Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020), pages 634–643
Marseille, 11–16 May 2020

Dialogue Act Annotation in a Multimodal Corpus of First Encounter Dialogues

Costanza Navarretta[1], **Patrizia Paggio**[1,2]
[1]University of Copenhagen, [2]University of Malta
costanza@hum.ku.dk, paggio@hum.ku.dk

Abstract

This paper deals with the annotation of dialogue acts in a multimodal corpus of first encounter dialogues, i.e. face-to- face dialogues in which two people who meet for the first time talk with no particular purpose other than just talking. More specifically, we describe the method used to annotate dialogue acts in the corpus, including the evaluation of the annotations. Then, we present descriptive statistics of the annotation, particularly focusing on which dialogue acts often follow each other across speakers and which dialogue acts overlap with gestural behaviour. Finally, we discuss how feedback is expressed in the corpus by means of feedback dialogue acts with or without co-occurring gestural behaviour, i.e. multimodal vs. unimodal feedback.

Keywords: Multimodal corpus, dialogue acts, multimodal feedback

1. Introduction

A dialogue act is a speech segment, or utterance, that has a communicative function in a conversation, and is thus a type of speech act (Searle, 1969). Dialogue acts are essential to the understanding of the dynamics and semantics of dialogues and therefore to the construction of dialogue systems.

This paper deals with the annotation of dialogue acts in a multimodal corpus of first encounter dialogues, i.e. face-to-face dialogues in which two people who meet for the first time talk with no particular purpose other than just talking. Contrary to most existing corpora displaying dialogue act tags, our corpus is not shaped by a domain specific task which the participants have to solve together, but it is guided by their willingness to talk to each other exchanging general information about themselves. Therefore, the distribution of dialogue act labels in the corpus is bound to be different from that of dialogues normally classified as task-oriented.

Furthermore, the corpus is multimodal. The dialogues have been video-recorded, and the gestural behaviour has been carefully annotated with respect to form, dynamics and functions of head movements, facial expressions and body posture. Therefore, the annotation of dialogue acts provides an additional layer of functional linguistic analysis which will enable analyses of how multimodal signals contribute to the structure and content of the dialogues.

The annotation of dialogue acts in a multimodal corpus is also relevant for the implementation of embodied conversational agents since the annotations allow not only to model the dialogue structure, but also the gestures together with the linguistic context.

The first encounters corpus dealt with in our work is particularly interesting, since it shows how subjects from a specific culture address each other when they are not acquainted and then exchange information about themselves. For this reason, first encounters have been collected and studied in projects investigating social conventions in different cultures, e.g. (Rehm et al., 2009). This type of information is crucial for behavioural models of culturally aware virtual agents and robots, particularly when they meet people for the first time.

In this paper we describe the method used to annotate dialogue acts in the corpus, including a report of inter-agreement test results. We provide descriptive statistics of the annotation, particularly looking at which dialogue acts tend to follow each other across speakers. We analyse the way head movements and facial expressions reinforce some of the most frequent dialogue acts occurring in the corpus. Finally, we discuss how feedback is modelled in the corpus by means of utterances and gestural behaviour considered together.

2. Background Studies

Various classifications have been proposed the past forty years to annotate and analyse dialogue acts in different domains, e.g. HCRC MapTask (Anderson et al., 1991), DAMSL (Allen and Core, 1997), Verbmobil (Alexandersson et al., 1997), and SWBD-DAMSL (Jurafsky et al., 1997). Common to all these classifications is the fact that they were proposed in order to construct specific dialogue systems and to provide the necessary background for the implementation of dialogue management strategies (Allen and Core, 1997; Alexandersson et al., 1997; Jurafsky et al., 1997). Finally, a few schemes were proposed in order to provide large annotated data for training and testing dialogue models. This was the case of the multimodal AMI corpus (Carletta, 2007), which provided many types of linguistic annotations, including dialogue acts, as well as gestural annotations in the domain of project meetings. The interest for dialogue act annotation is also reflected in the many efforts dedicated to automatic dialogue act labelling in different corpus types and languages by means of machine learning techniques (Verbree et al., 2006; Purver et al., 2007; Milajevs and Purver, 2014; Amanova et al., 2016).

In order to facilitate the interoperability between the various annotation frameworks, an effort has been made to create a standard for dialogue act annotation that would take into account and combine categories from the previous schemes. The result of that work is the ISO 24617 standard (Bunt et al., 2010; Bunt et al., 2017). Guidelines have also been provided to implement the ISO dialogue act annotation framework in the annotation tool for multimodal behaviour ANVIL (Bunt et al., 2012). According to this implementation, each functional dimension of the dialogue acts is annotated in a different ANVIL track. Finally, Bunt et al. (2019) have applied the ISO 24617 standard to re-annotate dialogue acts in a number of dialogue corpora which were previously coded according to different dialogue act schemes.

We decided to apply the ISO standard in our work to take advantage of the interoperability it offers. Furthermore, some of the categories implemented in the standard are similar to those used in the functional annotation of the gestural behaviour in our corpus, as will be explained below.

3. The corpus

The annotated corpus described here consists of 12 video-recorded dyadic conversations in Danish between six male and six female participants who meet each other for the first time. The corpus is about one hour long and was collected and annotated at the University of Copenhagen within the Nordic project NOMCO (Paggio et al., 2010). The participants were asked to stand in front of each other and talk freely for about five minutes to get acquainted with one another. Three different cameras and two cardioid microphones were used to video-record the conversations.

The already existing annotations follow the MUMIN framework (Allwood et al., 2007) and comprise, in addition to the orthographic transcription, labels referring to the form and functions of head movements, facial expressions and body postures, and their relation to the speech segment they are associated with (Paggio and Navarretta, 2017). Dialogue act labels were added as described in the next section.

The annotations can be made available for research purposes on request.

4. The annotation of dialogue acts

The annotation of dialogue acts was performed using the ANVIL tool (Kipp, 2004) as was also done previously for the annotation of gestural behaviour. A set of specifications was defined based on the ISO 24617-2:2012 standard[1].

The following dialogue act categories were included in the specifications:

- General Purpose Dialogue Acts:
  Accept-Offer, Accept-Request, Accept-Suggest, Address-Offer, Address-Request, Address-Suggest, Agreement, Answer, Check-Question, Choice-Question, Confirm, Correction, Decline-Offer, Decline-Request, Decline-Suggest, Disagreement, Disconfirm, Inform, Inform-Answer, Instruct, Offer, Promise, Propos-Question, Question, Request, Set-Question, Suggest.

- Interaction Structuring Dialogue Acts:
  AcceptApology, AcceptThanking, Apology, Confratulate, InitialGoodbye, InitialGreeting, InitialSelfIntroduction, ReturnGoodbye, ReturnGreeting, ReturnSelfIntroduction, Thanking.

- Feedback-related Dialogue Acts:
  AlloFeedbackGive, AlloFeedbackElicit, AutoFeedback.

- Related to Own Communication Management:
  OwnCommManagement, Retraction

It must be noted that in the current phase of the dialogue act annotation work, we do not distinguish between the various dimensions of dialogue acts and annotate them all on the same track by only picking one of the labels at a time. Furthermore, we focus on the types from the first three dimensions, and do not, as done in the standard, distinguish between *Own Communication Management* and *Time Management*. All relevant instances were annotated using *OwnCommManagement*. The reason for this is that the main annotator found it difficult to apply the *Time Management* label since in a non task-oriented corpus, pauses mostly signal that the speakers are planning their own speech (Maclay and Osgood, 1959; Chafe, 1974; Allwood et al., 2007). Relevant examples are pauses that occur after a retraction and precede the new speech as in the following exchange:

(1) *Det lyder da* - retraction
(It sounds actually)
PAUSE øhm - planning pause
(pause uhm)
Jeg synes det det lyder som om du er da på skinner
(I think it sounds as if you are actually on track)

These pauses are the most common in the corpus, while there are only few stalling examples in which the speakers explicitly tell the interlocutors that they are looking for the correct wording as in the following example:

(2) *øh hvad hedder det PAUSE jeg arbejder*
(uh what is it called PAUSE I work)

For the segmentation of dialogue acts, we followed the definition of a functional segment in the ISO standard: "A functional segment is a minimal stretch of communicative behaviour that has a communicative function" (ISO/DIS 24617-2, p.3). Examples of functional segments and the categories that have been assigned to them are shown below. Note that a dialogue act can be as short as a filler, and as long as a complex sentence:

- *Øhm* (filler) → OwnCommunicationManagement
- *Ja* (yes) → AlloFeedbackGive
- *Af en eller anden grund så var der ikke nogen der var hjemme* (for some reason, nobody was at home) → Inform

The annotation procedure was as follows. First a single coder annotated the dialogue acts in one conversation, and then the annotations were checked by two other coders. All coders were experts who were well-acquainted with the annotation methodology. Disagreements and problematic cases were discussed, and a pre-final version was produced based on the common understanding of the dialogue labels achieved through discussions of the first trial conversation.

[1] https://www.iso.org/standard/51967.html

Dialogue and speaker	Segmentation	Category	Overall (average)
F4-F1 speaker F4	Cohen's κ = 0.985 Corrected's κ = 0.989	Cohen's κ = 0.92 Corrected's κ = 0.944	Cohen's κ = 0.956 Corrected's κ = 0.98 Krippendorff's α = 0.956
F4-F1 speaker F1	Cohen's κ = 0.98 Corrected's κ = 0.98	Cohen's κ = 0.78 Corrected's κ = 0.89	Cohen's κ = 0.903 Corrected's κ = 0.938 Krippendorff's α = 0.903
M1-M5 speaker M1	Cohen's κ = 0.962 Corrected's κ = 0.965	Cohen's κ = 0.84 Corrected's κ = 0.88	Cohen's κ = 0.897 Corrected's κ = 0.941 Krippendorff's α = 0.9
M1-M5 speaker M5	Cohen's κ = 0.97 Corrected's κ = 0.97	Cohen's κ = 0.758 Corrected's κ = 0.824	Cohen's κ = 0.857 Corrected's κ = 0.899 Krippendorff's α = 0.857

Table 1: Inter-annotator agreement results

In this pre-final version, all dialogues were annotated by the same coder.

The inter-coder agreement tests were run on two dialogues, which were independently annotated by two expert coders. The inter-coder agreement figures for segmentation, category assignment and overall agreement as provided by the ANVIL tool for the four speakers in the two dialogues are in Table 1. The results for segmentation and category assignment are given in terms of Cohen's κ (Cohen, 1960) and corrected κ (Brennan and Predinger, 1981), while the overall agreement is also expressed in terms of Krippendorff's α (Krippendorff, 1970).

As can be seen, the results demonstrate high agreement. Nevertheless, the few cases of disagreement were inspected and the entire annotation revised by the second annotator.

5. The Dialogue Acts in the First Encounters

Dialogue act	Count	%
Inform	1135	29.84
AlloFeedbackGive	994	26.13
Inform-Answer	302	7.94
OwnCommManagement	261	6.86
Retraction	251	6.60
AutoFeedback	146	3.84
Confirm	136	3.58
Check-Question	112	2.94
Set-Question	111	2.92
Propos-Question	105	2.76
AlloFeedbackElicit	77	2.02
Agreement	41	1.08
Disconfirm	39	1.03
Choice-Question	14	0.37
ReturnGreeting	12	0.32
Address-Suggest	11	0.29
InitialGreeting	10	0.26
ReturnSelfIntroduction	10	0.26
Other (frequency $<$ 10)	37	0.96
Total	3804	100

Table 2: Frequency of dialogue acts in the NOMCO corpus

A total of 3804 dialogue acts were identified and annotated in the corpus using 31 different labels, which is a subset of the categories made available in the implemented specifications described in Section 4. Table 2 shows absolute and relative frequency counts for the most frequent types (with frequency of at least 10).

The most frequent dialogue act types are *Inform* and *AlloFeedbackGive*, followed at some distance by *Inform-Answer*, *OwnCommManagement*, *Retraction*, *AutoFeedback*, *Confirm* and various types of question.

The frequency figures of the dialogue act labels reflect the nature of the first encounter conversation type, in which the participants introduce themselves and exchange a lot of information about themselves in a rather compressed stretch of time. The frequency figures also indicate that many common dialogue acts are feedback-related. This again reflects the conversation type and the physical settings, with the participants standing in front of each other, and showing kindness and interest towards their interlocutor.

To delve deeper into the way dialogue acts are used in the first encounter corpus, we extracted all pairs of dialogue acts which follow each other at the onset by no longer than 0.5s. Using a 0.5s window is essentially an empirical choice. However, there is experimental evidence of the fact that human subjects are sensitive to temporal gaps of at least 0.5s (Giorgolo and Verstraten, 2008; Leonard and Cummins, 2010). Extending the window beyond this measure, on the other hand, would yield a larger number of associated dialogue act pairs with the consequent danger of capturing pairs which are not directly adjacent to each other.

A total of 1391 such dialogue act pairs were extracted. Figure 1 shows which interlocutor responses (on the x axis) follow at the onset dialogue acts from the other speaker. In the legend of the figure, only the dialogue acts that are most frequently followed by a response within 0.5s are mentioned explicitly, while less frequent categories are joined together under the 'Other' label. The distribution shows that there is a significant dependence between the adjacent categories (χ^2 = 1870.3, df = 208, p-value $<$ 2.2e-16). The clearest pattern is the one showing the inter-dependency between *AlloFeedbackGive* and *Inform*, which tend to follow each other.

Tables 3 and 4 give frequency counts and percentages of the dialogue acts that follow *AlloFeedbackGive* and *Inform*, respectively. Interestingly, the second most frequent dialogue act type following *AlloFeedbackGive* is another feedback giving utterance, probably showing an alignment of the two speakers along this dimension. Following an *Inform* act, on the other hand, we see not only *AlloFeedbackGive* as already noted and as expected, but also *Agreement*, and dif-

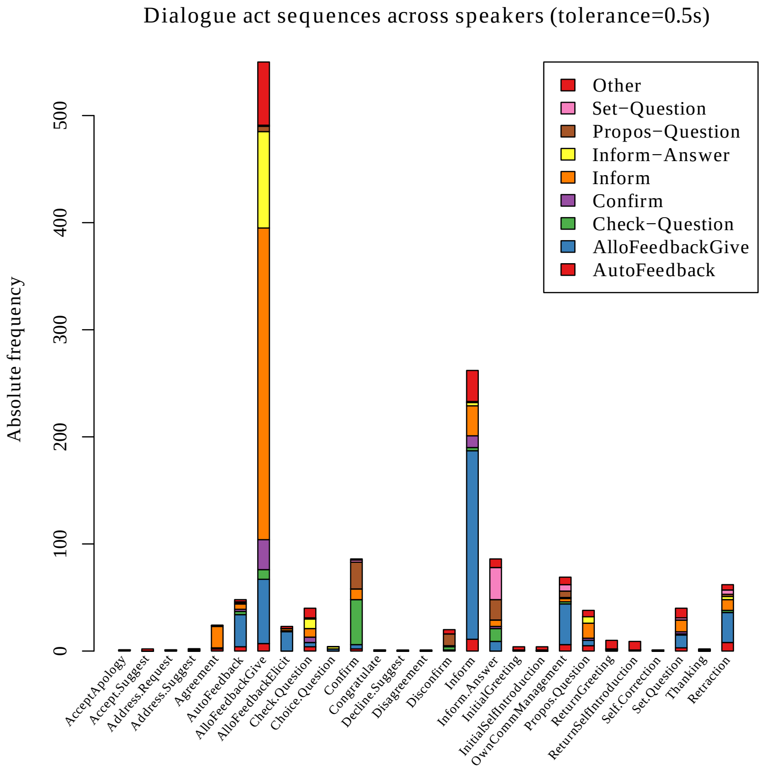

Figure 1: Dialogue act succession in the corpus (lag = 0.5s)

Dialogue act	Count	%
Inform	176	45.24
AlloFeedbackGive	60	15.42
OwnCommManagement	38	9.77
AutoFeedback	30	7.71
Retraction	28	7.20
AlloFeedbackElicit	18	4.63
SetQuestion	12	3.08
Others (< 10)	27	6.95
Total	389	100

Table 3: Dialogue acts following *AlloFeedbackGive* (tolerance = 0.5s)

Dialogue act	Count	%
AlloFeedbackGive	291	70.46
Inform	28	6.78
Agreement	20	4.84
ProposQuestion	14	3.39
SetQuestion	11	2.66
Retraction	10	2.42
Confirm	10	2.42
Others (< 10)	29	7.03
Total	413	100

Table 4: Dialogue acts following *Inform* (tolerance = 0.5s)

ferent types of question by which the interlocutor asks for more information.

These associations confirm the main characteristic of the dialogues, which is for the two speakers to create common knowledge about each other, as well as rapport through continuous feedback.

6. Relation between dialogue acts and gestural behaviour

To investigate the relation between dialogue acts and gestural behaviour, we extracted all the overlaps between dialogue acts and head movements on the one hand, and dialogue acts and facial expressions on the other. An overlap is defined by having at least a time point in common. We did this within as well as across speakers to look at how speakers combine utterances with gestural behaviour, but also how interlocutors respond non-verbally to the utterances of the other conversation participant.

Starting with overlaps between overlaps and head movements, there are a total of 3686 overlaps when we look at the same speaker (Figure 2), and 2389 overlaps across different speakers (Figure 3). While the plots show that all types of head movements occur together with many different dialogue act types, we also note that *AlloFeedbackGive* is mostly accompanied by *Nod* by the same speaker, whereas across speakers *Inform* mostly overlaps with *Nod*. These two overlap types show in fact how feedback works from two different perspectives. If we focus on multimodal expressions by one speaker, linguistic feedback expressions are often accompanied by nodding, while looking at the dynamics of the dialogue across speakers, feedback is often given while the other person is providing information.

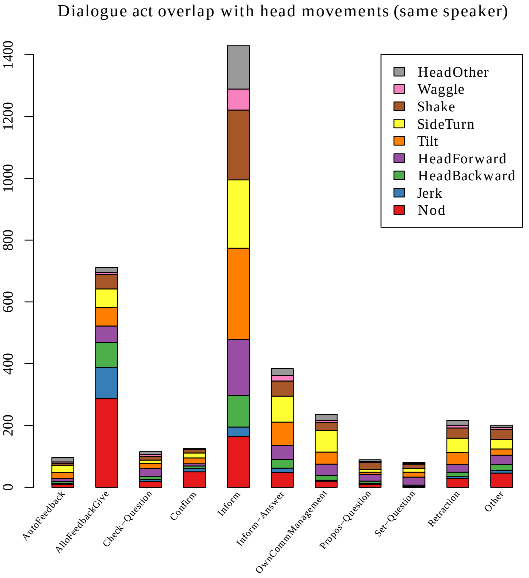

Figure 2: Overlaps between dialogue acts and head movements of the same speaker (n=3686). 'Other' collects dialogue acts which overlap with head movements less than 50 times.

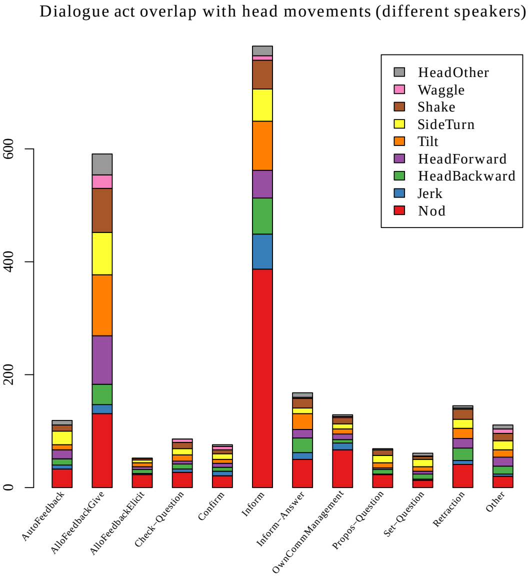

Figure 3: Overlaps between dialogue acts of one speaker and head movements of the other (n=2389)

As for the overlaps between overlaps and facial expressions, there are 1677 for the same speaker (Figure 4), and 1731 across speakers (Figure 5). There does not seem to be much difference in how facial expressions overlap with dialogue acts in the two plots with the noticeable case of *Retraction*, which appears in the plot showing overlaps for the same speaker. This type of overlap reflects the situation in which a speaker abandons what they were saying and starts a new utterance. This is often accompanied by a facial expression, which can be considered a kind of self feedback.

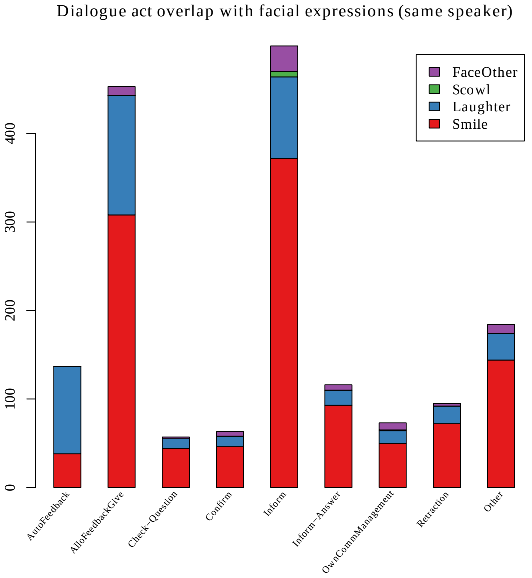

Figure 4: Overlaps between dialogue acts and facial expressions of the same speaker (n=1677)

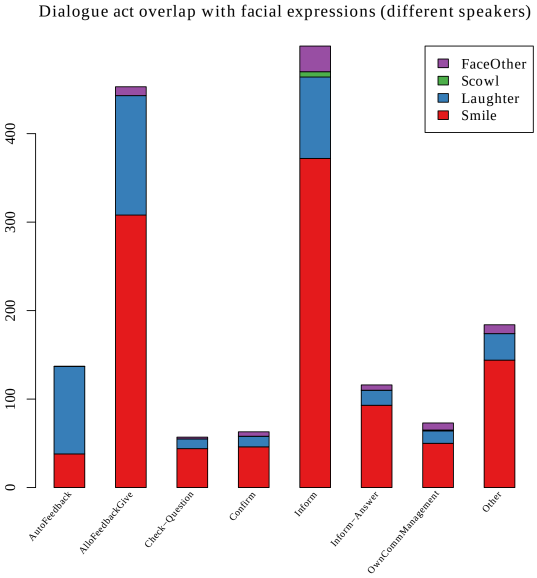

Figure 5: Overlaps between dialogue acts of one speaker and facial expressions of the other (n=1731)

7. Analysis of Feedback

In the preceding section, we described how different dialogue acts are associated with head movements and facial expressions focusing on the type of the movement. Here, we discuss the relation between the functional annotation of the data at the level of dialogue act on the one hand, and gestural behaviour on the other. For this analysis, we have chosen to focus on the feedback function, which was in fact annotated in the speech and the gestural channels separately. The gestures which are considered together with dialogue acts are head movements and facial expressions.

Feedback is the phenomenon according to which speakers exchange information about being able and willing to continue having *Contact* with each other, to have *Perception* of each other, and *Understanding* the message that is being communicated (*CPU*) (Allwood et al., 1993; Bunt et al., 2010). Feedback can be given or elicited and participants can express *agreement* or *disagreement* with what has been said. Feedback is expressed through unobstructive unimodal or multimodal signals. Examples of spoken feedback expressions are e.g. *yes* and *no*, while head nods and shakes as well as smiles are examples of common gestural feedback.

Feedback head movements and facial expressions have been addressed in numerous studies (Yngve, 1970; Duncan, 1972; Hadar et al., 1984; McClave, 2000; Cerrato, 2007). Moreover, the annotations in the Danish NOMCO corpus have been previously used to test whether feedback categories can be predicted automatically using the shape descriptions of head movements and facial expressions (Paggio and Navarretta, 2013), or to annotate feedback automatically in different corpora (Navarretta, 2013). Feedback speech expressions in a Swedish dialogue corpus were investigated by Allwood et al. (1993) while *yes/no* feedback speech expressions in the Danish NOMCO corpus were analysed in Paggio and Navarretta (2017). Finally, feedback *yes/no* expressions and nods in the Danish NOMCO corpus have been compared to the corresponding expressions in the Swedish NOMCO corpus and to the occurrences of nods in the Finnish first encounter corpus (Navarretta et al., 2012).

In this study we analyse the occurrence of feedback categories in speech and gestural behaviour, more specifically head movements and facial expressions, using the dialogue act annotations for speech and the MUMIN functional categories for gestures. We do this by looking at how often a feedback dialogue act of a certain type co-occurs with a head movement or a facial expression encoded with the same feedback category (see Table 5), another feedback category or a different functional category than feedback. The same is then done starting from feedback head movements and feedback facial expressions and extracting the co-occurring dialogue acts. Only the first co-occurring event is extracted in cases where there are multiple overlaps. Moreover, we only consider co-occurrence of speech and gestures produced by the same speaker. Multiple overlaps and co-occurrences of events across speakers are left for future investigation.

Even though the definition of feedback in MUMIN and in the ISO 24617 standard is the same, the names of the feedback categories vary slightly. Table 5 shows the correspondence between the feedback categories in the two annotation frameworks. Since both frameworks have a feedback category of auto or self feedback describing cases in which speakers expresses feedback to their own speech, we also included the phenomenon in our study. The major differences between the two annotations are the following. Firstly, MUMIN distinguishes between contact and perception (*CP*) and Contact, Perception and understanding (*CPU*), while both categories in the dialogue act annotations are annotated with the label *AlloFeedback*. This distinction, however, is not important when comparing the annotations. Secondly, the two categories *Agreement* and *Disagreement* belong to another dimension than feedback in the Dialogue Act annotation. However, for this study we considered them equivalent since they refer to the same kind of behaviours.

Dialogue Act	MUMIN
AlloFeedback	CP/CPU
AlloFeedbackGive	FeedbackGive
AlloFeedbackElicit	FeedbackElicit
Agreement	FeedbackAgree
Disagreement	FeedbackDisagree
AutoFeedback	SelfFeedback

Table 5: Feedback Categories in the annotations of dialogue acts and gestures MUMIN

Table 6 shows the number of occurrences of feedback dialogue acts of a) a certain feedback type which b) occur alone (FBDA), c) co-occur with a gesture (head movement and/or facial expression) encoded with the same feedback category (+=FBGesture), d) co-occur with a feedback gesture of another feedback type (+otherFBGesture), e) co-occur with a non feedback gesture (+notFBGesture), f) the total number of feedback dialogue acts (All), g) the percentage of feedback dialogue acts that co-occur with a gesture annotated with the same feedback category (%FB same), h) the percentage of feedback dialogue acts that co-occur with a gesture, independently of its function (%Multimodal). We do not show when feedback head movements and facial expressions co-occur since we are focusing on co-occurrences of feedback dialogue acts and gestures. As can be seen, there are 1263 feedback dialogue acts in the corpus and 85% of them co-occur with a head movement and/or a facial expression. Moreover, the majority of *Agreement*, and most of the *FeedbackGive* (back-channelling) dialogue acts are accompanied by a gesture annotated with the same feedback category, meaning that the two signals reinforce each other. Finally, 44% of *AutoFeedback* dialogue acts co-occur with a gesture that has exactly the same function. From the table it is also evident that many feedback dialogue acts co-occur with a feedback gesture of different type or a gesture with another function than feedback. In these cases speech and gesture convey different information at the same time.

In Table 7 we look at the occurrences of feedback expressed by facial expressions and the dialogue acts that co-occur

Category	FBDA	+=FBGesture	+otherFBGesture	+ notFBGesture	All	%FB same	%Multimodal
Agree	3	35	3	0	41	85	93
FBElicit	23	17	25	12	77	22	70
FBGive	156	743	48	47	994	75	84
Disagree	1	2	1	0	4	50	75
AutoFB	9	64	56	17	147	44	94
All	192	861	133	77	1263	68	85

Table 6: Feedback dialogue acts and co-occurring gestures

with them. The construction of the table is parallel to that of Table 6. The table indicates that only 34% of feedback facial expressions co-occur with a feedback dialogue act of the same type, while nearly all feedback facial expressions co-occur with a dialogue act. An exception is the single case of *Disagreement* which is always expressed by the facial expression alone. *FeedbackGive* and *Agreement* are the categories which are most often expressed by both a facial expression and a dialogue act. The most common dialogue acts co-occurring with feedback facial expressions (co-occurrence $>=$ 10) are given in Table 8. It is not surprising that feedback giving co-occurs most often with the corresponding dialogue act as seen previously, while it is interesting that feedback eliciting facial expressions co-occur often with an *Inform* dialogue act. In such cases, the speaker asks for feedback to what they have said. Similarly, a facial *AutoFeedback* signal often co-occurs with *Inform* dialogue acts and with dialogue acts related to own communication management and time management, such as self corrections, retractions and speech pauses.

In Table 9 we investigate the co-occurrences of feedback head movements with feedback dialogue acts and other dialogue acts. This table is constructed like Table 6 and Table 7. As also noted in Paggio and Navarretta (2017), the most frequent feedback gestures are head movements. Table 9 shows that 80% of feedback head movements co-occur with a dialogue act, and that *Agreement* and *FeedbackGive* are the categories which are most frequently expressed by both speech and a head movement (64% and 54% of the cases, respectively). *AutoFeedback* and *FeedbackElicit* are only seldom expressed by both speech and head signals. This is the same tendency which we saw earlier when analysing co-occurring feedback facial expressions and dialogue acts. In all these cases, the gesture and the co-occurring dialogue act have different functions, in other words they complement each other. This is not surprising since feedback elicit signals indicate that a speaker is asking for feedback, while self feedback signals express the speaker's comments to their own utterances. On the contrary, when providing feedback to the interlocutor (feedback giving and agreeing), speech and gesture often reinforce each other.

The most common dialogue acts co-occurring with feedback head movements (co-occurrence $>=$ 10) are given in Table 10. The table shows that the dialogue acts that co-occur with the various categories of feedback head movements are mostly the same as those co-occurring with feedback facial expressions of the same type. In particular, and in addition to feedback dialogue acts, feedback-related head movements co-occur especially with *Inform* and *Inform-Answer*, notably in the case of self feedback. Interestingly, *Check-Question* also co-occurs with *FeedbackGive* indicating cases in which the speaker is giving feedback at the same time as checking they have understood.

8. Conclusions and Future Work

In this paper we have presented the annotation of dialogue acts in the multimodal NOMCO corpus, a collection of naturally occurring first encounter dialogues in Danish. The dialogue act annotation is based on the ISO 24617 standard, and the validation of the annotations shows high agreement between two coders. The only difficulty in applying the standard was the distinction between *Own Communication Management* and *Time Management* categories, which does not seem relevant in conversations that are not task oriented, and where time management appears to be always functional to the management of communication.

We then presented descriptive statistics showing first how different dialogue act types follow one another, and then how they overlap with gestural behaviours, in particular head movements and facial expressions. Both analyses point at ways in which the dialogue participants create common ground and rapport by constantly checking mutual understanding and giving or eliciting feedback to and from each other. Finally, we focus on feedback behaviour by examining the functional annotation of the gestural behaviour and how it relates to the dialogue act annotation, and show that feedback gestures can constitute not only a reinforcement of the co-occurring spoken utterance, but also a complement to it.

In future work, we plan to distinguish the various dimensions of dialogue acts, and to add turn management dialogue acts to the annotations. It would also be useful to compare the dialogue act annotations in this corpus with dialogue act annotations in corpora addressing domain specific tasks, such as map task dialogues, or in corpora involving people who know each other well. This would complete a preceding study of feedback expressions in first encounters and in conversations between family members and friends that confirmed the hypothesis that the number and type of feedback signals is partly related to the degree of familiarity of the participants (Navarretta and Paggio, 2012). The role played in the expression of dialogue acts by other types of movement, such as hand gestures or body posture, is also an interesting topic for further research. Finally, we would like to experiment with the automatic annotation

Category	FBFace	+=FBDA	+otherFBDA	+ notFBDA	All	%FB same	%Multimodal
Agree	2	8	1	3	14	57	86
FBElicit	0	10	11	74	95	11	100
FBGive	93	276	23	91	483	57	81
Disagree	1	0	0	0	1	0	0
AutoFB	0	49	12	352	413	12	100
All	96	343	47	520	1006	34	90

Table 7: Feedback facial expressions and co-occurring dialogue acts

Face Feedback Give	
Dialogue act	Count
FeedbackGive	268
Inform	19
Agreement	16
Confirm	15
Check-Question	13
OwnCommunicationM	12
Face Feedback Elicit	
Inform	30
FeedbackGive	10
FeedbackElicit	10
Face AutoFeedback	
Inform	265
Inform-Answer	50
AutoFeedback	49
OwnCommunicationM	48
Confirm	10

Table 8: Dialogue acts that co-occur at least 10 times with feedback facial expressions

of dialogue acts using the NOMCO annotations as training and testing data.

9. References

Alexandersson, J., Buschbeck-Wolf, B., Fujinami, T., Kipp, M., Koch, S., Maier, E., Reithinger, N., Schmitz, B., and Siegel, M. (1997). Dialogue acts in verbmobil-2. Technical report, DFKI.

Allen, J. and Core, M. (1997). Damsl: Dialogue act markup in several layers (draft 2.1). Technical report, University of Rochester.

Allwood, J., Nivre, J., and Ahlsén, E. (1993). On the semantics and pragmatics of linguistic feedback. *Journal of Semantics*, 9(1):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 phenomena. In Jean-Claude Martin, et al., editors, *Multimodal Corpora for Modelling Human Multimodal Behaviour*, volume 41 of *Special issue of the International Journal of Language Resources and Evaluation*, pages 273–287. Springer.

Amanova, D., Petukhova, V., and Klakow, D. (2016). Creating annotated dialogue resources: Cross-domain dialogue act classification. In Calzolari et al., editor, *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, Paris, France, may. European Language Resources Association (ELRA).

Anderson, A., Bader, M., Bard, E., Boyle, E., Doherty, G., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H., and Weinert, R. (1991). The hcrc map task corpus. *Language and Speech*, 34:351–366.

Brennan, R. and Predinger, D. (1981). Coefficient kappa: Some uses, misuses, and alternatives. *Educational and Psychological Measurement*, 41(3):687–699.

Bunt, H., Alexandersson, J., Carletta, J., Choe, J.-W., Fang, A. C., Hasida, K., Lee, K., Petukhova, V., Popescu-Belis, A., Romary, L., Soria, C., and Traum, D. (2010). Towards and iso standard for dialogue act annotation. In *Proceedings 7th international conference on language resources and evaluation (LREC 2010)*, pages 2548–2555.

Bunt, H., Kipp, M., and Petukhova, V. (2012). Using DiAML and ANVIL for multimodal dialogue annotations. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 1301–1308, Istanbul, Turkey, May. European Language Resources Association (ELRA).

Bunt, H., Petukhova, V., and Fang, A. (2017). Revisiting the ISO standard for dialogue act annotation. In *Proceedings 13th joint ISO-ACL workshop on interoperable semantic annotation (ISA-13)*, pages 37–50, Montpellier, France.

Bunt, H., Petukhova, V., Malchanau, A., Fang, A., and Wijnhoven, K. (2019). The DialogBank: dialogues with interoperable annotations. *Language Resources and Evaluation*, 53(2):213–249.

Carletta, J. (2007). Unleashing the killer corpus: experiences in creating the multi-everything ami meeting corpus. *Language Resources and Evaluation Journal*, 41(2):181–190.

Cerrato, L. (2007). *Investigating Communicative Feedback Phenomena across Languages and Modalities*. Ph.D. thesis, Achool of Speech and Music Communication, Stockholm, KT.

Chafe, W. (1974). Language and Consciousness. *Language*, 50:111–133.

Cohen, J. (1960). A coefficient of agreement for nom-

Category	FBHead	+=FBDA	+otherFBDA	+ notFBDA	All	%FB same	%Multimodal
Agree	3	53	8	19	83	64	96
FBElicit	15	6	13	115	149	4	90
FBGive	286	504	23	116	929	54	69
Disagree	1	1	0	1	3	33	76
AutoFB	17	18	17	416	468	4	89
All	322	582	61	667	1632	36	80

Table 9: Feedback head movements and co-occurring dialogue acts

Head Feedback Give	
Dialogue act	Count
FeedbackGive	504
Confirm	26
Inform	25
Check-Question	19
AutoFeedback	17
OwnCommunicationM	13
Inform-Answer	11
Head Feedback Elicit	
Inform	61
Check-Question	16
FeedbackElicit	13
Inform-Answer	13
Head AutoFeedback	
Inform	268
Inform-Answer	56
OwnCommunicationM	58
AutoFeedback	18
FeedbackGive	17
Check-Question	10
Confirm	10
Head Feedback Agree	
Agreement	53
OwnCommunicationM	12

Table 10: Dialogue acts that co-occur at least 10 times with feedback head movements

inal scales. *Educational and Psychological Measurement*, 20(1):37–46.

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

Giorgolo, G. and Verstraten, F. A. (2008). Perception of ‘speech-and-gesture’ integration. In *Proceedings of the International Conference on Auditory-Visual Speech Processing 2008*, pages 31–36.

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

Jurafsky, D., Shriberg, E., and Biasca, D. (1997). Switchboard swbd-damsl shallow-discourse-function annotation: Coders manual (draft 1.3). Technical report, University of Colorado.

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

Krippendorff, K. (1970). Estimating the reliability, systematic error, and random error of interval data. *Educational and Psychological Measurement*, 30(1):61–70.

Leonard, T. and Cummins, F. (2010). The temporal relation between beat gestures and speech. *Language and Cognitive Processes*, 26(10):1457–1471.

Maclay, H. and Osgood, C. E. (1959). Hesitation phenomena in spontaneous english speech. *Word*, 15:19–44.

McClave, E. Z. (2000). Linguistic functions of head movements in the context of speech. *Journal of Pragmatics*, 32(7):855–878.

Milajevs, D. and Purver, M. (2014). Investigating the contribution of distributional semantic information for dialogue act classification. In *Proceedings of the 2nd Workshop on Continuous Vector Space Models and their Compositionality*, pages 4a–47. ACL.

Navarretta, C. and Paggio, P. (2012). Verbal and Non-Verbal Feedback in Different Types of Interactions. In *Proceedings of LREC 2012*, pages 2338–2342, Istanbul Turkey, May.

Navarretta, C., Ahlsén, E., Allwood, J., Jokinen, K., and Paggio, P., (2012). *Feedback in Nordic First-Encounters: a Comparative Study*, pages 2494–2499. European language resources distribution agency.

Navarretta, C. (2013). Transfer learning in multimodal corpora. In IEEE, editor, *In Proceedings of the 4th IEEE International Conference on Cognitive Infocommunications (CogInfoCom2013),*, pages 195–200, Budapest, Hungary, December.

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*, 7(1-2):29–37.

Paggio, P. and Navarretta, C. (2017). The Danish NOMCO Corpus of Multimodal Interaction in First Acquaintance Conversations. *Language Resources and Evaluation*, 51:463–494.

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

Purver, M., Niekrasz, J., and Ehlen, P. (2007). Automatic annotation of dialogue structure from simple user interaction. In A. Popescu-Belis, et al., editors, *MLMI 2007*, volume 4893 of *LNCS*, pages 48–59. Springer-Verlag, Berlin Heidelberg.

Rehm, M., André, E., Bee, N., Endrass, B., Wissner, M., Nakano, Y., Akhter Lipi, A., Nishida, T., and Huang, H.-H. (2009). Creating standardized video recordings of multimodal interactions across cultures. In Michael Kipp, et al., editors, *Multimodal Corpora: From Models of Natural Interaction to Systems and Applications*, pages 138–159. Springer Berlin Heidelberg, Berlin, Heidelberg.

Searle, J. (1969). *Speech Acts*. Cambridge University Press.

Verbree, D., Rienks, R., and Heylen, D. (2006). Dialogue-act tagging using smart feature selection; results on multiple corpora. In *IEEE Spoken Language Technology Workshop*, page 4 pages. IEEE.

Yngve, V. (1970). On getting a word in edgewise. In *Papers from the Sixth Regional Meeting of the Chicago Linguistic Society*, page 568.