

Abstracting a Dialogue Act Tagset for Meeting Processing

Andrei Popescu-Belis

ISSCO/TIM/ETI, University of Geneva
40, bd du Pont d'Arve – CH-1211 Geneva 4 – Switzerland
andrei.popescu-belis@issco.unige.ch

Abstract

This paper analyses three existing tagsets for dialogue acts, i.e., the function of utterances in dialogue. Then, a new tagset is proposed, named MALTUS, designed for the annotation of meeting recording transcripts. Several criteria for tagset definition are discussed, along with the possible theoretical inspiration for dialogue act tagsets. The DAMSL, SWBD-DAMSL, and ICSI-MR tagsets are analyzed with respect to the previous considerations. The definition of MALTUS is followed by quantitative data from the conversion and validation of ICSI-MR data, and then by perspectives on automatic tagging using MALTUS, and on further user-based studies of its relevance.

1. Introduction

The understanding of human dialogues is the key to many natural language engineering applications, among which we focus here on automatic *meeting processing and retrieval* (MPR). This application enables people who did not attend a meeting (e.g. a staff or a business meeting), or people who want to review a past meeting, to search for a particular piece of information connected to the meeting (Armstrong et al., 2003).

Within the framework of an MPR application, the understanding of interpersonal dialogue often requires that a dialogue function be assigned to each utterance, beyond its semantic content. For instance, to find unanswered questions in a meeting, a system must first detect which utterances correspond to ‘questions’ and which ones to ‘answers’.

In this paper, we provide a formal analysis of three existing tagsets for dialogue act annotation (DAMSL, SWBD-DAMSL, and ICSI-MR), and then derive a new tagset named MALTUS. After discussing its compatibility and differences with previous tagsets, we explain the conversion of annotated corpora to MALTUS, and the resulting validation. We finally outline research directions showing the relevance of the MALTUS set.

2. Annotation of Dialogue Acts

Many studies have assigned possible functions to *utterances* in dialogue, often depending on the type of the dialogue and on the goal of the study, without general agreement on a unique set, as the discussion by Levinson (1983, ch. 4) shows. In computational linguistics, many DA tagsets have been developed (Klein et al., 1998).

An utterance is a coherent, contiguous series of words from a given speaker, which serves a precise function in the dialogue (or sometimes more than one); or, in other words, carries a dialogue act. An utterance can often be equated with a proposition or a sentence, but in spoken language, utterances do not always correspond to well-formed or completed propositions. Utterances are the building blocks of dialogue structure, the minimal units that are of interest for dialogue retrieval in an MPR application.

Based on evidence from several linguistic theories, we distinguish six functional dimensions in which utterances

can play a role; examples of roles (hence dialogue act ‘tags’) are given for each dimension.

1. **Speech acts** (Searle 1969; Vanderveken, 1990): assertion, request, question, promise, apology, thanking, etc.
2. **Turn management**: backchannel; floor-holder, floor-grabber; hold.
3. **Adjacency pairs**: request or invite / accept or refuse, question / answer, etc.
4. **Overall thematic organization**: opening, closing, change-topic, continue-topic.
5. **Politeness management**: face-threatening, face-saving, neutral (Brown and Levinson, 1983).
6. **Rhetorical role**, e.g. in the RST (Rhetorical Structure Theory) frame: elaboration, purpose, restatement, etc.

3. Constraints for Defining DA Tagsets

The following constraints govern the definition of a tagset for dialogue acts – see also (Traum, 2000) for a more theoretical approach. These constraints could probably be extended to tagsets for other linguistic properties of written or spoken/transcribed data. These constraints can be used to analyze and criticize existing tagsets.

1. **Theory**: the tagset should be related to a theory of the ‘functions’ that it annotates.
2. **Insights from the data**: the tagset should be compatible with observations on actual utterances, in a given domain.
3. **Empirical validation**: the DA set should be reliably tagged by human annotators (high inter-annotator agreement, e.g. using *kappa*).
4. **Possibility of automatic tagging** using the tagset, at a reasonable performance level.
5. **Role of the application**: the tagset should be designed depending on the targeted NLP application (mark relevant ‘functions’ instead of all ‘functions’).
6. **Mapping to existing DA sets**: the DA set should be reasonably compatible with previous tagsets (or at least compared to them) so that useful insights are preserved, and data can be reused.

4. Formal Analysis of existing DA tagsets

Three related tagsets for dialogue acts are analyzed below – see also (Klein et al., 1998) for references to other tagsets. The main criteria, from the list above, are theoretical and empirical validity (1 and 3), and most of all tractability (4), i.e. ease of use for automated annotations of (transcribed) dialogues. Existing resources annotated with the respective tagsets will be mentioned too.

4.1. DAMSL

DAMSL, or Dialogue Act Markup in Several Layers (Allen and Core, 1997), is a four-dimensional tagset, using almost independent tags: the guidelines state that “all labels that apply” should be used for an utterance. We expressed this formally, using rewriting rules (Popescu-Belis, 2003, p. 11-12). An utterance can have zero, one or more labels in each of the following dimensions:

- communication status (e.g., uninterpretable, abandoned, self-talk)
- illocutionary force (e.g., task management, communication management, other)
- forward-looking function (e.g., statement, info-request, explicit-performative, exclamation)
- backward-looking function (e.g., agreement: accept, reject, other; understanding, answer)

DAMSL can be used to annotate arbitrary combinations of functions for each utterance, in the domain of goal-directed dialogues (TRAINS corpus). Several theories are conflated in the tagset. The main problem is that there are over 4 million possible combinations, which make a huge search space for automatic annotation.

4.2. SWBD-DAMSL

The application of DAMSL to the Switchboard (SWBD) data (telephone conversations) was accompanied by the derivation of a new, reduced tagset (Jurafsky et al., 1997). Utterances were annotated with DAMSL, yielding only 220 combinations of tags occurring in ca. 200,000 utterances (Jurafsky et al., 1998). These 220 labels were then clustered according to their similarity into 42 mutually exclusive tags. Examples of tags with their frequencies are: statement (36%), continuer (19%), opinion (13%), agree/accept (5%), abandoned (5%), appreciation (2%), yes-no-question (2%), non-verbal (2%), yes-answer (1%), etc. This tagset is of course well-adapted to automatic annotation, with only 42 possible tags. However, it is less expressive than DAMSL, and must be adapted for multiparty or goal-oriented conversations.

4.3. ICSI-MR

The ICSI-MR tagset was defined for the dialogue act annotation of data from the Meeting Recorder project at ICSI (Morgan et al., 2003). The tagset uses the SWBD-DAMSL tags, but allows the combination of several tags into a label for an utterance (Dhillon et al., 2004; Shriberg et al., 2004). The tagset also extends SWBD-DAMSL with disruption marks such as ‘interrupted’, ‘abandoned’, etc., and the ‘undecipherable’ label. We have expressed the syntax of the ICSI-MR tagset using rewriting rules (Popescu-Belis, 2003, p. 19-20), beyond the following

generic form that was provided by ICSI (Dhillon et al., 2004):

```
g-tag [^s-tag1 ... ^s-tagN].dis-tag
```

An ICSI label is made of one general tag, followed by zero or more specific tags, followed or not by a disruption tag (which may also appear alone). There is very little explicit dependence between tags, such as mutual exclusiveness. Our formalization shows that very few tags are mutually exclusive, and that the number of possible combination reaches several millions.

In a preliminary, empirical study, we found out that in six hours of meetings (ca. 7000 utterances), there were about 400 different tags (Clark and Popescu-Belis, 2004). Moreover, 13 tags had four different second-level tags.

The inter-annotator agreement reaches an acceptable level only when it is measured on a “reduction” of the DA labels to a much smaller number of classes, e.g. *kappa* of 0.8 when only five classes are used (Dhillon et al., 2004). The final dialogue act annotation of the ICSI-MR data was fixed after discussions among annotators.

5. Abstraction of the MALTUS Tagset

Our definition of MALTUS, a Multidimensional Abstract Layered Tagset for Utterances, has several goals: to reduce the number of possible labels by assigning exclusiveness constraints among tags; to remain compatible with ICSI-MR in order to reuse the data; to remain compatible with theories of dialogue structure; and to be informative enough for an MPR application. MALTUS is *abstract* in the sense that the assigned tags encompass broad meanings and could be refined further on, and *layered* since the labels have one principal component followed by a number of secondary components. This is close to ICSI-MR, but MALTUS sets many more constraints on mutual exclusiveness between tags in a label.

Formally, MALTUS is defined as follows (the same kind of rules was also used to describe the previous tagsets). The tags in boldface are terminal tags, and the other ones do not appear in actual labels. The carets (‘^’) are simply dialogue act separators, while ‘|’ means ‘or’ and ‘?’ denotes an optional tag.

DA	→	(U T1 (^T2?) (.D)?)
T1	→	S Q B H
T2	→	(RP RN RU)? ^AT? ^DO? ^ (RIC RIR)? ^PO?

An utterance is either marked U (undecipherable) or it has a level 1 (T1) tag, and zero or more level 2 tags (T2). In addition, it can bear a disruption mark: the fact that the disruption mark is independent from the undecipherable mark reflects compatibility with the ICSI-MR tagset. The level 1 tag can be: statement, question, backchannel or hold. Level 2 offers non-exclusive options: positive/negative/ other answer, attention, command/performative, restated information, politeness. Therefore, a significant number of functions can be annotated on each utterance, without compromising the size of the search space: there are only 770 possible MALTUS labels (combinations of tags).

The glosses of the tags, generally inspired from ICSI-MR and SWBD-DAMSL, 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)
- RIC = restated information with correction
- RIR = restated information with repetition
- DO = command or other performative (this can be refined into: command, commitment, suggestion, open-option, explicit performative)
- AT = the utterance is related to attention management (this can be refined into one of the following: acknowledgement, rhetorical question backchannel, understanding check, “follow me”, tag question)
- PO = the utterance is related to politeness (this can be refined into sympathy, apology, down-player, “thanks”, “you’re welcome”)
- D = the utterance has been interrupted or abandoned

More details, such as an annotation guide, are provided in (Popescu-Belis, 2003, ch. 4).

6. Validation and Conversion of ICSI-MR Annotations to MALTUS

The MALTUS tagset was designed so that an existing resource of about 75 hours of meeting conversations tagged with ICSI-MR (Shriber et al., 2004) can be reused. An explicit correspondence table and conversion procedure were designed. In this process, the consistency of the ICSI-MR data was checked, and feedback was provided to the maintainers of the ICSI-MR guidelines, in particular on the observed use of the disruption marks.

The conversion of ICSI-MR to MALTUS yielded a much smaller number of occurring labels than the original. A more abstract tagset thus greatly reduces the number of possible labels. In the process of conversion to MALTUS (see below), we validated the ICSI-MR data, detecting incoherent combinations of tags (e.g., two general tags in a label) and sending feedback to ICSI. The following analysis was carried on only 50 hours of data, but will be updated to all 75 hours in the near future.

We first separate prosodic utterances into functional utterances, a separation marked by ‘|’ in the original data, so that each utterance has only one DA label. We also discard the disruption marks, to focus on the DA labels only (about 8,900 labels out of ca. 69,000 are, or contain, disruption marks). We are left with 65,188 utterances with DA labels, with 685 observed types of labels. More precisely, there are 11 types of tags with 1 label, 135 with two labels, 361 with 3 labels, 131 with 4 labels, 42 with 5 and 5 with 6 labels. The maximum observed in the available data is five specific tags in a label (hence six tags in all). The numbers of occurrences are of course the

highest for labels with few tags: there are about 40,000 one-tag labels, and about 21,000 two-tag labels.

We also defined a correspondence between MALTUS and other tagsets (Popescu-Belis, 2003), as summarized for instance in the MATE Deliverable 1.1. Note however that the “mappings” between tagsets are imperfect for two reasons: first, since MALTUS is a rather abstract tagset, the “mapping” works only in one direction, from the more specific (ICSI-MR / SWBD / DAMSL) to the more abstract tagset (MALTUS). Indeed, a more abstract tagset cannot be mapped towards a more detailed one. Second, the problem of dimensionality makes a mapping incomplete, if one does not state which tags are mutually exclusive according to the guidelines. For instance, a conversion from SWBD to MALTUS would generate for each utterance only one tag (or sometimes two) from the abstract set, while up to six tags can be combined for an utterance.

7. Perspectives

7.1. Results on Automatic Annotation

The scores of the automatic detection (annotation) of dialogue acts are influenced by the size of the tagset. For instance, using the 42 SWBD-DAMSL tags and the SWBD data, statistical methods achieve ca. 70% accuracy (Stolcke et al., 2000). Preliminary studies in automatic tagging were conducted on the ICSI-MR data, using only six dialogue act tags: statement, question, backchannel, floor holder/grabber, and disruption (Clark and Popescu-Belis, 2004). On such a task, the baseline performance is 20%, one tag out of five, or 61% if ‘statement’, the most frequent tag, is always used. The automatic tagging in these preliminary studies reached a score of about 78%. The score using MALTUS reached 70% in preliminary tests, but further experiments with the full MALTUS set are under way.

7.2. Validation of Tagset via Query Analysis

Another experimental study attempts to derive user requirements for a meeting processing and retrieval application, by analysing queries elicited from potential users (Lisowska et al., 2004). These queries (about 500 for the moment) show that among the elements most frequently required by users, some dialogue acts such as questions, requests and offers, play a significant role. Therefore, the MALTUS set is suitable to an MPR application, though it could contain also other tags for theoretical completeness. Further experiments should show whether MALTUS could still be simplified overall, while adding some of the tags that users frequently search for in an MPR application.

7.3. A Principled Tagset

Another perspective is the development from scratch of a new tagset that follows more closely the theoretical dimensions of dialogue function outlined in section 1, thus departing from the DAMSL / ICSI-MR flavour. Criteria in section 2 should be used to define a tagset relevant to our application, and to provide compatibility tables with previous tagsets, so that existing resources can be reused.

Acknowledgments

The work presented here is part of the Swiss NCCR on “Interactive Multimodal Information Management” (IM2, <http://www.im2.ch>), funded by the Swiss National Science Foundation. The work pertains specifically to the IM2.MDM module, “Multimodal Dialogue Management” (<http://www.issco.unige.ch/projects/im2/mdm>). We would like to thank the authors of the ICSI-MR corpus, in particular Barbara Peskin and Liz Shriberg, for providing valuable resources and advice.

References

- Allen, J. F., and Core, M. G. (1997). “DAMSL: Dialogue act markup in several layers (draft 2.1)”. Technical report, Multiparty Discourse Group, Discourse Research Initiative, September/October, 1997.
- Armstrong, S., Clark, A., Coray, G., Georgescu, M., Pallotta, V., Popescu-Belis, A., Portabella, D., Rajman, M. and Starlander, M. (2003). “Natural Language Queries on Natural Language Data: a Database of Meeting Dialogues”. Proceedings of NLDB’03, Burg/Cottbus, Germany.
- Brown, P., and Levinson, S. C. (1983). *Politeness*. Cambridge University Press, Cambridge, UK.
- Clark, A. and Popescu-Belis, A. (2004) - Multi-level Dialogue Act Tags. Proceedings of SIGDIAL’04, Cambridge, MA, USA.
- Dhillon, R., Bhagat, S., Carvey, H., and Shriberg, E. (2004). Meeting Recorder Project: Dialogue Act Labeling Guide. ICSI Technical Report TR-04-002, Berkeley, CA, February 9, 2004.
- Jurafsky, D., Shriberg, E., and Biasca, D. (1997). “Switchboard SWBD-DAMSL shallow discourse-function annotation (coders manual, draft 13)”. Technical Report 97-02, University of Colorado, Institute of Cognitive Science.
- Jurafsky, D., Shriberg, E., Fox, B., and Curl, T. (1998). “Lexical, prosodic, and syntactic cues for dialogue acts”. In Proceedings of ACL/COLING-98 Workshop on Discourse Relations and Discourse Markers, p. 114-120.
- Klein, M., and Soria, C. (1998). “Dialogue acts”. In Klein, M., Bernsen, N. O., Davies, S., Dybkjær, L., Garrido, J., Kasch, H., Mengel, A., Pirrelli, V., Poesio, M., Quazza, S., and Soria, C. MATE Supported Coding Schemes. MATE Project LE4-8370, Deliverable D1.1.
- Levinson, S. C., (1983). *Pragmatics*. Cambridge University Press, Cambridge, UK.
- Lisowska, A., Popescu-Belis, A., Armstrong, S. (2004). “User Query Analysis for the Specification and Evaluation of a Dialogue Processing and Retrieval System”. Proceedings of LREC 2004, Lisbon, Portugal.
- Morgan, N., Baron, D. Bhagat, S., Carvey, H., Dhillon, R., Edwards, J. A., Gelbart, D., Janin, A., Krupski, A., Peskin, B., Pfau, T., Shriberg, E., Stolcke, A., and Wooters, C. (2003). “Meetings about meetings: research at ICSI on speech in multiparty conversations”. In Proceedings of ICASSP’03, Hong Kong, China.
- Popescu-Belis, A. (2003). Dialogue act tagsets for meeting understanding: an abstraction based on the DAMSL, Switchboard and ICSI-MR tagsets. Technical report, IM2.MDM-09, v1.1.
- Searle, J. R. (1969). *Speech Acts*. Cambridge University Press, Cambridge, UK.
- Shriberg, E., Dhillon, R., Bhagat, S., Ang, J., Carvey, H. (2004). “The ICSI Meeting Recorder Dialogue Act (MRDA) Corpus”. Proceedings of SIGDIAL’04, Cambridge, MA, USA.
- Stolcke, A., Ries, K., Coccaro, N., Shriberg, E., Bates, R., Jurafsky, D., Taylor, P., Martin, R., Meteer, M., and Van Ess-Dykema, C. (2000). “Dialogue Act Modeling for Automatic Tagging and Recognition of Conversational Speech”, *Computational Linguistics*, 26(3), p. 339–371.
- Traum, D. R. (2000). “20 Questions for Dialogue Act Taxonomies”. *Journal of Semantics*, 17(1), p. 7–30.
- Vanderveken, D. (1990). *Meaning and speech acts*. Cambridge University Press, Cambridge, UK.