Proceedings Chapter

Multi-level Dialogue Act Tags

CLARK, Alexander, POPESCU-BELIS, Andréi

Abstract

In this paper we discuss the use of multi-layered tagsets for dialogue acts, in the context of dialogue understanding for multi-party meeting recording and retrieval applications. We discuss some desiderata for such tagsets and critically examine some previous proposals. We then define MALTUS, a new tagset based on the ICSI-MR and Switchboard tagsets, which satisfies these requirements. We present some experiments using MALTUS which attempt to compare the merits of integrated versus multi-level classifiers for the detection of dialogue acts.

Multi-level Dialogue Act Tags

Alexander Clark and Andrei Popescu-Belis
ISSCO / TIM / ETI
University of Geneva
UNI-MAIL, Boulevard du Pont-d'Arve 40
CH-1211 Geneva 4
Switzerland
asc@aclark.demon.co.uk andrei.popescu-belis@issco.unige.ch

Abstract

In this paper we discuss the use of multi-layered tagsets for dialogue acts, in the context of dialogue understanding for multi-party meeting recording and retrieval applications. We discuss some desiderata for such tagsets and critically examine some previous proposals. We then define MALTUS, a new tagset based on the ICSI-MR and Switchboard tagsets, which satisfies these requirements. We present some experiments using MALTUS which attempt to compare the merits of integrated versus multi-level classifiers for the detection of dialogue acts.

1 Introduction

The processing of dialogues by computers serves two main applicative goals: understanding of human dialogues, for information extraction or summarization, and human-computer dialogue management, for language-based or multimodal interfaces. Whether the computer takes part in a dialogue or only attempts to monitor a recorded one, it is important to detect the functions of each of the human utterances that constitute the dialogue. In addition, when the computer must generate an utterance as a reply, this must also bear some of the functions expected by the hearer in return.

In this article, we focus on dialogue understanding for a dialogue storage and retrieval application, developed in the (IM)2 project[1]. The goal of the application is the multimodal recording of meetings (such as staff or business meetings), the processing and storage of the recordings into a database, and the possibility of querying the dialogue database (Armstrong et al., 2003). The query interface and the processing of the dialogue must therefore meet the needs of the potential users of the system, who will attempt to retrieve various types of information from the meeting recordings. While the result of the query is in general a chunk of recorded dialogue (preferably with multimedia rendering), the criteria used to query the database can vary from trivial ("who attended the meeting?") to very abstract ("what were the main decisions?"). Some form of understanding of the dialogue structure is thus required for a significant proportion of potential queries (more about requirements in subsection 2.3).

The utterance functions with which we deal in this paper are *dialogue acts*. Although dialogue acts (DA) tags are commonly used as a simple representation of the function of an utterance in dialogue, there is little consensus amongst researchers about what set of DA tags is appropriate in a particular situation. Our own application domain, meeting recording, is comparatively open-ended and we do not yet have a clear understanding of precisely what features will be most useful. In section 2, we will try to understand the multiplicity of DA tagsets, then we will analyze (section 3) the dialogue data and annotations on which we work. These considerations prompted us to abstract a new DA tagset, of which we explain the merits in section 4. Experiments on the automatic annotation of DAs using the MALTUS tagset are described in section 5; the results (subsection 5.2) are followed by a brief discussion.

2 Understanding Dialogue Structure: Dialogue Acts

2.1 The Concepts behind Dialogue Acts

Dialogues are series of speaker turns. Utterances can be defined as the atomic subparts of a turn that accomplish one or more "functions" with respect to speaker interaction. Utterances are in general sig-

[1] (IM)2 stands for Interactive Multimodal Information Management, a project sponsored by the Swiss Government (see http://www.im2.ch).

nalled by syntactic and/or prosodic means, but the specificity of their "function" belongs to pragmatics (Levinson, 1983, ch. 4). Linguists have identified several dimensions for the role of sentences uttered in a dialogue. These dimensions are not mutually exclusive, and there are certainly correlations between some of them (e.g. "question" as a speech act and as a member of an adjacency pair).

- Speech acts (Searle, 1969; Vanderveken, 1990): (1) representatives, such as assertions or conclusions; (2) directives, such as requests, questions, suggestions; (3) commissives, such as promises, threatenings, offers; (4) expressives such as thanks, apologies, congratulations; (5) declarations, such as excommunications, declarations of war, christening, firing from employment, etc.
- Turn management: backchannel, floor holder, floor grabber, hold;
- Adjacency pairs: utterances can be the first part or the second part of exchange pairs such as request / accept (or refuse); offer / accept; assess / (dis)agree; question / answer; etc.
- Overall organization and topics: openings, closings, topic-changers, topic-continuers, etc.
- Politeness management: face-threatening, face-saving, neutral;
- Rhetorical role: elaboration, purpose, restatement, etc.

2.2 Dialogue Acts in Computational Linguistics

There is not much agreement, within the CL/NLP community, on the definition of a dialogue act. The term denotes some function of an utterance in a dialogue, not reducible to its syntactic or semantic content. The function is selected, in general, among a set of possible dialogue acts (a DA tagset) that depends on the goals of its creator (Traum, 2000). One of the main inspiration sources for DA tagsets are speech acts, but the original repertoire (Searle, 1969; Vanderveken, 1990) has been gradually enriched with other possible functions. From the numerous DA tagsets (Klein and Soria, 1998), the following are particularly relevant to a general-domain meeting recording application.

The DA tags in DAMSL (Allen and Core, 1997) are nearly all independent: the DAMSL guidelines state that all tags (i.e. all "functions") that characterize an utterance should be associated with it. The DAMSL tags are grouped in four dimensions: communicative status, information level, forward-looking function and backward-looking function. In fact, several theories are conflated in DAMSL, which was initially designed as a shared resource with a focus primarily on task-oriented dialogs (Core and Allen, 1997). There are about 4 million possible combinations of DAMSL tags, which make a huge search space for automatic annotation.

The application of DAMSL to the Switchboard data (two-party telephone conversations) lead to SWBD-DAMSL (Jurafsky et al., 1997), a smaller tagset than DAMSL. About 200,000 SWBD utterances were first annotated with DAMSL tags: it was observed that only 220 combinations of tags occurred (Jurafsky et al., 1998). These 220 labels were then clustered into 42 tags, such as: statement (36%), opinion (13%), agree/accept (5%), yes-no-question (2%). The resulting search space (42 mutually exclusive tags) was well adapted to the initial goals, viz., the automatic annotation of dialogue acts and the use of dialogue act specific language models in speech recognition (Stolcke et al., 2000).

2.3 Requirements for the Definition of a DA Tagset

In this paper, our goal is to design a new DA tagset for our application, with the following constraints in mind (see also the analysis by D. Traum (2000)):

- Relation to one or more existing theories (descriptive, explanatory, etc.).
- Compatibility with the observed functions of actual utterances in context, in a given domain.
- Empirical validation: reliability of human application of the tagset to typical data (high inter-annotator agreement, at least potentially).
- Possibility of automatic annotation (this requirement is specific to NLP).
- Relevance to the targeted NLP application: there are numerous possible functions of utterances, but only some of them are really useful to the application. Within our IM2.MDM project, a study has been conducted on the relevance of dialogue acts (in particular) to typical user queries on meeting recordings (Lisowska, 2003)[2].
- Mapping (at least partially) to existing tagsets, so that useful insights are preserved, and data can be reused.

[2]Many other potential uses of dialogue act information have been hypothesized, such as their use to increase ASR accuracy (Stolcke et al., 2000), or to locate "hot spots" in meetings (Wrede and Shriberg, 2003).

3 Available Data and Annotations: ICSI Meeting Recorder

The volume of available annotated data suffers from the diversity of DA tagsets (Klein and Soria, 1998). One of the most significant resources is the Switchboard corpus mentioned above, but telephone conversations have many differences with multi-party meetings. Apart from the data recently available in the IM2 project, results reported in this paper make use of the ICSI Meeting Recording (MR) corpus of transcribed and annotated dialogues (Morgan et al., 2003; Shriberg et al., 2004)[3].

3.1 Overview of ICSI MR Corpus

The ICSI-MR corpus consists of 75 one-hour recordings of staff meetings, each involving up to eight speakers on separate mike channels. Each channel was manually transcribed and timed, then annotated with dialogue act and adjacency pair information (Shriberg et al., 2004). Following a preliminary release in November 2003 (sound files, transcriptions, and annotations), the full corpus was released in February 2004 to IM2 partners.

The dialogue act annotation makes use of the pre-existing segmentation of each channel into (prosodic) utterances, sometimes segmented further into functional utterances, each of them bearing a separate dialogue act. There are about 112,000 prosodic utterances, and about 7,200 are segmented into two functional utterances (only one is segmented in three).

3.2 Discussion of the ICSI-MR DA Tagset

Each functional utterance from the ICSI-MR corpus is marked with a dialogue label, composed of one or more tags from the ICSI-MR tagset (Dhillon et al., 2004). The tagset, which is well documented, is based on SWBD-DAMSL, but unlike SWBD-DAMSL, it allows one utterance to be marked with multiple tags. Also, the SWBD-DAMSL tagset was extended, for instance with disruption tags such as 'interrupted', 'abandoned', etc. Utterances can also be marked as 'unintelligible' or 'non-speech'. An ICSI-MR label is made of a general tag, followed by zero or more specific tags, followed or not by a disruption tag:

```
gen_tag [^spec_tag_1 ... ^spec_tag_n] [.d]
```

Our formalization of the guidelines using rewriting rules (Popescu-Belis, 2003) shows that few tags are mutually exclusive. The number of possible combinations (DA labels) reaches several millions. For instance, even when not considering disruption marks, the labels are a combination of one general tag out of 11, and one or more specific tags out of 39. If up to five specific tags are allowed (as observed empirically in the annotated data), there are more than 7,000,000 possible labels; if specific tags are limited to four, there are about 1,000,000 possible labels.

Some studies acknowledge the difficulties of annotating precisely with ICSI-MR, but also the fine-grained distinctions it allows for, e.g. between the possible functions of four related discourse particles ('yeah', 'right', 'okay', and 'uhhuh'): agreement/acceptance, acknowledgment, backchannel, floor grabber (Bhagat et al., 2003). Conversely, inter-annotator agreement on such fine-grained distinctions (specific tags) is lower than agreement on major classes, though the kappa-statistic normally used to measure agreement adjusts to a certain extent for this. In fact, ICSI-MR also provided a set of five 'classmaps' that indicate how to group tags into categories which reduce the number of possible labels. For instance, the simplest one reduces all DA labels to only five classes: statement, question, backchannel, floor holder/grabber, disruption. Our MALTUS proposal (see 4.1 below) could be viewed as a classmap too: it preserves however more ICSI-MR tags than the existing classmaps, and assigns in addition conditions of mutual exclusiveness.

We also note that, while SWBD-DAMSL was an attempt to reduce the dimensionality of the DAMSL tagset (which had a clear theoretical base), the ICSI-MR tagset allows SWBD tags to be combined again instead of going back to DAMSL tags. Although our proposal that we proceed to describe (MALTUS) remains close to ICSI-MR for reusability reasons, we are also working on a more principled DA tagset that departs from ICSI-MR (Popescu-Belis, 2003).

3.3 Some Figures for the ICSI-MR Data

In the process of conversion to MALTUS (see 4.2 below), we validated the ICSI-MR data and made several observations. Detected incoherent combinations of tags (e.g., two general tags in a label) and other remarks have also been sent back to ICSI.

We first separate prosodic utterances into functional utterances, so that each utterance has one DA label (and not two, separated by '|'), thus obtaining 120,205 utterances. Also at this stage, we split utterances that correspond to reported speech (marked with ':'). We then discard the disruption marks to focus on the DA labels only – about 12,000 labels out of ca. 120,000 are disruption marks, or contain one. We are left with 113,560 utterances with DA labels, with 776 observed types of labels. An important parameter is the number of occurring vs. possible labels,

[3] See `http://www.icsi.berkeley.edu/Speech/mr/`

Nb. of tags in label	Nb. of theoretical comb.	Nb. of occurring comb.	Nb. of tokens
1	11	11	68,213
2	429	129	37,889
3	8,151	402	5,054
4	100,529	176	2,064
5	904,761	49	326
6	6,333,327	9	14
7	...	0	0
Total:	7,347,208	776	113,560

Table 1: Number of possible labels (combinations of tags): theoretical vs. actual.

Maximal nb. of tags	Maximal theoretical accuracy on ICSI-MR
1	0.601
2	0.934
3	0.979
4	0.997
5	0.999
6	1

Table 2: Maximal accuracy of DA tagging of the ICSI-MR data that could be reached using a limited number of tags per label.

which depends a lot on the number of specific tags in a label, as summarized in table 1. The maximum observed in the available data is five specific tags in a label (hence six tags in all).

There is no guarantee that meaningful labels cannot have more than six tags. However, such labels are probably very infrequent, and a reasonable option for automatic tagging is to limit the number of tag combinations, which is the main goal of the MALTUS tagset. The maximal accuracies that could be obtained on the available ICSI-MR data if the number of tags in a label was limited to 1, 2, etc. are shown in Table 2. In computing the accuracy we consider here only perfect matches, but scores could be higher if partial matches count too. Two or three tags per label already allow very high accuracy, while considerable reducing the search space.

4 The MALTUS DA Tagset

4.1 Definition

We defined MALTUS (Multidimensional Abstract Layered Tagset for Utterances) in order to reduce the number of possible combinations by assigning exclusiveness constraints among tags, while remaining compatible with ICSI-MR (Popescu-Belis, 2003). MALTUS is more abstract than ICSI-MR, but can be refined if needed. An utterance is either marked U (undecipherable) or it has a general tag and zero or more specific tags. It can also bear a disruption mark. More formally (? means optional):

```
DA -> (U | (gen_tag (spec_tags)?)) (.D)?
   gen_tag   ->  S | Q | B | H
   spec_tags -> (RP | RN | RU)? AT? DO? PO?
```

The glosses of the tags, generally inspired from ICSI-MR, are:

- U = undecipherable (unclear, noisy)
- S = statement
- Q = question
- B = backchannel
- H = hold (floor holder, floor grabber, hold)
- RP = positive answer (or positive response)
- RN = negative answer (or negative response)
- RU = other answer (or undecided answer or response)
- RI = restated information
- DO = command or other performative (can be refined into: command, commitment, suggestion, open-option, explicit performative)
- AT = the utterance is related to attention management (can be refined into: acknowledgement, rhetorical question backchannel, understanding check, follow me, tag question)
- PO = the utterance is related to politeness (can be refined into sympathy, apology, downplayer, "thanks", "you're welcome")
- D = the utterance has been interrupted or abandoned

4.2 Conversion of ICSI-MR to MALTUS

There are only about 500 possible MALTUS labels, but observations of the converted ICSI-MR data show again that the probability distribution is very skewed. An explicit correspondence table and conversion procedure were designed to convert ICSI-MR to MALTUS, so that the considerable ICSI-MR resource can be reused.

Correspondences between MALTUS and other tagsets (Klein and Soria, 1998) were also provided (Popescu-Belis, 2003). Such "mappings" are imperfect for two reasons: first, they work only in one direction, from the more specific tagset (ICSI-MR / SWBD / DAMSL) to the more abstract one (MALTUS). Second, a mapping is incomplete if one does not state which tags must be mutually exclusive.

For MALTUS too, the idea to use at most three tags per label in an automatic annotation program might reduce the search space without decreasing the accuracy too much. Another idea is to use only the labels that appear in the data that is, only 50 labels. An even smaller search space is provided by the 26 MALTUS labels that occur more than 10 times each. If only these are used for tagging, then only 70 occurrences (only 0.061% of the total) would be incorrectly tagged, on the ICSI-MR reference data. Occurring labels ordered alphabetically and their frequencies (when greater than 10) are listed below.

```
B (15180)
H (12288)
Q (5320)
  Q^AT (3137)
    Q^AT^RI (69)
  Q^DO (239)
  Q^RI (60)
  Q^RN (19)
S (51304)
  S^AT (8280)
    S^AT^RI (273)
  S^DO (3935)
    S^DO^RI (32)
    S^DO^RN (38)
    S^DO^RP (41)
    S^DO^RU (16)
  S^PO (791)
    S^PO^RI (13)
    S^PO^RU (61)
  S^RI (765)
    S^RI^RN (46)
    S^RI^RP (436)
    S^RI^RU (18)
  S^RN (2219)
  S^RP (7612)
  S^RU (1298)
```

Further analysis will tell whether this list should be enriched with useful labels that are absent from it. Also, a comparison of MALTUS to the SWBD set (26 labels vs. 42) should determine whether the loss in informativeness in MALTUS is compensated by the gain in search space size and in theoretical grounding.

5 Automatic Classification

As discussed above, one of the desiderata for a tagset in this application domain is that the tags can be applied automatically. A requirement for annotations that can only be applied manually is clearly unrealistic except for meetings of very high importance. The ICSI-MR corpus on the other hand is concerned with producing a body of annotated data that can be used by researchers for a wide range of different purposes: linguists who are interested in particular forms of interaction, researchers in acoustics and so on. It is by no means a criticism of their work that some of the distinctions that they annotate or attempt to annotate cannot be reliably automated.

Here we report some preliminary experiments on the automatic annotation of meeting transcripts with these tagsets. Our focus here is not so much on evaluating a classifier for this task but rather evaluating the tagsets: we are interest in the extent to which they can be predicted reliably from easily extracted features of the utterance and its context. Additionally we are interested in the multi-level nature of the tagsets and exploring the extent to which the internal structure of the tags allows other options for classifiers. Therefore, our goal in these experiments is not to build a high performance classifier; rather, it is to explore the extent to which multi level tagsets can be predicted by classifying each level separately – i.e. by having a set of "orthogonal" classifiers – as opposed to classifying the entire structured object in a single step using a single multi-class classifier on a flattened representation. Accordingly there are a number of areas in which our experimental setup differs from that which would be appropriate when performing experiments to evaluate a classifier.

Since in this paper we are not using prosodic or acoustic information, but just the manual transcriptions, there are two sources of information that can be used to classify utterances. First, the sequence of words that constitutes the utterance, and secondly the surrounding utterances and their classification. generally in prior research in this field, some form of sequential inference algorithm has been used to combine the local decisions about the DA of each utterance into a classification of the whole utterance. The common way of doing this has been to use a hidden Markov model to model the sequence and to use a standard decoding algorithm to find either the sequence with maximum a posteriori (MAP) likelihood or to select for each utterance the DA with MAP likelihood. In the work here, we will ignore this complexity and allow our classifier access to the gold standard classification of the surrounding utterances. This will make the task substantially easier, since in a real application, there will be some noise in the labels.

5.1 Feature selection

There are two sorts of features that we shall use here – internal lexical features derived from the words in the utterance, and contextual features derived from

the surrounding utterances. At our current state of knowledge we have a very good idea about what the lexically derived features should be, and how they should be computed – namely n-grams or gappy n-grams including positional information. Additionally, there are ways of computing these efficiently. However, with regard to the contextually derived features, our knowledge is much less complete. (Stolcke et al., 2000) showed that in the Switchboard corpus there was little dependence beyond the immediately adjacent utterance, but whether this also applies in this multi-party domain is unknown. Thus we find ourselves in a rather asymmetric position with regard to these two information sources. As we are not here primarily interested in constructing a high performance classifier, but rather identifying the predictable elements of the tag, we have resolved this problem by deliberately selecting a rather limited set of lexical features, together with a limited set of contextual features. Otherwise, we feel that our experiments would be overly biased towards those elements of the tag that are predictable from the internal lexical evidence.

We used as lexical features the 1000 most frequent words, together with additional features for these words occurring at the beggining or end of the utterance. This gives an upper bound of 3000 lexical features. We experimented with a variety of simple contextual features.

Preceding same label (SL) the immediately preceding utterance on the same channel has a particular DA tag.

Preceding label (PL) a preceding utterance on a different channel has a particular DA tag. We consider an utterance to be preceding if it starts before the start of the current utterance.

Overlapping label (OL) an utterance on another channel with a particular DA tag overlaps the current utterance. We anticipate this being useful for identifying backchannels.

Containing label (CL) an utterance on another channel with a particular DA tag contains the current channel – i.e. the start is before the start of the current utterance and the end is after the end of the current utterance.

Figure 1 shows an artificial example in a multiparty dialog with four channels. This illustrates the features that will be defined for the classification of the utterance that is shaded. In this example we will have the following features `SL:C1`, `PL:B1`, `PL:D1`, `CL:D1`, `OL:A1`, `OL:B1`, `OL:B2`, `OL:D1`. We have found

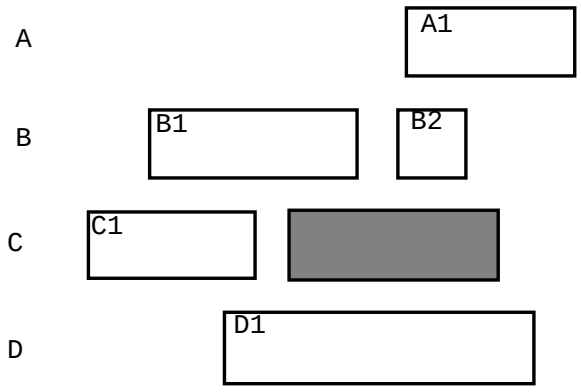

Figure 1: Artificial example illustrating contextual features defined for a particular utterance (shaded). There are four channels labelled A to D; each box represents an utterance, and the DA tag is represented by the characters inside each box.

that the overlapping label feature set does not help the classifiers here, so we have used the remaining three contextual feature sets. Note the absence of contextual features corresponding to labels of utterances that strictly follow the target utterance. We felt that given the fact that we use the gold standard tags this would be too powerful.

The data made available to us was preprocessed in a number of ways. The most significant change was to split utterances that had been labelled with a sequence of DA labels (joined with pipes). We separated the utterances and the labels at the appropriate points and realigned. The data was provided with individual time stamps for each word using a speech recognizer in forced recognition mode: where there were errors or mismatches we discarded the words.

5.2 Results

We use a Maximum Entropy (ME) classifier (Manning and Klein, 2003) which allows an efficient combination of many overlapping features. We selected 5 meetings (6771 utterances after splitting) to use as our test set and 40 as our training set leaving a further five for possible later experiments. As a simple baseline we use the classifier which just guesses the most likely class. We first performed some experiments on the original tag sets to see how predictable they are.

We started by defining a simple six-way classification task which classifies disruption forms, and undecipherable forms as well as the four general tags defined above. This is an empirically very well-founded distinction: the ICSI-MR group have provided some

inter-annotator agreement figures(Carletta et al., 1997) for a very similar task and report a kappa of 0.79. Our ME classifier scored 77.9% (baseline 54.0%).

We also tested a few simple binary classifications to see how predictable they are. Utterances are annotated for example with a tag J if they are a joke. As would be expected, the Joke/Non-Joke classification produced results not distinguishable from chance. The performance of the classifiers on separating disrupted utterances from non disrupted forms scored slightly above chance at 89.9% (against baseline of 87.0%). We suspect that more sophisticated contextual features could allow better performance here. A more relevant performance criterion for our application is the accuracy of classification into the four general tags. In this case we removed disrupted and undecipherable utterances, slightly reducing the size of the test set, and achieved a score of 84.9% (baseline 64.1%).

With regard to the larger sets of tags, since they have some internal structure it should accordingly be possible to identify the different parts separately, and then combine the results. We have therefore performed some preliminary experiments with classifiers that classify each level separately. We again removed the disruption tags since with out current framework we are unable to predict them accurately. The baseline for this task is again a classifier that chooses the most likely tag (S) which gives 41.9% accuracy. Using a single classifier on this complex task gave an accuracy of 73.2%.

We then constructed six classifiers as follows

Primary classifier S, H, Q or B

Politeness classifier PO or not PO

Attention classifier AT or not AT

Order classifier DO or not DO

Restatement classifier RI or not RI

Response classifier RP, RN, RU or no response

These were trained separately in the obvious way and the results combined. This complex classifier gave an accuracy 70.5%. This mild decrease in performance is rather surprising – one would expect the performance to increase as the data sets for each distinction get larger. This can be explained by dependences between the classifications. There are a number of ways this could be treated – for example, one could use a sequence of classifiers, where each classifier can use the output of the previous classifier as a feature in the next. It is also possible that these dependencies reflect idiosyncracies of the tagging process: tendencies of the annotators for whatever reasons to favour or avoid certain combinations of tags.

6 Conclusion

We have discussed some issues concerning the design and use of dialogue act tagsets. It is too early to draw firm conclusions from this preliminary study. We can note the obvious point that simplified smaller tagsets are easier to predict accurately than larger ones. There appear to be non-trivial dependencies between the tags for reasons that are not yet clear. We expect the performance of a final, fully automatic classifier to be substantially higher than the results presented here, owing to the use of more powerful classifiers and, more importantly, larger and richer feature sets. Finally we note that an important point of tagset design has not been addressed empirically here: the question of whether particular distinctions in the tagset are actually useful in our application. Future studies will address this point by studying the queries formulated by potential users of meeting processing and retrieval systems.

Acknowledgments

We are grateful to the ICSI MR group for sharing with us the data as part of the IM2/ICSI agreement – in particular to Barbara Peskin and Liz Shriberg. This research is part of the Multimodal Dialogue Management module (see `http://www.issco.unige.ch/projects/im2/mdm`) of the IM2 project.

References

James F. Allen and Mark G. Core. 1997. DAMSL: Dialog act markup in several layers (draft 2.1). Technical report, Multiparty Discourse Group, Discourse Research Initiative, September/October 1997.

Susan Armstrong, Alexander Clark, Giovanni Coray, Maria Georgescul, Vincenzo Pallotta, Andrei Popescu-Belis, David Portabella, Martin Rajman, and Marianne Starlander. 2003. Natural language queries on natural language data: a database of meeting dialogues. In *NLDB'2003 (8th International Conference on Applications of Natural Language to Information Systems)*, Burg/Cottbus, Germany.

Sonali Bhagat, Hannah Carvey, and Elizabeth Shriberg. 2003. Automatically generated prosodic cues to lexically ambiguous dialog acts in multiparty meetings. In *ICPhS 2003*, Barcelona.

Jean Carletta, Amy Isard, Stephen Isard, Jacqueline C. Kowtko, Gwyneth Doherty-Sneddon, and Anne H. Anderson. 1997. The reliability of a dialogue structure coding scheme. *Computational Linguistics*, 23:13–31.

Mark G. Core and James F. Allen. 1997. Coding dialogues with the DAMSL annotation scheme. In David Traum, editor, *Working Notes: AAAI Fall Symposium on Communicative Action in Humans and Machines*, pages 28–35, Menlo Park, CA. American Association for Artificial Intelligence.

Rajdip Dhillon, Sonali Bhagat, Hannah Carvey, and Elizabeth Shriberg. 2004. Meeting recorder project: Dialog act labeling guide. Technical Report TR-04-002, ICSI (International Computer Science Institute), Berkeley, CA.

Daniel Jurafsky, Elizabeth Shriberg, and Debra Biasca. 1997. Switchboard SWBD-DAMSL shallow-discourse-function annotation (coders manual, draft 13). Technical Report 97-02, University of Colorado, Institute of Cognitive Science.

Daniel Jurafsky, Elizabeth Shriberg, Barbara Fox, and Traci Curl. 1998. Lexical, prosodic, and syntactic cues for dialog acts. In *ACL/COLING-98 Workshop on Discourse Relations and Discourse Markers*, pages 114–120.

Marion Klein and Claudia Soria. 1998. Dialogue acts. In Marion Klein, Niels Ole Bernsen, Sarah Davies, Laila Dybkjaer, Juanma Garrido, Henrik Kasch, Andreas Mengel, Vito Pirrelli, Massimo Poesio, Silvia Quazza, and Claudia Soria, editors, *MATE Deliverable 1.1: Supported Coding Schemes*, MATE (Multilevel Annotation, Tools Engineering) European Project LE4-8370.

Stephen C. Levinson. 1983. *Pragmatics*. Cambridge University Press, Cambridge, UK.

Agnes Lisowska. 2003. Multimodal interface design for the multimodal meeting domain: Preliminary indications from a query analysis study. Technical report, IM2.MDM, 11/2003.

Christopher Manning and Dan Klein. 2003. Optimization, maxent models, and conditional estimation without magic. In *Tutorial at HLT-NAACL 2003 and ACL 2003*. ACL, Edmonton, Canada.

Nelson Morgan, Don Baron, Sonali Bhagat, Hannah Carvey, Rajdip Dhillon, Jane A. Edwards, David Gelbart, Adam Janin, Ashley Krupski, Barbara Peskin, Thilo Pfau, Elizabeth Shriberg, Andreas Stolcke, and Chuck Wooters. 2003. Meetings about meetings: research at ICSI on speech in multiparty conversations. In *ICASSP 2003 (International Conference on Acoustics, Speech, and Signal Processing)*, Hong Kong, China.

Andrei Popescu-Belis. 2003. Dialogue act tagsets for meeting understanding: an abstraction based on the DAMSL, Switchboard and ICSI-MR tagsets. Technical report, IM2.MDM, v1.1, 09/2003.

John R. Searle. 1969. *Speech Acts*. Cambridge University Press, Cambridge, UK.

Elizabeth Shriberg, Raj Dhillon, Sonali Bhagat, Jeremy Ang, and Hannah Carvey. 2004. The ICSI meeting recorder dialog act (MRDA) corpus. In *Proceedings of SIGDIAL '04 (5th SIGdial Workshop on Discourse and Dialog)*, Cambridge, MA.

Andreas Stolcke, Klaus Ries, Noah Coccaro, Elizabeth Shriberg, Rebecca Bates, Daniel Jurafsky, Paul Taylor, Rachel Martin, Marie Meteer, and Carol Van Ess-Dykema. 2000. Dialogue act modeling for automatic tagging and recognition of conversational speech. *Computational Linguistics*, 26(3):339–371.

David R. Traum. 2000. 20 questions for dialogue act taxonomies. *Journal of Semantics*, 17(1):7–30.

Daniel Vanderveken. 1990. *Meaning and speech acts*. Cambridge University Press, Cambridge, UK.

Britta Wrede and Elizabeth Shriberg. 2003. The relationship between dialogue acts and hot spots in meetings. In *IEEE Speech Recognition and Understanding Workshop*, St. Thomas, U.S. Virgin Islands.