Classifying Multimodal Turn Management in Danish Dyadic First Encounters

Costanza Navarretta[1] *Patrizia Paggio*[1,2]

(1) University of Copenhagen, Njalsgade 140 Buil. 25,4. Copenhagen Denmark
(2) University of Malta, Valletta Malta

costanza@hum.ku.dk, paggio@hum.ku.dk

ABSTRACT

This paper deals with multimodal turn management in an annotated Danish corpus of video recorded dyadic conversations between young people who meet for the first time. Conversation participants indicate whether they wish to give, take or keep the turn through speech as well as body behaviours. In this study we present an analysis of turn management body behaviours as well as classification experiments run on the annotated data in order to investigate how far it is possible to distinguish between the different types of turn management expressed by body behaviours using their shape and the co-occurring speech expressions. Our study comprises body behaviours which have not been previously investigated with respect to turn management, so that it not only confirms preceding studies on turn management in English but also provides new insight on how speech and body behaviours are used together in communication. The classification experiments indicate that the shape annotations of all kinds of body behaviour together with information about the gesturer's co-occurring speech are useful to classify turn management types, and that the various behaviours contribute to the expression of turn features in different ways. Thus, knowledge of the different cues used by speakers in face-to-face communication to signal different types of turn shift provides the basis for modelling turn management, which is in turn key to implement natural conversation flow in multimodal dialogue systems.

KEYWORDS: Multimodal Communication, Turn Management, Multimodal Corpora, Machine Learning.

1 Introduction

This article deals with turn management in the Danish NOMCO first encounters corpus. Turn management concerns the regulation of the conversation flow (Allwood et al., 2007) and is achieved through verbal and non verbal behaviours. Determining these cues is important for implementing conversational systems which can interact with users in a natural way.

Since conversations are social activities, they are influenced by many factors including the communicative situation, the cultural, physical and social setting, the number and role of the participants, their age, gender, and relation. These factors also play a role in the conversation flow, which turn management is an important part of (Du-Babcock, 2003; Allwood et al., 2007; Tanaka, 2008).

According to the Conversation Analysis (CA) tradition, starting with (Sacks et al., 1974), the conversation flow is regulated according to a set of so-called turn-taking rules consisting of a turn-construction and a turn-allocation component. According to these rules, the conversation participants alternate their speech smoothly. The turn-taking system has been criticised because it presupposes that the interaction flow is determined by pre-defined rules, thus it does not account for the fact that turn management as conversations also depend on social and cultural factors, inter alia (O'Connell et al., 1990; Cowley, 1998). Furthermore, more analyses of conversations indicate that the ideal smooth turn-taking system described by (Sacks et al., 1974) does not occur: in conversations participants often speak at the same time. Also long pauses occur and are acceptable in many conversations (O'Connell et al., 1990; Cowley, 1998; Campbell, 2008, 2009a). Recognising that many speech overlaps can occur in conversations, (Schegloff, 2000) defines rules describing how overlaps are managed. Furthermore, he distinguishes between non problematic and problematic overlap cases. Only in the latter, overlap management is needed. Differing from this CA-related view, other researchers suggest that co-occurring speech, and co-occurring body behaviours, are a natural aspect of conversations, signalling that people communicate in synchrony (Campbell, 2009b).

Independently of the research tradition, however, most studies point out that the interaction, flow, comprising both sequential and overlapping behaviours, is regulated by multimodal cues, thus involving both speech and non-verbal behaviours, inter alia (Kendon, 1967; Yngve, 1970; Ford and Thompson, 1996; Duncan, 1972; Allwood et al., 2007; Hadar et al., 1984b).

In this paper we present an analysis of how head movements, hand movements, facial expressions and body postures relate to turn management in a corpus of dyadic Danish first encounter conversations, and we compare these findings to results from preceding studies of the same topic.

Furthermore, we investigate the relation between body behaviours and turn management types by applying supervised machine learning to the annotated corpus data to test how far the shape of the body behaviours and the co-occurring speech can be used to carry out automatic classification of the turn management type of the behaviours. The focus in the paper is on the investigation of the contribution of each type of behaviour to turn management, thus both in the analysis and classification each behaviour type is considered separately. Although the expression of a turn management behaviour is probably a combination of speech and different gestural cues, at this stage of our research we wanted to find out to what degree and in what way each channel (head, face, body) contributed to the phenomenon of turn management. For each channel, however, we consider the whole range of movement types

available in the corpus. The rest of the paper is organised as follows. In section 2 we discuss relevant background literature while in section 3 we describe our data. Section 4 contains the analysis of turn management in the data and in section 5 we describe and discuss the classification experiments conducted on the corpus annotations. Finally, we conclude and present future work in section 6.

2 Background Literature

Many researchers (Kendon, 1967; Argyle and Cook, 1976; Duncan and Fiske, 1977; Goodwin, 1981) recognise the importance of body behaviours, especially gaze, head movements and hand gestures in turn management. (Kendon, 1967) and (Argyle and Cook, 1976) focus on the role of gaze direction and of mutual gaze, respectively, while turn management multimodal cues comprising intonation, syntax and hand gesture are described in (Duncan, 1972). In (Duncan and Fiske, 1977), also the behaviours of the listener are analysed, and backchannelling signals by the listeners are distinguished from regular turns. (Hadar et al., 1984b) find that linear movements of the head ("postural shifts") tended to occur after "grammatical" pauses (between clauses or sentences) and towards the initiation of speech, both between speaking turns and between syntactic boundaries inside speaking turns. Their analysis is based on the automatically collected head movements of four subjects engaged in conversations. The authors conclude that head movements are involved in regulating turn taking and marking syntactic boundaries inside speaking turns. Smaller and quicker movements tended to occur after dysfluencies inside grammatical boundaries, especially after short pauses (Hadar et al., 1984a).

(Gravano and Hirschberg, 2009) examine the relation between particular acoustic and prosodic turn-yielding cues and turn taking in a large corpus of task-oriented dialogues.

Machine learning has been applied to annotated multimodal data in numerous studies. For example, feedback nods and shakes are successfully predicted from speech, prosody and eye gaze in a multimodal corpus of conversations (Morency et al., 2005, 2007, 2009), while (Jokinen and Ragni, 2007) train machine learning algorithms on manually annotated multimodal data in order to recognise some of the communicative functions of head movements and facial expressions. They achieve promising results although their data is small.

In the rest of the paper, we discuss to what extent body behavioural cues are also present in our data in connection to turn management and, in line with preceding machine learning experiments we train classifiers on speech and body behaviours to predict the communicative function of these behaviours, in this case their turn management function.

3 The Corpus

The corpus which we have used in our analysis is a Danish corpus of first encounter conversations which is freely available for research. The corpus was collected and annotated during the NOrdic Multimodal Corpora (NOMCO) project, and it is part of a Nordic corpus of comparable recorded conversations in Danish, Finnish and Swedish (Paggio et al., 2010; Navarretta et al., 2011, 2012). The Danish first encounter dialogues comprise 12 five-minutes conversations between two young people who did not know each other in advance. The participants are 6 female and 6 male university students or university educated young people, aged 19-36 years. They were instructed to talk in order to get acquainted, as if they met at a party. The interactions were recorded by three cameras in a studio at the University of Copenhagen. Each person participated in two conversations, one with a

female and one with a male. The two conversations were recorded on two different days (Paggio and Navarretta, 2011). In Figure 1 and 2 snapshots from one of the recordings are given.

Figure 1: Snapshot from the corpus: frontal camera views

Figure 2: Snapshot from the corpus: side camera view

3.1 The Annotations

The corpus is orthographically transcribed with time stamps at the word level, and the body behaviours are annotated with pre-defined shape and function features according to the MUMIN annotation scheme (Allwood et al., 2007). Body behaviours can be assigned more functions at the same time, and they can be linked to speech segments produced by the gesturer or the interlocutor if the coders judge that they are semantically related.

The data were annotated by a coder and than corrected by a second coder. Disagreement cases were resolved by a third coder. An agreed upon version of the data was created. The coders were instructed to take into account the whole context when annotating. This comprises the multimodal behaviours of both speakers. With respect to turn management, the coders were instructed to annotate signals of turn management by the participants. Note that an actual turn change is not considered a necessary prerequisite for the assignment of a turn management feature. For instance, a turn elicit or a turn take signal can be coded independently of its success. Likewise, a turn hold signal can be coded if the speaker is attempting to keep the turn even though the interlocutor may take it from them.

Inter-coder agreement tests on the annotations of facial expressions and head movements resulted in kappa scores (Cohen, 1960) between 0.6-0.9 depending on the categories. The scores comprise segmentation and classification. The lowest scores were due to differences in the segmentation of facial expressions. The highest scores were achieved in the annotation of head movements. Agreement on the classification of functional categories was 0.82. More details about the corpus annotation and inter-coder agreement experiments are given in (Paggio and Navarretta, 2011; Navarretta et al., 2011). In this study, we used the annotations of head movements, facial expressions and body postures related to turn management. The shape features describing these behaviours are in Table 1. Head movements are

Shape attribute	**Shape values**
HeadMovement	Nod, Jerk, HeadForward, HeadBackward,Tilt, SideTurn, Shake, Waggle, HeadOther
HeadRepetition	Single, Repeated
General face	Smile, Laugh, Scowl, FaceOther
Eyebrows	Frown, Raise, BrowsOther
BodyDirection	BodyForward,BodyBackward, BodyUp, BodyDown, BodySide, BodyTurn, BodyDirectionOther
BodyInterlocutor	BodyToInterlocutor, BodyAwayFromInterlocutor
Shoulders	Shrug, ShouldersOther

Table 1: Shape attributes and values

annotated with features describing the form of the movement and an indication of whether the movement is performed once or more times. A general face attribute is used to annotate facial expressions together with the form of the eyebrows. Finally, body postures are annotated with information on direction, whether the body is facing the interlocutor, and what the movement of the shoulders is.

We distinguish six types of turn related behaviours:

- TurnTake: the speaker signals that she wants to take a turn that wasn't offered, possibly by interrupting;
- TurnHold: the speaker signals that she wishes to keep the turn;
- TurnAccept: the speaker signals that she is accepting a turn that is being offered;
- TurnYield: the speaker signals that she is releasing the turn under pressure;
- TurnElicit: the speaker signals that she is offering the turn to the interlocutor;
- TurnComplete: the speakers signals that she has completed the turn.

In the ANVIL tool, each modality is coded in a so called track, while links joining an annotation from a track to the other indicate that the two annotations are semantically related.

4 The Analysis

The corpus consists of 18000 speech tokens, 3117 head movements, 1448 facial expressions and 982 body postures. Out of these, 738 head movements, 247 facial expressions and 223 body postures have been labelled with a turn management function. Thus, 24% of the head movements, 17% of the facial expressions and 23% of the body postures have a turn

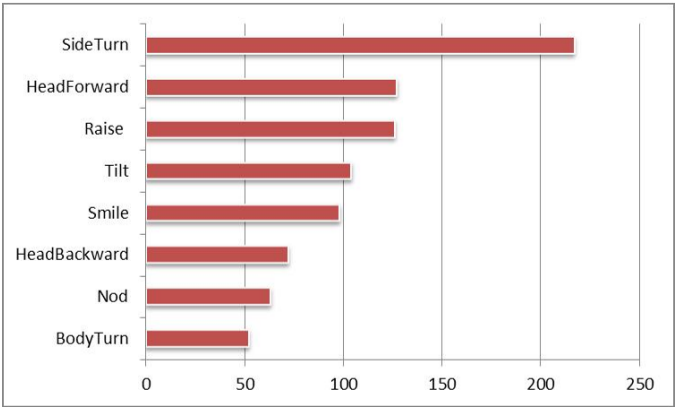

Figure 3: Most frequently occurring turn management behaviours

management function in the corpus. Table 2 shows, conversely, how the turn management associated with body behaviours is distributed across the three different types.

Body behaviour	**%**
Head movements	61
Facial expressions	20.5
Body postures	18.5

Table 2: Turn management distribution across body behaviours

Figure 3 shows the body behaviours which are most frequently related to a turn management function in this corpus. It is interesting to notice that all three body parts are related to turn management, and not only head movements, hand gestures and gaze which have been studied in preceding studies. Furthermore, more types of head movement than those indicated in the literature are relevant to turn management.

The most frequently assigned turn management categories are TurnHold, TurnAccept and TurnElicit, while TurnYield is rare in the corpus. This is not surprising given the type of social activity and the setting: people who meet for the first time and have to get acquainted are in general both friendly and polite, and they avoid interrupting their interlocutor. Figure 4 shows the turn management categories which are assigned more frequently to the three main types of body behaviour in the first encounter conversations.

The most common turn management functions of head movements are TurnHold, TurnAccept and TurnElicit, while TurnElicit, TurnAccept and TurnTake are the most frequent turn management functions of the facial expressions. Finally, TurnAccept, TurnElicit and TurnHold are the functions must frequently assigned to body postures. It must be noted that here we report each body part independently, but in conversations often several body behaviours co-occur.

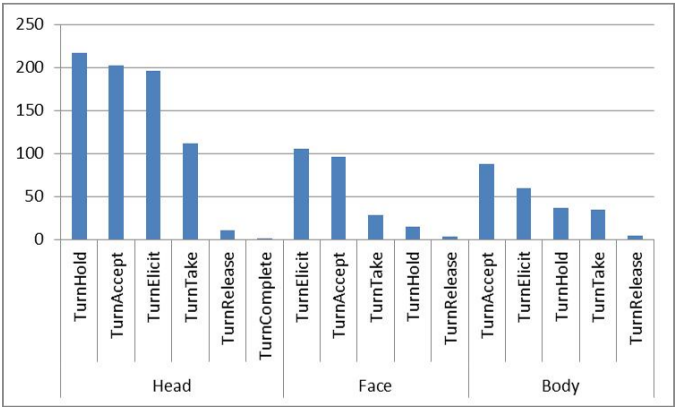

Figure 4: Turn management related types and body behaviours

4.1 Discussion

Our study of body behaviours and turn management indicates that not only head movements, but also facial expressions and body postures are often used in turn management in the Danish first encounter dialogues. The data show that the proportion of different turn management features varies depending on the body part involved. Especially the proportion of TurnHold changes a lot from being rather large in connection with head movements to being low for face and body, whereas TurnElicit is the feature that has the largest proportion only in connection with facial expressions. In the corpus, behaviours related to the three body parts are not explicitly linked to each other. Therefore, in the present stage of our research, we do not yet know whether the relations we see between body behaviours and turn management are additive, as could be expected for features associated with all three body parts, e.g. TurnAccept, or whether different body parts are associated with different turn behaviours.

Another observation relates to the fact that various types of head movements are involved in turn management. This finding is in line with other studies of Danish conversations which have shown that the participants not only use nods and shakes to give or elicit feedback, but also tilts, side movements and forward and backward movements (Paggio and Navarretta, 2011; Navarretta, 2011).

Approximately 20% of the communicative head movements, facial expressions and body postures have a turn management function in the conversations. The frequency of the turn management values assigned in the corpus, in particular the fact that very few occurrences of turn release under pressure are found, reflect the type of social interaction. The participants meet for the first time, thus they are kind and avoid interrupting each other.

5 Classification of Turn Management Types

In this section we describe classification experiments on the annotated corpus. We investigate to which extent it is possible to carry out automatic classification of the turn management types expressed by body behaviours on the basis of their shape, together with the co-occurring

words. In these experiments we again consider each body part independently, thus the features modelling the shape of each body behaviour and the words they are semantically connected with were extracted and used as different datasets.

More specifically, we trained classifiers on datasets containing descriptions of the shape of each body behaviour and its turn management function. Then we run the classifiers on data where only shape values were considered, in order to predict the turn management function of the behaviours. In succeeding experiments we also included the gesturer's or interlocutor's words linked to each body behaviour with the purpose of investigating whether the speech-related information improved the classification.

All experiments were run in WEKA (Witten and Frank, 2005). As baseline in each experiment, we use the results from WEKA's ZeroR algorithm which always chooses the most frequently occurring nominal class in the dataset. The most frequently occurring nominal class is a reasonable baseline for this data, given the fact that there are no other studies of multimodal turn management involving the same types of behaviour which we could compare our results with. An alternative possibility one could consider is taking actual turn change as a baseline measure. However, since the assignment of turn management features, as was explained earlier, does not necessarily imply a change of turn (or keeping the turn in the case of a TurnHold feature), such a baseline would make little sense. The classifier run in all experiments is WEKA's SMO algorithm, which implements John Platt's sequential minimal optimization algorithm for training a support vector classifier. SMO was chosen because it is the best performing algorithm on this type of data (Navarretta and Paggio, 2010)[1]. We report the results of classification in term of Precision (**P**), Recall (**R**) and F-measure (**F**) figures. The results were evaluated using ten-fold cross-validation.

Table 3 shows the results obtained for head movements, while Table 4 and Table 5 contain the results for facial expressions and body postures, respectively. In all cases, the classifiers are first trained on shape features for each body behaviours, then on shape features and the related gesturer's words (*gest-words* in the tables), on shape features and the related interlocutor's words (*interl-words*) and, finally, on shape features and gesturer's or interlocutor's words (*all-words*).

Algorithm	**Dataset**	**P**	**R**	**F**
ZeroR		0.09	0.29	0.13
SMO	shape	0.36	0.43	0.39
SMO	shape+gest-words	**0.47**	**0.5**	**0.48**
SMO	shape+interl-words	0.36	0.43	0.39
SMO	shape+all-words	0.46	0.5	0.46

Table 3: Identification and classification of turn management from head movements and related speech

The shape features of the head movements, and especially shape information enriched with the co-occurring gesturer's words, considerably improve the identification of the relevant turn management types of the head with respect to the baseline. In this case the improvement in terms of F-score with respect to the baseline is of 0.35%.

[1]The performance of various algorithms on the data was tested according to the strategy proposed in (Daelemans et al., 2003) and previously tested in (Navarretta, 2009). The other classifiers tested on these data are Naive Bayes, KStar and BFTree.

```
  a   b   c   d   e   f   <-- classified as
123  74   7  13   0   0 |   a = TurnHold
 41 137   2  16   0   0 |   b = TurnElicit
 40  53   2  17   0   0 |   c = TurnTake
 45  49   5 103   0   0 |   d = TurnAccept
  4   6   0   0   0   0 |   e = TurnYield
  0   1   0   0   0   0 |   f = TurnComplete
```

Figure 5: Confusion matrix for head movement

As can be seen in the confusion matrix relevant to the third dataset, and shown in (Fig. 5), the best results are obtained by the classifier on the three features TurnHold, TurnElicit and TurnAccept, which are also the most represented in the data. Performance drops dramatically with TurnTake, which is the fourth most represented category with 202 tokens. TurnYield and TurnComplete are almost not represented, so the results for these two features are not interesting.

Algorithm	**Dataset**	**P**	**R**	**F**
ZeroR		0.18	0.43	0.25
SMO	shape	0.41	0.51	0.45
SMO	shape+gest-words	**0.42**	**0.51**	**0.46**
SMO	shape+interl-words	0.41	0.51	0.45
SMO	shape+all-words	0.42	0.50	0.45

Table 4: Identification and classification of turn management from face and related speech

Turning now to facial expressions, we see again an improvement with respect to the baseline, this time of 0.21%. However, it makes very little difference whether words are added to the training. As it was to be expected given the distribution of the classes shown earlier in Figure 4, the classifier does best at identifying TurnElicit and TurnAccept, whilst it cannot recognise the other three turn behaviours.

Algorithm	**Dataset**	**P**	**R**	**F**
ZeroR		0.16	0.4	0.22
SMO	shape	0.42	0.4	0.35
SMO	shape+gest-words	0.42	0.44	0.39
SMO	shape+interl-words	0.38	0.41	0.35
SMO	shape+all-words	**0.43**	**0.45**	**0.4**

Table 5: Identification and classification of turn management from body postures and related speech

Finally, for body postures, using shape information and all words (both the gesturer's and the interlocutor's) gives the best classification results (F-measure improves with 18% with respect to the baseline), although the contribution of words is not as important as it is for head movements. Once again, TurnElicit and TurnAccept are the features that the classifier identifies with the highest accuracy (both are classified correctly 63% of the time).

5.1 Discussion of experimental results

All our classification experiments show that the shape of body behaviours can be used for the classification of the turn management type associated with the behaviours obtaining an F-score of around 0.4. The best results are obtained for head movements and body postures if also the gesturer's co-occurring words are considered. The words of the interlocutor, on the other hand, do not improve the results, but it must be noted that body behaviours related to turn management are related to the gesturer's own words more often than to the interlocutor's. As far as facial expressions are concerned, on the other hand, considering the words that co-occur with the expression makes little difference to the results. This is probably due to the fact that facial expressions are often quite long and may co-occur with long stretches of speech, so that the word tokens provide rather diverse and therefore potentially confusing information.

The best classification results are obtained using head movements. This is not uprising since head movements provide for the largest number of body behaviours in the data. However, not all classes are identified with the same accuracy. Whilst TurnHold, TurnElicit and TurnAccept are classified with an accuracy in the 0.5-0.6 range, TurnTake, TakeYield and TurnComplete seem very difficult to identify. It is somewhat puzzling that TurnTake should perform so poorly, given its frequency. It remains to be seen whether turn taking behaviour can be classified by means of different features, e.g. prosody patterns, or whether more data are necessary to achieve higher accuracy.

Turning now to the results achieved with facial expressions and body posture, it is again not surprising that TurnElicit and TurnAccept, the most frequent classes in this case, are those for which the classification achieves reasonable accuracy. However, it is interesting to note that TurnElicit performs a little better than TurnAccept in both sets of experiments, even though it is less frequently represented in conjunction with body postures. Presumably, the words associated with TurnElicit help optimise the classification.

In general, the results of these experiments are promising given the restricted size of the corpus. However, the F-score we obtain is not really high enough to allow for automatic annotation of the turn management of body behaviours. However, we believe there is scope for improvement in that co-occurring body behaviours, as was explained earlier, were not considered together in this study. Furthermore, additional cues, such as prosody patterns or actual turn change, could be added to the datasets to improve classification accuracy especially for the TurnTake class, which the classifiers we have trained in this study cannot identify on the basis of body features and associated words.

6 Conclusions and Future Work

In this paper we have presented an analysis of turn management behaviours in the Danish NOMCO first encounters corpus. Our study both confirms preceding studies on turn management, and provides new insights. In particular, we found that all kinds of head movement, facial expression and body posture are involved in turn management. Furthermore, the turn management types frequently occurring in this corpus depend on the type of social activity in which the participants are involved. Since a number of studies of multimodal conversations have found that there is a relation between factors such as social activity, cultural environment, number of participants and communicative body behaviours (Lu et al., 2011; Maynard, 1987; Navarretta et al., 2012), we also plan to compare turn management

behaviours in this corpus with corresponding behaviours in other types of corpora.

Acknowledgements

The collection and annotation of the NOMCO corpus was 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). We would like to thank Anette Luff Studsgård, Sara Andersen, Bjørn Wessel-Tolvig and Magdalena Lis for their codings. Special thanks also go to the NOMCO project's partners, Elisabeth Ahlsén, Jens Allwood and Kristiina Jokinen.

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. *Multimodal Corpora for Modelling Human Multimodal Behaviour. Special Issue of the International Journal of Language Resources and Evaluation*, 41(3–4):273–287.

Argyle, M. and Cook, M. (1976). *Gaze and mutual gaze*. Cambridge University Press, Cambridge, UK.

Campbell, N. (2008). Multimodal processing of discourse information; the effect of synchrony. In *Second International Symposium on Universal Communication*, pages 12–14.

Campbell, N. (2009a). An audio-visual approach to measuring discourse synchrony in m ultimodal conversation data. In *Proceedings of Interspeech 2009*, pages 12–14.

Campbell, N. (2009b). Multimodal Processing of Discourse Information; The Effect of Synchrony. In *ISUC - Second International Symposium on Universal Communication*, pages 12–15, Osaka.

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

Cowley, S. J. (1998). Of Timing, Turn-Taking, and Conversations. *Journal of Psycholinguistic Research*, 27(5):541–571.

Daelemans, W., Hoste, V., Meulder, F. D., and Naudts, B. (2003). Combined Optimization of Feature Selection and Algorithm Parameter Interaction in Machine Learning of Language. In *Proceedings of the 14th European Conference on Machine Learning (ECML-2003)*, pages 84–95, Cavtat-Dubrovnik, Croatia.

Du-Babcock, B. (2003). A comparative analysis of individual communication processes in small group behavior between homogeneous and heterogeneous groups. In *Proceedings of the 68th Association of Business Communication Convention*, pages 1–16, Albuquerque, New Mexico, USA.

Duncan, S. Junior (1972). Some Signals and Rules for Taking Speaking Turns in Conversations. *Journal of Personality and Social Psychology*, 23(2):283–292.

Duncan, S. Junior. and Fiske, D. (1977). *Face-to-face interaction*. Erlbaum, Hillsdale, NJ.

Ford, C. and Thompson, S. (1996). Interactional Units in Conversation: Syntactic, Intonational, and Pragmatic Resources for the Management of Turns. In Ochs, E., Schegloff, E., and Thompson, S., editors, *Interaction and Grammar*, pages 134–184. Cambridge University Press, Cambridge.

Goodwin, C. (1981). *Conversational organization: Interaction between speakers and hearers*. Academic Press, New York.

Gravano, A. and Hirschberg, J. (2009). Turn-yielding cues in task-oriented dialogue. In *Proceedings of SIGDIAL 2009: the 10th Annual Meeting of the Special Interest Group in Discourse and Dialogue, September 2009*, pages 253–261, Queen Mary University of London.

Hadar, U., Steiner, T., and Rose, F. C. (1984a). The Relationship Between Head Movements and Speech Dysfluencies. *Language and Speech*, 27(4):333–342.

Hadar, U., Steiner, T., and Rose, F. C. (1984b). The timing of shifts of head postures during conversation. *Human Movement Science*, 3(3):237–245.

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

Kendon, A. (1967). Some functions of gaze-direction in social interaction. *Acta Psychologica*, 26:22–63.

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 Heylen, D., Kipp, M., and Paggio, P., editors, *Proceedings of the workshop on Multimodal Corpora at ICMI-MLMI 2011*, Alicante, Spain.

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

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

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

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

Navarretta, C. (2009). Automatic recognition of the function of third-person singular pronouns in texts and spoken data. In S. Lalitha Devi, A. B. and Mitkov, R., editors, *Anaphora Processing and Applications. 7th Discourse Anaphora and Anaphor Resolution Colloquium - DAARC 2009, Goa, India Proceedings*, volume 5847 of *LNAI*, pages 15–28. Springer Verlag., Berlin/Heidelberg.

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

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

Navarretta, C., Ahlsén, E., Allwood, J., Jokinen, K., and Paggio, P. (2012). Feedback in nordic first-encounters: a comparative study. In *Proceedings of LREC 2012*, pages 2494–2499, Istanbul Turkey.

Navarretta, C. and Paggio, P. (2010). 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, Upssala, Sweden.

O'Connell, D., Kowal, S., and Kaltenbacher, E. (1990). Turn-Taking: A Critical Analysis of the Research Tradition. *Journal of Psycholinguistic Research*, 19(6):345–373.

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

Paggio, P. and Navarretta, C. (2011). Head movements, facial expressions and feedback in Danish first encounters interactions: a culture-specific analysis. In Stephanidis, C., editor, *Universal Access in Human-Computer Interaction. Users Diversity. Proceedings of 6th International Conference, UAHCI 2011, Held as Part of HCI International 2011*, pages 583–590, Orlando, FL, USA. Springer.

Sacks, H., Schegloff, E., and Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4):696–735.

Schegloff, E. A. (2000). Overlapping talk and the organization of turn-taking for conversation. *Language in Society*, 29:1–63.

Tanaka, H. (2008). Communication strategies and cultural assumptions: An analysis of French-Japanese business meetings. In Tietze, S., editor, *International Management and Language*, pages 154–170. Routledge, New York, NY.

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

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