Transfer learning of feedback head expressions in Danish and Polish comparable multimodal corpora

Costanza Navarretta, Magdalena Lis

University of Copenhagen

Njalsgade 140, Build. 25,4th floor

Copenhagen, Denmark

E-mail: costanza@hum.ku.dk, magdalena@hum.ku.dk

Abstract

The paper is an investigation of the reusability of the annotations of head movements in a corpus in a language to predict the feedback functions of head movements in a comparable corpus in another language. The two corpora consist of naturally occurring triadic conversations in Danish and Polish, which were annotated according to the same scheme. The intersection of common annotation features was used in the experiments. A Naïve Bayes classifier was trained on the annotations of a corpus and tested on the annotations of the other corpus. Training and test datasets were then reversed and the experiments repeated. The results show that the classifier identifies more feedback behaviours than the majority baseline in both cases and the improvements are significant. The performance of the classifier decreases significantly compared with the results obtained when training and test data belong to the same corpus. Annotating multimodal data is resource consuming, thus the results are promising. However, they also confirm preceding studies that have identified both similarities and differences in the use of feedback head movements in different languages. Since our datasets are small and only regard a communicative behaviour in two languages, the experiments should be tested on more data types.

Keywords: Comparable Multimodal Corpora, Feedback Head Movement, Machine learning

1. Introduction

Face-to face communication is multimodal since it involves both speech and body behaviours. The study and automatic treatment of multimodal communication data require annotated audio and video-recorded data of different type reflecting the many factors that influence communication such as language (Rehm et al.2009, Navarretta et al. 2012), task (deKok and Heylen 2010), number, age, and familiarity degree of the participants (Navarretta and Paggio 2012).

The past decade, a number of multimodal corpora have been collected and annotated such as AMI, CALLAS, NOMCO and SPONTAL. However, the need for annotated and freely available multimodal data is still huge also because not all multimodal data are freely available. Furthermore, the annotation of multimodal corpora is extremely time-consuming. It is therefore important to address the reusability of existing annotated resources.

Reusing resources has been a central theme in NLP and in other research communities for a long time. One of the strategies adopted has been using existing annotated corpora as training data for classification algorithms in order to annotate data in other domains among many (Blitzer et al. 2007, Moore and Lewis 2010, Pan and Yang 2010, Saenko et al. 2010).

In the present work, we follow this line of research. More specifically, we want to investigate to what extent the multimodal annotations of spontaneously occurring conversations in one language can be used to train classifiers to recognise the function of feedback head movements in comparable conversations in another language.

The paper is organized as follows. First, we discuss related studies (section 2) and present our data (section 3). Secondly, we describe the machine learning experiments and evaluate them (section 4). Finally, we conclude and discuss possible extensions of this work (section 5).

2. Related studies

Several studies of multimodal communication have focussed on the feedback function of head movements, which are the most frequently occurring body behaviour in many types of conversation.

Feedback is a unobtrusive behaviour by which conversation participants give or elicit signals of whether they are perceiving, understanding and accepting the current message (Allwood et al. 1992). Eliciting feedback is also known as backchannelling. The feedback functions of head movements have been studied in both monolingual (Yngve 1970, Maynard 1987, McClave 2000, Cerrato 2007, Paggio and Navarretta 2011) and multilingual corpora (Maynard 1987, Rehm et al 2009, Liu and Allwood 2011, Navarretta et al. 2012, Navarretta and Lis 2013). These studies indicate that there are both similarities and differences in the way people express feedback via head movements and co-speech, and the differences may depend inter alia on the language (Navarretta et al. 2012, Navarretta and Lis 2013), the task (de Kok and Heylen 2010) or the degree of familiarity between the participants (Navarretta and Paggio 2012). In particular, Navarretta and Lis (2013) analyzed multimodal feedback expressions, speech and head movements in manually annotated Danish and Polish triadic conversations which involved well acquainted people of the same age and gender. They found that the Danish and Polish participants used similar types of feedback head movements and verbal expressions, but they also discovered significant differences in the use of repeated feedback speech tokens and head movement in the two datasets.

In order to classify, predict, model or generate the functions of multimodal behaviors, machine learning has been applied with success on the annotations of multimodal corpora. This is also the case for feedback head movements. Examples are the studies by Ragni and Jokinen (2007) and Jokinen et al. (2008) who test the application of unsupervised and supervised machine learning algorithms on the manual annotations of the shape and functions of head movements in more languages in order to recognize some of their functions. Fujie et al. (2004) and Morency et al.(2005, 2009) apply machine learning to model and/or generate head nods and shakes in interactions between humans or between humans and robots. They use multimodal information comprising speech, head movements and gaze, while Navarretta and Paggio (2010) train a classifier to identify the semantics of yes/no feedback expressions on speech and head movement annotations in a Danish map task corpus. Finally, Paggio and Navarretta (2012) apply a support vector machine on the annotations of head movements and co-occurring facial expressions in Danish first encounters to predict the feedback functions of head movements.

Navarretta (2013) investigates to what extent the multimodal annotations in a corpus can be used to predict the feedback functions of head movements in a corpus of different type. A number of supervised machine learning experiments were performed on the data. Two Danish corpora were used in the experiments: the DKCLARIN/MOVIN corpus, which is also used in the present study, and the Danish NOMCO first encounters. The results of the experiments confirm that multimodal behavior partly depends on the conversation type, but they also indicate that multimodal annotations which follow the same theoretic model can be used in different domains after having been adapted to each other if they only vary in granularity. Furthermore, Navarretta (2013) found out that using one corpus as training data to identify feedback head movements in the other gives better results than the majority baseline.

The experiments also indicated that supplying the annotations of a small conversational corpus with more data from other corpora can result in better classification results.

In the present work, we follow the same strategy used as in (Navarretta 2013), but we do not work with conversations of different type in the same language, but with conversations of the same type in two different languages.

3. The data

Our data are comparable triadic natural occurring conversations collected in private homes. The Danish data were part of the DKCLARIN/MOVIN database (MacWhinney and Wagner 2010) collected by researchers at the University of Southern Denmark. The conversations which we have used were transcribed and multimodally annotated as part of the Danish CLARIN project by researchers at the University of Copenhagen (Navarretta

2011). In Figure 1 is a screenshot from the Danish corpus.

Figure 1: Screenshot from the Danish data

The Polish data were collected, transcribed and annotated under the on-going European CLARA project.
Both the Danish and Polish corpora are triadic and the participants were native speakers of Danish and Polish, respectively. The participants in each conversation were well acquainted (family members or friends) women aged 50+ who were recorded in their private homes while they were eating and talking freely. There was no restriction on the subjects to be addressed, but the participants were aware of being recorded. The Danish participants were filmed by one video camera, while two cameras were used to record the Polish subjects.
In Figure 2, two screenshots of the Polish data are shown.

Figure 2: Snapshots from the Polish data

The length of the interactions included in this study is of approximately 20 minutes conversations per language. A more detailed description of the two corpora is in (Navarretta 2011) and (Navarretta and Lis 2013), respectively.
The two corpora are, thus, comparable under different

aspects such as the communicative settings, the familiarity degree of the participants, their age and gender. An aspect that is important in the context of the present work is the comparability of the two corpora's annotations. In fact the two corpora have been annotated according to the same annotation framework (Allwood et al. 2007).

3.1 The Annotations

Both corpora are orthographically transcribed in PRAAT (Boersma and Weenik, 2013) with speech token time stamps. The transcriptions have been then imported in the ANVIL tool (Kipp 2004) and the shape and the functions of various types of body behavior have been annotated according to the MUMIN scheme (Allwood et al. 2007). In the scheme, predefined attribute and value pairs describe the shape and the communicative functions of body behaviors. Shape and function features are independent. Furthermore, body behaviors are linked to speech tokens via the ANVIL multilink facility when the annotators judge that they are semantically related. Figure 3 shows snapshots of the annotations in the two corpora.
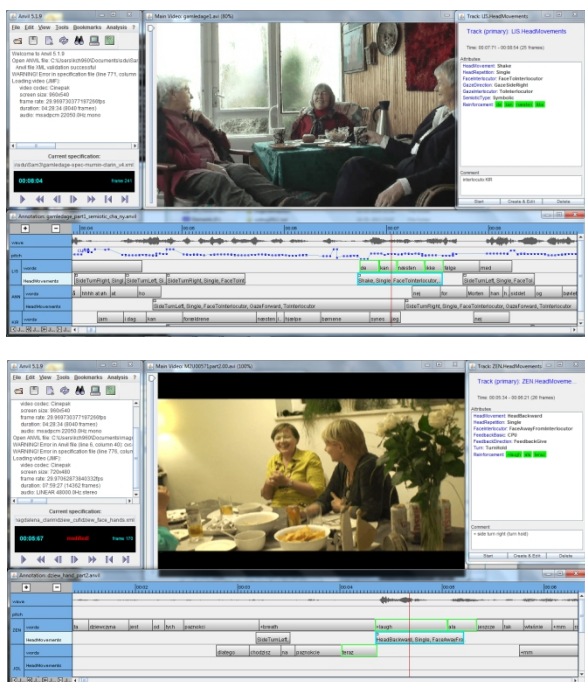

Figure 3: Snapshots from the ANVIL tool

In some cases, the annotations in the two corpora have different levels of granularity. In order to make the annotations compatible, we have taken the intersection of the attributes and values used in the two corpora, and in some cases have joined values which were distinguished in one corpus and not in the other.

The annotations that are relevant to this study are those of the shape and feedback function of head movements. The behavioral attributes and values used in this study are shown in Table 1.

Head movements are annotated via three attributes, which indicate the type of movement (HeadMovement), whether the movement is repeated or not (HeadRepetition) and whether the person is facing the interlocutor (FaceToInterlocutor). Feedback is annotated via three attributes. The first attribute, FeedbackBasic is used to

annotate if there is feedback and whether it involves all three aspects of Contact, Perception and Understanding, CPU, (Allwood et al. 1992) or only one or two of them (FeedbackOther).

Attribute	Value
HeadMovement	Nod, Jerk, HeadBackward, HeadForward, Tilt, SideTurnLeft, SideTurnRight, Shake, Waggle, HeadOther
HeadRepetition	HeadSingle, HeadRepeated
FaceToInterlocutor	FaceToInterlocutor, FaceAwayFromInterlocutor
FeedbackBasic	CPU, FeedbackOther, NONE

Table 1: MUMIN attributes and values for Head Movements and Feedback Basic

In DKCLARIN, the FeedbackBasic attribute also includes a value (SelfFeedback) indicating that the participants were providing feedback to their own contributions. Since this aspect was not annotated in the Polish data, we did not include it. Finally, we decided not to consider in this preliminary work other aspects of feedback which are accounted for in the MUMIN scheme and are annotated in both corpora. These comprise the indication of whether a behavior signals that a person is giving or eliciting feedback and whether he is agreeing or disagreeing with the interlocutors.

3.2 Head movements in the two corpora

This study regards 1535 head movements. More specifically, the Danish corpus contains 687 communicative head movements of which 476, that is 69.2 % are related to feedback. In the Polish data, there are 848 head movements of which 489 (57%) signal feedback. Thus, there are more head movements in the Polish corpus than in the Danish corpus, but the percentage of feedback head movements is higher in the latter corpus. All types of head movements are used to express feedback in both corpora, but the most frequent feedback movements are nods. Our preceding study of this data (Navarretta and Lis 2013) has also shown that there are significantly more repeated unimodal feedback head movements and spoken expressions in the Polish than in the Danish data. Furthermore, we observed that the Polish participants in these interactions used significantly more repeated multimodal feedback head movements and spoken expressions than the Danish participants. Thus the value HeadRepeated is assigned much more frequently to the annotations of feedback head movements in the Polish data than in the Danish data.

4. Machine learning experiments

The main aim of our classification experiments was to investigate the reusability of the annotations of one corpus as training data for a classifier to identify the feedback function of head movements in a comparable corpus in the other language. We reproduce here some of the experiments described in (Navarretta 2013) who used different conversational Danish corpora. Since we work

with data from different languages, we have not included speech in the present work.

The machine learning experiments were performed in the WEKA platform for data mining (Witten and Frank 2005). The features used in the training data are the descriptions of the shape of head movements and their feedback function as described in section 3. Feedback is then predicted in the test corpus.

The main classification algorithm in the experiments is the WEKA Naive Bayes classifier using estimator classes. We also tested on our data KStar and SMO, a support vector machine, but Naïve Bayes produced the best results in most cases.

It must be noted that in the transfer learning experiments performed by Navarretta (2013) on corpora of different conversation types but belonging to the same language a support vector machine algorithm was used because it performed better on those data.

The different performance of the various types of classifiers in the two studies is mainly due to the fact that more classes had to be identified in the monolingual experiments (Navarretta 2013) than in the present experiments.

The results of a majority classifier, ZeroR, are used as one of the baselines in the evaluation of the experiments.

As in Navarretta (2013) we also use a second baseline to evaluate the transfer learning experiments. The results of the Naïve Bayes classifier trained and tested on the monolingual data are the second baseline which is used to compare the performance of the classifier in the transfer learning experiments and in experiments where training and test data belong to the same dataset. Ten-fold cross validation was used in these monolingual experiments.

In the evaluation, the results of all experiments are given in term of Precision, Recall and F-measure scores (P, R, and F respectively in the tables which show the experiment results).

In a first preliminary experiment, we wanted to measure the differences between the data in the two corpora. Following the strategy proposed by Blitzer et al. (2007), we added to the items of both corpora the name of the corpus to which they belong, and then we joined them. Finally, we trained the majority and Naïve Bayes classifiers on the shape annotations of the head movements to identify the corpus name (a binary classification problem).

In these experiments the classifier was trained on each feature independently and then on all the shape features.

The results of these experiments are in Table 2 and indicate that the Naïve Bayes classifier can identify the corpus in which head movements occur from the head shape features better than the majority baseline. This is the case in all experiments independently from the shape features used in the training data. That is the improvement with respect to the results obtained by the majority classifier is significant in all cases. Significance was calculated as paired t-test with significance threshold (p) < 0.05. However, the best results were obtained when all shape features were used. The improvement in terms of F-measure with respect to the majority baseline in this case was of 0.31. The worse results were obtained when only annotations of the type of head movement were used as training data. In this case the F-measure improvement was 0.22.

Classifier	Data	P	R	F
Baseline ZeroR		0.3	0.55	0.39
Naïve Bayes	Hmovement	0.61	0.61	0.61
Naïve Bayes	HRepetition	0.75	0.56	0.66
Naïve Bayes	FacetoIntelocutor	0.73	0.64	0.69
Naïve Bayes	Hmove+Hrep+FacetoInt	0.7	0.7	0.7

Table 2: Identifying the corpus from the annotations of head movements

4.1 The transfer learning experiments

We run two groups of symmetric experiments. In the first group of experiments, we trained the Naive Bayes on the Polish annotations and tested the resulting model on the Danish data. In the second group of experiments, we reversed training and test data, thus the classifier was trained on the Danish data and the resulting model was tested on the Polish data. We tested all possible combinations of head attributes to see whether some attributes contributed to classification more than others. In particular, we expected that we would obtain the best results when training the data on the type of head movement alone. However, the performance of the Naïve Bayes classifier did not change significantly using different combinations of shape attributes in the training and test data (one-tailed t-test with significance threshold 0.05).

For simplicity, we only provide the results of the experiments in which all the attributes describing the shape of head movements (Table 1) were used since the classifier also performed best with these data. The results of the experiments in which we tested the use of Danish data for identifying feedback head movements in the Polish corpus are given in Table 3.

Classifier	Training set	P	R	F
ZeroR baseline 1	Polish	0.31	0.56	0.4
NaiveBayes baseline 2	Polish all	0.74	0.74	0.74
Naïve Bayes	Danish	0.61	0.53	0.57

Table 3: Danish training data and Polish test data

The results of the experiments in which the Polish data were used to train the Naïve Bayes classifier to identify feedback head movements in the Danish data are in Table 4.

Classifier	Training set	P	R	F
ZeroR baseline 1	Danish	0.39	0.62	0.5
NaiveBayes baseline 2	Danish	0.78	0.78	0.78
Naïve Bayes	Polish	0.66	0.55	0.61

Table 4: Polish training data and Danish test data

The results of both experiments indicate that training the Naïve Bayes classifier on the shape annotations of head movements in one corpus to identify feedback head movements in the comparable corpus in the other language gives better results than the majority baseline. The improvement of the F-measure with respect to the majority classifier is of 0.17 when the Danish annotations

are the training data and the Polish corpus is the test data. The improvement of the F-measure with respect to the majority classifier when the Naïve Bayes classifier is trained on the Polish data and tested on the Danish data is of 0.11. The results in both experiments are significant. Also in this case significance of results was calculated as paired t-test with $p < 0.05$.

The decrease in performance with respect to the second baseline (the results of the classifier trained and tested on the same dataset via ten-fold cross-validation) is 0.14 in the first experiment that is in the case when the classifier is trained on the Danish data and tested on the Polish data. The decrease of performance with respect to the monolingual baseline in the case when the Naïve Bayes classifier is trained on the Polish data and tested on the Danish corpus is of 0.17. The decrease in performance of the classifier in both cases is significant. Thus, the results of transfer learning are in between the two baselines.

5. Discussion

Our transfer learning experiments indicate that it is possible and useful to train classifiers on the annotations of head movements in a Danish corpus to annotate the feedback function of the head movements in a comparable Polish corpus. Similarly, training a classifier on the Polish annotations of head movements to identify those which signal feedback in the Danish comparable corpus gives better results than the majority baseline. Furthermore, the increase in performance with respect to the majority baseline is significant in both experiments. The results of the first group of experiments in which the Danish language model was tested on the Polish data were better than those obtained in the second group of experiments in which the Polish model was tested on the Danish data. The difference in the performance of Naïve Bayes in the two experiment groups can be due to the fact that the Polish dataset, although it is larger, contains a lower percentage of feedback head movements than the Danish data.

The results are promising because they indicate that the annotations of body behaviours in a multimodal corpus can be used to annotate the functions of the same behaviours in other corpora even when the corpora are in different languages. Since the manual annotation of multimodal behaviours is resource consuming, the reusability of multimodal annotations should be investigated further and possibly combined with automatic annotations.

The decrease in performance with respect to the results obtained by the classifier when tested and trained on data from the same corpus is also significant both when the Danish corpus is the training data and the Polish corpus is the test data and when training and test data are reversed. The results with respect to the first and second baseline are compatible with our previous analysis of feedback head movements in these two corpora (Navarretta and Lis 2013). This study showed that there are both similarities and differences in the way the Danish and Polish participants express feedback via head nods and speech in these conversation types. However, the fact that repetition of feedback head behaviours is an important discriminative feature between the two corpora was only partially confirmed by the experiments in which we measured the similarity degree of the head shape

annotations in the two corpora according to the strategy proposed in (Blitzer et al. 2007) and tested on these types of data in (Navarretta 2013)

The machine learning results in this study are also in line with those obtained by Navarretta (2013) who made similar machine learning experiments using Danish corpora of different types of conversation as training and test data. One of the corpora, the DK-CLARIN corpus has been used both in Navarretta (2013) and in this study. In Navarretta (2013) the improvements of the SMO classifier with respect to the majority baseline were 0.05 when the annotations in a first encounters corpus were the training data and the DK-CLARIN conversations were the test data. The improvement of the F-measure with respect to the majority baseline was 0.18 when the annotations from the DK-CLARIN corpus were the training data and the NOMCO first encounters were the test data.

The decrease of performance with respect to the second baseline that is when the SMO classifier was trained and tested on the annotations of the same corpus was of 0.26 when the DK-CLARIN annotations were tested on the NOMCO corpus and of 0.3 when the NOMCO annotations were tested on the DK-CLARIN annotations. Even though the present transfer learning experiments have a better performance, the results of Navarretta's experiments and those in this study are only partially comparable since the feedback annotations in (Navarretta 2013) also included information on self-feedback. Moreover, the attributes and values describing the shape of head movements in that study were fewer than those used in the present experiments. Including speech tokens to the data did not improve the classification results in the monolingual experiments.

Since the size of our corpora is small, our results should be tested on larger datasets. Furthermore, we do not know to what extent the results obtained depend on the similarity of the two corpora and on the fact that feedback head movements in these comparable datasets are quite similar. However, the results of these experiments as well as those described in (Navarretta 2013) indicate the utility of using standardized annotations as those provided by the MUMIN annotation scheme (Allwood et al. 2007) and that transfer learning is possible on these types of data. Thus, these kinds of experiment should be repeated in more language, conversation and behaviour types.

6. Conclusion

The results of our classification experiments show that the annotations of head movements in Danish conversations can be used to identify feedback head movements in comparable conversations in Polish. Similarly, the annotations of the Polish conversations are useful for the identification of feedback head movements in Danish comparable conversations.

Since the size of our data is not large, the experiments should be repeated on more and different types of corpora and other behaviour types should be included. However, these results are interesting especially if we consider that only few shape features were used in our experiments and that the shape descriptions were coarse-grained.

In general, the results of our experiments are promising because the manually annotation of multimodal corpora is expensive and extremely time consuming,.

These results also confirm the utility of producing formally annotated data according to a common annotation scheme as well as the appropriateness of the MUMIN framework for producing annotations that can be used as training data of classifiers.

In future, we will test transfer learning on more phenomena and data. We should also look at whether taking into account the annotations of more co-occurring behaviours improves classification and whether it is possible to use transfer learning on corpora annotated according to a different annotation model.

7. Acknowledgements

8. References

Allwood, J. Nivre, J. and E. Ahlsén, E. (1992) On the semantics and pragmatics of linguistic feedback, *Journal of Semantics*, vol. 9, pp. 1–26.

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 *Multimodal Corpora for Modelling Human Multimodal Behaviour. Special Issue of the International Journal of Language Resources and Evaluation*, vol. 41, no. 3–4, pp. 273–287, 2007.

Blitzer, J. Dredze, M. and Pereira, F. (2007) Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification," in *Proceedings of the 2007 Conference of Association for Computational Linguistics*. ACL 2007. Association for Computational Linguistics.

Boersma, Paul and David Weenink (2013). Praat: doing phonetics by computer. Retrieved 2013, from http://www.praat.org/.

Cerrato, L. (2007). *Investigating communicative feedback phenomena across languages and modalities*, Ph.D. dissertation, Stockholm, KTH, Speech and Music Communication,.

de Kok, I.A., Heylen, D.K.J. (2010) Differences in Listener Responses between Procedural and Narrative Tasks. In: *Proceedings of the 2nd International Workshop on Social Signal Processing*, SSPW '10, Florence, Italy, pp. 5–10. ACM, New York.

Fujie, S. Ejiri, Y., Nakajima, K., Matsusaka, I. and Kobayashi, T. (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.*

Jokinen, K, . Navarretta, C. and Paggio, P. (2008) Distinguishing the communicative functions of gestures, in *Proceedings of the 5th MLMI*, ser. LNCS 5237 Utrecht, The Netherlands: Springer, September, pp. 38–49.

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

Lu, J. and Allwood, J. Unimodal and multi-modal feedback in Chinese and Swedish monocultural and intercultural interactions (a pilot study). In P. Paggio, E. Ahlsén, J. Allwood, K. Jokinen, and C. Navarretta (2011), (eds), *Proceedings of the 3rd Nordic Symposium on Multimodal Communication*, pages 40-47. NEALT, May 27-28.

MacWhinney, B. and Wagner, J. (2010) Transcribing, searching and data sharing: The CLAN software and the TalkBank data repository, *Gespraechsforschung*, vol. 11, pp. 154–173.

Maynard, S. (1987) Interactional functions of a nonverbal sign: Head movement in Japanese dyadic casual conversation. *Journal of Pragmatics*, vol. 11, pp. 589–606,.

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

Moore, R. C. and Lewis, W. (2010) Intelligent selection of language model training data, in *Proceedings of the ACL 2010 Conference*. Uppsala, Sweden: Association for Computational Linguistics, 11-16 July, pp. 220–224.

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

Navarretta, C. (2011) Annotating non-verbal behaviours in informal interactions, in A. Esposito, A. Vinciarelli, K. Vicsi, C. Pelachaud, and A. Nijholt, (Eds) *Analysis of Verbal and Nonverbal Communication and Enactment: The Processing Issues*, ser. LNCS, Springer Verlag, vol. 6800, pp. 317–324.

C. Navarretta. Transfer Learning in Multimodal Corpora. In *Proceedings of the 4th IEEE International Conference on Cognitive Infocommunications (CogInfoCom2013)*, Budapest, Hungary 2-5 December 2013, pp. 195-200.

Navarretta, C. and Lis, M. (2013) Multimodal feedback expressions in Danish and Polish spontaneous conversations. In Allwood, Ahlsén, Paggio, Navarretta and Jokinen (Eds). *NEALT Proceedings. Northern European Association for Language and Technology, Proceedings of the fourth Nordic Symposium of Multimodal Communication*, November 2012, Gothenburg Sweden, pp. 55-62, Linköping Electronic Conference Proceedings.

Navarretta, C., Ahlsén, E., Allwood, J., Jokinen, K., Paggio, P. (2012) Feedback in Nordic First-Encounters: a Comparative Study. In: *Proceedings of LREC 2012*, pp. 2494–2499. Istanbul Turkey.

Paggio, P. and Navarretta, C. (2013) "Head movements,

facial expressions and feedback in conversations - empirical evidence from Danish multimodal data," *Journal on Multimodal User Interfaces - Special Issue on Multimodal Corpora*, vol. 7, no. 1–2, pp. 29–37.

Pan, S. J. and Yang, Q. (2010) A survey on transfer learning, *Knowledge and Data Engineering, IEEE Transactions on*, vol. 22, no. 10, pp. 1345–1359.

Rehm, M., Andre, E., amd Birgit Endrass, N.B., Wissner, M., Nakano, Y., Lipi, A.A., Nishida, T., Huang, H.H. (2009)Creating Standardized Video Recordings of Multimodal Interactions across Cultures. In: M. Kipp, J.C. Martin, P. Paggio, D. Heylen (Eds.) *Multimodal Corpora. From Models of Natural Interaction to Systems and Applications*, no. 5509 in LNAI, pp. 138–159, Springer.

Saenko, K., Kulis, B., Fritz, M. and Darrell, T. (2010) Adapting visual category models to new domains, in *ECCV Proceedings*, Heraklion, Greece.

Yngve, V. (1970) On getting a word in edgewise," in *Papers from the sixth regional meeting of the Chicago Linguistic Society*, pp. 567–578.

Witten, I.H. and Frank, E. (2005) *Data Mining: Practical machine learning tools and techniques*, 2nd edition San Francisco: Morgan Kaufmann.
.