal corpora of face-to-face interaction, based on the same principles as the GSLC have also been collected in a number of other countries, which makes interlinguistic and intercultural comparisons of sub-corpora possible.

ACA Multimodal Analysis - An Example

We will now consider an example of the use of a multimodal activity-based corpus – a study of feedback in the physical examination sub-activity/phase of a typical doctor-patient interaction. This example of how an activity-based multimodal corpus can be used illustrates that even if, as in this case, spoken output was in focus, a multimodal corpus provides information on what goes on in the activity, which is important for determining what is actually said and why. In this case, linguistic vocal verbal feedback in three types of sub-activity in doctor-patient interactions was analyzed.

The results of counting utterances, words, feedback words (absolute numbers, showing the amount of speech) and the relative share of feedback words out of the total number of words as well as a classification of the type of feedback (given as the percentage of the total number of feedback words), based on the transcription and coding of a specific doctor-patient interaction is shown in table 1. The numbers are given for each of the following three phases: case history (case hist), physical examination (phys ex) and ordination (ordin) totally and separately for the doctor (D) and patient (P) in each of the phases. The *share of feedback* is the share of feedback word tokens out of the total number of word tokens – it indicates how much feedback is used in relation to other words. The *share of utterances containing initial feedback units* (i.e. a feedback word like *yes, no* or *m* at the beginning of an utterance) and the *share of utterances containing only feedback words* indicate the role of feedback and the type of utterances dominating an activity. The *share of totally overlapped feedback units* can tell us if there is a great deal of back-channeling from one participating during long utterances or narratives produced by the other participant. The *share of interrupting feedback* shows if participants interrupt each other frequently, e.g. because the interaction is fast.

Sub-activities	Number of utte-rances	Number of words	Total Number of feed-back words	Feed-back share of speech
Case hist	711	5680	530	9.3
Case hist D	373	2315	249	10.8
Case hist P	338	3365	281	8.4
Phys ex	492	3317	358	10.8
Phys ex D	251	2166	176	8.1
Phys ex P	241	1151	182	15.8
Ordin	831	7473	667	8.9
Ordin D	410	5344	268	5.0
Ordin P	421	2129	399	18.7

Sub-activities (% of feeback words)	Initial FB	Only FB	Inter-rupting FB	Over-lap-ped FB
Case hist	23.6	27.4	2.8	8.0
Case hist D	18.2	32.7	2.4	12.6
Case hist P	29.5	21.6	3.3	3.0
Phys ex	21.1	24.4	3.3	4.9
Phys ex D	20.7	16.7	2.8	4.4
Phys ex P	21.6	32.4	4.4	5.4
Ordin	23.6	30.2	3.5	11.0
Ordin D	22.4	14.6	2.7	6.3
Ordin P	24.7	45,4	4.3	15.7

Table 1. Feedback measures, utterances and words for doctors and patients in three subactvities of patient-doctor consultation.

Why do we find the numbers related to spoken feedback that appear in the table for the different sub-activities, i.e. what do they reflect in terms of influencing factors and typical patterns of interaction in the three phases of the doctor-patient interaction? If we take a look at the physical examination, it can be distinguished by the physical conditions of the examination being different from that of the case history and the ordination and by the focus of action rather than speech. Some feedback characteristics are that the physical examination contains fewer utterances, words and feedback

expressions totally, but a higher share of feedback from the patient and fewer overlapped feedback utterances than the other two sub-activities. The lack of overlap reflects a slower and more structured turn management. There is more focus on instruments and body parts, which also leads to less eye contact between the participants. What do the case history and physical examination have in common? They both have similar purposes, i.e. the doctor collects information, but this is done in different ways, in the case history by listening to the patient' s story with the goal of obtaining information through dialog, and by the doctor's own examination, using observation more than dialog. What do the physical examination and ordination have in common? Both these phases contain considerably more totally overlapped utterances consisting only of feedback from the patients than from the doctor. This shows that the doctor speaks the most in both these sub-activities. This is so even more in the ordination phase than in the physical examination. (In the case history, on the other hand, the patient speaks the most.)

In order to see what characterizes typical exchanges in the physical examination and how this relates to the quantitative findings, we can look at example 1 below. (English translations of the Swedish utterances are given in italics, ( ) encloses quiet speech, < > encloses comments about what happens, [ ] encloses overlap, /// = long pause).

Example 1.

D: ja ska ta de stående också om du ställer dej där borta
I will take it standing too if you stand over there
P: mm (där nej) <patient gets up> de ä så stelt å resa sej <doctor measures blood pressure>
mm (there no) <patient gets up> if is so stiff to get up <doctor measures blood pressure>
D: men du blir inte yr när du reser dej
but you don't get dizzy when you get up
P. joo ibland
yes sometimes
D: just när du [reser dej ur sängen]
right when you [get out of bed]
P: [joo ja kan inte]
resa mej hastigt [utan]tar de
[yes I can't] can't get up fast [but] take it
D: [nähä] <doctor measures blood pressure>*[no] <doctor measures blood pressure>*
D: /// hundrasjutti sjutti
/// a hundred seventy seventy
P: jo då
yes then
D: jaa
yes

In a typical sequence of the physical examination, the doctor has initiative. This means that he does not have to start his utterances with feedback, he can change topic, ask questions and often gives feedback to events. Feedback as reactions to events in the interaction is common in both participants. The sequence can evolve as follows: The doctor asks a yes/no question or another question requiring only a short answer in relation to a specific part of the examination. The patient answers the question and the doctor gives feedback to the answer. However, the examination continues after this feedback and during the silence, the patient quite often makes a short comment. In example 1, the doctor's first three utterances contain no initial feedback. The doctor questions the patient, while he measures his blood pressure and the patient answers. The last two feedback utterances are reactions to the result of the measurement (in this case by the doctor).

Conclusions – The creation and use of multimodal corpora

The collection and use of the GSLC and other related corpora that have been briefly described here have been on-going for more than 30 years. The idea of a variety of social activities in different and mainly naturalistic settings has all the time been in focus and this has made possible a number of observations over the years that have become increasingly relevant for applications related to human-computer interaction, including the design of Embodied Communicative Agents and avatars for use in different types of activities and cultures. The corpus has not originally been collected with these applications in mind, but some of the more recent additions to it have been directly related to this domain. Activities related to different types of service provision, such as information about merchandise, tourism information, travel agency etc. are areas of application, which are represented in the corpus and more of this type of material can be included in the future. The possibility to study how different activity related factors interact is relevant for questions of what can be kept fairly stable and what should be varied in the behavior of human-human like interfaces.

References

Ahlsén, E., Allwood, J. & Nivre, J. (2003). Feedback in Different Social Activities. In P. Juel-Henrichsen (ed.) *Nordic Research on Relations between Utterances. Copenhagen Working Papers in LSP, 3*. 2003, pp.9-37.

Allwood, J. (1976) *Linguistic Communication as Action and Cooperation,* GML 2, Univ of Gothenburg, Dept of Linguistics.

Allwood, J. (1999). "The Swedish Spoken Language Corpus at Göteborg University". In Proceedings of Fonetik 99, *GPTL 81*, Univ of Göteborg, Dept of Linguistics.

Allwood, J. (2000). An Activity Based Approach to Pragmatics". In Bunt, H., & Black, B. (Eds.) *Abduction, Belief and Context in Dialogue: Studies in Computational Pragmatics*. Amsterdam, Benjamins, pp. 47-80.

Allwood, J. (2008). Multimodal Corpora. In Lüdeling, A. & Kytö, M. (eds.) *Corpus Linguistics. An International Handbook*. Mouton de Gruyter, Berlin. 207-225.

Nivre, J. (1999). Modified Standard Orthography, Version 6 (MSO6). Univ of Gothenburg, Dept of Linguistics.

Nivre, J. (2004) Gothenburg Transcription Standard (GTS) V.6.4. Univ of Gothenburg, Dept of Linguistics.