A Multidimensional Approach to Multimodal Dialogue Act Annotation

Volha Petukhova, Harry Bunt,
Department of Language and Information Science,
Faculty of Arts, Tilburg University
P.O. Box 90153, 5000 LE Tilburg, Netherlands
E-mail: v.petukhova@uvt.nl; harry.bunt@uvt.nl

Abstract

This paper investigates the benefits of multidimensional approaches to dialogue act annotation, and the advantages of using layered multidimensional 'open' dialogue act taxonomies. We performed a comparative analysis of a one-dimensional and two multidimensional dialogue act annotation schemes, and concluded that not only does a multidimensional approach support a more accurate analysis of human communication, but contrary to what is often believed it also facilitates dialogue annotation. This conclusion was supported by an analysis of nonverbal expressions in multiparty dialogue. We found that, like verbal ones, nonverbal communicative acts may also serve several communicative functions simultaneously within different dimensions. Finally, we showed that the multiple communicative functions which a dialogue utterance may have, and which form the basis of the multidimensional approach, are automatically learnable in a data-oriented way.

1 Introduction

In order to understand and to describe what is happening in dialogue, it has become common to analyse dialogues in terms of dialogue acts. Recent years have witnessed a growing interest in the annotation of dialogue corpora in terms of dialogue acts, and several efforts have been undertaken toward the development of dialogue act annotation schemes. Existing dialogue act schemes differ not only in their precise sets of tags, but more importantly with respect to (1) the underlying approach to dialogue modelling; (2) the definition of the related concepts; and (3) the level of granularity of the defined tag set. Generally, annotation schemes can be divided into one- and multidimensional ones.

One-dimensional schemes allow coding dialogue utterances with only one tag, and their tag sets are kept as a rule very simple. Because of their simplicity, they are thought to be reliable and to take less efforts to apply consistently by annotators. Some researchers, e.g. [13], [14], note, however, that one-dimensional annotation schemes also have serious disadvantages. Allen and Core in [1], [9], [10] criticise Searle's theory saying that his taxonomy of illocutionary acts is a set of mutually exclusive categories and does not allow utterances to perform multiple actions simultaneously, as it is actually the case in a real conversation.

Studies of human dialogue behaviour indicate that natural dialogue involves several activities beyond those strictly related to performing the task or purpose for which the dialogue is instrumental (such as obtaining certain information, instructing another participant, negotiating an agreement, etc.). In natural conversation, among other things, dialogue participants constantly 'evaluate whether and how they can (and/or wish to) continue, perceive, understand and react to each other's intentions' [2]. They share information about the processing of each other's messages, elicit feedback, manage the use of time, taking turns, contact and attention, etc. Communication is therefore a complex, multi-faceted activity, and this is reflected in the fact that dialogue utterances are therefore most of the time multifunctional. A dialogue act taxonomy should therefore contain the concepts needed to cover all these aspects of dialogue.

Mutidimensional approaches to dialogue act annotation, which incorporate a multifunctional view on dialogue behaviour, have been recognised by many researchers as empirically better motivated, and allowing a more accurate modelling of theoretical distinctions. While the multifunctionality of dialogue utterances has been widely recognised, computationally oriented approaches to dialogue generally see multifunctionality as a problem, both for the development of annotation schemes and for the design of dialogue systems [21]. Information that may be obtained through a multifunctional analysis is therefore often sacrificed for simplicity in computational modelling. As a consequence, the actual multifunctionality of dialogue utterances and related phenomena are still understudied, and have so far escaped extensive description and formalisation. This paper aims to analyse the advantages and possible drawbacks of one- and multidimensional approaches to dialogue annotation by presenting empirical evidence. For this purpose, the comparative analysis of one- and multidimensional dialogue annotation schemes tested on dialogue corpus data has been carried out.

Since multimodality is significant for enabling multifunctionality in utterances, we also studied nonverbal behaviour in dialogue and considered its possible characterisation in terms of dialogue acts. A lot of work has been carried out in studying human nonverbal behaviour, e.g. [12]; [8]; [16]. To the best of our knowledge, however, little research has been done to reveal relations between observable utterance nonverbal features and the intended multiple functions of these utterances. Taking a context-change (or information-status update) approach to the interpretation of dialogue acts, we tried to identify those features of communicative behaviour in different modalities that trigger updates of the dialogue context, i.e., have an intended communicative function, and to indicate what (multiple) aspects of the dialogue context they address.

Finally, every communicative function is required to have some reflection in observable features of communicative behaviour, i.e. for every communicative function there are devices which a speaker can use in order to allow recognition by the addressee. We investigated the automatic learnability of the multiple communicative functions on the basis of such features as linguistic cues, intonation properties and dialogue history.

2 Comparing Schemes

In this section we discuss the results of the comparative analyses of three dialogue act annotation schemes: the one-dimensional AMI scheme and two multidimensional schemes, DIT and DAMSL. We targeted the better understanding of the range of communicative functions that an utterance can have and to show empirically whether a multidimensional approach provides a better account of dialogue phenomena. The approaches were compared with respect to the following properties:

1. adequate coverage of relevant dialogue phenomena;
2. granularity of the defined tag set;
3. annotation costs in terms of time spent on manual annotation by human annotators;
4. flexibility, e.g. extensibility and transformation possibilities;
5. reusability, e.g. task and domain dependency.

Four AMI pilot video-recoded meetings (totally 1819 utterances) were manually annotated using these three schemes, and the results were processed and compared. We paid particular attention to the co-occurrences and dependencies between dialogue acts in different dimensions, using the formal notion of interactive 'dimension' provided by DIT [6].

The AMI project assumes that the relevant information about meeting aspects and phenomena can be automatically extracted from manually or automatically annotated data, e.g. through a meeting browser. For this purpose, a dialogue act annotation scheme has been designed[1] which is one-dimensional, allowing an utterance to be coded with only one tag. Additionally, the relations (POSitive, NEGative, PARTial, UNCertain and ELAboration) can be coded in adjacency pairs. This is based on the observation that dialogue acts are often performed in response to a specific previous dialogue act. Coding is also possible in the category of 'reflexivity' which indicates whether the content of the dialogue act is about the communication itself or about the process of decision making, rather than about the task. Each utterance thus has one dialogue act tag, possibly extended also with a relation and/or a reflexivity tag. Figure 1 shows the AMI dialogue act hierarchy.

[1] We used for our analysis the version of the 21st of April, 2005. The scheme has been revised several times in an later stage of the project. The latest version of AMI dialogue act annotation scheme is available at http://mmm.idiap.ch/private/ami/annotation/dialogue_acts_manual_1.0.pdf

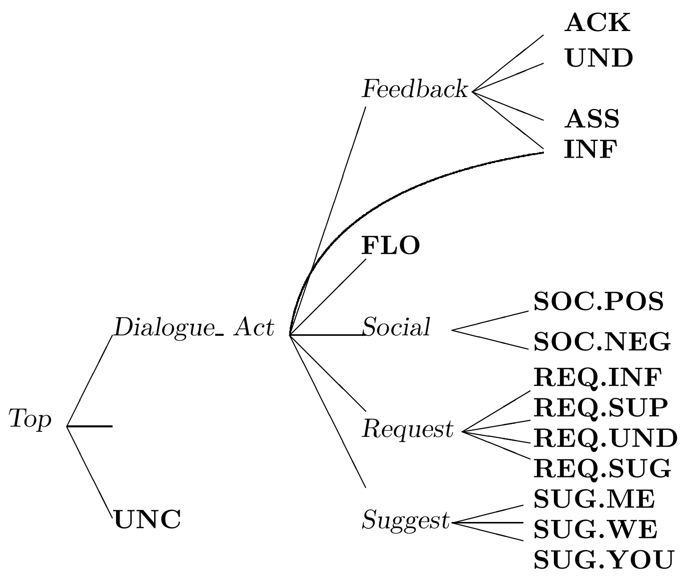

Figure 1: *The AMI Dialogue Act Hierarchy*

By contrast, the DIT dialogue act taxonomy (named after Dynamic Interpretation Theory [3], which has spurred its design), is a multilayered and multidimensional scheme, where the layers form convenient groupings of functions and the dimensions in every dimension (like questions and informs)[2]. The scheme distinguishes 11 dimensions, as indicated in Figure 2 (dimensions are given in bold), and allows a dialogue utterance to be annotated with zero or one functional tag per dimension.

The DAMSL annotation scheme [1] distinguishes four layers: Forward-looking function, Backward-looking function, Communicative status and Information level, and allows an utterance to be annotated with one or more labels in each layer. An utterance may thus simultaneously perform functions such as responding to a question, confirming understanding, promising to perform an action, and informing.

For all three taxonomies, definitions of the dialogue act types are intention-based. Definitions in AMI and in some cases in DAMSL, however, are informal and descriptive, and in AMI are often given in the form of instructions for the annotator. DIT provides clear, unambiguous and formal definitions, and fine-grained distinctions between different types of dialogue acts. This facilitates dialogue act ascription by human annotators, and the effective computation of dialogue act tags. Despite some differences between DAMSL and DIT (e.g. feedback and other dialogue control functions are better defined in DIT; tags

[2]For the latest version of the DIT dialogue act taxonomy, also called 'DIT^{++}, visit http://ls0143.uvt.nl/dit/

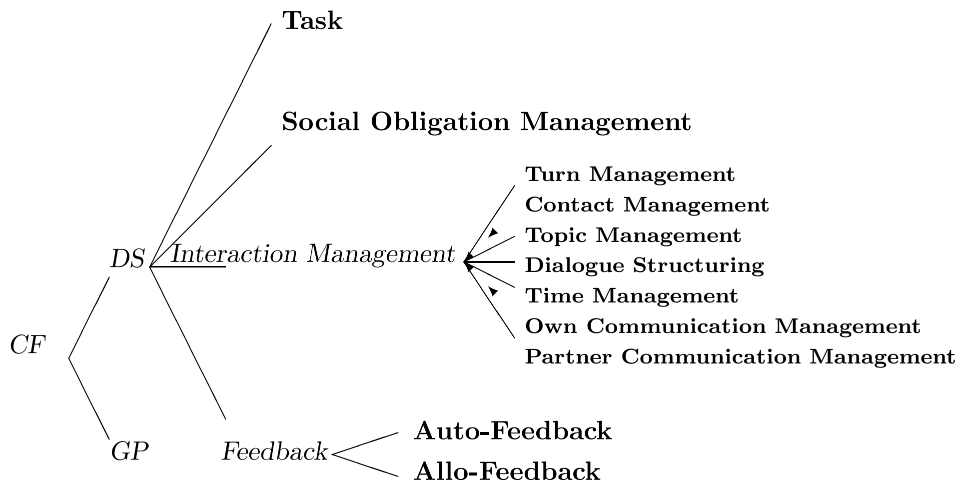

Figure 2: *The DIT Dimensions*

for indirect dialogue acts are not presented in DAMSL), they have conceptually much in common: both taxonomies have a well-worked out theoretical background, and are multidimensional. We found good matches for the majority of communicative functions from both tag sets. Some relevant phenomena are not covered by the AMI annotation scheme (about 14,2% of all corpus utterances were UNCODED); there are for example no communicative functions in AMI for aspects of interaction management such as time, topic, contact, own and partner communication management, and discourse structuring. Our corpus analyses showed that these functions need to be included because this information is (1) a significant part of natural human conversation in general, and meetings in particular; and (2) important for understanding the functions of nonverbal acts (see Section 3).

We observed that dialogue participants use linguistic and nonverbal elements virtually all the time to address several aspects of the communication simultaneously; the vast majority of dialogue utterances have more than one communicative function. Consider, for example, utterance (1), produced at the opening of a dialogue:

(1) Are we going?

The speaker in (1) makes a suggestion, so we have a dialogue act with the communicative function SUGGEST. In addition, with utterance (1) the speaker signals that he wants to start the dialogue, which gives the utterance the communicative function OPENING. It also has the DIT communicative function CONTACT-CHECK, because the speaker wants to establish whether the partner is ready for communication. According to DAMSL, we would label this utterance as Open-Option + Opening + a Communication management (Information level layer) function. Following the AMI scheme, the most suitable tag is SUGGEST WE, which stands for the suggestion that the group do some action. Being allowed to assign only one communicative function, this fails to capture the interaction management phenomena that the multidimensional schemas do cover.

Multidimensional schemes such as DIT and DAMSL make more fine-grained, theoretically and empirically motivated distinctions between dialogue act types. For example, DIT distinguishes 19 Information Providing communicative functions (DAMSL has 11) based on accurate definition of differences in the speaker's motivation for providing the information; in different additional beliefs about what the addressee knows; and in differences in strength of the speaker's trust in the correctness of the information that he provides. The AMI scheme, by contrast, defines only one communicative function INFORM which can be combined with 5 relation tags: POS, NEG, PART, UNC and ELA. Figure 3 illustrates the approximate correspondence between some DIT and AMI Information Providing functions.

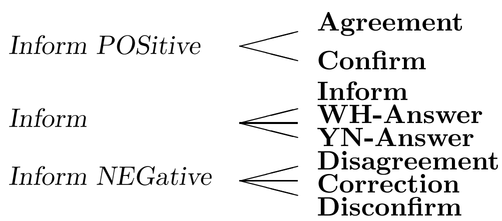

Figure 3: *Correspondence between AMI and DIT Information Providing functions*

Manual annotations are time consuming, and it is generally thought that tagging dialogue utterances according to a multidimensional scheme costs more annotation time than with a one-dimensional one. The analyses showed that the ratio of annotation time to real dialogue time ratio was approximately 25:1 when coding with the AMI scheme[3], and approximately 19:1 when coding with DIT or DAMSL. This can be explained by the fact that a one-dimensional annotation scheme like AMI poses quite a challenge for annotators, because it is often hard to judge what phenomena have been merged in one tag.

Multidimensional annotations schemes like DIT and DAMSL are 'open' in the sense that they are not restricted to a particular task or domain and

[3]Annotation time for the AMI scheme has been measured by annotators at the University of Twente and reported in an internal report.

are easily adapted to various purposes. Initially designed for two-agent task-oriented dialogues, they perfectly fit the AMI meeting data. The multidimensional schemes can be reused relatively easily, as when we need a dialogue to be annotated with a view to studying the turn-taking behaviour during the conversation; the roles of participants and their dominance relations; or what utterances elicited what kind of feedback, which of them were rejected or accepted in order to measure the efficiency of the discussion, and so on. Moreover, a multidimensional dialogue act taxonomy can be extended if needed with new dimensions and with new elements within dimensions. Our experience was that one-dimensional annotation schemes are very hard to update; even small corrections may lead to significant changes in the hierarchy and/or other types of dialogue acts and their descriptions. The multidimensional schemes could on the other hand be easily converted to a single-dimension tag set with complex tags if needed, according to the goal of the analysis or application requirements. For this purpose, a multidimensional scheme could be applied to a large corpus and the combinations/co-occurrences of tags could be analysed for possible merging into one tag[4]. For example, comparable experiments have been done by Andrei Popescu-Belis, who analysed 113,560 dialogue utterances (ISCI-MRDA corpus) according to the MRDA-annotation scheme, which has the theoretical number of possible combinations at about 7 million, and observed empirically that about 760 tags combination occur in the studied corpus, mostly composed of 1, 2 or 3 tags [18]. Further estimation of the frequency of these labels and dependencies between tags can reduce the search space significantly, resulting in a one-dimensional tag set, which is theoretically and empirically well-motivated.

An argument that is sometimes used against a multidimensional scheme is that dialogue act annotation using such schemes is not reliable and inter-annotator agreement scores are low. For measuring inter-annotator agreement, the standard kappa statistics is often used. When considering inter-annotator agreement for the use of multidimensional tags, this statistics is not an appropriate measure, because in the case of multidimensional annotation there can be *partial* agreement [11][5].

3 Annotation of Nonverbal Acts

In natural communication, the participants use all modalities and media that are available in the communicative situation; nonverbal behaviour is an essential part of human communication. This includes the use of gestures, facial expressions, pauses, gazes, posture shifts, etc; communicative resources which make the communication richer in many ways. We analysed nonverbal expressions (gaze, posture shifts, facial expressions, head movements and hand/arm

[4] We should notice here that the annotation scheme as a result of such an experiment would not be truly one-dimensional, because a complex label would still refer to multiple aspects of dialogue and its context, but the tag set would be a one-dimensional one allowing coding dialogue utterances with a single label.

[5] We do not go into details here; please consult the paper that is referenced for more details.

gestures) using the same method as when analysing verbal expressions, namely we investigated the communicative functions of nonverbal acts using the same annotation scheme. We used the DIT framework, since it contains well-defined layers of dialogue control functions, which are the distinctive functions of nonverbal acts observed in AMI meetings. In total, we studied 35 communicative functions that nonverbal expressions may have in dialogue.

Our observations showed that, generally, nonverbal communicative acts closely related to feedback or other interaction management dimensions. For example, short, not deep head nods usually have a feedback function indicating that a contribution has been understood well enough to allow the conversation to proceed. They also have a TURN GIVING function, encouraging the partner to continue with his utterance, and a communicative function in the dimension of Contact Management, ensuring the partner that he/she is ready to receive messages from him.

Repetitive head nods, lip movements, raising a finger, and beginning gesticulation may indicate that the previous contribution has been understood and the participant would like to grab the turn to add or correct something. Nonverbal expressions with Turn Management functions may also have functions in Own Communication Management. It was noticed, for example, by Butterworth [7] that an excessive amount of gaze aversion when the speaker is having difficulty formulating the message may lead a listener to interfere. Here, also expressions of uncertainty (e.g. lip compression, curving the mouth downwards, lowering eyebrows and eyelids, constricting the forehead muscles, head waggle, etc.) may invite the partner to take the turn and assist.

Speech-focused movements accompanying relatively unpredictable content words (e.g. iconic gestures during lexical search), body-focused movements (e.g. searching for elusive words or expression in the memory) normally indicate that the speaker needs some time to gather his/her thoughts or to formulate the utterance and therefore is stalling for time (STALLING), but he/she would like to keep the turn (TURN KEEPING). Sometimes pauses can increase the pressure on other participants to say something (TURN GIVING). The longer the pause, the more pressure builds on the other person to respond. Pauses near the beginning of an utterance can have the function of CONTACT CHECK, requesting attention. Speakers often make short pauses until the gaze of a recipient has been obtained and secured.

Finally, we observed that participants in a dialogue employ a broad range of social affective nonverbal expressions. Facial expression, body orientation, eye contact, hand/arm gestures and head movements are equally important for a comfortable and pleasant interaction. With respect to multidimensionality these expressions also perform several functions in a dialogue. All initiative utterances in the dimension of Social Obligation Management, e.g. INITIAL GREETING, put a so-called 'reactive pressure' on the addressee to reply, and therefore have a TURN GIVING function. Such utterances also play a role in the dimensions of Dialogue Structuring and Topic Management, e.g. opening and closing a conversation, apologies for unexpected topic shifts, etc.

The results of our investigation support our analyses of verbal utterances reported in this paper. The same dialogue act taxonomies can be used to interpret both verbal and nonverbal dialogue behaviour.

Nonverbal behaviour adds:

- communicative functions in feedback and other dimensions of Interaction Management;
- information for the interpretation of the verbal utterances.

Therefore, nonverbal features can be useful for the recognition (including automatic recognition) of utterance functions in context. We hypothesize that the use of relevant features from different sources makes it possible to design more refined strategies for interpreting the communicative functions of dialogue utterances and aim to prove this in our further research.

4 Automatic Recognition of Multiple Communicative Functions

To determine communicative functions of dialogue utterances is a complex task. In order to allow successful recognition of the intended communicative functions by the addressee the speaker can use several communicative devices. Such observable communicative features may be linguistic cues, intonation properties, accompanying facial expressions, head movements, etc. Since computer dialogue systems do not have the rich experience and background knowledge that human participants have, it is important that they operate on the basis of automatically learnable features of utterances and the dialogue context. In this section we report the results of experiments carried out to investigate the automatic learnability of multiple communicative functions.

For this purpose we trained a Naive Bayes classifier on the OVIS (Dutch Public Transport Information System) corpus[6]. The OVIS dialogues are task-oriented human-computer dialogues where the user is expected to require information about train connections and schedule. Our training set contained 1971 instances corresponding to the user's utterances, manually tagged according to the DIT dialogue act annotation scheme. The classes in the training set corresponded to the DIT communicative functions plus one 'all class', which combines the information about all applied communicative functions in one label (e.g. INF;Au_F_P_O;T_TAKE which stands for Inform about overall positive auto-feedback by taking the turn). A Naive Bayes classifier is a simple probabilistic classifier based on the so-called independent feature model. Naive Bayes classifiers are known to often work well for many complex real-world situations, and are particularly suited when the dimensionality of the inputs is high. Moreover, this classifier exhibits high accuracy and speed when applied to a large database, and can be efficiently trained.

[6]OVIS corpus data is available on http://www.let.rug.nl/ vannoord/Ovis/

As we pointed out above, features play a very important role in supporting accurate recognition of communicative functions. The use of features from different sources makes it possible to design more refined strategies for interpreting the functions of dialogue utterances [15]. The OVIS training set contains automatically extracted features from the dialogue history (e.g. communicative function of the preceding utterance), prosody (e.g. pitch, energy, duration and tempo) and the wording of the users utterance[7].

For example, the length of an initial pause could indicate the user's stalling for time in order to gather his thoughts, or a hesitation indicating an understanding problem. Tempo, which corresponds to the number of syllables per second, may indicate the user's awareness of the system's understanding problem which normally forces the user to speak slower than usual. When replying to the system's questions the user may also show considerable variation in pitch (measured, for example, by standard deviation in pitch). Especially information from the dialogue history combined with prosodic information is a good indicator of the user's awareness of system understanding problems.

The learnability of assigning multiple communicative functions to utterances was tested in experiments where the learning material ranged from 250 utterances to the full database of 1971 words (250 utterances were added to the learning material, and from 500 items onwards 500 utterances were added to the learning material). The results were compared with the baseline (prediction of the user input classes solely on the basis of one single feature: the most recently asked system's question [15]). The results were obtained using 10-fold cross-validation for data partition. The database was partitioned ten times, each time a different 10% of the database was used as test set and the remaining 90% as training set. The cross-validation was stratified, i.e. the 10 folds contained approximately the same proportions of instances with relevant tags as the entire database.

In Figure 4 shows the overall learning curve representing the system's generalisation accuracy for increasing sizes of training sets as the mean percentage over the ten folds. With a training set of 250 items, a score of 41.6% is reached. This scoreincreases to 57.7% in the final experiment with 1971 utterances.

Table 1 gives an overview of success scores expressed as the percentage correctly predicted all-class labels in all training experiments in comparison to baseline scores.

Number of Training Items	250	500	1000	1500	1971
Baseline score	50.8	54.1	43	42.7	41.9
Accuracy (%)	41.6	48.6	50.9	55.3	57.7

Table 1: *Baseline and simple learning scores on 'all-class' classification task*

The generalisation performance of the system with different amounts of examples provides insight into the relative learnability of the different classes in

[7]Prosodic and linguistic features were derived from both the output of the Automatic Speech Recognition (ASR) module of the OVIS system as well as the raw audio[15].

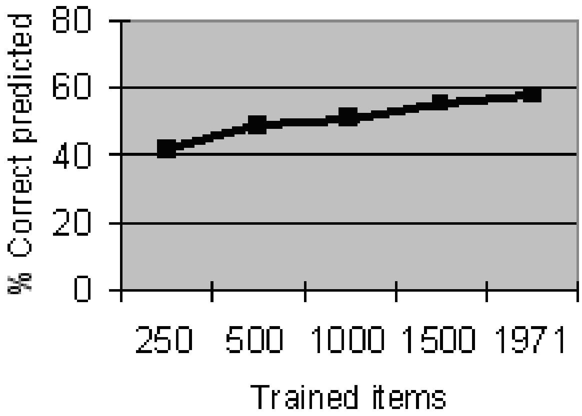

Figure 4: *Global learning curve expressed as the average percentage correctly predicted all-class labels with increasing number of examples*

isolation. We noticed that turn (accuracy: 98.7%; baseline: 98.1%), time (accuracy 99.3%; baseline: 98.1%), contact (accuracy: 93.3%; baseline: 87.8%), topic (accuracy: 98.2%; baseline: 97.8%) and social obligation management functions (accuracy: 96.9% (baseline: 96.2%) are almost perfectly learned, even when the system is provided with only a relatively limited set of examples. However, we should admit here that the OVIS corpus in general does not show rich variation with respect to these particular functions. For example, turns with few exceptions are distributed according to the following pattern TURN GIVING (prompt of the system) - TURN ACCEPT (prompt of the user), which causes no learning problems for the system. The most frequently used function for contact management as CONTACT INDICATION (first prompt of the user in dialogue) is also very easily learned by the system.

The other functions are not so easily learned: general purpose, feedback and discourse structuring functions eventually reach a success score of approximately 70.1% (baseline: 60.8%), 72.1% (baseline: 66.2%) and 82.9% (baseline: 76.4%) respectively. These scores should be evaluated in the light of the relatively high degree of granularity of these functions and the relatively low frequency of each of them in the training sets.

These experiments indicate that the multifunctionality of dialogue utterances is learnable in a data-oriented way; all scores obtained in experiments on the whole training set were higher than those of the baseline. The tendency is quite clear that the more examples are added to the training set, the higher the accuracy that is reached, which is promising for future work on larger corpus data. The training set we had is obviously too small (only 1971 utterances) to get better results, and some functions are underrepresented.

Conclusions

In this paper we argued, first, that since most utterances have more than one communicative function, dialogue act annotation schemes should support annotation in several dimensions simultaneously. The annotation of utterances in multidimensional space can help to represent the meaning of dialogue contributions more accurately, and was found to facilitate annotation.

This was backed up by a comparative analysis of a one-dimensional and multidimensional dialogue act taxonomies, which was helpful to obtain a good understanding of the advantages of the multidimensional approach.

We supported our analysis of spoken utterances by the analysis of nonverbal behavior in meetings. We showed that nonverbal communicative acts may serve several communicative functions simultaneously within different dimensions, especially those concerned with Feedback, Interaction and Social Obligation management.

Finally, we showed that the multifunctionality of dialogue utterances is automatically learnable in a data-oriented way, having trained a Naive Bayes classifier on all dimensions in isolation (tag learning) and on all classes together (label learning). The automatic learnability should be proved on larger annotated corpus data from a different domain. Since the DIT dialogue act taxonomy is not task- or domain dependent, it could be applied to any dialogue data without any adjustments. The labels we used in our OVIS training set could be used for the annotation of other dialogues, e.g. the AMI or IDIAP corpus (see http://www.idiap.ch/mmm/corpora).

Acknowledgments

This research was partly conducted within the project 'Multidimensional Dialogue Modelling', sponsored by the Netherlands Organisation for Scientific Research (NWO), under grant reference 017.003.090. This paper is largely based on earlier work performed in collaboration with Dirk Heylen and Rieks op den Akker within the AMI Training Programme at the University of Twente in the Human Media Interaction (HMI) Research Group, to whom the authors are thankful for their comments and provided corpus materials.

References

[1] Allen, J., Core, M. (1997). *Draft of DAMSL: Dialog Act Markup in Several Layers.* Available at http://www.cs.rochester.edu/research/cisd/resources/damsl/.

[2] Allwood, J. (2000). *An activity-based approach to pragmatics.* In: Abduction, Belief and Context in Dialogue, H. Bunt and W. Black (eds.), Amsterdam: John Benjamin, 47-81.

[3] Bunt, H. (2000). *Dynamic interpretation and dialogue theory.* In: M. Taylor, F. Neel, and D. Bouwhuis (eds.), *The structure of multimodal dialogue II*, Amsterdam: Benjamins, 139-166.

[4] Bunt, H. (2000). *Dialogue pragmatics and context specification.* In: H. Bunt and W. Black (eds) Abduction, Belief and Context in Dialogue. Amsterdam: Benjamins, 81-150.

[5] Bunt, H. (2005). *A Framework for Dialogue Act Specification.* ISO-SIGSEM workshop, Tilburg. Available at: http://let.uvt.nl/research/TI/sigsem/wg/meeting4/fdas-orig.pdf

[6] Bunt, H. (2006). *Dimensions in Dialogue Act Annotation.* In: Proceedings LREC 2006, Genova.

[7] Butterworth, B. (1980). *Evidence from pauses in speech.* In: Language Production: Speech and Talk, vol.1, B.Butterworth(eds.), London: Academic Press, 155-177.

[8] Cassell, J., Torres, O., Prevost, S.(1999). *Turn Taking vs. Discourse Structure: How Best to Model Multimodal Conversation.* In: Machine Conversations, Wilks, Y.(eds.), The Hague: Kluwer, 143-154.

[9] Core, M. and Allen, J. (1997). *Coding Dialogs with the DAMSL Annotation Scheme.* AAAI Fall Symposium on Communicative Action in Humans and Machines, Boston, MA, November 1997.

[10] Core, M. (1998). *Analyzing and Predicting Patterns of DAMSL Utterance Tags.* AAAI Spring Symposium on Applying Machine Learning to Discourse Processing, Stanford, CA, March 1998.

[11] Geertzen, J., Bunt, H. (2006). *Measuring annotator agreement in a complex hierarchical dialogue act scheme.* In: Proceedings of SIGDIAL 2006, Sydney.

[12] Kendon, A. (2004). *Gesture: visible action as utterance.* Cambridge: University Press.

[13] Klein, M. (1999). *Standardization efforts on the level of dialogue act in the MATE project.* Available at http://acl.ldc.upenn.edu/W/W99/W99-0305.pdf.

[14] Larsson, S. (1998). *Coding Schemas for Dialogue Moves.* Technical report from the S-DIME project.

[15] Lendvai, P., Bosch, van den A., Krahmer, E. (2003). *Machine Learning for Shallow Interpretation of User Utterances in Spoken Dialogue Systems.* In: Proceedings of EACL-03 Workshop on Dialogue Systems: interaction, adaptation and styles of management, Budapest.

[16] McNeill, D. (1992). *Hand and Mind: what gestures reveal about thought.* Chicago and London: The University of Chicago Press.

[17] Poesio, M, Traum, D. (1998). *Towards an axiomatization of dialogue acts.* In: Proceedings of Twente Workshop on the Formal Semantics and Pragmatics of Dialogues, 207-222.

[18] Popescu-Belis, A. (2005). *Dialogue Acts: One or More Dimensions?.* ISSCO working paper.

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

[20] Larsson, S, Traum, D. (2000). *Information state and dialogue management in the Trindi dialogue move engine toolkit.* In: Natural Language Engineering 6(3-4), 323-340.

[21] Traum, D. (2000). *20 questions on dialogue act taxonomies.* In: Journal of Semantics, 17(1), 7-30.