Feedback in Nordic First-Encounters: a Comparative Study

Costanza Navarretta*, Elisabeth Ahlsén†, Jens Allwood†, Kristiina Jokinen‡♠, Patrizia Paggio*♦

University of Copenhagen, Centre for Language Technology*, University of Gothenburg†, University of Helsinki‡, University of Tartu♠, University of Malta♦

Nialsgade 140, build.25, 2300 Copenhagen S*

University of Gothenburg 412 96 Göteborg†

Siltavuorenpenger 1A, 00014 Helsinki‡

Msida MSD 2080, Malta♦

E-mail: costanza@hum.ku.dk, elisabeth.ahlsen@ling.gu.se, jens@ling.gu.se, kristiina.jokinen@helsinki.fi, paggio@hum.ku.dk

Abstract

The paper compares how feedback is expressed via speech and head movements in comparable corpora of first encounters in three Nordic languages: Danish, Finnish and Swedish. The three corpora have been collected following common guidelines, and they have been annotated according to the same scheme in the NOMCO project. The results of the comparison show that in this data the most frequent feedback-related head movement is Nod in all three languages. Two types of Nods were distinguished in all corpora: Down-nods and Up-nods; the participants from the three countries use Down- and Up-nods with different frequency. In particular, Danes use Down-nods more frequently than Finns and Swedes, while Swedes use Up-nods more frequently than Finns and Danes. Finally, Finns use more often single Nods than repeated Nods, differing from the Swedish and Danish participants. The differences in the frequency of both Down-nods and Up-Nods in the Danish, Finnish and Swedish interactions are interesting given that Nordic countries are not only geographically near, but are also considered to be very similar culturally. Finally, a comparison of feedback-related words in the Danish and Swedish corpora shows that Swedes and Danes use common feedback words corresponding to yes and no with similar frequency.

Keywords: comparable multimodal corpora, feedback-related head movements, multimodal annotation

1. Introduction

This paper is about feedback-related communicative body movements, gestures henceforth, in comparable corpora of first encounters in three Nordic languages, Danish, Finnish and Swedish. The three corpora have been collected and annotated under the Nordic NOMCO project, which aims at: i) creating comparable annotated resources for the languages involved in the project, ii) comparing the annotated data in order to investigate specific communicative phenomena, iii) developing, extending and adapting models of interactive communication management that can be applied in interactive systems, iv) applying machine learning techniques to automatically recognize or predict the use of gestures with different interactive communication functions in various settings (Paggio et al. 2010).

The important role played by head movements in the expression of feedback has been investigated in numerous studies focusing on their function as signals from the addressee to the speaker, as well as feedback elicitation from the speaker to the addressee (Duncan 1972, McClave 2000, Maynard 1987). The phenomenon has also been studied for Nordic languages from a monolingual perspective (Boholm and Allwood 2010, Cerrato 2007, Paggio and Navarretta 2011).

First encounters interactions have previously been used in intercultural studies, see inter alia (Argyle, 1975; Rehm et al., 2009; Allwood and Lu 2010) because they allow an investigation of how varying degrees of familiarity and liking as well as social status and norms are dealt with in communication by persons of different cultural background.

In the present study we investigate how feedback-related head movements are used in Nordic first encounters. The Nordic countries are generally assumed to be quite similar culturally and socially, and they are also related linguistically, both through the Nordic languages Danish, Norwegian, Swedish, Faeroese and Icelandic and through the Finno-Ugric languages Sami, Finnish and Estonian. This study aims to explore whether this alleged similarity extends to the way people give or elicit feedback when they meet each other for the first time.

The abstract is organized as follows: first the three corpora are presented, and the annotations relevant to the present study are described; then the data from the three corpora are analyzed and compared. Finally, we present on-going and future work.

2. The NOMCO First Encounters

All corpora contain dyadic meetings between young people who do not know each other in advance. The interactions are video-recorded in a studio with the two participants facing each other against a light background. Two or three cameras are used in order to record the two participants individually and also to get a general view of them together. The Danish and Swedish corpora have been annotated by one coder and then checked by a second coder. In case of disagreement a third coder made the final decision. In Finnish, the annotation was also done by one coder and checked by a second one, but there was also an expert coder who monitored the

annotations, and the differences were settled by mutual agreement between the coders.

2.1 The Danish First Encounters

The Danish corpus consists of twelve interactions of approximately 5 minutes each for a total of about one hour. Twelve participants take part in the meetings, six males and six females. They are all native Danish speakers, and they are university students or have an academic degree. Their age is between 21 and 36 years.
Each participant is involved in two interactions, one with a male and one with a female. The two interactions are not recorded in sequence.
Subjects are standing opposite each other, and are recorded by three cameras, one taking a long shot of their bodies from the side, and the other two taking mid shots of them from the front. The three views are shown in Figure 1.
The corpus has been orthographically transcribed and the words have been time aligned using PRAAT (Boersma and Weenik, 2009). Pauses and information about offset and onset hesitation are coded. Expressions such as *hm* and *øh* are transcribed and word stress is marked.
Figure 1 shows a print screen of the three views of one Danish video recording.

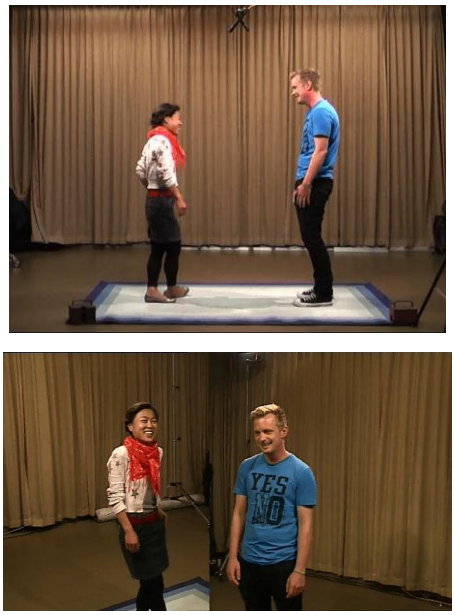

Figure1: The three views of a Danish video

2.2 The Finnish First Encounters

The Finnish corpus consists of 16 interactions with the following gender distribution: 8 female-female, 6 female-male and 2 male-male. The participants are all native Finnish speakers and are gathered from the student mailing lists of the University of Tampere, Tampere University of Technology and Tampere University of Applied Sciences. Their age ranges between 21-40 with one participant being over 50.
Each participant took part in two encounters with different partners. The interactions are 6-10 minutes long. Each recording started when one of the participants entered the recording space where the other one was already waiting. Three cameras were used to record the two participants separately and both together. From the three video recordings a mosaic version was also produced so that it is possible to follow each participant individually and both together on one video. Figure 2 shows the mosaic view of one of the videos.

Figure 2: Mosaic view of one Finnish video

2.3 The Swedish First Encounters

The Swedish first encounters corpus consists of 39 video recordings of interactions in Swedish, each approximately 8-10 minutes long, in total about 5 hours.
The participants are university students from the University of Gothenburg between 19 and 34 years old (mean age: 25). The gender distribution is as follows: 19 interactions are male-female, 11 are male-male and 9 are female-female. Figure 3 shows the total view of one of the Swedish videos.
So far 16 recordings of the Swedish video corpus have been transcribed using GTS, Gothenburg Transcription Standard (Nivre, 2004) and MSO 6, Modified Standard Orthography for Swedish (Nivre, 1999).

Figure 3: The total view of a Swedish video

3. The Multimodal Annotations

All three corpora have been annotated according to the MUMIN annotation scheme (Allwood et al. 2007). The scheme provides predefined attribute value pairs which

describe the shape and dynamics of communicative gestures as well as their functions. Table 1 contains the shape description of head movements which we have included in the present study.
Thus head movements are described in the schema in terms of the movement type and an indication of whether the gesture occurs once or is repeated. The distinction between up-nods and down-nods depends on the direction of the initial head movement.

Behavior feature	Behavior value
HeadMovement	Down-nod
	Tilt
	Up-nod (Jerk)
	Shake
	Waggle
	SideTurn
	HeadBackward
	HeadForward
	Other
Head Repetition	Single
	Repeated

Table 1: Attributes and values for head movements.

Feedback is coded in terms of three features (see Table 2). Basic, indicates whether the gesture has a feedback function It can take two values *CPU* (ContactPerceptionUnderstanding) and *Other*. The former value is the default in our data, and indicates that the participant is willing and capable of interacting, perceiving and understanding what is being communicated. The latter value is assigned in cases where only Contact or Contact and Perception are seen. Direction, indicates whether Feedback is given (*Given*), elicited (*Elicit*) or both (*GiveElicit*). Finally, Agreement indicates whether the participants agree with the interlocutor (*Agree*) or not (*Disagree*).
The Danish, Finnish and part of the Swedish corpora have been annotated in the ANVIL tool (Kipp 2004).

Behavior feature	Behavior value
FeedbackBasic	Contact/ Perception/Understanding
	Other
FeedbackDirection	Give
	Elicit
	Give-Elicit
FeedbackAgreement	Agree
	Disagree

Table 2: Attributes and values for feedback

3.1 Inter-annotator agreement experiments

The inter-annotator agreement was measured in the annotation of the Danish data, and the result for head movements was 0.70-0.80, which is quite good for this type of annotation (Navarretta et al. 2010). The figures refer to Cohen's kappa (1960) values for both segmentation and recognition.
We also made an inter-annotator agreement experiment between coders from the three countries. In this experiment a Danish coder, a Finnish one and a Swedish one independently annotated feedback-related head movements in two minutes of a Swedish First Encounter video. A Swedish interaction was chosen because Swedish is understood by the coders from all the three languages. The comparison study only included gestures and features relevant to the present study.
The results of the experiment indicate an agreement on feedback of approx. 0.60. This figure comprises segmentation and classification. Most disagreement figures are due to segmentation. Although the coders in most cases recognized the same feedback-related gestures, the length of the gestures varies, due to different segmentation practices followed in the three countries: gesture's length was often highest in the Danish annotation and lowest in the Finnish one. Furthermore, there are cases where a complex gesture is coded as one gesture by a coder and as two gestures by another coder. It must be noted that also in the Danish inter-coder agreement experiment, the most problematic aspect was segmentation, showing the importance of developing automatic recognizers, and the project has in fact developed a head tracker that can distinguish head movements with an acceptable reliability (Jongejan 2012).
Considering only classification, figures from the inter-annotator agreement experiment are much better. In particular, agreement on the assignment of feedback features is high (kappa over 0.90) while agreement is approx. 0.67 for gesture shape. Distinguishing gestures on the same axis, such as Up- and Down-nods seems to be difficult, and the chosen segmentation strategy may also have influenced the assignment of shape attributes.
In conclusion, the inter-annotator agreement experiment between the three countries shows that the three groups have followed slightly different annotation practices with respect to segmentation, and this has some consequences on the description of the shape of head movements. The experiment, on the other hand, also indicates high agreement between coders in the three countries in the task of identifying head movements that are related to feedback and in assigning feedback-related attributes and values.

4. Comparing the Data

4.1 Head movements

For the present study we extracted from the three corpora the head movements which have been coded with feedback features.
Head movements are extracted from 10 Danish interactions and 16 Swedish interactions. In the Swedish data, Shakes and Tilts have so far only been annotated in four interactions. For Finnish, the data is from 14

interactions, i.e. from 7 individuals taking part in two interactions each.
In the Danish corpus 1174 feedback-related head movements have been recognized, in the Swedish data 1115, and in the Finnish data 652.
In general more gesture types are recognized as being related to feedback in the Danish and Finnish first encounters than in the Swedish corpus, but in all three languages Nods are the most common feedback-related head movements, confirming preceding studies on many languages.
Since Nods are the most common feedback-related head movements, we compare the frequency of these gestures in the three corpora. Since the inter-coder agreement experiments showed different strategies in the way in which gestures have been segmented, we have abstained from comparing gestures' length using these annotations.
The frequency, which is given in Table 3, is calculated as number of gesture type per second. Both single and repeated Down- and Up-nods are included in the comparison. Note that the Finnish data in the table have been extracted from only four interactions.

	Danish no/sec	Finnish no/sec	Swedish no/sec
Nod (Down-nod + Up-nod)	0.17	0.16	0.14
Single	0.08	0.12	0.05
Repeat	0.09	0.04	0.09
Down-nod	0.14	0.11	0.07
Down-nod single	0.05	0.08	0.02
Down-nod repeated	0.09	0.03	0.05
Up-nod	0.03	0.05	0.07
Up-nod single	0.03	0.04	0.03
Up-nod repeated	0.00	0.01	0.04

Table 3: Frequency of single and repeated feedback Nods in the three corpora

This comparison indicates that in general the Finnish and Danish participants used Nods more frequently than the Swedish participants. However, there are large differences between individual persons: some have very strong and easily perceivable nods, while others perform small-scale nodding which is challenging to perceive. In the Finnish data, strong nodding was typical to acknowledge greetings.
According to the data in the table, Finns use repeated Nods less frequently than single Nods, while the opposite holds for Danes and Swedes. In fact, Finns use single nods three times more frequently than repeated nods and Swedes use single nods twice as frequently as single nods. The same pattern holds if we consider Down-nods. For Up-nods, single movements are more common in the Danish and Finnish data, while in the Swedish data there are slightly more repeated movements than single ones. Finally, the Swedish participants use Up-nods more frequently than the Danish and Finnish do.

4.1.1. Significance tests

We believe that Down- and Up-nods have different functions and occur in different contexts, thus we have run significance tests on the frequencies of these gestures in the three languages in the form of a two-tailed t-test with two-sample unequal variance.
The zero-hypotheses we wanted to test on our data are the following:

1. Concerning single Nods, the differences between the three datasets are statistically significant, i.e. the Finnish participants use single nods more often than the Danish ($p<0.01$) or Swedish participants ($p<0.001$). Similarly, the Danish participants use single nods more often than the Swedish ($p<0.01$).
2. Concerning repeated Nods, differences between Danish and Finnish and Finnish and Swedish are statistically significant, i.e. the Danes and the Swedes use repeated Nods much more frequently than the Finns ($p<0.001$), while the differences between Danish and Swedish data are not statistical significant ($p>0.1$).
3. The differences between the frequencies of feedback-related Down-nods in the three datasets are statistically significant. Danes use Down-nodes slightly more frequently than Finns ($p<0.1$), while they use them much more often than Swedes do ($p<0.001$). Also Finns use feedback-related Down-nods more frequently than Swedes ($p<0.01$).
4. Also concerning single Down-nods, the differences between the three datasets are statistically significant, i.e. Finns use more single Down-nods than Danes ($p<0.1$) and Swedes ($p<0.01$). Danes use single Down-nods more often than Swedes ($p<0.1$).
5. Concerning repeated Down-nods, the differences between Danish and Swedish and between Danish and Finnish are statistically significant, i.e. Danish use more repeated Down-nods than Swedish ($p<0.01$) and Finnish ($p<0.001$). Repeated Down-nods are slightly more common in the Swedish data ($p<0.1$) than in the Finnish data.
6. The differences in frequencies of Up-nods between all three datasets are statistically significant, i.e. Swedes use more Up-nods than Danes ($p<0.001$) or Finns ($p<0.1$), and Finns use more Up-nods than Danes ($p<0.01$).
7. Concerning single Up-nods, the differences between the three datasets are not statistically significant ($p>0.1$).
8. Concerning repeated Up-nods, differences between Swedish and Danish and between Swedish and Finnish are statistically significant, i.e. Swedes use repeated Up-nods more frequently than Danes ($p<0.01$) or Finns ($p<0.1$). The differences between Danish and Finnish data are not statistically significant.

In conclusion, Danes use Down-nods much more frequently than Swedes and slightly more often than Finns. On the other hand, Swedes use Up-nods significantly more often than Danes and slightly more often than Finns. Finally, Finns use Up-nods significantly more often than Danes in these data, while they use Down-nods slightly less than the Danish participants.

These differences are interesting because they occur in the same type of interaction between participants having similar age and educational background from Nordic countries that are traditionally considered to be near not only geographically but also culturally.
Since the data compared in this study are of limited size, these results must be tested on more interactions of the same type and from other communicative situations.

4.2 Words

In what follows, we compare the occurrences of the most frequent feedback words in Danish and Swedish. The two languages are very similar with respect to their lexicon and grammar. The results of this comparison are in Table 5.
The words which we have included are those corresponding to yes (*ja*), no (*nej/næ*), okay (*okey/okay*), and yes/well (*jo*).

	Danish n/sec	**Swedish n/sec**
yes (ja)	0.18	0.18
no (nej)	0.04	0.03
okay (okey)	0.07	0.04
yes/well (jo)	0.02	0.01

Table 5: Most frequently used feedback words

The table shows that Danes and Swedes use *ja* and *nej* with the same frequency. However, Danes use the word *okay* more frequently than Swedes.
In future, we will investigate whether similar head movements co-occur with the same expressions, and we will extend the comparison of feedback expressions to Finnish.

5. Conclusions

In this paper, we presented a first comparison of the use of feedback expressions and feedback-related head movements in three comparable corpora of video-recorded first-encounters in Danish, Finnish and Swedish. The corpora are not only comparable in the setting and the social background of the involved participants, but also in the annotation framework.
The comparisons of the frequency with which various types of head movement are used in the three languages show that the head movement which most often expresses feedback is Nod. This confirms the claims put forward in studies on many languages.
The comparison of frequent feedback words such as *yes* and *no* in Swedish and Danish indicate that the two languages use these words with similar frequency. In Finnish, feedback words, especially the function of *joo* and *niin* (‘yeah’) have been extensively studied by Sorjonen (2001). In the future, we plan to compare the function and frequency of feedback words in Danish and Swedish with Finnish, which differs in lexicon and grammar from the other two Nordic languages. We will also analyze multimodal feedback expressions on more data and communicative settings.
Although the compared data are not large and only concern a single communicative activity, the differences we have discovered are interesting, because of the cultural background of the counties involved in this study. Denmark, Sweden and Finland are in many respects culturally similar, being connected by a long history of collaboration, competition and conflict, leading to similar religions, similar customs, similar welfare states and similar legal systems. In spite of all these similarities, there are also differences, some of which may be reflected in the frequency differences we observe in our data.
In all three Nordic countries, Nods are important in giving and eliciting feedback, but for some reason, yet to be determined, this is done with different frequency in the three languages. Down-Nods are used more frequently by the Danes than by Finns and especially Swedes. On the other hand Up-Nods are more common in the Swedish data than in the Finnish and especially the Danish data. In Swedish, the distinction between Up-nod and Down-nod seems to be that Up-nods express uptake of new information more clearly than Down-nods. Possibly, this functional difference is not noticed in an inter-Nordic context and instead Up-nods could be associated with other features such as arrogance.
Another interesting observation is that compared to Swedes and Danes, Finns seem to use more single Nods than repeated Nods, making the Finnish nodding behavior less “vigorous” than in the neighboring countries. This difference in the form of nodding behavior may give rise to the impression that Finns give less feedback to the partner since single nods are less noticeable than repeated ones.
In general, the small differences in communicative behavior our comparative study is pointing to are somewhat unexpected and need more investigation to be confirmed and more clearly understood.

6. Acknowledgements

The NOMCO project is funded by the NORDCORP program under the Nordic Research Councils for the Humanities and the Social Sciences (NOS-HS).
We would like to thank the Danish, Finnish and Swedish annotators

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 Martin et al. (eds) *Multimodal Corpora for Modelling Human Multimodal Behaviour*.
Argyle, M. (1975) Bodily Communication. Methuen&Co. Ltd. London.
Boholm, M. and Allwood, J. (2010). Repeated head movements, their function and relation to speech. In Kipp et al. (eds.) *Workshop on Multimodal Corpora: Advances in Capturing, Coding and Analyzing Multimodality*. LREC 2010.
Allwood, J. and J. Lu (2010). Chinese and Swedish multimodal communicative feedback. Presented at the 5th International Conference on Multimodality, Sydney.
Boersma, Paul and David Weenink (2009). Praat: doing phonetics by computer (version 5.1.05). Retrieved May 1, 2009, from http://www.praat.org/.
Duncan, S. (1972). Some signals and rules for taking

speaking turns in conversations, *Journal of Personality and Social Psychology*, Volume 23, Issue 2, 283–292.

Jokinen, K. (2011). Turn taking, Utterance Density, and Gaze Patterns as Cues to Conversational Activity. In *Proceedings of the ICMI-MMC Workshop*, November, Spain.

Jongejan, B. (2012). Automatic annotation of head velocity and acceleration in Anvil. *Proceedings of LREC 2012*.

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.

Maynard, Senko K. (1987). Interactional functions of a nonverbal sign Head movement in Japanese dyadic casual conversation, *Journal of Pragmatics*, 11,5. 589–606.

Navarretta, C., E. Ahlsén, J. Allwood, K. Jokinen, P. Paggio (2011) Creating Comparable Multimodal Corpora for Nordic Languages. In *Proceedings of the 18th Nodalida*., 153-160.

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

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

Rehm, M., E. André, N. Bee, B. Endrass, M. Wissner, Y. Nakamo, A. Akhter Lipi, T. Nishida and H.H. Huang (2009). The Intercultural Dimension of Multimodal Corpora. In Kipp et al. (eds) *Multimodal Corpora*. LNAI 5509. Springer,138–159.

Sorjonen, M. (2001). *Responding in conversation: A study of response particles in Finnish*. Pragmatics & Beyond, New Series 70. John Benjamins, Amsterdam.