

Designing an Open, Multidimensional Dialogue Act Taxonomy

Harry Bunt and Yann Girard
Tilburg University
{bunt|y.girard}@uvt.nl

Abstract

This paper discusses issues in the design of a rich taxonomy of dialogue acts that is hierarchically structured in such a way that a notion of ‘dimension’ is reflected, corresponding to the intuition that communication is a complex activity, with multiple aspects that can be addressed simultaneously. The taxonomy is also ‘open’ in the sense that it is based on clear criteria for including dialogue act types and for how they fit into the taxonomy, allowing easy addition of other act types.

1 Introduction

To describe what is happening in a dialogue from semantic and pragmatic points of view, it has become common to consider dialogues in terms of communicative actions, called ‘communicative acts’ or ‘speech acts’ or ‘dialogue acts’. In recent years the term ‘dialogue act’ has become particularly popular among researchers involved in the design of computer dialogue systems or in dialogue annotation, see e.g. Jurafsky & Martin (2000).

This paper is concerned with the definition of dialogue acts and especially with the

definition of taxonomies of dialogue acts, inspired by the goal to build a public registry of dialogue act specifications, as undertaken by the Task Domain Group on Semantic Content Representation within the International Standards Organisation ISO (ISO/TC 37/SC 4/TDG 3).¹ We outline a number of fundamental and practical issues that have to be addressed in developing a repository of dialogue acts, focusing on the following issues:

- How are dialogue acts defined? How do dialogue acts relate to speech acts, communicative acts, utterances, turns, etc.?
- What uses of dialogue acts do we envisage, that should be supported by a repository of dialogue acts? What requirements on dialogue act specification follow from potential uses of dialogue acts, such as manual or automated annotation?
- What exactly does it mean for a dialogue act annotation system to be ‘multidimensional’ and/or ‘layered’? How are ‘dimensions’ (and ‘layers’) defined, and why?

¹Some of the ideas presented in this paper have been introduced in a presentation at the 4th Joint ISO-SIGSEM Workshop on the Representation of Multimodal Semantic Information, Tilburg, January 10-11, 2005, and appear in an unpublished discussion paper prepared for that workshop - see Bunt (2005).

- What criteria are relevant for identifying a particular class of dialogue acts? In designing a system of dialogue act types, what are the criteria for structuring the system?

2 The dialogue act concept

2.1 Defining dialogue acts

The term ‘dialogue act’ is sometimes used in a rather loose sense, to mean ‘speech act, used in dialogue’. There are also more formal approaches, where dialogue acts are considered as concepts in the description or annotation of dialogue utterance meanings, and have a well-defined formal semantics. For instance, Bunt and Romary (2002) have proposed to view the meaning of an utterance as *the way in which the utterance is meant to change the information state of an interpreting system upon understanding the utterance*.

When analysing the meaning of a dialogue utterance, we can distinguish two fundamental aspects: (1) the semantic (or ‘referential’, ‘propositional’) content: the objects, events, situations, properties, relation, etc. that the utterance is about; and (2) the communicative function or purpose that the utterance has in the communication. Using these two aspects, a formal interpretation can be given to a dialogue act by viewing the combination of a communicative function and a semantic content as an operation that updates the information states of the dialogue participants in a certain way. This approach is known as the *information-state* or *context-change* approach to dialogue acts (see e.g. Traum and Larsson, 2003; Bunt, 2000; Cooper et al. 2003).

The use of update operations on information states (or contexts) does not mean that any logically possible type of update operation corresponds to a dialogue act. The whole idea of dialogue acts is that they are a way to characterize dialogue behaviour; therefore, dialogue acts should have an empirical basis: every dialogue act type should have some

reflection in observable features of communicative behaviour. In other words, for every dialogue act type there are behavioural (linguistic) devices which a speaker can use in order to indicate the communicative function(s) of his contribution. This means that we have two criteria for distinguishing a particular type of dialogue act: (1) it corresponds to a specific context-changing effect; (2) the intended context-changing effect can be indicated by means of certain observable features of communicative behaviour.

2.2 Dialogue act types

There are often alternative possible ways to characterize the type of dialogue act performed by a given utterance. For example, the utterance *What did you say?*, can be characterized either as a feedback act, providing information about the speaker’s understanding of the previous utterance, or as a question, and as such as different from the statement *I didn’t hear what you said*, which may also be characterized as a feedback act.

Characterizations as a question or an inform relate more closely to the surface form of the utterance than the characterization as a FEEDBACK ACT. Characterizing these utterances as feedback acts takes into account what the question and the statement are about. Rather than choosing between these alternative characterizations, it seems more attractive to combine the two and characterize these utterances as FEEDBACK QUESTION and FEEDBACK INFORM, respectively.

It is common to speak of dialogue act *types* (or speech act types) as synonymous with: dialogue (speech) acts with a certain communicative function (illocutionary force); the case just considered shows that this may be inaccurate, for characterizing utterances as feedback acts is saying something about the type of their semantic content, rather than about their communicative function. Also, characterizing an utterance as a feedback

question says something both about semantic content type and communicative function.

Indirect speech acts may also be considered as allowing more than one characterization. An utterance such as *It's rather chilly in here* can be seen as intended to inform the addressee of something, but also as a request - to lit a fire, for instance. On the standard view, an indirect speech act occurs when a speaker uses an utterance to perform an additional speech act to the one that is 'directly' associated with the utterance in view of its appearance, as illustrated by *Do you know what time it is?* (as a request to tell what time it is) or *What time do you think it is?* as a reproach for being late.

To understand an utterance as being used to perform an indirect speech act, the addressee must reason with his understanding of the utterance as 'surface speech act', including its semantic content, and his knowledge of the context in order to construe an indirect interpretation as a speech which is appropriate in the given context.

When dialogue acts are viewed as context-changing operations, however, the notion of an indirect dialogue act comes to stand in a different light. Consider, for example, the direct and indirect questions *What time is it?* and *Do you know what time it is?* In both cases we may assume that the speaker wants to know what times it is, but when using the direct question the speaker makes the assumption that the addressee knows the answer to the question, whereas the indirect question does not carry this assumption - the utterance in that case expresses precisely that the speaker does not know whether the addressee knows the answer. If we follow the traditional analysis of indirect speech acts where the speaker is taken to perform the same speech act as an *extra* act, in addition to what is expressed directly, then we have to say that the indirect question creates in the addressee, among other things, the effect of

the 'direct' question where the speaker wants to know whether the addressee knows what time it is, plus the effects of the indirectly expressed question where the speaker wants to know what time it is. This combination of beliefs would clearly be inconsistent, however. It would therefore be wrong to analyse the indirect question as the direct question plus an additional question. Instead, the indirect question associated with should be analysed as expressing the speaker's wish to obtain the information what time it is, *without* also expressing the expectation that the addressee is able to tell that. This makes the indirect question a (slightly) different type of dialogue act than the direct question.

Similar analyses apply to other indirect dialogue acts, such as indirect requests.

Theories of dialogue acts or communicative acts often emphasize the multifunctionality of dialogue utterances, i.e., the phenomenon that an utterance can have several functions at the same time (see e.g. Allwood, 2000). This is also reflected in some dialogue act annotation schemas, such as DAMSL (Allen and Core, 1997), which allow the assignment of multiple dialogue act tags to an utterance. One of the reasons for the multifunctionality of utterance is that it can have an effect related to various dimensions of the communication process, such as exchanging task-related information, giving feedback, and managing the interaction.

2.3 Uses of dialogue acts

Dialogue acts (DAs) have been used for several different purposes: to support conceptual analysis of natural human dialogue; as building blocks in the interpretation and generation of utterances in a dialogue system; to annotate dialogues, either manually or automatically; or to define the inter-agent communication between software agents; see e.g. FIPA (2002). Each of these applications brings specific constraints and requirements. Here, we

only consider the use of dialogue acts for tagging, and its implication for to the design of a well-structured system of dialogue acts.

When very small sets of tags are used, such as the LINLIN tag set (Ahrenberg, Dahlbäck & Jönsson, 1995) or the HCRC tag set (Carletta et al., 1996; Isard & Carletta, 1995), then there is little need to be concerned with its organization, but larger tag sets, such as those of DAMSL or DIT (see e.g. Keizer, 2003), call for a well-motivated structure to support annotators' work. For the ISO effort to develop a registry of standardized concepts for semantic annotation, it is moreover worth taking into account that the specification an *exhaustive* tag set for all domains and all purposes is hard to imagine. Explicit performatives, for instance, form an open class of dialogue communicative functions. Also, degrees of granularity in dialogue act distinctions are often possible. It therefore seems best to design a structured set of tags, with a clear, well-motivated structure, containing a number of obviously needed instances in the various categories, and with clear principles for how to add tags to the set as may be needed for specific domains or specific purposes. Such a system is what we suggest to call an 'open taxonomy'. Moreover, we propose to structure such a taxonomy according to the intuitive notion of 'dimensions of communication', mentioned above in relation to the multifunctionality of dialogue contributions.

3 Dimensions in dialogue act assignment

3.1 Formal concepts

Dimensions of communication are, intuitively, different aspects of the communication process that can be addressed independently and simultaneously by means of dialogue acts.

As an example of a dimension, consider the turn-taking system. For a dialogue participant

A, the following situations may arise:

1. A has the turn, i.e. he is in a position to make a contribution to the dialogue. The following cases may arise:
 - (a) A uses the turn and makes his contribution. In this case he does not have to perform any turn management action.
 - (b) His turn is contested: dialogue partner B is trying to get the turn. The following situations may occur:
 - i. A wants to keep the turn. The efforts that he makes in order to achieve that, constitute a TURN KEEPING act.
 - ii. A is willing to concede the turn. The act of indicating to B that B may take the turn, constitutes a TURN GIVING act.
2.
 - (a) B has the turn and is using it. If A is happy with that, he does not have to perform any turn management action.
 - (b) B has the turn and is using it. If A wants to get the turn, without waiting until B concedes it, A's efforts to get it constitute a TURN GRABBING act.
 - (c) B is offering A an opportunity to take the turn.
 - i. If A seizes the opportunity and takes the turn, then that constitutes a TURN TAKING act.
 - ii. If A is not willing to accept the turn, his behaviour that indicates that is a TURN REFUSAL act.

This example shows that a dialogue agent may perform one of five possible turn management acts, but never more than one: the alternatives within a dimension are mutually exclusive.

In general, dimensions are independent sets of features such that per dimension only one value may be assigned for an object that is characterized in the multidimensional space. To formalize this notion, we clearly need a formal device for assigning values to the objects to be characterized; in the case of dialogue annotation, that is a formal device of assigning annotation tags to the ‘markables’ to be annotated. We therefore introduce a *dialogue act assignment system* as follows.

Definition 1: A Dialogue act assignment system is a 4-tuple $A = \langle D, f, C, T \rangle$ where D is a set of (simple) dialogue act tags, f is a function assigning tags to utterances (which may be simple elements of D , or complex structures built from D elements), C is a set of constraints on admissible combinations of tags, and T is a set of additional labels that f may assign to utterances – T contains such labels as *inaudible* and *abandoned*.

It may be noted that the DAMSL annotation system speaks of ‘layers’ in annotations as well as of multidimensionality, and seems to use these terms as synonyms. One of these layers/dimensions is called Communicative Status, and contains such tags as *uninterpretable* and *abandoned*, which seems better modelled as part of the annotation system than as a dimension in a set of dialogue act tags. (And perhaps DAMSL’s ‘Other Level’ tags are best treated in this way as well.)

To reflect the multifunctionality of dialogue contributions, the DA assignment function should be allowed to assign sets of tags to utterances, where the elements of the set correspond to different dimensions of communication. To this end, the DA tag set may be organized as a taxonomy, i.e. as partitioned into named subsets such that the assignment function associates at most one tag per dimension with any given utterance. More formally:

Definition 2: A multidimensional dialogue act assignment system is a 4-tuple $A = \langle$

$D, f, C, L \rangle$ where $D = \{D_1, D_2, \dots, D_m\}$ is a dialogue act taxonomy with ‘dimensions’ D_1, D_2, \dots, D_m and where the combination constraints C allow a dialogue utterance to be assigned a tag in each of the dimensions, but never more than one tag per dimension.

We consider this definition as capturing the essence of a multidimensional system. Another aspect is the independence of the assignment of a tag in one dimension from the tags in other dimensions. This is captured by the following definition of independence:

Definition 3: Two dimensions in a multidimensional annotation system are **independent** if any pair of tags from the two dimensions is admissible.

Definition 4: If any two dimensions in a multidimensional dialogue act assignment system are independent, then the system is called **orthogonal**.

Orthogonality is not to be taken as a strictly necessary requirement of a multidimensional system (it does not seem realistically feasible for DA tagging), but it is desirable to be as much orthogonal as possible (and thus to keep the set of constraints C as simple as possible).

It may be noted that we defined a dialogue act taxonomy as simply a partitioned set of tags, thereby excluding the possibility of a taxonomy to have several levels. The reason for this choice is that a set of dimensions is itself not a dimension, according to Definition 2, since it would give rise to multiple tags from that dimension set. Still, it is convenient to have more than one level in a DA taxonomy, for grouping a number of dimensions under a more general name, like ‘interaction management’. To distinguish such a grouping from dimensions proper, we propose to use the term *layer* with this definition: as a set of dimensions or, recursively, a set of layers, thereby making a clear distinction

between layers and dimensions. We will incorporate this notion of layer in Definition 5 below.

3.2 Multidimensional dialogue act tags

We noted above that an attractive way to characterize an utterance may be as a pair like FEEDBACK QUESTION, consisting of the name of a dimension (FEEDBACK) and the name of a communicative function (QUESTION). This suggests that DA tags may be pairs. On the other hand, characterizing an utterance as a TURN KEEPING act does not require a second element, since the turn keeping function is necessarily concerned with the dimension of turn management. A question, by contrast, can be about any type of information and therefore relate to any interaction dimension. We therefore propose to classify communicative functions as being either *general-purpose* or *dimension-specific*. A DA tag is then either a pair, consisting of a general-purpose function and a dimension. or a single dimension-specific function. This leads to the following modified definition of a multidimensional dialogue act assignment system, to which we have also added the notion of layers:

Definition 5: A **layered multidimensional dialogue act assignment system** is a 7-tuple $A = \langle GP, DS, D, f, C, L, T \rangle$ where GP is a set of general-purpose communicative function names, DS is a taxonomy of dimension-specific communicative function names, D is the taxonomy of dimension names that mirrors the DS taxonomy, L is a set of layers (i.e., set of (sets of...) dimensions of D , and where f , C and T are as before, except that f is now a function from utterances to sets of tags (or labels from T), each tag being either an element from DS or a pair $\langle g, d \rangle$ with $g \in GP$ and $d \in D$.

3.3 The DIT taxonomy

We have applied the concepts defined here and redesigned the DA taxonomy of DIT, adding some of the dialogue types distinguished in DAMSL. It should also be noted that some of the DIT categories of communicative functions for feedback and interaction management have been inspired by the work of Allwood et. al. (1994). (For the complete resulting taxonomy see <http://pil294.uvt.nl/dit>). Slightly simplified, the taxonomy of dimension-specific functions in DIT looks as follows:

Task-Oriented Functions

Task/Domain-Specific Functions: Hire, Fire, Ap-
point,...; Acquit, Condemn, Appeal,...

Task Management Functions: ...

Dialogue Control Functions

Feedback Functions

Auto-Feedback Functions: Overall Positive,
Execution Negative, Evaluation Positive,
..., Perception (= Overall) Negative

Feedback Elicitation Functions: Evaluation,
Execution

Allo-Feedback Functions: Allo-Overall
Positive, Allo-Execution Negative, Allo-
Evaluation Positive, ..., Allo-Perception (= Overall) Negative

Interaction Management Functions

Turn Management: Turn accepting, Turn giving,
Turn grabbing, Turn keeping, Turn refusal

Time Management: Stalling, Pausing

Contact Management: Contact check, Contact
indication

Topic Management: Topic shift, Topic shift announcement,...

Own Communication Management: Error signaling,
Retraction, Self-correction

Partner Communication Management: Completion,
Partner correction

Dialogue structuring: Opening, Closing, DA
announcement

Social Obligations Management Functions

Greeting: Init-greeting, React-greeting

Self-introduction: Init-self-introduction,
React-self-introduction

Apology: Apologising, Apology-downplay

Gratitude: Thanking, Thanking-downplay

Valediction: Init-goodbye, React-goodbye

It may be noted that general-purpose communicative functions can also be put into a

(partial) hierarchy, but the hierarchical relation in this case has a different significance from that between dimension-specific ones, namely as an expression of degree of specificity. For example, a confirmation is more specific than an answer, and a check is more specific than a question.

The DIT taxonomy is being used for annotation in the DIAMOND project (see <http://pi1294.uvt.nl/diamond/>), and in the PARADIME project (PARAllel Agent-based Dialogue Management Engine) as part of the Dutch national IMIX project on interactive multimodal information extraction (see http://www.nwo.nl/nwohome.nsf/pages/NWOP_653H9J). Inter-annotator agreement data are not yet available, and are not easy to obtain for multidimensional annotation, but are one of the aims of these activities. Another major aim is the establishment of annotation guidelines, of which there is only a beginning, and annotation tools.

4 Multidimensional dialogue act annotation

Using a layered multidimensional DA assignment system for annotation raises several issues, some of which have been discussed by Larsson (1998), such as the consequences of multidimensional tags for measuring inter-annotator agreement. One obvious suggestion, that follows from the intended orthogonality of the various dimensions, is to consider calculating inter-annotator agreement *per dimension*. But even within a single dimension the issue of inter-annotator agreement is not a simple one in a DA system with hierarchical relations among communicative functions. If one annotator marks an utterance as a YES/NO-QUESTION concerned with domain information, and another as a CHECK, these annotators do not agree completely but cannot be said to disagree completely either. A more dramatic inter-annotator disagreement occurs for instance when one annotator thinks that an

utterance does not have a function in a certain dimension, while another annotator thinks it has.

This brings us to another issue that deserves further study: should it be assumed that every utterance in principle has a function in every dimension, if only implicitly? Every utterance could conceivably be said to have a feedback function, for instance, since it can always be taken to provide some information about the processing of previous utterances. Similarly, if we assume the existence of a topic management function that corresponds to continuing the dialogue without a change of topic, so 'TOPIC CONTINUATION' WOULD BE A DEFAULT VALUE IN THIS DIMENSION, then every utterance could be said to have a topic management function. So it seems that one consistent strategy for multidimensional tagging could be to assume the existence of default values for every dimension (except the domain and task management dimension) and to annotate each utterance with an 11-tuple of functions in the dialogue control dimensions. This is to be contrasted with the alternative of only annotating non-default values, and assuming a variable multiplicity of the tags to be assigned to utterances.

5 Related and future work

Most closely related to the work discussed in this paper is the effort of the Discourse Research Initiative that has resulted in the DAMSL annotation scheme (Dialogue Act Markup in Several Layers; see Allen & Core, 1997). While presented as a layered, multidimensional scheme, the DAMSL scheme is not based on clearly defined notions of dimension and layer.

In the communicative functions that it contains, the DAMSL scheme has much in common with the DIT taxonomy. An important difference is the much more elaborate and fine-grained set of functions for feedback and other aspects of dialogue control functions

that is available in DIT. For a more detailed comparison of the contents of DAMSL and DIT see Keizer (2003). Other surveys and comparative discussions of dialogue act annotation schemes and taxonomies include Larsson (1998); Lendvai (2004) and the MATE survey (Mengel et al., 2000); discussions of issues in the definition and use of dialogue acts include, in particular, Core & Allen (1997); Traum (1999); Stolcke et al. (2000) and Popescu-Belis (2005).

The latest version of the DIT taxonomy has been designed to include most of what is found in DAMSL, organized in a more systematic way. This should make it possible to develop annotation tools that are simpler than those of DAMSL, since the (approximate) orthogonality of the DIT dimensions allows annotators to more freely assign combinations of tags in various dimensions than is the case in DAMSL.

Bibliography

- Ahrenberg, L., N. Dahlbäck & A. Jönsson (1995) Codings Schemes for Studies of Natural Language Dialogue. In: *Working Notes from the AAAI Spring Symposium*, Stanford.
- Allen, J. et al. (1995) The TRAINS project: A case study in building a conversational planning agent. *J. of Experimental and Theoretical Artificial Intelligence* 7, 7–48.
- Allen, J. & M. Core (1997) Draft of DAMSL: Dialogue Act Markup in Several Layers.
- Allwood, J., J. Nivre & E. Ahlsén (1994) Semantics and Spoken Language Manual for Coding Interaction Management. Report from the HSFR project Semantik och talsprak.
- Allwood, J. (2000) An activity-based approach to pragmatics. In H. Bunt & W. Black (eds.) *Abduction, Belief and Context in Dialogue. Studies in Computational Pragmatics*. Amsterdam: Benjamins, 47–80.
- Bunt, H. (2000) Dialogue pragmatics and context specification. In H. Bunt & W. Black (eds.) *Abduction, Belief and Context in Dialogue. Studies in Computational Pragmatics*. Amsterdam: Benjamins, 81–150.
- Bunt, H. (2005) A Framework for Dialogue Act Specification. Paper presented at the 4th Joint ISO-SIGSEM Workshop on the Representation of Multimodal Semantic Information, Tilburg, 10-11 January 2005. Available at <http://let.uvt.nl/research/ti/sigsem/wg>
- Bunt, H. & L. Romary (2002) Towards multimodal content representation. In K. Lee & K.S. Choi (eds.) *Proc. LREC 2002 Workshop on International Standards in Terminology and Linguistic Resources Management*. Paris: ELRA, 54–60.
- Carletta, J., A. Isard, S. Isard, J. Kowtko & G. Doherty-Sneddon (1996) HCRC dialogue structure coding manual. Technical Report HCRC/TR-82.
- Cooper, R., S. Ericsson, S. Larsson & I. Lewin (2003) An information state approach to collaborative negotiation. P. Kuhnlein, H. Rieser & H. Zeevat (eds.) *Perspectives on Dialogue in the new Millennium*. Benjamin, Amsterdam, 271–287.
- Core & J. Allen (1997) Coding dialogues with the DAMSL annotation scheme.
- FIPA (2002) FIPA SL Content Language Specification. Geneva: Foundation for Intelligent Physical Agents, Document No. SC000081.
- Isard, A. & J. Carletta (1995) Transaction and action coding in the Map Task Corpus. Research Paper HCRC/RP-65.
- Jurafsky, D., E. Shriberg & D. Biasca (1997) Switchboard SWBD-DAMSL Shallow-Discourse-Function-Annotation Coders Manual, Draft 13.
- Jurafsky, D. and J.H. Martin (2000) *Speech and Language Processing*. Prentice-Hall.
- Keizer, S. (2003) *Reasoning under Uncertainty in Natural Language Dialogue using Bayesian Networks*. PhD Thesis, Twente University, Enschede.
- Larsson, S. (1998) Coding Schemas for Dialog Moves. Unpublished paper; see <http://www.ling.gu.se/> sl
- Lendvai, P. (2004) Extracting information from spoken input. PhD Thesis, Tilburg University.
- Mengel, A., L. Dybkjaer, L. Garrido, J.M. Heid, V. Pirelli, M. Poesio, S. Quazza, A. Schiffrin & C. Soria (2000) MATE Dialogue Annotation Guidelines. <http://www.ims.uni-stuttgart.de/projekte/mate/mdag/>
- Popescu-Belis, A. (2005) Dialogue Acts: One or More Dimensions? ISSCO Working Paper 62, ISSCO, Geneva.
- Stolcke, A. et al. (2000) Dialogue act modeling for automatic tagging and recognition of conversational speech. *Computational Linguistics* 26:3, 339 – 373.
- Traum, D. (1999) Twenty Questions for Dialogue Act Taxonomies. *Proc. of Amstelogoe '99*.
- Traum, D. & S. Larsson (2003) The Information State Approach to Dialogue Management. In R. Smith & J. van Kuppevelt (eds.) *Current and New Directions in Discourse and Dialogue*. Dordrecht: Kluwer, 325–353.