

A Framework for Dialogue Act Specification

Harry Bunt
Tilburg University
`harry.bunt@uvt.nl`

23 December 2004

1 Introduction

This is a discussion document, intended to prepare the ground for the construction of a registry of dialogue act specifications, as one of the activities in the Task Domain Group on Semantic Content Representation within the International Standards Organisation (ISO/TC 37/SC 4/TDG 3),

This document outlines a number of fundamental and practical issues that have to be addressed in developing a repository of dialogue acts, focusing on the following:

- How do we define dialogue acts? How do dialogue acts relate to speech acts, communicative acts, and other such concepts?
- 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?
- How do dialogue act concepts relate to linguistic and multimodal communicative behaviour? What about language- and culture- dependence?
- 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? How do existing dialogue act taxonomies relate to such criteria?
- What information types are useful in characterising dialogue acts in a repository?
- What has already been done that we can reuse?

2 The Dialogue Act Concept

In order to understand and to describe what is happening in a dialogue, or how a dialogue is structured, or how a contribution to a dialogue can be generated, it has become common to consider dialogues in terms of communicative actions, called “communicative acts” or “speech acts” or “dialogue acts”. The term “dialogue act” has in recent years become most popular among researchers involved in the design of computer dialogue systems or in dialogue annotation, (see e.g. Jurafsky & Martin, 2000, ch. 19).

The term “dialogue act” is sometimes used in a rather loose sense, to mean “speech act, used in dialogue”. But there is also a more formal notion of dialogue act, which attempts to go beyond the traditional informal notions from speech act theory. In a more formal approach, dialogue acts are considered as concepts in the analysis and description of the meaning of dialogue utterances, and have a well-defined formal semantics, which Bunt and Romary (2002) mention as one of the fundamental requirements of any repository of concepts for the representation of semantic information.

A contribution in a dialogue has a *sender*, who makes the contribution, and an *addressee*, who is intended to receive and to process the contribution. (In addition, there may be ‘overhearers’ or ‘witnesses’.) 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:

- A** What is the utterance about? What objects, events, situations, substances,... does it refer to? What propositions involving these elements are considered, using what properties, relations,...?
- B** What are the purposes of considering the information **A**? Is it to inform the addressee of **A**, or to ask him about **A**, or to warn him about **A**, or to explain **A**, or ...?

The information (A) is the *semantic* (propositional or referential) *content* of the utterance; component (B) is often called the *communicative function* (“locutionary force” in speech act theory). The formal approach to dialogue act concepts defines the combination of a semantic content and a communicative function as a dialogue act and gives a formal semantics to these notions by viewing a dialogue act as an operation that changes the state of the addressee in a certain way. The semantic content corresponds to information that obtains a certain place in the state of the addressee, and the communicative function describes the precise place which that information obtains in the information state when the addressee has correctly understood the utterance. This approach is known as the information-state or the context-change approach to dialogue.

Note that this makes dialogue acts analytical aspects of the interpretation of communicative behavior, rather than units of communicative behaviour. To say that a speaker performs a certain type of dialogue act is to say that he produces an utterance (possibly linguistic, or gestural, or multimodal) of which the analysis of its meaning involves an intended type of change of state/context which can be described by the communicative function and the semantic content of that dialogue act.

In sum, we propose the following definition of a dialogue act:

<p>A dialogue act is a unit in the semantic description of communicative behaviour produced by a sender and directed at an addressee, specifying how the behaviour is intended to influence the context through understanding of the behaviour.</p>
--

3 Uses of dialogue acts

Dialogue acts (DAs) have been used for the following purposes:

- To support conceptual analysis of natural human dialogue.
This is historically the first use of dialogue acts or illocutionary acts, going back to language philosophers such as Austin (1962) and Searle (1969), with later linguistically oriented analytical work such as that of Gazdar (1981), Isard (1975) and Allwood (1976).
- As building blocks in the interpretation and generation of utterances in a dialogue system.
This is the AI application of dialogue act concepts, which goes back to studies in the early eighties (Allen and Perrault, 1980; Appelt, 1981), with applications in dialogue system engineering such as Alexandersson et al. (1998), Allen et al. (1995), Black et al. (1991), Bunt et al. (1984), Hinkelman & Spackman (2000), Jekat et al. (1995).
- To annotate (a corpus of) dialogues.
This is a relatively new application of dialogue acts, applied to human-human as well as to (simulated) human-machine dialogue, that has become important since transcriptions of dialogues of various sorts have become available resources, and can provide an empirical basis for dialogue research. In a way, this is the continuation of the traditional analytical study of dialogue with the help of modern technology and resources. See e.g. Core & Allen (1997); Mengel et al. (2000), Stolcke et al. (2000).
- To define the inter-agent communication between software agents; See e.g. FIPA (2002).

The different uses to which dialogue acts can be put, bring different desiderata, requirements and constraints for the specification of a repository of dialogue acts. Since these desiderata, requirements and constraints often apply more to the design of a *system* of dialogue acts rather than to the definition of individual dialogue acts, we will return to this below when we consider dialogue act taxonomies. We will disregard the use of dialogue acts for describing the communication between software agents in the rest of this document.

One important difference between the usual views on dialogue acts and those on speech acts, is that utterances are considered to be *multifunctional*, i.e. to be used to perform multiple dialogue acts, whereas speech act theory assumes that an utterance encodes a single speech act (see e.g. Allwood, 2000; Bunt, 2000).

4 Aspects of Dialogue Act Definitions

4.1 Defining Criteria

The above definition of a dialogue act, as an update operation on information states (or contexts), was noted to be a formal, analytical concept. This does not mean that we should feel free to take just any logically possible type of update operation and call it a dialogue act. In order to be appropriate for the analysis or annotation of human dialogue, dialogue acts should have an empirical basis. Formal state-changing operations of which no traces can be found in empirical dialogue behaviour should thus be ruled out, or formulated more positively: every type of dialogue that is to be distinguished should have some reflection in observable communicative behaviour - which does not mean that the speaker's intended context-changing effect

is *always* reflected in his behaviour (rather, it means that there are devices which a speaker can use in order to signal his intention). On the other hand, different forms of behaviour that correspond to the same state-changing operation should correspond to the same dialogue act. This means that we have two criteria for distinguishing a particular type of dialogue act:

1. a specific context-changing effect is distinguished;
2. the intended context-changing effect can be reflected in observable features of communicative behaviour.

The criteria that for every type of dialogue act that should be linguistic or other features of dialogue behaviour that provide evidence to an addressee about the speaker's communicative intentions, means that different languages and different communities may be expected not to use exactly the same types of dialogue acts.

4.2 Dialogue Act Types

There are often alternative possible ways to characterise the type of dialogue act that has been performed. For example, when a speaker performs the utterance (1a),

- (1) a. *What did you say?*
b. FEEDBACK ACT
c. QUESTION

we can either say as in (1b) that he performs a FEEDBACK ACT, providing information about his understanding of the previous utterances just like in (2a), or that he asks a QUESTION – and as such does something different from the INFORM in (2c):

- (2) a. *I didn't hear what you said.*
b. FEEDBACK ACT
c. INFORM

Both kinds of characterisations are correct and informative. Characterisations as a QUESTION or an INFORM may be said to relate more closely to the surface form of the utterance than the characterisation as a FEEDBACK ACT. The characterizations (1c) and (2c) go beyond the surface form, and take the semantic content of the utterances into account.

Since inherent in the concept of a dialogue act is the distinction of communicative function and semantic content as its two principal components, it seems obvious that an utterance's semantic content should not be used to determine its communicative function. Using speech act terminology, it is customary to speak of “different types of speech act” as synonymous with “speech acts with different illocutionary functions” – the above examples show that this may be dangerous, for characterising utterances like (1) and (2) as being of the type FEEDBACK ACT is *not* saying something about their communicative functions, but rather something about the type of their semantic content. We will therefore use the terms “dialogue act type” and “communicative function” as different ways of saying something about the meaning of a dialogue contribution.

4.3 Indirect speech acts

Indirect speech acts are often also discussed as presenting an ambiguity for their characterisation of an utterance as to what type of dialogue act is performed. As an example, consider:

(3) *It's rather chilly in here*

Here the speaker can be said to inform the addressee of something, but also to perform a REQUEST - to light a fire, for instance.

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 which is appropriate in the given context.

4.4 Information sources for dialogue act assignment

Examples (1), (2) and (3) show that an utterance may be assigned different dialogue act interpretations, depending on what information is taken into account:

1. utterance surface features only
2. utterance surface features and semantic content
3. utterance surface features, semantic content, and context information

The full characterisation of an utterance as a dialogue act obviously calls for characterising both its communicative function and its semantic content, so only using utterance features cannot be sufficient for a full characterisation. However, from a methodological point of view, the semantic content should not go into the determination of communicative function. What about the role of context?

4.5 Use of context in dialogue act assignment

Different from indirect speech acts, another situation where context information is used to assign a dialogue act interpretation to an utterance, is when the contribution of utterance surface features to the computation of the communicative function is dependent on context information.

So-called *declarative questions* are a case in point. These are sentences that look and sound exactly like indicative sentences, but that are nonetheless intended as a kind of questions, such as:

(4) *So you're going home for Christmas.*

Why does this utterance function as a kind of question? The sentence *So you're going home for Christmas.*, which has virtually identical surface features, does not function as a question. The explanation is in the way information is distributed over the participants, plus the fact that both participants mutually know this distribution. If speaker *S* believes that addressee *A* already knows the semantic content of the utterance which, upon first sight, appears to express an INFORM, and if *S* knows that he and *A* mutually know this, then *S* may use this utterance form to *verify* the semantic content. This can be captured by the following interpretation rule:

- (5) *An utterance with the semantic content p and the features of an inform has the communicative function INFORM unless the speaker believes that it is mutually believed that the addressee knows that p , in which case the utterance has the communicative function CHECK.*

This is the essence of the analysis of declarative questions that was given by Beun (1989). Similar analyses can be given for example for rhetorical questions, which are actually INFORMS, and for exam questions. Notice that, given a rule like (5), it is not permitted to interpret the utterance features that might lead to an INFORM interpretation without checking the assumed distribution of information about the semantic content; there simply is no interpretation of these utterance features without taking the information distribution into account. This checking of contextual features is fundamentally different from the general reasoning with contextual knowledge that is needed for dealing with indirect speech acts, and that leads to *additional* interpretations.)

5 The Design of Dialogue Act Taxonomies

The development of a repository of dialogue acts not only raises a number of issues concerning the definition of individual dialogue act types and communicative functions, but also issues in designing a coherent set of types and functions where conceptually similar types and functions are assembled in certain classes – i.e., where the set is organised as a taxonomy.

We mentioned above three different purposes that dialogue acts are used for: conceptual analysis of natural human dialogue, dialogue management (interpretation and generation of utterances) in a dialogue system, and the annotation of dialogue corpus material. Each of these applications brings specific constraints and requirements:

- To support conceptual analysis of dialogues:
 - definitions of DA types and communicative functions (CFs) should be conceptually clear to the analyst
 - the terminology used for DA types and CFs should be self-explanatory
 - conceptually similar DA types should be near neighbours in the taxonomy
 - the assignment of DA types and CFs may make use of all sorts of contextual knowledge and of look-ahead in the dialogue
- To annotate dialogue corpus material
 - to support manual annotation, definitions of DA types and CFs should:
 - * be in such terms that they facilitate human dialogue act recognition
 - * be clear enough to give consistency in annotations and to enable inter-annotator agreement
 - to support automatic annotation, DA types and CFs should be defined in such terms as to facilitate the effective computation of dialogue act tags
 - the assignment of DA types and CFs may make use of look-ahead in the dialogue
 - the (hierarchical) organisation of the taxonomy should support underspecification in annotations by reflecting different degrees of underspecification
 - conceptually similar DA types should be near neighbours in the taxonomy
 - the terminology used for DA types and CFs should follow formal, informal, or de facto standards

- To be used for dialogue management in a dialogue system
 - the assignment of CFs to utterances should occur through an algorithm that uses only information effectively available to or computable by the system
 - look-ahead in the dialogue is not possible
 - the taxonomy should support (reflect) the process of computing CFs, in particular the (hierarchical) organisation of the taxonomy should support underspecification in CF recognition
 - the terminology used for DA types and CFs is hardly relevant

Ideally, a dialogue act taxonomy should support all its potential uses, and should therefore fulfill all the requirements mentioned here and should satisfy all constraints...

Whatever it is one wants to use dialogue acts for, in all cases one wants to assign dialogue acts to utterances. It may therefore be useful to formalise a notion of a *Dialogue Act Assignment System*. We do this in the next section.

5.1 Layers and dimensions in dialogue act analysis

Definition 1: A **Dialogue act assignment system** is a pair $A = \langle T, C \rangle$ where T is a set of dialogue act tags, to be assigned to utterances, and C is a set of constraints on admissible combinations of tags.

A dialogue act tag may be a communicative function, a dialogue act type, or a combination of those. A dialogue act tag set may be a taxonomy, which means that the tag set is partitioned into named subsets:

Definition 2: A **Dialogue act taxonomy** is a set $T = \{C_1, \dots, C_n\}$ of disjoint sets of dialogue act tags.

As we mentioned above, when utterances are interpreted in terms of dialogue acts, they are considered to be potentially multifunctional, i.e. several dialogue acts can be assigned to a given utterance. One reason for this is the fundamental assumption that communication processes have several dimensions, such as exchanging task-related information, giving feedback, and managing the interaction. This means that a dialogue act assignment system should allow a dialogue act assignment *function* to assign sets of tags to utterances, observing the constraints of the system. A dialogue act taxonomy should reflect this idea of multidimensionality. Several existing taxonomies, such as the DAMSL taxonomy and the DIT taxonomy, do reflect this to some extent, but could be redesigned in order to be more satisfactory in this respect.

In terms of dialogue act tagging, a dimension is, informally speaking, a subset D of the total tag set T , characterised by the fact that all dialogue acts corresponding to D are concerned with a particular type of information (such as task information, or feedback information, or turn management information). Moreover, the tags within a dimension should be mutually exclusive, and the assignment of a tag in one dimension should be independent of the assignment of a tag in another dimension. More formally:

Definition 3: A **multidimensional** dialogue act assignment system is a system $A = \langle T, C \rangle$ where $T = D_1, D_2, \dots, D_m$ (with ‘dimensions’ D_i) with any pair D_i ,

D_j being disjoint, and where the combination constraints C allows a dialogue utterance to be assigned a tag in each of the dimensions, i.e., to simultaneously have a function in each of the dimensions. Moreover, the combination constraints C do not allow the assignment of more than one tag per dimension.

Definition 4: Two dimensions in a multidimensional annotation system are **independent** if any pair of tags from the two dimensions is admissible.

Definition 5: If any two dimensions in a multidimensional dialogue act assignment system are independent, then the system is called ‘**orthogonal**’.

These definitions capture some of our intuitions, but they are not quite accurate, since they do for instance not take into account that the CFs in a domain may be partially ordered by degree of specificity (e.g. CONFIRM is more specific than YN-ANSWER, CHECK is more specific than YN-QUESTION), so that an utterance may have multiple CFs within a given dimension. Also, the definition does not take into account that an utterance may have multiple functions within a given dimension but with different semantic content, as in the case of indirect interpretations.

A well-designed dialogue act taxonomy should contain dialogue act classes which correspond to dimensions as defined here, with as much orthogonality as possible, while observing the constraints and requirements listed above.

6 Information types for a dialogue act registry

From the above considerations we may conclude that the characterisation of elements in a dialogue act repository should focus on the definition of communicative functions as outlined in sections 2 and 4, which means that the intended context-changing effect of a CF should be specified, plus the features of communicative behaviour that provide evidence for recognising the CF. These features will to some extent be language-dependent. Moreover, the place of a CF within a dialogue act taxonomy is important for using the CF in dialogue analysis, dialogue annotation, and dialogue management.

In addition to specifying communicative functions, it seems useful to also include data categories for *dialogue act types* in a repository, since these reflect the idea of multidimensionality in a dialogue act assignment system.

7 Reusable Work

Previous work on the definition of dialogue acts and dialogue act taxonomies that we would like to mention as relevant for the ISO TC 37/SC 4/TDG 3 activity includes at least the following:

- Proposed taxonomies of dialogue acts: DAMSL ((Allen & Core, 1997) and several variants (see e.g. Larsson, 1998); the TRAINS system of dialogue acts (Allen et al., 1995); the DIT taxonomy (Bunt 1995; 2000); the GBG-IM schema (Allwood et al., 1994).
- Critical surveys of dialogue act taxonomies: Larsson (1998), Keizer, 2003, MATE (2000).
- Discussions of issues in the definition and use of dialogue acts: Core & Allen (1997), Larsson (1998), Traum (1999).
- A format for the specification of dialogue acts in a registry: FIPA (2002).

Bibliography

- Alexandersson, J., B. Buschbeck, T. Fujinami, M.Kipp, S. Koch, E. Maier, N. Reithinger, B. Schmitz & M. Siegel (1998) Dialogue acts in Verbmobil-2, Second Edition. Verbmobil Report No. 226, Saarbrücken: DFKI.
- Allen, J. et al. (1995) The TRAINS project: A case study in building a conversational planning agent. *Journal of Experimental and Theoretical Artificial Intelligence* 7, 7–48.
- Allen, J. & M. Core (1997) Draft of DAMSL: Dialogue Act Markup in Several Layers.
- Allen, J. & R. Perrault (1978) Analyzing intention in utterances. *Artificial Intelligence* 15, 143–178.
- Allwood, (1976) Linguistic communication as action and cooperation. Gothenburg Monographs in Linguistics 2, Gothenburg University.
- 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.
- Appelt, D. (1981) *Planning natural language utterances to satisfy multiple goals*. PhD Thesis, Stanford University.
- Austin (1962) *How to do things with words*. Oxford: Clarendon Press.
- Beun, R.J. (1989) *The recognition of declarative questions in information dialogues*. PhD Thesis, Tilburg University.
- Black, W.J., J. Allwood, H. Bunt, F. Dols, C. Donzella, G. Ferrari, J. Gallagher, R.Haidan, B.Imlah, K.Jokinen, J.-M.Lancel, J. Nivre, G. Sabah & T. Wachtel (1991) A pragmatics-based language understanding system. In *Information processing systems and software: results and progress of selected projects*. Brussels: Esprit.
- Bunt,H. (1995) Dynamic Interpretation and Dialogue Theory. in M.Taylor, D. Bouwhuis & F. Nél (eds.) *The Structure of Multimodal Dialogue, Vol. 2*. Amsterdam: Benjamins, 139–166.
- 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. & L. Romary (2002) Towards multimodal content representation. *Proc. LREC 2002 Workshop*.
- Bunt, H. & L. Romary (2004) Standardization in Multimodal Content Representation: Some Methodological Issues. *Proc. LREC 2004*, 2219–2222.
- Bunt, H., R.J. Beun, F.Dols, J. van der Linden & G. Schwartzberg (1984) The TENDUM dialogue system and its theoretical basis. *IPO Annual Progress Report 19*, 105–113.
- 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.
- Gazdar, G. (1981) Speech act assignment. In A.Joshi, I. Sag & B. Webber (eds.) *Elements of discourse understanding*. Cambridge University Press.
- Hinkelman, E.& G. Spackman (2000) Abductive speech act recognition, corporate agents, and the COSMA system. In H. Bunt & W. Black(eds.) *Abduction, Belief and Context in Dialogue. Studies in Computational Pragmatics*. Amsterdam: Benjamins, 429–356.
- Jekat, S., a. Klein, E.Maier, I.Maleck, M. Mast, J.J. Quantz (1995) Dialogue Acts in Verbmobil. Verbmobil Report No.65, Saarbrücken: DFKI.
- 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, available from <http://www.ling.gu.se/sl>
- 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/>
- Searle, J.R. (1969) *Speech Acts*. Cambridge University Press.
- 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 Amsteloque '99*.