Multimodal Behaviour and Feedback in Different Types of Interaction

Costanza Navarretta*, Patrizia Paggio*±
University of Copenhagen*, University of Malta±
Njalsgade 140, build. 25, 4th floor
2300 Copenhagen S – DK*
Msida MSD 2080, Malta±
E-mail: costanza@hum.ku.dk, paggio@hum.ku.dk

Abstract

In this article, we compare feedback-related multimodal behaviours in two different types of interactions: first encounters between two participants who do not know each in advance, and naturally-occurring conversations between two and three participants recorded at their homes. All participants are Danish native speakers. The interactions are transcribed using the same methodology, and the multimodal behaviours are annotated according to the same annotation scheme. In the study we focus on the most frequently occurring feedback expressions in the interactions and on feedback-related head movements and facial expressions. The analysis of the corpora, while confirming general facts about feedback-related head movements and facial expressions previously reported in the literature, also shows that the physical setting, the number of participants, the topics discussed, and the degree of familiarity influence the use of gesture types and the frequency of feedback related expressions and gestures.

Keywords: multimodal corpora, dyadic and group interactions, feedback

1. Introduction

This paper compares verbal and non-verbal feedback expressions in two Danish multimodal corpora. Although the setting and the communicative situation of the two corpora are quite different, both corpora have been annotated according to the same multimodal annotation scheme. This allows us to compare how feedback is expressed by means of linguistic expressions and gestures in different communicative situations. In this study we use gestures as a general term comprising non-verbal behaviours in general, although our analysis focuses on head movements and facial expressions.

The expression of feedback through head movements has been investigated in interactions in different languages. Studies have dealt with various aspects such as the relation between head movement type and type of communicative function (Argile 1975, McClave 2000), culture-specific characteristics of feedback-related head movements (Maynard 1987, Cerrato 2007, Jokinen et al. 2008, Rehm et al. 2009, Lu et al. 2010, Jokinen and Allwood 2010, Paggio and Navarretta 2011), the relation between verbal feedback expressions, prosody, facial expressions and head movements (Paggio and Navarretta 2010, Navarretta and Paggio 2010), the prediction of feedback head gestures from linguistic features and eye gaze (Fujie et al. 2004).

In the present study we look at how feedback is expressed by gestures and speech in the same language, but in two different communicative situations involving people who have different degrees of familiarity. The effect of familiarity on speech flow in Japanese is discussed e.g. in Campbell (2007), where it is claimed that as familiarity between speakers increases over time, so does the speech flow. In other words, Japanese speakers who become familiar tend to be less silent when chatting to each other. We wanted to see whether we could observe a similar tendency in a multimodal environment. Another interesting issue that our corpora allow us to investigate is whether multimodal behaviour, in particular the way speakers give each other feedback, changes with the number of participants while keeping the situation type stable. The participants in the interactions in our corpora are Danish native speakers. In the remainder of the paper, first we present our corpora (Section 2), and we describe the annotation methodology (Section 3). Then, we analyse and discuss the feedback data from the two corpora (Section 4) and finally, we conclude (Section 5).

2. The corpora

The first corpus is the Danish NOMCO corpus of first encounters (Paggio et al. 2010; Navarretta et al. 2011). It contains 12 dyadic (two participants) meetings of the duration of approximately 5 minutes each. The two participants in each meeting do not know each other in advance and have been told to try to get acquainted through the conversation. All the participants are university students or people with a university education aged between 21 and 36. Half of them are males and half are females. Each person participated in two meetings, one with a person of the same gender and one with a person of the opposite gender.

The interactions were video-recorded in a studio. The participants were standing in front of each other while being engaged in the conversation. The participants were recorded by three cameras, one taking a panorama view of them from the side (Figure 1), and the other two taking mid shots of each of them (Figure 2).

A more detailed description of the corpus collection can be read in Paggio and Diderichsen (2010) and Paggio and Navarretta (2011).

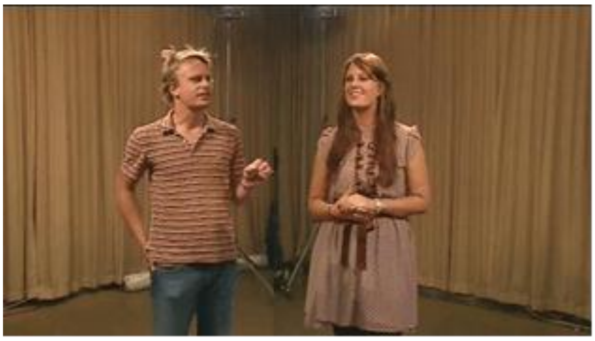

Figure 1: A combined frontal view of one First encounter

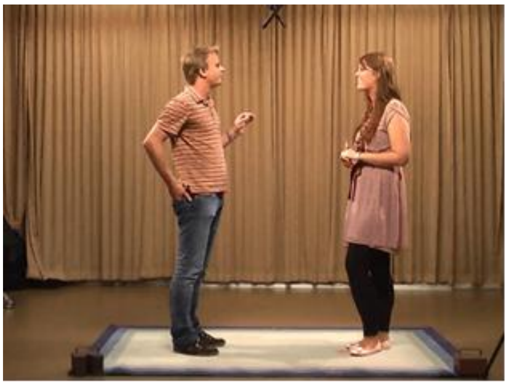

Figure 2: Panorama view from a NOMCO first encounter interaction

The second corpus used in this study consists of video-recordings of spontaneous interactions between two or three persons who are well acquainted. The participants were recorded in their private homes sitting around a table, drinking, eating and talking freely about various subjects comprising soccer, family relations and the current economic crisis. The recordings are part of a larger database, the MOVIN database, and have been collected and transcribed according to conversation analysis (CA) conventions by researchers at the University of Southern Denmark (MacWhinney and Wagner 2010). Part of the MOVIN video recordings and CA transcriptions are freely available from the talkbank homepage[1]. In the present study, four interactions are included for a total of 25 minutes. Five women, all aged 50+ were involved in the interactions

Figure 3 shows a snapshot from one of the triadic interactions.

3. The annotations

Both corpora are transcribed orthographically and time-aligned at the word level. In the first encounters word stress and pauses are included in the transcription. In the conversations between acquainted persons the orthographical transcription and alignment at the word level was added reusing the existing CA transcriptions (Navarretta 2011b).

In both corpora expressions such as *øhm* and *hm* as well as laughter have been transcribed.

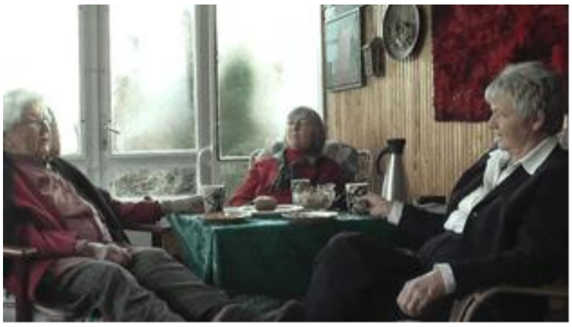

Figure 3: A snapshot from a triadic MOVIN interaction

The transcriptions of both corpora were imported into the ANVIL tool (Kipp 2004), which was used to annotate the gestures and their relation with speech (Paggio and Navarretta 2011, Navarretta 2011a, Navarretta2011b).

In figure 4 a print screen from the ANVIL tool with a dyadic MOVIN interaction is shown.

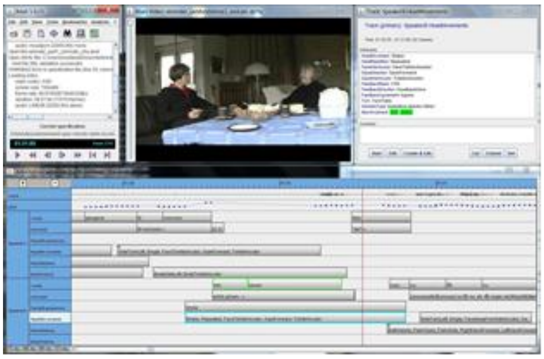

Figure 4: Print screen from the ANVIL tool

The multimodal annotations follow the MUMIN annotation scheme (Allwood et al. 2007), which provides pre-defined feature-value pairs accounting for different gesture shapes and functions. Here, we focus on feedback-related words, head movements and facial expressions.

Table 1 shows the features and values used to annotate the shape of head movements and facial expressions used in this study.

Head movements are described with two features, one indicating the type of movement and the other explaining whether the movement is performed once or more times.

Facial expressions include here only a feature for the general expression of the face.

Feedback is annotated with three features, Basic, Direction and Agreement (Table 2).

[1] http://talkbanken.org

Gesture feature	Gesture value
HeadMovement	Nod, Tilt, Up-nod, Shake, Waggle, SideTurn, HeadBackward, HeadForward, Other
HeadRepetition	Single, Repeated
Face	Smile, Laughter, Scowl, FaceOther

Table 1: Head movement and facial expressions

Feedback feature	Feedback value
Basic	CPU (Contact, Perception, Understanding) Other (Contact, Perception) or (Contact)
Direction	Give, Elicit, GiveElicit
Agreement	Agree, Disagree

Table 2: Feedback features

The first feature indicates whether there is feedback. In this study the only relevant value is CPU, which indicates that the gesturer's behaviour shows signs of contact, perception and understanding. The second feature says whether the participant is giving, eliciting, or giving as well as eliciting feedback. Finally, the Agreement feature indicates whether the participant agrees with the interlocutor or not.

4. Analysis of the annotated data

The present analysis is based on 10 of the videos from the first encounters corpus (about 50 minutes in total), 2 dyadic interactions from the MOVIN corpus (about 10 minutes) and 3 triadic interactions still from MOVIN (about 25 minutes in total).

4.1 Feedback-related words

There are 13,735 tokens (words, laughs and *øhm* expressions) in the first encounters corpus, while there are 2216 tokens in the dyadic MOVIN data, and 3170 in the triadic interactions.

In these study, we account for the most commonly occurring expressions related to feedback, in other words *yes* and *no* expressions, yes/no expressions henceforth. These expressions comprise the following words: *ja* (yes), *jo* (yes), *jamen* (well), *vel* (well), *okay*, *nej* (no), *næh* (no).

In the first encounters corpus, there are 1051 yes/no expressions, corresponding to 0.78% of the tokens. In the dyadic MOVIN data there are 152 yes/no expressions (0.68% of the tokens) and in the triadic MOVIN data there are 376 (1.18% of the tokens). Thus, there are more yes/no expressions in the first encounters than in the dyadic conversations between well-acquainted speakers, but there are significantly more yes/no expressions in the triadic interactions than in the dyadic ones.

There are probably two explanations for these differences. On the one hand, people who are getting acquainted may be prone to providing linguistic feedback to each other more than speakers who know each other, as can be seen from the two sets of dyadic interactions. The difference between dyadic and group interactions, on the other hand, may partly be explained by the fact that sometimes in triadic interactions two participants express feedback both verbally and through gestures simultaneously. The difference may also be partly due to the fact that yes/no expressions are also used to regulate turn taking, which is obviously more complex in the group situation.

The most frequently occurring feedback word is *ja* in all three interaction types.

4.2 Feedback-related Head Movements

The number of head movements annotated in the three datasets and their frequency per second are in Table 3.

These figures indicate that the participants in the first encounters moved their heads less frequently than those in the interactions between well-acquainted people, and that well-acquainted people produced head movements with similar frequency regardless of the number of speakers. If number of gestures is taken to be indicative of interaction flow, these figures would seem to parallel Campbell's claims in the gestural modality. Well-acquainted people both gesture and talk more fluently than people who do not know one another. The difference in gesture frequency, however, is probably also determined by the different physical settings, by the participants' age and the discussed topics.

Gesture	Nomco	no/sec	Movin2	no/sec	Movin3	no/sec
Nod	582	0.18	36	0.10	146	0.20
Tilt	427	0.13	21	0.05	34	0.04
SideTurn	356	0.11	126	0.33	265	0.35
Head Forward	286	0.09	44	0.12	62	0.08
Shake	269	0.08	41	0.11	32	0.04
Head Backward	207	0.06	32	0.08	32	0.04
HeadOther	161	0.05	47	0.12	146	0.19
Jerk	133	0.04	4	0.01	6	0.01
Waggle	67	0.02	2	0.01	0	0
Head Total	2488	0.77	353	0.91	723	0.95
Smile	544	0.17	31	0.08	20	0.03
Laughter	199	0.06	21	0.05	21	0.03
Other	53	0.02	4	0.01	1	0.01
Scowl	5	0.01	0	0	1	0.01
Face Total	801	0.25	56	0.14	43	0.05

Table 3 Head Movements in the Corpora

In general, participants in the first encounters moved their heads less than the participants in the MOVIN data. On the contrary, the first group used more facial expressions than the second group. The participants in the triadic conversations were those that moved their heads more frequently, but produced fewer facial expressions.

In all three corpora, Nod is the most frequently occurring head movement and Smile is the facial expression which is produced more often by the involved subjects. The frequency of the other head movements in the three corpora is not the same, and it is partially influenced by the physical settings in the three interaction types.

Table 4 shows the subset of the head movements that are used to express feedback.

The data in this table confirm that nods and, to a lesser extent, shakes are frequently used head movements in the expression of feedback (McClave 2000). However, other head movements are also relevant for feedback confirming previous studies (Paggio and Navarretta, 2011).

Also in this case, the most frequently occurring head movement is Nod in all three data sets, while other types of head movement occurred with different frequency in the three corpora.

Feedback Gesture	Nomco	no/sec	Movin2	no/sec	Movin3	no/sec
Nod	434	0.13	30	0.08	144	0.20
Tilt	137	0.04	8	0.02	16	0.02
Shake	115	0.04	22	0.06	19	0.03
Head Backward	114	0.04	6	0.02	7	0.01
Head Forward	109	0.03	17	0.04	18	0.02
Up-nods	103	0.03	3	0.01	6	0.01
SideTurn	89	0.03	82	0.21	198	0.26
Waggle	19	0.01	0	0.01	0	0
Other	42	0.01	14	0.04	62	0.08
Head Total	1265	0.39	182	0.47	476	0.62
Smile	258	0.09	28	0.07	15	0.02
Laughter	92	0.03	17	0.04	9	0.01
Other	39	0.01	1	0.01	1	0.00
Scowl	0	0	0	0	0	0
Face Total	409	0.13	46	0.12	25	0.03

Table 4 Feedback Related Head Movements and Facial Expressions in the Corpora

Again, feedback-related head movements are more frequent in the MOVIN corpora than in the NOMCO corpus. The frequency of feedback-related gestures in the dyadic interactions in both MOVIN and NOMCO is not as high as in the triadic interactions.

As it was the case for feedback words, also in the case of head movements it is expected that three persons produce more gestures than two because they often give feedback to each other simultaneously.

The influence of the interaction setting is also, again, evident. There are more tilts in the NOMCO corpus, where speakers are facing each other, while side turns are much more frequent in the MOVIN data, where participants are sitting around a table. The fact that Up-nods are more frequent in the first encounters than in the corpora with well-acquainted participants, on the other hand, does not seem to depend on the setting. Given the relative infrequency of the movement, however, we will not try to give an explanation for the difference.

Facial expressions occur in feedback, but less frequently than head movements. Smile and Laughter are the most frequently used facial expression, both in general and as feedback signs in all three data sets.

The frequency of feedback-related facial expressions is the same in the dyadic interactions independently of the degree of acquaintance of the participants. On the other hand, participants in the triadic interactions used facial expressions less frequently than in the dyadic data to give or elicit feedback.

A possible explanation can be that participants in the triadic interactions did not face one another directly, and thus they did not have eye contact in the same way as the participants in the dyadic interactions. The lack of direct eye contact is likely to have meant a less pronounced production of facial expressions.

The content of the conversations might also have played a role. However, this is an aspect we have not studied in detail, and which it would take a larger number of different conversations to investigate.

5. Conclusions

We have compared the transcriptions and the head movement annotations in Danish multimodal corpora of interactions between participants with different degrees of familiarity and in different settings. In one of the settings there are both dyadic and triadic interactions. We have reported on data concerning verbal feedback, head movements and facial expressions in general, and specifically with a feedback function.

The number of videos that record interactions between well-acquainted subjects is limited, and therefore the results we get from those must be considered indicative. That said, the data suggest that subjects that are familiar with one another move their heads more than people who do not know each other: this effect seems parallel to the increased flow of speech in relation to familiarity that has been observed in the literature.

Our data also indicate that subjects in dyadic interactions use their facial expressions more frequently than those in the triadic interactions, probably due to the fact that eye contact is easier in the former context than in the latter.

Our comparative data, while confirming general facts about feedback-related head movements previously reported in the literature, also shows that the physical setting, the number of participants, the topics discussed, and the degree of familiarity influence the use of gesture types and the frequency of feedback related expressions and gestures.

6. Acknowledgements

The work described in this paper is funded by the NORDCORP program under the Nordic Research Councils for the Humanities and the Social Sciences (NOS-HS) and by the Danish Research Council for the Humanities (FKK). The Danish NOMCO corpus has been annotated by Sara Andersen, Josephine B. Arrild, Anette Studsgård and Bjørn N. Wesseltolvig.

7. References

Allwood, J., Cerrato, L., Jokinen, K., Navarretta, C. and Paggio, P. (2007) The MUMIN Coding Scheme for the Annotation of Feedback, Turn Management and Sequencing. In J. C. Martin et al. (eds) *Multimodal Corpora for Modelling Human Multimodal Behaviour*. Special issue of the International Journal of Language.

Argyle, M. (1975) *Bodily Communication*. Methuen & Co. Ltd. London.

Campbell, Nick (2007) Individual Traits of Speaking Style and Speech Rhythm in a Spoken Discourse. In *Proceedings of the COST 2102 Workshop*, 107-120.

Cerrato, L. (2007). *Investigating Communicative Feedback Phenomena across Languages and Modalities*. PhD. Thesis Stockholm, KTH, Speech and Music Communication.

Fujie, S., Y. Ejiri, K. Nakajima, Y Matsusaka, and T. Kobayashi. (2004). A conversation robot using head gesture recognition as para-linguistic information. In *Proceedings of the 13th IEEE International Workshop on Robot and Human Interactive Communication*, 159-164.

Jokinen, K. and Allwood, J. (2010). *Hesitation in Intercultural Communication: Some observations on Interpreting Shoulder Shrugging*. Proceedings of the International Workshop on Agents in Cultural Context, The First International Conference on Culture and Computing 2010. Kyoto, Japan. pp.25-37.

Jokinen, K., Navarretta, C. and Paggio, P. (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.

Kipp, M.(2004) *Gesture Generation by Imitation - From Human Behavior to Computer Character Animation*. Ph.D. thesis, Saarland University, Saarbruecken, Germany, Boca Raton, Florida, dissertation.com.

Lu, J., Allwood, J. and Ahlsén, E. (2011). A Study on Cultural Variations of Smile Based on Empirical Recordings of Chinese and Swedish First Encounters. In Proceedings of ICMI 2011 Workshop Multimodal Corpora for Machine Learning: Taking Stock and Road mapping the Future, Alicante, Spain November, 8 pages.

Maynard, S. K. (1987). Interactional functions of a nonverbal sign head movement in Japanese dyadic casual conversations, *Journal of Pragmatics*, Volume 11, Issue 5, 589–606.

MacWhinney, B. and Wagner, J. (2010) Transcribing, searching and data sharing: The CLAN software and the TalkBank data repository. I: *Gespraechsforschung*, Vol. 11, 2010, s. 154-173.

McClave, E. (2000). Linguistic functions of head movements in the context of speech. *Journal of Pragmatics*, 32:855–878.

Navarretta, C. (2011a) Annotating Non-verbal Behaviours in Informal Interactions. To appear in Esposito et al. (Eds.) *Analysis of Verbal and Nonverbal Communication and Enactment: The Processing Issues*, LNCS 6800, Springer Verlag, 317-324.

Navarretta, C. (2011b) Anaphora and gestures in multimodal communication. In Hendrickx et al. (Eds.) *Proceedings of DAARC 2011*, Ediçoes Colibri, 171-181.

Navarretta, C., Ahlsén, E., Allwood, J., Jokinen, K., Paggio, P. (2011) Creating Comparable Multimodal Corpora for Nordic Languages. In *Proceedings of the 18th Nordic Conference of Computational Linguistics (NODALIDA 2011)*. Riga, Latvia, May 11-13, pp. 153-160.

Navarretta, C. and Paggio, P. (2010) Classification of Feedback Expressions in Multimodal Data. *Proceedings of ACL 2010*, Uppsala, Sweden, pp. 318-324.

Paggio, P. and Diderichsen, P. (2010) Information structure and communicative functions in spoken and multimodal data . In Henrichsen (Ed.) *Linguistic Theory and Raw Sound*. Copenhagen Studies in Language, 149-168.

Paggio, P. and Navarretta, C. (2010). Feedback in Head Gestures and Speech. In M. Kipp et al. (Eds.) LREC 2010 Workshop *Multimodal Corpora: Advances in Capturing, Coding and Analyzing Multimodality*, 1-4.

Paggio, P. and Navarretta, C. (2011) Head Movements, Facial Expressions and Feedback in Danish First Encounters Interactions: A Culture-Specific Analysis. In C. Stephanidis (Ed.) *Universal Access in Human-Computer Interaction- Users Diversity*, LNCS 6766, Springer Verlag, 583-590.

Paggio, P., Allwood, J., Ahlsén, E., Jokinen, K., Navarretta, C. (2010). The NOMCO multimodal Nordic resource - goals and characteristics. In *Proceedings of LREC 2010*, 2968-2973.

Rehm, M. and Nakano and Y., Andre and E., Nishida, T. (2008) Culture-Specific First Meeting Encounters between Virtual Agents. In *Proceedings of the 8th international conference on Intelligent Virtual Agents (IVA '08)*, Prendinger et al. (Eds.). Springer-Verlag, Berlin, Heidelberg, 223-236.