# Learning to classify the feedback function of head movements in a Danish corpus of first encounters

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

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

## ABSTRACT

This paper deals with the automatic classification of feedback by head movement in the Danish NOMCO corpus, a collection of dyadic interactions in which speakers get to know each other for the first time. The results show that by using a combination of features related to the shape of head movements and facial expressions, together with features of the words these gestures are related with, good results (an F-score of 0.72) are achieved in distinguishing head movements used to express feedback from those that serve a different communicative function. Moreover, the distinction between feedback give and feedback elicit, can also be learnt with very good accuracy (an F-score of 0.913), although this result should be taken with caution due to the fact that one of the behaviours is much more dominant than the other in the corpus.

## Keywords

Feedback, backchanneling, head movements, automatic classification

## 1. INTRODUCTION

The fact that head movements play an important role in regulating conversational exchange has been observed in many earlier studies. For example, Yngve in [27] and Duncan in [6] consider head nods examples of backchannels, that is feedback signals given by the listener without trying to take the floor. Hadar et al. [8] monitor head movements in five subjects during conversation, and find that properties such as amplitude, frequency and cyclicity of the movement can be associated with the different conversational functions of agreeing, wanting to take the turn, aligning with the interlocutor's stressed syllables and pauses. Maynard [12] studies head nods in dialogues between Japanese speakers. The most frequent function is found to be feedback by listeners, but speakers also nod a lot in different contexts. McClave, in a qualitative study of head movements in dialogues between two pairs of American speakers [13], observes that head movements occur together with a whole array of functions and senses, one of which is linked to what she calls backchanneling requests: the speaker nods to ask the listener for feedback, and the listener in turn nods. Finally, Cerrato [4] has looked at the communicative function of head movements in Swedish, which is like Danish a Scandinavian language, and found that in a subset of the Swedish GSLC corpus [19], 70% of all head movements are related to feedback, and that most of these are nods and up-down movements.

Other studies have looked at the application of machine learning algorithms to multimodal corpora with the purpose of predicting various aspects of gestural behaviour. For example, Jokinen and Ragni [10] and Jokinen et al. [9] found that machine learning algorithms could be trained to recognise some of the functions of head movements in Estonian and Danish. Reidsma et al. [26] show that there is a dependence between focus of attention (a combination of head, gaze and body features) and the assignment of dialogue act labels. Related are also the studies by Akker and Schultz [20] and Murray and Renals [17], both of which achieve promising results in the automatic segmentation of dialogue acts using the annotations in a large multimodal corpus. Finally, feedback expressions (head nods and shakes) are successfully predicted from speech, prosody and eye gaze in interactions with embodied agents as well as human communication by Fuije et al. and Morency et al. [7, 15, 16, 14].

In two earlier studies [25, 18], we looked at the relation between head movements and facial expressions on the one hand, and the dialogue act functions of linguistic feedback expressions on the other, in a corpus of map-task dialogues in Danish, and showed that head gestures, where they occur, contribute to the semantic interpretation of feedback expressions in a significant way. In this paper, on the contrary, we investigate the extent to which the communicative function of head movements can be classified automatically based on the shape features of the movements themselves, on the shape and function features of the associated facial expressions (if present) and on the words that are semantically associated with the head movements. In other words, we consider the whole multimodal sign, consisting of overlapping gestures and related speech, to classify the meaning of the head movements. The empirical data we use here do not come from map-task dialogues. They are extracted instead from a corpus of free dyadic conversations, and contain a much larger number as well as a larger variety of gestures.

The corpus is described in section 2. In section 3 we describe the classification tasks, the datasets on which the classification is performed and the results achieved. In section 4 we discuss our results and indicate ideas for further research.

# 2. THE DANISH NOMCO CORPUS

The corpus used in this study is a collection of first encounters dyadic interactions in Danish recorded and annotated within the NOMCO (Nordic Multimodal Communication) project, and is one of a number of parallel multimodal corpora showing different types of interaction in Swedish, Danish, Finnish and Estonian [23].

## 2.1 The recordings

The Danish NOMCO first encounters corpus consists of a set of 12 dialogues of an average duration of 5 minutes each, for a total of about an hour of interaction. The participants, six males and six females, all native speakers of Danish, did not know each other beforehand. Each subject participated in two interactions, one with a female and one with a male. Subjects were standing opposite each other, and were recorded by three cameras, one taking a long shot of their entire bodies from the side, and the other two taking mid shots of them from different angles. The two views are shown in Figure 1. After the recordings, the participants were given a questionnaire in which they rated their experience on a number of parameters. The results are quite homogeneous for all 12 participants, and indicate that the subjects were not too affected by the artificial setting even though they were aware of it (see [24] for more details).

**Figure 1: Recordings from the Danish NOMCO dialogues: total and split views**

## 2.2 The annotation

Both speech and gestures have been annotated. An orthographic speech transcription with time stamps for each word was done in Praat [3]. The transcription also contains an indication of whether words are stressed, as well as whether there is onset or offset hesitation. Furthermore, each word is tagged with a focus or a topic tag if it belongs to the focus or the topic of the sentence, following the method of annotation described in [21, 22]. Pauses and so-called filled pauses, i.e. words like *hmm* or vocal behaviours like *laugh* are included in the transcription. Finally, tags corresponding to syntactic clause boundaries were added in a semi-automatic way. Summing up, the speech transcription contains a variety of information which will allow us in future to investigate the relation between gestures and various linguistic features. In this study, however, we have only used the word tokens, including those for pauses and filled pauses.

The gesture annotation was done using the ANVIL annotation tool [11], and following the coding conventions defined in the MUMIN coding scheme [1]. In this study, we only consider the annotation of head movements and facial expressions, which includes attributes for the shape of the gesture, as well as a reduced set of function attributes.

It should be noted that head movements and facial expressions are annotated in different tracks, so their mutual correspondence is given by temporal overlap.

The gesture annotation features are shown in Table 1. Those related to the shape of the gestures are self-explanatory. The functional features only concern feedback and are inspired by the framework advocated by Allwood et al. [2], where feedback is described as unobtrusive behaviour that has the purpose of either signalling or eliciting signals of *contact*, *perception* and *understanding*. FeedbackBasic is thus used to annotate if the feedback involved includes all three aspects (CPU), or maybe only one of them (FeedbackOther). In practice, the only value used is CPU. FeedbackDirection indicates whether the gesture is a feedback signal or is being used to elicit a feedback signal.

**Table 1: Annotation features for gestural behaviour**

| Attribute | Value |
|---|---|
| HeadMovement | Nod, Jerk, HeadForward, HeadBackward,Tilt, SideTurn, Shake, Waggle, HeadOther |
| HeadRepetition | Single, Repeated |
| General face | Smile, Laugh, Scowl, FaceOther |
| Eyebrows | Frown, Raise, BrowsOther |
| FeedbackBasic | CPU, FeedbackOther |
| FeedbackDirection | FeedbackGive, FeedbackElicit |

Finally, each gesture is explicitly linked to a sequence of semantically related words (one or more) uttered by the speaker who is producing the gesture. There is often temporal overlap between gesture and corresponding words, but sometimes the gesture may precede or follow the words. In some cases, the gesture is produced without speaking, and it is then linked to a pause. In others, it co-occurs with and is linked to a filled pause.

An inter-coder agreement test was run after some initial training in order to test to what extent three coders identified the same gestures and assigned the same categories to the recognised gestures. The results in terms of Cohen's kappa [5] were in the range 0.5-0.6 for face attributes and 0.6-0.8 for head movements. The highest disagreement values for facial expressions were mainly due to disagreement on segmentation. The coders found it often harder to decide where facial expressions start and end, than doing the same for head movements. After five videos had been annotated, the two coders who had shown the most disagreement repeated the test. Results improved of about 10% for both face and head gestures.

## 2.3 Corpus analysis

So far, 9 of the 12 videos in the corpus have been analysed. The same 9 videos are also used for the machine learning experiments described below. The total duration of the videos is 3,027 seconds, they contain 10,800 words and a total of 3,391 gestures. Their distribution is shown in Table 2, where their frequency of occurrence (gesture per word) is also indicated.

**Table 2: Gestures in the Danish NOMCO corpus**

| Gesture type | # | g/w |
|---|---|---|
| All gestures | 3391 | 0.31 |
| Head | 2293 | 0.21 |
| Face | 1098 | 0.10 |

The average number of head movements per person is 127.28, with a standard deviation of 34.61, whilst the average number of facial expressions per person is 61, with a standard deviation of 26.80.

Head movements come in many different types. Nods (either single or repeated), are the most common, followed by tilts and side turns. Table 3 shows the counts for each type. Face expressions are either smiles, laughs or scowls (very few), accompanied or not by eyebrow raises or frowns. Eyebrow raises and frowns may also occur on their own.

**Table 3: Gestures in the Danish NOMCO corpus**

| Head movement | # |
|---|---|
| Tilt | 387 |
| SideTurn | 328 |
| Repeated Nod | 317 |
| HeadBackwards | 264 |
| Simple Nod | 205 |
| Repeated Shake | 199 |
| HeadForward | 199 |
| HeadOther | 148 |
| Jerk (UpNod) | 122 |
| Waggle | 66 |
| Simple Shake | 58 |
| Total | 2293 |

If we look at the function of the gestures we see that 47% of all gestures express feedback, in other words they have been annotated with the feature FeedbackBasic CPU. Head is the preferred modality when it comes to feedback (60% of all feedback gestures), which is not surprising since the head is the preferred modality of gestural expression in general. If we look at the distribution of feedback gestures among the various specific types, however, we see that repeated nods, smiles, single nods and tilts are the most frequent types, in that order. The exact distribution among the 11 most frequent types, making up for 92% of all the feedback gestures, is shown in Table 4.

Finally, feedback gestures fall into different types depending on the feedback direction. FeedbackGive is by far the most frequent type, followed by FeedbackElicit. There are also some cases of mixed direction, labelled FeedbackGiveElicit, and a few in which the annotators could not choose a specific value and assigned the label FeedbackUnderspecified (exact figures are given below in Section 3).

**Table 4: Most frequent feedback gestures in the Danish NOMCO corpus**

| Gesture type | # | % |
|---|---|---|
| Repeated Nod | 250 | 16 |
| Smile | 248 | 16 |
| Single Nod | 134 | 8 |
| Tilt | 125 | 8 |
| Raise | 117 | 7 |
| Shake | 112 | 7 |
| HeadBackwards | 110 | 7 |
| HeadForward | 99 | 6 |
| Jerk (UpNod) | 92 | 6 |
| Laughter | 91 | 6 |
| SideTurn | 84 | 5 |
| Total | 1462 | 92 |

To conclude this section, feedback is well represented in the Danish NOMCO corpus especially in terms of gesture type variation, but also in terms of different feedback direction types. The data should provide sufficient material to investigate ways in which gestural feedback can be learnt automatically.

# 3. CLASSIFICATION EXPERIMENTS

## 3.1 The tasks

The following two tasks were defined for the study:
– classification of head movement function;
– classification of feedback types.

For the first task, our hypothesis was that, based on formal and temporal characteristics of the head movements and the related facial expressions, as well as the word tokens that the movements are linked with, it would be possible to distinguish feedback gestures from non-feedback gestures.

As for the second task, we wanted to investigate to what extent, given the same features used in task one, plus knowledge of whether the gestures are feedback gestures, it would be possible to distinguish between feedback giving and feedback eliciting gestures. However, since one of the types, FeedbackGive, is the dominant one in the corpus, we considered this an explorative task.

## 3.2 The dataset

Preparing the dataset consisted in extracting head movement and facial expression data from the two relevant tracks in the ANVIL annotation and combining them in an appropriate way.

Head movement and facial expression features were combined if the corresponding gestures are performed by the same participant, and they overlap temporally. Temporal overlap was calculated by taking into account start and end points of each gesture. The time points are given in milliseconds, and no restriction was posed on the temporal overlap.

Since overlapping gestures can have different durations, one gesture type can overlap with more gestures of the other type and viceversa. Head movements are more frequent than facial expressions in our data, thus we have extracted the annotations of all head movements, and we have added to them the annotations of the overlapping facial expressions. Note that if a head movement overlaps with more than one facial expression, this will result in more than one classification instance. For example, if a head movement overlaps with two facial expressions one after the other, it is represented in the dataset as two instances with the same head movement features, but with different facial expression features. This may be seen as somewhat artificial. On the other hand, since this situation only happens a limited number of times, it was decided to live with the chosen representation format. In the opposite case, in which a head movement has no overlapping facial expression, the value *None* is assigned to all facial features with nominal value, whilst *0* is assigned to all those that have a numeric value (start point, end point and duration).

The resulting dataset consists of 2,437 elements, that is 144 elements more than the observed head movements. Approximately half (1,281) of them have an overlapping facial expression. As regards feedback, 1,125 head movements contain feedback-related features, and they overlap with a facial expression in 587 cases. In 77% of these cases (451 instances) the facial expression is also assigned a feedback related function. Concerning the different feedback types, 77% of the feedback movements are annotated with the feature FeedbackGive, 20% with FeedbackElicit, and 3% with either FeedbackGiveElicit, meaning that both functions are active at the same time, or with FeedbackUnderspecified, meaning that the annotator could not decide. The counts for the various types are shown in Table 5.

**Table 5: Head movement instances in the dataset**

| Head movement function | # |
|---|---|
| FeedbackGive | 872 |
| FeedbackElicit | 221 |
| FeedbackGiveElicit | 30 |
| FeedackUnderspecified | 2 |
| Total Feedback | 1125 |
| Feedback None | 1312 |
| Total head movement | 2437 |

## 3.3 The results

In the first experiment we investigate which features are most useful to the automatic classification of head movements which have a FeedbackBasic attribute, i.e. our first task. The features that we have considered are the form of head movements, the form of the overlapping facial expressions, and the words and phrases to which the head movements are linked. These make up 1385 different expressions. The most frequent one (254 occurrences) is in fact the pause, "+", indicating that the head movement is done while the other person is speaking. The second most frequent expression (171 occurrences) is the stressed "ja" (*yes*), followed by *breath*, *okay* with different stress patterns, *laugh* and *mm*. Words and phrases were considered in two different ways. First, all word tokens that could be categorised as feedback words (*yes, no* etc.), were replaced by the attribute *FBWord*. Filled pauses were included in this group. In this way, the most frequent word token became *FBWord*, with 661 occurrences, followed by the pause. Then, all word tokens were considered without any further semantic grouping. In both cases, the topic and focus attributes the words were coded with were also kept. We experimented with adding these features one at the time. In all cases, we also included the duration of the movement, either the head movement alone, or both movements if facial expressions are used.

**Table 6: Classification of FBBasic using different features. Baseline obtained with ZeroR, all others with NaiveBayes**

| Classifier | Precision | Recall | F-Measure |
|---|---|---|---|
| Baseline | 0.29 | 0.538 | 0.377 |
| HeadNoFace | 0.657 | 0.658 | 0.655 |
| HeadFace | 0.66 | 0.661 | 0.658 |
| HeadFaceFBWordsFocus | 0.717 | 0.712 | 0.712 |
| HeadFaceAllWordsFocus | 0.721 | 0.72 | 0.72 |

The results of the first set of experiments (Table 6) show that using the form of the head movements improves the classification of feedback-related head movements with respect to the baseline (where Feedback *None* is always chosen as the most frequent type). The F-Measure improvement is statistically significant (chi-square = 7.078, p = 0.008). Adding to these data the form of overlapping facial expressions, however, does not contribute to the classification at all. On the other hand, taking into consideration all the words which are semantically related to the gestures improves the classification up to an F-measure of 0.72%. The improvement, however, is not statistically significant. It does not make any difference whether the words are added as individual tokens or whether specific feedback words are grouped together.

In the second set of experiments, corresponding to our second task, we investigated to what degree feedback direction could be classified automatically. In addition to the features used to solve the first task, for both head movements and facial expressions we also kept the FeedbackBasic and FeedbackAgree features. In addition, for the face expressions FeedbackDirection was also kept. In other words, the classifier is not trying to understand the direction of feedback solely based on formal attributes, but it has access to the information of whether the gestures have a feedback function to start with, and whether it expresses agreement and disagreement.

**Table 7: Classification of FBDirection using different features. Baseline obtained with ZeroR, all others with NaiveBayes**

| Classifier | Precision | Recall | F-Measure |
|---|---|---|---|
| Baseline | 0.29 | 0.538 | 0.377 |
| HeadNoFace | 0.87 | 0.888 | 0.876 |
| HeadFace | 0.894 | 0.902 | 0.897 |
| HeadFaceFBWordsFocus | 0.915 | 0.919 | 0.911 |
| HeadFaceAllWordsFocus | 0.916 | 0.92 | 0.913 |

Table 7 shows the results of the second set of experiments. Also in this case using the head features (form features as well as basic feedback and agreement features) results in an improvement with respect to the baseline, and in this case the F-measure improvement is in fact quite dramatic and highly statistically significant (chi-square = 19.84, p <

0.001). Adding face and word features slightly improves the results, although not in a significant way. Again, there is no difference between considering all word tokens and grouping feedback words in one group. If we look at the best results, we can see that 96% of the FeedbackGive instances were classified correctly, whilst the percentage for FeedbackElicit was 46%.

## 4. DISCUSSION AND CONCLUSIONS

The results obtained in the classification tasks were generally positive in the sense that the classifiers performed better than the baseline in both tasks. However, several issues merit discussing.

Concerning the first basic task of distinguishing head movements expressing feedback from others, our expectation that the formal features of the head movements would play a significant role, is confirmed. On the other hand, the formal features of the corresponding facial expressions do not appear relevant. This may be due to the fact that there are too few of them in general, but it may also reflect the fact that feedback expressed by head movements may or may not be accompanied by a facial expressions, i.e. a smile or a laugh, with no systematic pattern. At the moment, our annotators are going through all the smiles in the corpus while re-considering their communicative function, and adding labels for other behaviours, such as turn and sequencing.

It may also be that the overlapping function we have used to pair head movements with facial expressions is too unrestricted. However, taking a more restrictive approach would result in even less overlapping expressions, thus the results of the classifications would likely not improve.

We were also expecting a clearer effect from using the words linked to the gestures. There is an indication that words are useful, although the improvement they bring is not significant. Perhaps, a different grouping of words into fewer categories, e.g. one for feedback words, one for pauses, one for filled pauses and one for the rest, may provide a less sparse dataset and thus a better basis for the classification. Alternatively, we need a larger annotated corpus.

It must also be noted that the task of distinguishing between gestures that have a feedback function from those that don't, is of course based on a very coarse-grained distinction. Once the full set of functional attributes from the MUMIN coding scheme is annotated in the corpus, we will be able to experiment with learning several communicative classes, e.g. feedback, turn and sequencing.

As for the second task, the results are in fact better than we expected. The fact that the F-measure value is as high as 0.913, and so much better than the baseline, however, is misleading because the classifier trivially discards false positives for the most frequent category, *None*, every time the Feedback value of the gesture is CPU. On the other hand, one could have expected that the next most frequent value, FeedbackGive would be wrongly assigned most of the time since it is much more frequent than FeedbackElicit. However, the classifier correctly identifies FeedbackElicit in 46% of the cases. It is difficult to assess from the model it builds, which combinations of features are most effective in distinguishing between the two types (here we are ignoring the other two values, which only occur sporadically). We can see, for instance, that the words *ja* and *okay* are predominantly used in connection with giving rather than eliciting feedback. The relative frequency of the different head movement types (*nod, tilt*, etc.) is also different depending on the feedback direction. In future studies, we would like to investigate how the directly preceding and following context, words as well as gestures, can be used to predict the occurrence of the two classes.

Another dimension we have not considered here, and which the data support, is individual variation. Thus it may well be that gestural feedback behaviour along the lines described here, can be learnt with even larger accuracy for specific individuals.

## 5. ACKNOWLEDGMENTS

The NOMCO project is funded by the NORDCORP programme under the Nordic Research Councils for the Humanities and the Social Sciences (NOS-HS). The Danish NOMCO corpus has been annotated by Sara Andersen, Josephine B. Arrild, Anette Studsgård and Bjørn N. Wesseltolvig.

## 6. REFERENCES

[1] J. Allwood, L. Cerrato, K. Jokinen, C. Navarretta, and P. Paggio. The mumin coding scheme for the annotation of feedback, turn management and sequencing. *Multimodal Corpora for Modelling Human Multimodal Behaviour. Special Issue of the International Journal of Language Resources and Evaluation*, 41(3–4):273–287, 2007.

[2] J. Allwood, J. Nivre, and E. Ahlsén. On the semantics and pragmatics of linguistic feedback. *Journal of Semantics*, 9:1–26, 1992.

[3] P. Boersma and D. Weenink. Praat: doing phonetics by computer (version 5.1.05) [computer program]. Retrieved May 1, 2009, from http://www.praat.org/, 2009.

[4] L. Cerrato. *Investigating Communicative Feedback Phenomena across Languages and Modalities*. PhD thesis, Stockholm, KTH, Speech and Music Communication, 2007.

[5] J. Cohen. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1):37–46, 1960.

[6] S. Duncan. Some signals and rules for taking speaking turns in conversations. *Journal of Personality and Social Psychology*, 23:283–292, 1972.

[7] S. Fujie, Y. Ejiri, K. Nakajima, Y. Matsusaka, and T. Kobayashi. 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*, pages 159 – 164, september 2004.

[8] U. Hadar, T. J. Steiner, and F. C. Rose. Head movement during listening turns in conversation. *Journal of Nonverbal Behavior*, 9(4):214–228, December 1985.

[9] K. Jokinen, C. Navarretta, and P. Paggio. Distinguishing the communicative functions of gestures. In *Proceedings of the 5th MLMI*, LNCS 5237,

pages 38–49, Utrecht, September 2008. Springer.

[10] K. Jokinen and A. Ragni. Clustering experiments on the communicative properties of gaze and gestures. In *Proceeding of the 3rd. Baltic Conference on Human Language Technologies*, Kaunas, October 2007.

[11] M. Kipp. *Gesture Generation by Imitation – From Human Behavior to Computer Character Animation*. Boca Raton, Florida: Dissertation.com, 2004.

[12] S. Maynard. Interactional functions of a nonverbal sign: Head movement in Japanese dyadic casual conversation. *Journal of Pragmatics*, 11:589–606, 1987.

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

[14] L.-P. Morency, I. de Kok, and J. Gratch. A probabilistic multimodal approach for predicting listener backchannels. *Autonomous Agents and Multi-Agent Systems*, 20:70–84, Springer 2009.

[15] L.-P. Morency, C. Sidner, C. Lee, and T. Darrell. Contextual recognition of head gestures. In *Proceedings of the International Conference on Multi-modal Interfaces*, 2005.

[16] L.-P. Morency, C. Sidner, C. Lee, and T. Darrell. Head gestures for perceptual interfaces: The role of context in improving recognition. *Artificial Intelligence*, 171(8–9):568–585, 2007.

[17] G. Murray and S. Renals. Detecting action meetings in meetings. In *Proceedings of the 5th MLMI*, LNCS 5237, pages 208–213, Utrecht, The Netherlands, September 2008. Springer.

[18] C. Navarretta and P. Paggio. Classification of feedback expressions in multimodal data. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL 2010)*, pages 318–324, Uppsala, Sweden, Juli 11-16 2010.

[19] J. Nivre, J. Allwood, J. Holm, D. Lopez-Kästen, K. Tullgren, E. Ahlsén, L. Grönqvist, and S. Sofkova. Towards Multimodal Spoken Language Corpora: TransTool and SyncTool. In *Proceedings of the Workshop on Partially Automated Techniques for Transcribing Naturally Occurring Speech at COLING-ACL '98*, Montreal, Canada, August 1998.

[20] H. op den Akker and C. Schulz. Exploring features and classifiers for dialogue act segmentation. In *Proceedings of the 5th MLMI*, pages 196–207, 2008.

[21] P. Paggio. Annotating information structure in a corpus of spoken Danish. In *Proceedings of the 5th International Conference on Language Resources and Evaluation LREC2006*, pages 1606–1609, Genova, Italy, 2006.

[22] P. Paggio. Information structure and pauses in a corpus of spoken Danish. In *Conference Companion of the 11th Conference of the European Chapter of the Association for Computational Linguistics*, pages 191–194, Trento, Italy, 2006.

[23] P. Paggio, J. Allwood, E. Ahlsén, K. Jokinen, and C. Navarretta. The NOMCO multimodal nordic resource - goals and characteristics. In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, Valletta, Malta, 2010. European Language Resources Association (ELRA).

[24] P. Paggio and P. Diderichsen. Information structure and communicative functions in spoken and multimodal data. In P. J. Henriksen, editor, *Linguistic Theory and Raw Sound*, volume 49 of *Copenhagen Studies in Language*, pages 149–168. Samfundslitteratur, 2010.

[25] P. Paggio and C. Navarretta. Feedback in head gesture and speech. In K. et al., editor, *Proceedings of LREC-2010*, pages 1–4, Malta, May 17 2010.

[26] D. Reidsma, D. Heylen, and R. op den Akker. On the Contextual Analysis of Agreement Scores. In M. Kipp, J.-C. Martin, P. Paggio, and D. Heylen, editors, *Multimodal Corpora From Models of Natural Interaction to Systems and Applications*, number 5509 in LNAI, pages 122–137. Springer, 2009.

[27] V. Yngve. On getting a word in edgewise. In *Papers from the sixth regional meeting of the Chicago Linguistic Society*, pages 567–578, 1970.