Creating Comparable Multimodal Corpora for Nordic Languages

Costanza Navarretta
University of Copenhagen,
Centre for Language Technology
costanza@hum.ku.dk

Elisabeth Ahlsén
University of Gothenburg
elisabeth.ahlsen@ling.gu.se

Jens Allwood
University of Gothenburg
jens@ling.gu.se

Kristiina Jokinen
University of Helsinki and
University of Tampere
kristiina.jokinen@helsinki.fi

Patrizia Paggio
University of Copenhagen,
Centre for Language Technology
paggio@hum.ku.dk

Abstract

This paper describes the collection and annotation of comparable multimodal corpora for Nordic languages in a project involving research groups from Denmark, Estonia, Finland and Sweden. The goal of the project is to provide annotated multimodal resources to study communicative phenomena, such as feedback, turn-taking and sequencing in the languages involved in the project and to compare these phenomena. Studies so far include verbal expressions, head movements and facial expressions related to feedback.

1 Introduction

Human communication is multimodal, that is it involves speech and communicative body movements, such as facial expressions, head movements, body postures, gaze and hand gestures. All these behaviors occur naturally and have been claimed to be intertwined in communication (McNeill, 2002; Kendon, 2004). Investigating the characteristics of the various modalities and exploiting their interaction in various communicative and cultural situations has been the focus of a number of recent national and international projects and networks, such as AMI, CALLAS, CALO, CHIL, HUMAINE, ISLE, SPONTAL and SSPNET.

The present collaborative Nordic project is in line with these initiatives and involves research groups from Denmark, Estonia, Finland and Sweden. The main goals of the project are the following:

- providing comparative annotated multimodal data;
- using these data to investigate specific communicative phenomena such as feedback and turn-taking;
- developing, extending and adapting models of multimodal interactive communication management that can serve as a basis for interactive systems;
- applying machine learning techniques in order to test the possibilities for automatically recognizing or predicting hand gestures, head movements and facial expressions with different interactive communication functions.

In what follows we first present the data which we have collected so far (section 2), then we discuss the annotation model which is used and briefly describe annotation procedures and available annotations (section 3). In section 4 we present some of the data that have been extracted from the annotations until now and in section 5 we conclude and outline future work.

2 The corpora

The data we work with are video recordings of interactions from a number of social activities. These activities have different purposes and involve different numbers of participants with varying roles, degree of familiarity, position in the room etc. All these aspects can influence the participants' multimodal behaviors.

In the project, we will reuse existing resources, but we are also collecting new comparable data where the social activities recorded in the various languages are the same, and the recording settings are similar. Furthermore, the data are annotated following a common annotation model, which will allow a comparison of data and annotated phenomena. In this paper we will primarily focus on the new data, the annotation model and the studies carried out so far, differing from (Paggio et al., 2010) where we described the various corpora in the project.

The annotated data will be made available for research purposes through the project website (http://sskkii.gu.se/nomco/).

2.1 Corpora of first encounters

First encounters have been studied in intercultural studies (see i.a. Argyle, 1975; Kendon, 1999) because in these data it is possible to study central communicative aspects such as how different cultures deal with varying degrees of familiarity and liking as well as with social status and norms. A comparative multimodal study of first encounters in German and Japanese has been previously conducted in the CUBE-G project (Rehm et al., 2009) with the purpose of generating and testing behavioral models for virtual agents in the two cultures.

Our comparable corpora of first encounters are studio-recorded conversations and are presently available for Swedish and Danish, but a corresponding corpus for Finnish is being collected.

The first encounters corpora are interesting because Nordic cultures are generally regarded as relatively similar, and our data will provide us with empirical evidence for similarities as well as differences in a first-meeting scenario.

The interactions in both the Swedish and Danish first encounters corpora involve two subjects who are standing in front of a light background. The participants were instructed to get to know each other in a short interaction, as they might do at a party or a reception. After the recording they answered a questionnaire about their reactions to both the interlocutor and the interaction setting.

Additional first encounter data has also been collected to compare Swedish and Danish data with data from more distant cultures as well as intercultural communication situations. A number of Chinese-Chinese interactions in Chinese and a number of Swedish-Chinese interactions in English have been recorded. There is also a comparable dataset of first encounter recordings in German, recorded in Austria (Csokor, 2010).

The Swedish first encounters corpus

The Swedish first encounters corpus consists of 39 videorecordings of interactions in Swedish, each approximately 8-10 minutes long, in total about 5 hours. In terms of gender, 19 of the interactions are male-female, 11 are male-male and 9 are female-female. The age range is 19 to 34 with a mean age of 25.

The Chinese corpus consists of 6 videorecorded Chinese-Chinese first encounter interactions in Chinese, in total about 1 hour (with a mean duration about 10 minutes), containing 3 male-female, 2 male-male and 1 female-female encounters.

The intercultural Swedish-Chinese corpus contains 10 videorecorded Swedish-Chinese first encounters in English, in total 1½ hour (mean duration about 9 minutes). Four of these interactions are male-female, 3 are male-male and 3 are female-female.

The Danish first encounters corpus

The Danish corpus of first encounters consists of approximately one hour of video-recordings, comprising 12 interactions of approximately 5 minutes each and involves 12 speakers, six males and six females, all between 21 and 36 years old. Each speaker participated in two interactions, one with a male and one with a female.

The answers to the questionnaire show that the participants were in general positive about the interaction. They report that they felt well-liked and free to express their opinions. They judged the conversations as interesting although they were aware that the setting was not completely natural (Paggio et al., 2010).

The corpus has been orthographically transcribed and a set of gestures (i.e. communicative body movements) have been annotated as it will be described in section 4.

2.2 Corpora of group interactions

Besides two-person dialogues we have also video recordings of multiparty interactions. Some of these recordings have been collected under this project, while others were already available to the involved research groups.

When the number of participants increases, interaction management becomes more complex as the responsibility of smooth communication is divided among all of them: interlocutors have both pair-wise and shared interactions, and some of them can simply act as onlookers and not take an active role in the activity. The use of multimodal means in communication is thus expected to differ from two-party dialogues, and the observational studies in conversation analysis and sociolinguistic studies have indeed shown how different non-verbal signals and spatial proximity work in the coordination and control of group interactions (Goffman 1963; Hall 1966; Kendon 1990).

The group meeting corpora aim to provide comparable data for studying conversational activity in multiparty communications. However, we want to emphasise that our current group meeting corpora do not form a similar uniform set of corpora across the languages as the first encounters. We thus do not aim at the "sameness" of the group meeting corpora but regard similarity as an abstract concept which requires semantic interpretation of the actual context: similarity can be loosely characterized in terms of the number of participants, the activities that they are involved in and the view-points from which the events are looked at. Our goal is thus to collect a large variety of group meetings so as to provide as wide a basis for conversations studies as possible, and thus unravel comparable features of the group communication. We assume that this can be best achieved by using the same annotation scheme for the various group meeting corpora. In our case, we have used the MUMIN annotation scheme (section 3).

A Swedish corpus of group meetings in different social activities, which is a subcorpus of the Gothenburg Spoken Language Corpus (GSLC) (Allwood et al., 2000) is available for use in the project. The corpus consists of 82 video- or audiorecorded meetings of in total 122 hours, containing 636 268 word tokens, according to the GTS 6.4 Transcription Standard (Nivre, 2004). The corpus contains arranged and naturally occurring discussions, formal and informal meetings, and dinner discussions. The number of speakers range between 2 and 12 per recording, with a mean of 7-8 speakers. The total number of speakers is 502, with a total number of 255 males, 224 females and 23 participants unidentified for gender.

A Danish corpus of informal meetings between people that are well acquainted (friends or family members) are being annotated according to the annotation model described in section 3. The videos are collected and transcribed by the University of South Denmark, and will be available through the Danish CLARIN homepage[1].

They involve varying numbers of speakers of different age who are recorded while talking informally. In all the recordings the participants are sitting around a sofa table at private homes.

The Estonian corpus of group interactions contains two 30 minutes long conversations among three participants. The participants perform according to their designated roles in scenarios which concern the planning and inspection of a new school building. Despite the acted scenarios, the participants behave fairly naturally.

The Finnish group interactions consist of card-playing interactions among four participants and conversations between a Finnish teacher and an immigrant student. The Finnish interactions are collected by Minna Vanhasalo.

3 The annotation model

Data are annotated according to a common model which is an adaptation of the MUMIN model (Allwood et al. 2007). This model has been used to annotate communicative non-verbal behavior and its relation to speech in various languages, e.g. Greek (Koutsombogera et al. 2008), Danish (Paggio and Navarretta, 2010; Navarretta and Paggio, 2010), Estonian (Jokinen and Ragni, 2008) and Japanese (Jokinen et al. 2009). The model describes the shape and the communicative function of gestures, including head movements, facial expressions, hand gestures and body postures in terms of pre-defined behavior attributes and values.

The main focus in the model, according to Allwood et al. (2007), is on the communicative function of gestures. The description of the shape of gestures provided in the model is coarse-grained, but can be refined according to specific requirements in different studies.

[1] https://infra.clarin.dk/clarindk/forside.jsp.

The communicative functions which have been dealt with in the MUMIN model are feedback, turn management and sequencing. Furthermore, each gesture can be assigned a semiotic type following Peirce's (1931) classification, which distinguishes between indexical, iconic and symbolic signs.

Gestures can also be assigned a value indicating the attitude they show[2] and can be connected to a word or more words if the annotators judge that there is a semantic relation between the gestures and the words.

Gestures can be multifunctional, thus several categories can be assigned to the same gesture, e.g. a nod can indicate feedback-giving and turn taking at the same time.

We have slightly modified the MUMIN model to fit the project's specific goals, and the granularity of the attributes might change depending on the phenomena we are focusing on. For example, we have simplified the linking of gestures to words using a single link type, called *MMRelationSelf*, which connects a gesture produced by a participant to the word(s) produced by the same participant, while in MUMIN four relations were recognized following (Poggi and Caldognetto, 1996).

As an example of the annotation categories used in the project to describe the shape of gestures, we show the values and attributes defined for head movements in table 1. These gestures are annotated with two attributes: the first attribute indicates the type of movement while the second one records whether a movement occurs once (*Single*) or more times (*Repeated*).

Behavior attribute	Behavior value
HeadMovement	Nod
	Tilt
	Jerk (Up-nod)
	Shake
	Waggle
	SideTurn
	HeadBackward
	HeadForward
	Other
HeadRepetition	Single
	Repeated

Table 1: Attributes and values for head movements

Table 2 contains the attributes and values accounting for the communicative function of feedback. The first attribute in the table, *FeedbackBasic*, indicates whether there is feedback or not. The second attribute, *FeedbackDirection*, describes whether a subject is giving or asking for feedback. The last attribute, *FeedbackAgreement*, is used when an interaction participant agrees or disagrees with what stated by the interlocutors.

Behavior attribute	Behavior value
FeedbackBasic	Contact/ Perception/ Understanding(CPU)
	Other (C, CP)
FeedbackDirection	Give
	Elicit
	Give-Elicit
FeedbackAgreement	Agree
	Disagree

Table 2: Attributes and values for feedback

4 The Swedish Annotated Data

In what follows we describe the Swedish corpora currently annotated and the procedures used to perform the annotations. The Swedish corpora have all been transcribed using the GTS (Gothenburg Transcription Standard (Nivre, 2004) and MSO 6 (Modified Standard Orthography) for the Swedish data (Nivre, 1999).

4.1 The first encounters data

So far, 13 of the Swedish first encounters are fully transcribed and checked by an independent transcriber.

Coding of communication management oriented gestures (head gestures, facial expressions and hand gestures) will be done using a modified version of the MUMIN coding schema.

A small corpus of Swedish-Swedish, Chinese-Chinese and Swedish-Chinese interactions has been transcribed and given a preliminary coding of feedback related gestures.

For a number of the Swedish recordings, some of the basic prosodic features of feedback expressions (pitch, F0 shapes, timing and duration) have been analyzed with the purpose of investigating the relation between prosodic features of feedback and head movement as feedback. Experimental and naturalistic feedback data is also being analyzed with respect to emotional and attitudinal features.

[2] The list of attitudes and emotions is open-ended.

A study focusing on repeated head movements (head nods and head shakes) and the speech co-occurring with them in the Swedish first acquaintance corpus showed that the main function of such repeated head movements is communicative feedback. This is also the most frequent function of the speech co-occurring with the head movements. However, there is mostly no 1-1 relation between repetition in head movement and vocal words. Repeated head movements are more often accompanied by single than repeated words. Both repeated head movements and repeated vocal words can also occur without accompaniment in the other modality. Also in these cases, the most frequent function for the head movements is communicative feedback. However, the most frequent function of repeated words without accompaniment in the other modality is own communication management. Frequent functions of repeated head movements, besides feedback, are emphasis, self-reflection, citation, self-reinforcement and own communication management.

Other findings in the study are that affirmative repeated head nods mostly start with an upward movement and involve two repetitions (Boholm & Allwood, 2010).

First acquaintance recordings of 4 Chinese-Chinese, 4 Swedish-Swedish and 8 Chinese-Swedish recordings, where the Chinese-Swedish interactions took place in English, were analyzed.

Some of the preliminary results are (i) that in both the Swedish and the Chinese interactions, unimodal vocal feedback is more common than unimodal gestural feedback, (ii) that both the Swedes and the Chinese use gestural feedback more multimodally than unimodally. Some differences are that the Chinese do not have a special word which exactly corresponds to yes in vocal feedback. The most common vocal feedback is “n”. In gestural feedback, they use more laughter, “gaze around”, gaze sideways and covering their mouth with hands. The Swedes use more vocal “m” and ingressive feedback sounds and in gestural feedback only the Swedes have up-nods and tilts. Both Swedes and Chinese use more feedback gestures when they speak English in the intercultural interactions (Allwood & Lu, 2010).

4.2 The group interaction data

Parts of the Swedish group interaction data corpus have been coded, for example for communicative acts, main addressee and group decision processes in previous studies. Gestures are only coded when judged to be especially important for the interaction by the transcribers.

5 The Danish annotated data

The Danish data annotated so far are described below.

5.1 First encounters data

The Danish corpus of first encounters has been transcribed in PRAAT (Boersma and Weenik, 2009) following the guidelines provided by Grønnum (2006) for the DanPASS project. The transcriptions are orthographic and, in addition, contain information on word stress, pauses and filled pauses. They have been made by a coder and checked by a second coder and consist of approx. 17500 tokens, of which a 16150 are running words, 550 are onomatopoeic expressions such as “hmm” and “øh” and 800 are pauses.

The transcriptions are imported into the ANVIL tool (Kipp, 2004), which is used to create the multimodal annotations.

Three coders have annotated the communicative body movements and their relation to speech following a common annotation manual. So far, head movements and face expressions have been annotated, together with the communicative function of feedback and the links connecting gestures to words in the orthographic transcription.

The annotation procedure has been the following: each video is annotated by one coder and the annotation is then revised by a second coder. Disagreements are discussed and an agreed upon annotation version is created. In cases where it is not possible to reach an agreement, a third coder resolves the disagreement.

Two inter-coder agreement experiments have been run in order to test to which extent the three coders identified the same gestures and assigned the same categories to the recognized gestures. The first experiment was run in the beginning of the annotation process, and the second one when half of the data had been annotated. In both experiments a video was annotated independently by the three annotators and then the annotations were automatically compared in ANVIL, which tests both gesture segmentation and category assignment.

The results of the latest experiment in terms of Cohen’s kappa (Cohen, 1960) show an agreement in-between 60-80%. The agreement for head movements is in general higher than for face expressions. The highest disagreement values are mainly due to disagreement in the segmentation of facial expressions. Deciding where exactly a smile starts and ends, for example, is often more difficult than doing the same for a side turn.

The intercoder agreement figures improved for nearly all categories in the second experiment, partly because the coders had achieved more experience, partly because the annotation manual had been revised establishing clearer distinction criteria for problematic categories. The final agreement scores are in line with those achieved in similar annotation tasks, e.g. (Jokinen et al., 2008).

So far the first 5 annotated videos have been analyzed. The gestures annotated in the first five videos are approximately 2000, of which 40% have been judged to have a feedback function.

The direction of most feedback gestures is *Give* and there are only few feedback eliciting gestures. This is probably due to the type of social activity, but comparison with videos belonging to other types of activities will confirm this hypothesis.

The most used behavior for the expression of feedback is *HeadMovement* (61%), followed by *Face* (28%) and *Eyebrows* (11%). However, if we look at specific movement and expression types, we see that *Smile* is the type most often used to give feedback (17%), followed by *RepeatedNod* (13%). The frequency of all other types in conjenction with feedback is below 10%.

A comparative study of feedback in the Danish first encounters corpus and in similar Japanese data is being carried out aiming to investigate differences and similarities in the way Danish and Japanese people communicate feedback in this type of social interaction (Paggio et al., forthcoming).

5.2 The informal meetings data

So far, four videos with two and three participants have been orthographically transcribed in PRAAT and then imported into ANVIL. The transcriptions of these interactions consist of approx. 5,300 running words. The multimodal annotations comprise facial expressions, head movements, hand gestures and body postures. The following communicative functions have been included: feedback, turn management, sequencing and deixis. The multimodal annotations comprise the following types of communicative body movements: 110 facial expressions, 1,051 head movements, 368 hand gestures and 89 body postures. How often these behaviors have been judged to express feedback varies. Thus, a feedback function is assigned to 58% of the facial expressions, 60.5% of the head movements, 7.5% of the hand movements and 29% of the body postures.

6 The Estonian/Finnish data

About 20 minutes of the Estonian group conversations (10 minutes of each conversation) have been annotated using the MUMIN annotation scheme, which was adapted to three person interactions. The data has been used in comparing Estonian and Danish dialogue strategies (Jokinen et al., 2008), and in investigating meta-gesturing or conversation control, e.g. stand-up gestures (Jokinen and Vanhasalo, 2009).

Annotations were produced in several passes with kappa agreement ranging between 40-80%. The final annotations comprise 151 utterances, 657 facial display elements, 442 hand gesture elements, and 380 body posture elements. Facial display elements make about 44% of all non-verbal communication, confirming the importance and frequency of facial expressions in communication. The data indicate a clear correlation between speaking and non-verbal communication: the participant who talks most (produce most utterances) also seems to produce most nonverbal behaviors. Furthermore, facial displays seem to be evenly distributed while there are individual differences in the use of hand gestures and body posture.

The Finnish card-playing conversations have been analyzed with focus on gesturing. Salo (2002) studied pointing gestures as deictic elements but emphasized that the use of pointing gestures is richer and more complicated. In line with this research Jokinen and Vanhasalo (2009) show how pointing gestures also function as an effective means to control and coordinate the dialogue.

7 Conclusions and future work

In the paper we have described the first phase of the creation of comparable multimodal annotated corpora for Danish, Estonian, Finnish and Swedish. These corpora comprise video recordings of different types of social activities, such as the first encounter interactions, recorded in the same way for the different languages, but also group meetings in different contexts, which provide a rich variation of interaction data. We have also provided a preliminary analysis of how feedback is expressed through gestures and speech in the first encounter data, and how they compare with similar data for Chinese and Japanese. Further coding and analysis of the corpora will provide a basis for additional studies of multimodal interactive communication management on feedback, but also on other phenomena such as turntaking and sequencing.

Acknowledgments

This work is done under the NOMCO project, which is funded by the NORDCORP programme under the Nordic Research Councils for the Humanities and the Social Sciences (NOS-HS).

References

Allwood, Jens, Maria Björnberg, Leif Grönqvist, Elisabeth Ahlsé,n and Cajsa Ottesjö, (2000). The Spoken Language Corpus at the Dept of Linguistics, Göteborg University. *FQS - Forum Qualitative Social Research*, Vol. 1, No. 3. - Dec. 2000, pp 22.

Allwood, Jens, Loredana Cerrato, Kristiina Jokinen, ostanza Navarretta, and Patrizia Paggio (2007). The MUMIN coding scheme for the annotation of feedback, turn management and sequencing phenomena. In J.-C. Martin, P. Paggio, P. Kuehnlein, R. Stiefelhagen, and F. Pianesi (Eds.), *Multimodal Corpora for Modelling Human Multimodal Behaviour*, Volume 41 of *Special issue of the International Journal of Language Resources and Evaluation*, pp. 273–287. Springer.

Allwood, Jens and Jia Lu (2010). Chinese and Swedish multimodal communicative feedback. Paper presented at the 5:th International Conference on Multimodality. University of Technology, Sydney. Dec 1-3, 2010.

Boersma, Paul and David Weenink (2009). Praat: doing phonetics by computer (version 5.1.05). Retrieved May 1, 2009, from http://www.praat.org/.

Boholm, Max and Jens Allwood (2010). Repeated head movements, their function and relation to speech. In M. Kipp et al. (Eds.) *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC 2010) Workshop Multimodal Corpora: Advances in Capturing, Coding and Analyzing Multimodality*, Malta, 17 May 2010, http://www.lrec-conf.org/proceedings/lrec2010/index.html.

Cohen, Jacob A. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37-46.

Csokor, Laurentz (2010). *Feedback in Mixed-Sex Conversation Settings.* Master Thesis at the Faculty of Life Sciences, University of Vienna, Austria.

Grønnum, Nina (2006). DanPASS - A Danish Phonetically Annotated Spontaneous Speech Corpus. In N. Calzolari, K. Choukri, A. Gangemi, B. Maegaard, J. Mariani, J. Odijk and D. Tapias (Eds.), *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC-06)*, Genova, Italy, May.

Goffman, Erwin (1963). *Behaviour in public places: notes on the social order of gatherings*.The Free Press, New York.

Hall, Edward T. (1966). *The Hidden Dimension: man's use of space in public and private*, New York: Doubleday.

Kendon, Adam (1990). Spatial organization in social encounters: the F-formation system, In Kendon, A: *Conducting Interaction: Patterns of behavior in focused encounters*, Studies in International Sociolinguistics, Cambridge University Press.

Jokinen, Kristiina, Costanza Navarretta and Patrizia Paggio (2008). Distinguishing the communicative functions of gestures. In *Proceedings of the 5th Joint Workshop on Machine Learning and Multimodal Interaction*, 8-10 September 2008, Utrecht, The Netherlands.

Jokinen, Kristiina and Minna Vanhasalo (2009). Stand-up Gestures – Annotation for Communication Management. In *Proceedings of the NODALIDA 2009 Workshop Multimodal Communication: from Human Behaviour to Computational Models.* Odense, Denmark, May 2009, pp. 15-20.

Koutsombogera, Maria, Lida Touribaba and Harris Papageorgiou (2008) Multimodality in Conversation Analysis: A Case of Greek TV Interviews. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC 2008) Workshop on Multimodal Coorpora from Models of Natural Interaction to Systems and Applications*, Marrakesh, May 2008, pp. 12-15.

Navarretta Costanza and Patrizia Paggio. Classification of Feedback Expressions in Multimodal Data. *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics* (ACL 2010), Uppsala, Sweden, Juli 11-16, 2010, pp. 318-324.

Nivre, Joakim (1999). *Modifierad Standardortografi, Version 6 (MSO6)*. Department of Linguistics, University of Gothenburg.

Nivre, Joakim (2004). *Göteborg Transcription Standard. (GTS) V. 6.4*. Department of Linguistics, University of Gothenburg.

Paggio, Patrizia, Jens Allwood, Elisabeth Ahlsén, Jristiina Jokinen & Costanza Navarretta (2010). The NOMCO Multimodal Nordic Resource – Goals and Characteristics. In Calzolari, N, Choukri, K. Maegaard, B, Mariani, J., Odijk, J, Piperidis, S, Rosner, M. & Tapias, D. (Eds.) *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)* Valetta, Malta May 19-21. ELRA. http://www.lrecconf.org/proceedings/lrec2010/index.html

Paggio, Patrizia, Kristiina Jokinen and Costanza Navarretta (forthcoming). Head movements, facial expressions and feedback in first encounters interaction. To appear in the *Proceedings of HCI International 2011*, Orlando Florida, July 9-14.

Paggio Patrizia and Costanza Navarretta. Feedback in Head Gestures and Speech. In M. Kipp et al. (Eds.) In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC 2010) Workshop Multimodal Corpora: Advances in Capturing, Coding and Analyzing Multimodality*, Malta, May 17 2010, pp. 1-4.

Peirce, Charles S. (1931). *Collected Papers of Charles Sanders Peirce, 1931–1958*, 8 vols. Edited by C. Hartshorne, P. Weiss and A. Burks. Cambridge, MA: Harvard University Press.

Poggi, Isabella and Emanuela Magno Caldognetto (1996). A score for the analysis of gestures in multimodal communication. In *Proceedings of the Workshop on the Integration of Gesture and Language in Speech*. Applied Science and Engineering Laboratories. L. Messing, Newark and Wilmington, Del, pp. 235–244.

Rehm, Matthias, Elisabeth Andre, Nikolaus Bee, Birgit Endrass, Michael Wissner, Yukiko Nakano, Afia Akhter Lipi,Toyoaki Nishida, and Hung-Hsuan Huang (2008). Creating Standardized Video Recordings of Multimodal Interactions across Cultures. In Kipp et al (eds.) *Multimodal Corpora. From Models of Natural Interaction to Systems and Applications*. LNAI 5509. Springer, pp.138–159.

Salo, Minna (2002). Puhujan eleiden kuvailu: eleiden muoto ja merkitys sekä ajoitus. ("Description of the speaker's gestures: the form and meaning of gestures and their timing") Master's Thesis. Department of Finnish and General Linguistics. University of Tampere.