Multimodal Turn Management in Danish Dyadic First Encounters

Costanza Navarretta
University of Copenhagen
Copenhagen, Denmark
costanza@hum.ku.dk

Patrizia Paggio
University of Copenhagen and
University of Malta
Valletta, Malta
paggio@hum.ku.dk

Abstract

This paper presents studies of multimodal turn management in a Danish corpus of video recorded conversations between two young people who meet for the first time. More specifically, we investigate multimodal behaviours by which the conversation participants indicate whether they wish to give, take or keep the turn. In this study we present quantitative analyses of such cues, as well as an investigation of the length of the individual participants' speech contributions. The quantitative studies comprise 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 synchronously in communication. The investigation of the participants' vocal contributions shows gender differences in that male participants talked more when interacting with a female than they did when their interlocutor was a male, while female participants talked more when interacting with a female than when they interacted with a male.

Keywords: multimodal communication, multimodal corpora, turn management, annotation

1 Introduction

Human-human communication is multimodal by nature because people communicate through both speech and non-verbal behaviours. This article is about speech and non-verbal[1] cues with a turn management function in the Danish NOMCO first encounters corpus (Paggio and Navarretta, 2011).

Turn management is the process by which conversation participants regulate the interaction flow (Allwood et al., 2007). This is done by both verbal and body behavioural cues (Kendon, 1967; Yngve, 1970; Ford and Thompson, 1996; Duncan, 1972; Allwood et al., 2007; Hadar et al., 1984).

Sacks et al. (1974) propose pre-defined *turn-taking* rules to model the way in which the participants regulate their turn flow smoothly, in other words avoiding too long pauses and speech overlaps. Schegloff (2000) adds to the turn-taking system an overlap-management system in order to account for the many overlaps which have been found in real life conversations, see inter alia (O'Connell et al., 1990; Cowley, 1998). Overlap management is only needed, according to Schegloff, when overlap is problematic, that is when it occurs for longer time.

Several researchers disagree with the view of a pre-defined turn-taking system because turn-taking depends on many factors such as the conversation setting, the cultural environment, the degree of familiarity and the number of the participants (O'Connell et al., 1990; Cowley, 1998; Du-Babcock, 2003; Tanaka, 2008). Furthermore, silence and overlaps, in both speech and bodily behaviours, should be seen as a natural part of conversation, signalling that people communicate in synchrony (Campbell, 2009; Esposito et al., 2010).

In this paper, we present two studies of turn management cues expressed by various body behaviours comprising head movements, gaze, hand movements, facial expressions and body postures in a corpus of dyadic Danish first encounters, and we relate these findings to the literature. Then, we focus on how long the participants's turns are depending on their different interlocutors. The rest of the paper is organised as follows. In section 2 we discuss relevant background literature while in section 3 we describe our corpus and the relevant

---

[1] In this paper, we use non-verbal in the sense of pertaining to body behaviour rather than speech.

annotations. In section 4 we present our analyses of the data, section 5 contains the discussion, and in section 6 we conclude and present future work.

2   Background Studies

Different studies underline the central role of specific vocal and body behaviours in turn management. For example, Kendon (1967) and Argyle and Cook (1976) investigate the role of gaze direction and of mutual gaze, respectively, while Duncan (1972) annotates speech and body behaviours in dyadic English conversations and identifies verbal and non-verbal cues by which speakers signal that they want to give the turn to the interlocutor (*turn-yielding* in Duncan's terminology). The verbal cues which Duncan defines comprise the following phenomena: a) intonation, b) the use of hedges, such as *you know* and *I guess*, and c) syntax. Only one non-verbal cue indicating that the speaker is terminating the turn is identified: the completion of on-going hand gestures. In Duncan's conversations at least two cues co-occurred when speakers showed that they wanted to release the turn.

Hadar et al. (1984) analyse the occurrences of head movements in conversations between four participants. In these data, they find that linear movements of the head ("postural shifts") often occur after "grammatical" pauses both between clauses and sentences. Furthermore, postural shifts are also identified towards the initiation of speech, both between speaking turns and between syntactic boundaries inside speaking turns. Their conclusion is that head movements are involved in regulating turn taking and marking syntactic boundaries inside speaking turns. They also find that smaller and quicker movements tend to occur after dysfluencies inside grammatical boundaries, especially after short pauses.

Differing from the other researchers, Duncan and Fiske (1977) focus on the behaviours of the listeners. They propose to distinguish backchannelling signals which do not provide new semantic information from regular turns..

In the rest of the paper, we discuss to what extent speech cues and body behaviours are also present in our data in connection to turn management.

3   The Data

Our data is the Danish NOMCO corpus of first encounters[2]. The corpus was collected and annotated under the NOrdic Multimodal COrpora (NOMCO) project. Comparable conversations of first encounters were also collected and annotated in Finnish and Swedish (Paggio et al., 2010; Navarretta et al., 2011; Navarretta et al., 2012).

The Danish corpus consists of 12 dyadic conversations with the length of approximately 5 minutes each. The participants are all young people, university students or university educated, aged from 21 to 36. The participant population comprised 6 females and 6 males who had a common acquaintance, but did not know their interlocutors in advance. Each subject participated in two conversations, one with a female participant and one with a male participant, and the two conversations were recorded on two different days (Paggio and Navarretta, 2011). The participants were instructed to talk in order to get acquainted, as if they met at a party, and they were only told that they participated in a project on Danish.

The interactions were recorded by three cameras at the University of Copenhagen. Frontal views of each subject and a panorama view of the two participants standing in front of each other are available. The three camera views are shown in 1 and 2.

Figure 1: Snapshot from a conversation: frontal camera views

3.1   The annotations

The corpus was orthographically transcribed in PRAAT (Boersma and Weenink, 2009) with time stamps at the word level. Stress and phrasal information are also available. The transcriptions

---

[2]This section is almost identical to the description of the corpus provided in earlier papers.

Figure 2: Snapshot from a conversation: side camera view

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

Table 1: Shape Features of Head Movements, Facial Expressions and Body Postures

were imported in the multimodal annotation tool ANVIL (Kipp, 2004) and the body behaviours were annotated according to the MUMIN annotation scheme (Allwood et al., 2007). The scheme provides pre-defined features describing the shape and function of gestures and their semiotic type (Peirce, 1931). Since body behaviours are multifunctional, they can be assigned more functions at the same time. Body behaviours which are judged to be semantically related to speech segments produced by the gesturer or the interlocutor can be linked explicitly to these in the annotation (Allwood et al., 2007).

The body behaviours were annotated by a coder and then corrected by a second coder. Disagreement cases were resolved by a senior coder. The analyses discussed in this paper are based on the final concerted version (Paggio and Navarretta, 2011; P.Paggio and Navarretta, 2012).

Navarretta et al. (2011) give an account of intercoder agreement tests on the annotations which, depending on the categories, resulted in kappa scores (Cohen, 1960) between 0.60-0.90.

In this study, we use the annotations of head movements, facial expressions and body postures related to turn management. The features describing the shape of these behaviours are presented in Table 1. Body postures are annotated with information on direction, whether the body is facing the interlocutor, and what the movement of the shoulders is. Facial expressions are described with a general face attribute and an eyebrows attribute. Finally, head movements are described by the form of the movement and an attribute indicating whether the movement is performed one or more times.

The MUMIN scheme distinguishes the following six turn management behaviours:

- TurnTake: signals that the speaker wants to take a turn that wasn't offered, possibly by interrupting

- TurnHold: signals that the speaker wishes to keep the turn

- TurnAccept: signals that the speaker is accepting a turn that is being offered

- TurnYield: signals that the speaker is releasing the turn under pressure

- TurnElicit: signals that the speakers is offering the turn to the interlocutor

- TurnComplete: signals that the speaker has completed the turn.

4 Turn Management in the Corpus

4.1 Turn management and gesture types

The corpus contains 18000 speech tokens comprising filled pauses. Table 2 shows the body behaviours annotated in the corpus, the body behaviours with a turn management function, and their percentage.

Table 3 shows how the turn management associated with body behaviours is distributed across the three different behavioural types.

In table 4, the body behaviours which are most frequently related to a turn management function in this corpus are shown.

Behaviour	Total	Turn	%
Head	3117	738	24
Face	1448	247	17
Body	982	223	23
Total	5547	1208	100

Table 2: Turn management body behaviours

Behaviour	%
Head movements	61
Facial expressions	20.5
Body postures	18.5

Table 3: Turn management distribution across body behaviours

The tables show that all three body parts are related to turn management, and not only head movements, hand gestures and gaze on which preceding studies mainly have focused. The second table also shows that more types of head movement than those indicated in the literature are relevant to turn management in this corpus. Interesting are especially the occurrences of forward and backward movements of the head which have not been related earlier to turn management.

Some of these head movements may be accompanied by movements of the torso and the body. However, we have not considered whole body behaviour here.

The most frequently assigned turn management categories are TurnHold, TurnAccept and TurnElicit, while TurnYield, which in our coding scheme is used to code turn/releasing under pressure, is extremely rare in the corpus. This reflects the type of social activity and the communicative situation: people who meet for the first time are

Turn M. Behaviour	No.
SideTurn	217
HeadForward	127
EyebrowsRaise	126
Tilt	104
Smile Shake	98
HeadBackward	76
Nod	72
BodyTurn	63
	52

Table 4: Most frequent turn management behaviours

Behaviour	Turn M.Type	No.
Head	TurnHold	217
	TurnAccept	202
	TurnElicit	196
	TurnTake	112
	TurnYield	10
Face	TurnElicit	105
	TurnAccept	96
	TurnTake	28
	TurnHold	15
	TurnYield	3
Body	TurnAccept	88
	TurnElicit	60
	TurnHold	37
	TurnTake	34
	TurnYield	4

Table 5: Turn management related types and body behaviours

both friendly and polite and, in Denmark, this also implies not interrupting the interlocutor. Table 4.1 shows the turn management categories which are assigned more frequently to head movement, facial expressions and body postures.

The table indicates that each body behaviour type is mostly related to two or three specific turn management functions: for head movements these are TurnHold, TurnAccept and TurnElicit; for facial expressions they are TurnElicit and TurnAccept; and finally for body postures, they are TurnAccept and TurnElicit.

4.2 Multimodal turn shift cues

In the above analysis we have looked at each body part independently, however in the data several body behaviours often co-occur. A more truly multimodal approach is presented below in an analysis of part of the corpus. More specifically, we investigate in two first encounter conversations whether the turn offering cues observed in English conversations by Duncan (1972) and described earlier, also hold in our corpus. As already noted, Duncan finds that at least one of six cues connected to intonation, speech content and hand gestures occurred in turn eliciting situations. When more than a cue is present, they occur simultaneously or in tight sequences.

In this study, we only include turn alternations without overlapping speech since we want to verify on our data Duncan's results (1972) work. We

also exclude answers to direct questions because the question is an explicit turn eliciting cue. The turn shifts relevant to our study are 35% of the turn shifts in the two conversations. In 95% of these, the speaker concludes a syntactic phrase. The pitch level, on the other hand, only goes down at the end of the phrase in 10% of the cases. The speaker's head and body freeze in all cases, and similarly, the speaker and interlocutor always have eye contact. The participants in the two conversations analysed do not move their hands very frequently. Thus, the speaker finishes talking and moving their hands nearly simultaneously only in three of the turn shifts considered here. The syntactic completion of a phrase and a pitch different from an intermediate pitch level are two of the cues described by Duncan (Duncan, 1972), but we only found occurrences of the former in the two conversations. It can be argued that the freezing of the head and body are parallel to the termination of hand gestures by the speaker which Duncan mentions as the body cue associated with turn-yielding. Furthermore, also in our data the speaker ends the production of speech and co-occurring hand movement nearly simultaneously. As for the use of gaze, the fact that speakers look at their interlocutors to indicate that they want to release the turn confirms a tendency also found in other studies (Kendon, 1967; Argyle and Cook, 1976; Jokinen, 2011).

Duncan (1972) also finds that speakers signal that they are offering the turn to the interlocutor by using hedges such as *I think* and *I guess*.

We have only found few vowel lengthenings on the stressed syllable of a terminal clause in the two conversations, and there are no occurrences in the corpus of hedges in turn eliciting situations. In other words, the presence of hedges does not play a significant role in our corpus.

4.3   Turn length

The length of speakers' turns can be studied from different perspectives. In the third study reported here, we investigate whether there are differences in the amount of speech produced by the participants depending on the gender of the interlocutor. In particular, we look at the number of words uttered by each participant and the length of their turns. The pattern which results from this study is the following. Female participants utter more words and keep the floor for longer time when they talk with another female than they do when they interact with a male. Male participants, on the contrary, talk for longer time and utter more words when their interlocutor is a female than they do when the interlocutor is a male. This is true in all conversations with one exception in which the female participant speaks more than her male interlocutor. However, she also speaks more than her female interlocutor and is in fact the participant who talks more than all the other participants. In sum, she can be considered an outlier in the sample.

5   Discussion

Since more than 20% of the communicative head movements, facial expressions and body postures annotated in the corpus have a turn management function, our study shows that all these behaviours often have a turn management function. This confirms the fact that humans use all their body when they synchronise their contributions.

We also see that different body parts are often related to specific turn management functions, an aspect that ought to be investigated further.

The less frequently occurring turn management functions in the corpus are TurnYield and TurnComplete (the latter only occurs once). The fact that only few occurrences of turn release under pressure are found, reflects the type of social interaction as well as the culture. The participants meet for the first time, thus they want to be kind and friendly, and avoid interrupting each other as a consequence. As for TurnComplete, the phenomenon is more relevant in different communication situations, e.g. interviews, in which the speaker is asked for something and stops talking when the answer to the question is complete.

Many types of head movement are involved in turn management in addition to nods and shakes, which many earlier studies have treated as typical turn shift signals. This finding is similar to the conclusion drawn in another analysis of the same corpus, in which we found that all types of head movements are also used by the participants to give or elicit feedback (Paggio and Navarretta, 2011).

Whilst the role of body cues for turn shift was analysed quantitatively across the whole corpus, two of the conversations were studied with special regard to how various syntactic and prosodic features contribute to turn management. In particular,

the analysis of turn eliciting cues in the two files confirms Duncan's observation that a speaker's completion of a syntactic phrase signals that the speaker wants to relinquish the turn. The presence of hedges, vowel lengthening and pitch change, on the other hand, are either seldom or never found in the data. Our analysis also shows that the speaker can signal that they are offering the turn to the interlocutor not only by finishing off on-going hand gestures, but also by bringing the head and body to a standstill while keeping their gaze on the interlocutor.

The investigation of how long the participants talk shows that five out of six men talk more when interacting with a person of the opposite gender than with one of the same gender. Conversely, female participants tend to be more talkative with an interlocutor of the same gender. Thus, the data indicate that there is a gender difference when males and females participate in first encounters. Whether this is a general tendency or an idiosyncratic characteristic of our corpus should be investigated further.

6 Conclusions and Future Work

In this paper we have presented three studies of turn management behaviours in the Danish NOMCO first encounters corpus. The first study, where we look at the type of body behaviours involved in turn management, confirms preceding studies on turn management, but it also provides new insights. In fact, our data shows that all kinds of head movement, facial expression and body posture are used as turn management cues. We also found a relation between the turn management types which occur most frequently in this corpus and the type of social activity in which the participants are involved. Furthermore, the data show that each body behaviour is often related to specific turn management types.

The second study, which investigates vocal and non-verbal cues of turn shift in part of the corpus, shows both similarities and dissimilarities with the cues found by Duncan in English dyadic conversations. More specifically, the syntactic cues hold also in our data, while the intonation patterns seem not to be the same in the two languages. We found also that the head and the body (Duncan focuses on the hands) complete ongoing movements to signal that the speaker is finishing the turn and is prepared to give the floor to the interlocutor. This analysis of co-occurring behaviours will in future be extended to the entire corpus.

In the third study, where we looked at turn length and number of words in a turn, and how these varied for each participant depending on the conversation, we found an interesting gender difference, which to our knowledge has not been reported earlier. In our corpus, thus, males talk more when they interact with females, while females talk more when their interlocutor is another female. However, we would like to test whether this tendency can also be seen in other datasets and in different types of conversations before making any general claim about gender differences.

Several issues of relevance to the phenomenon of turn management were not touched upon here and could be studied either based on this corpus or adding datasets for different languages. For example, we did not consider mirroring or synchronised behaviours of the participants, see e.g. (Campbell, 2009). Nor did we investigate whether the degree of familiarity of the interlocutors affects turn management, or to what extent turn management cues vary from language to language. Especially the last issue, which is in line with preceding comparative studies of body behaviours and other communicative functions in different cultures and communicative situations (Lu et al., Under publication; Maynard, 1987; Navarretta et al., 2012; Navarretta and Paggio, 2012), can be studied by comparing the findings described here with similar turn management data from parallel multimodal corpora from the NOMCO collection.

References

Allwood, J., Cerrato, L., Jokinen, K., Navarretta, C. & 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. & Cook, M. (1976). *Gaze and mutual gaze*. Cambridge University Press, Cambridge, UK.

Boersma, P. & Weenink, D. (2009). *Praat: doing phonetics by computer*. Retrieved May 1, 2009, from http://www.praat.org/.

Campbell. N. (2009). An audio-visual approach to mea- suring discourse synchrony in m ultimodal conversa- tion data. In *Proceedings of Interspeech 2009*, pp. 12–14.

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.

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. Jr. and D.W. Fiske, D. W. (1977). *Face-to-face in- teraction*. Erlbaum, Hillsdale, NJ.

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

Esposito, A., Campbell, N., Vogel, C., Hussain, A. & Nijholt, A. (eds). (2010). *Development of Multimodal Interfaces: Active Listening and Synchrony*, volume 5967 of *LNCS*. Springer Verlag.

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

Hadar, U. Steiner, T. J. & Clifford Rose, F. (1984). The timing of shifts of head postures during conversation. *Human Movement Science*, 3(3), 237–245.

Jokinen, K. (2011). Turn taking, utterance density, and gaze patterns as cues to conversational activity. In *Proceedings of ICMI-MMI*, Alicante, Spain, November.

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

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

Lu, J., Allwood, J. & Ahlsén, E. Under publication. A study on cultural variations of smile based on empirical recordings of Chinese and Swedish first encounters. In D. Heylen, M. Kipp, and P. Paggio, editors, *Proceedings of the workshop on Multimodal Corpora at ICMI-MLMI 2011*, Alicante, Spain, Nov.

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

Navarretta, C. & Paggio, P. (2012). Verbal and non-verbal feedback in different types of interactions. In *Proceedings of LREC 2012*, pp. 2338–2342, Istanbul Turkey, May.

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

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, May.

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

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

Paggio, P., Ahlsén, E., Allwood, J., Jokinen, K. & Navarretta, C. (2010). The NOMCO multimodal Nordic resource - goals and characteristics. In *Proceedings of LREC 2010*, pp. 2968–2973, Malta, May 17-23.

Peirce., C. S. (1931). *Collected Papers of Charles Sanders Peirce, 1931-1958, 8 vols*. Harvard University Press, Cambridge, MA.

Paggio, P. & Navarretta, C. (2012). Head movements, facial expressions and feedback in conversations - empirical evidence from danish multimodal data. *Journal on Multimodal User Interfaces - Special Issue on Multimodal Corpora*.

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

Schegloff, E. (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 S. Tietze, editor, *International Management and Language*, pp 154–170. Routledge, New York, NY.

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