# Towards an ISO standard for dialogue act annotation

**Harry Bunt*, Jan Alexandersson, Jean Carletta, Jae-Woong Choe, Alex Chengyu Fang, Koiti Hasida, Kiyong Lee, Volha Petukhova, Andrei Popescu-Belis, Laurent Romary, Claudia Soria, and David Traum**

*TiCC: Tilburg Centre for Creative Computing
Department of Communication and Information Sciences
Tilburg University, The Netherlands harry.bunt@uvt.nl

**Abstract**

This paper describes an ISO project developing an international standard for annotating dialogue with semantic information, in particular concerning the communicative functions of the utterances, the kind of content they address, and the dependency relations to what was said and done earlier in the dialogue. The project, registered as ISO 24617-2 Semantic annotation framework, Part 2: Dialogue acts", is currently at DIS stage.

## 1. Introduction

The notion of a dialogue act plays a key role in studies of dialogue, in particular in the interpretation of the communicative behaviour of dialogue participants and in the design of spoken dialogue systems. Informally speaking, dialogue acts are such actions as providing information, requesting the performance of a certain action, apologizing for a misunderstanding, and providing feedback. More formally, dialogue acts can be viewed as corresponding to update operations on the information states of understanding participants in the dialogue; this approach is commonly known as the 'information-state update' or 'context-change approach' to the analysis of dialogue – see e.g. Traum & Larsson (2003); Bunt (2000).

A dialogue act has two main components: a communicative function and a semantic content. The semantic content specifies the objects, relations, actions, events, etc. that the dialogue act is about; the communicative function can be viewed as a specification of the way an addressee uses the semantic content to update his or her information state when he or she understands the corresponding stretch of dialogue.

Dialogue act annotation is the activity of marking up stretches of dialogue with information about the dialogue acts performed, and is usually focused on marking up the communicative functions of dialogue utterances. Utterances in dialogue are very often multifunctional, i.e., they express more than one dialogue act (see e.g. Allwood, 1992; Bunt, 1994; 2000; Popescu-Belis, 2005; 2008; Traum & Hinkelman, 1992; Traum, 2000). The following dialogue fragment illustrates this:

(1) 1. A: Henry, can you take us through these slides?
2. H: O..w..k..ay... just ordering my notes.

WIth the first utterance, A performs two dialogue acts: he assigns the next turn to the participant Henry, and formulates an indirect request. In his response, H (1) accepts the turn; (2) stalls for time; (3) accepts the request, and (4) informs A of the reason why he does not immediately fulfill the request.

The multifunctionality of dialogue utterances, which is indeed a virtually ubiquitous phenomenon (see Bunt 2009; 2010) has led to the development of 'multidimensional' dialogue act annotation schemas, which support the marking up of utterances with more than one functional tag. A particular feature of the approach to dialogue act annotation described in this paper is that a multidimensional schema is proposed which incorporates a well-motivated set of dimensions, which provides a principled account of the multifunctionality of dialogue units and provides optimal support for accurate functional tagging.

## 2. Background

Over the years a number of dialogue act annotation schemas has been developed, such as those of the TRAINS project in the US (Allen et al., 1994), the Map Task studies in the UK (Carletta et al., 1996), and the Verbmobil project in Germany (Alexandersson, 1998). These schemas were all designed for a specific purpose and a specific application domain; they contained overlapping sets of communicative functions and made use of often mutually inconsistent terminology. In the 1990s a group of dialogue researchers came together as the Discourse Research Initiative, and drafted the general-purpose schema for multidimensional dialogue act annotation called DAMSL: Dialogue Act Markup using Several Layers (Allen and Core, 1997; Core et al., 1998). With its focus on multidimensionality and domain-independence, this represented an important step forward in dialogue act annotation. Several variations and extensions of the DAMSL schema have been constructed for specific purposes, such as Switchboard-DAMSL (Jurafsky et al., 1997). The comprehensive DIT$^{++}$ schema (Bunt, 2006; Bunt, 2009) combines the multidimensional DIT schema, developed earlier (Bunt, 1994) with concepts from these various alternative schemas, and provides precise and mutually consistent definitions for its communicative functions and dimensions.

In the EU-funded LIRICS project, a set of core communicative functions from the DIT$^{++}$ schema has been defined using ISO standard 12620 for the specification of data categories (see LIRICS deliverable D4.3, 2006). The data categories have been tested for their usability and cover-

age in the manual annotation of a test suite of dialogues in English, Dutch and Italian (see LIRICS, 2007), as well as in machine learning experiments in automatic tagging (Geertzen, 2009).

Preparatory studies in the LIRICS project and in an expert group of the International Organization for Standards ISO have indicated that the current state of the art makes it feasible to develop an international standard for dialogue act annotation. This paper describes the ISO project that aims to develop such a standard, the project "Semantic annotation framework (SemAF) – Part 2: Dialogue acts"; see ISO DIS 24617-2 (2010). The basic notions and principles of the project are discussed in sections 3 - 5, including a metamodel of dialogue act annotation. Section 5 discusses the notion of a dimension, on which the multidimensional structure of the proposed annotation schema is based. Section 6 describes and motivates the core dimensions and dialogue act types distinguished in this standard. Section 7 outlines the Dialogue Act Markup Language (DiAML) defined for interoperable dialogue act annotation. Section 8 closes with some concluding remarks.

## 3. Basic concepts and metamodel

### 3.1. Functional segmentation

Viewing a dialogue act as a unit in the semantic description of communicative behaviour, the question arises what stretches of such behaviour are considered as corresponding to dialogue acts. Spoken dialogues are traditionally segmented into *turns*, defined as stretches of communicative behaviour produced by one speaker, bounded by periods of inactivity of that speaker.

Turns can be quite lengthy and complex, and are for most purposes too coarse as the stretches of behaviour to assign communicative functions to. Communicative functions can be assigned more accurately to smaller units, which we call *functional segments*, and which we define simply as the functionally relevant minimal stretches of communicative behaviour.

According to the semantic characterization of dialogue acts given above, a dialogue act has at least two participants: (1) an agent whose communicative behaviour is interpreted, usually called the "speaker", or "sender"; and (2) a participant to whom he is speaking and whose information state he wants to influence, called the "addressee" (also called "hearer" or "recipient"). There may of course be more than one addressee. Besides sender and addressee(s), there may be various types of side-participants who witness a dialogue without participating in it. Clark (1996) distinguishes between 'side-participants', 'bystanders', and 'overhearers', depending on the role that they play in the communicative situation.

Of the two most central aspects of a dialogue act, communicative function and semantic content, the former corresponds intuitively to the *type of action* that is performed, and as mentioned above, the term "dialogue act annotation" is commonly used to describe the assignment of communicative function labels to stretches of dialogue. A semantically more complete characterization of a functional segment also provides information about the *type of semantic content*. For example, in the DAMSL annotation schema the Information Level can be indicated using three possible values: Task, Task Management, and Communication. These values indicate whether the semantic content of the dialogue act is concerned with performing the task that underlies the dialogue, or with discussing how to perform the task, or with the communication. This is a coarse 3-way distinction of semantic content types. We propose a more fine-grained classification of content type by distinguishing communication-related information (DAMSL's 'Communication' type) into a number of subtypes, such as information about the processing of something that was said before (feedback information), about the allocation of turns (turn management information), or about the structuring of the dialogue (topic and dialogue structure information). These types of semantic content are also called 'dimensions', and are discussed in more detail below in Section 4.

### 3.2. Dependency relations

Many dialogue acts are semantically dependent on one or more dialogue acts that occurred earlier in the dialogue. This is for example the case for answers, whose meaning crucially depends on which question is being answered, and for the acceptance or rejection of offers, suggestions, invitations, and requests. For these dialogue acts, an important aspect that may be marked up is the relation to the 'antecedent' dialogue act(s) on which their meaning depends.

Feedback-providing and eliciting acts also relate to what happened earlier in the dialogue, but in a slightly different way. Feedback acts are concerned with the *processing* of what was said before - such as its perception or its interpretation. The following examples illustrate the difference.

(2) 1. A: Is this flight also available on Tursday?
2. B: Yes, it's available on Thursday as well.
3. B: On Thursday you said?

B's utterance 2 is used to give an answer to the question expressed in 1; its meaning depends on that of the question, and it responds to the *dialogue act* expressed in utterance 1. Utterance 3, by contrast, checks the understanding of utterance 1 and as such responds to the *utterance*, rather than to the dialogue act that it expresses. The former type of dependence relation we call a 'functional dependence relation'; the latter type a 'feedback dependence relation'.

Note that nonverbal feedback, for instance in the form of nodding, or backchannels like *m-hm*, may relate to what is *currently* said, rather than to what was previously said. This is also the case for speech editing acts like self-corrections (*on Tuesday I mean Thursday*) and completions of what the partner is trying to say.

### 3.3. Metamodel

The metamodel in Figure 1 shows a UML-like representation of the fundamental concepts that are involved in dialogue act annotation. Each dialogue act is related to one functional segment, but each functional segment is related to one or more dialogue acts, reflecting the possible multifunctionality of functional segments (see Bunt, 2009; 2010).

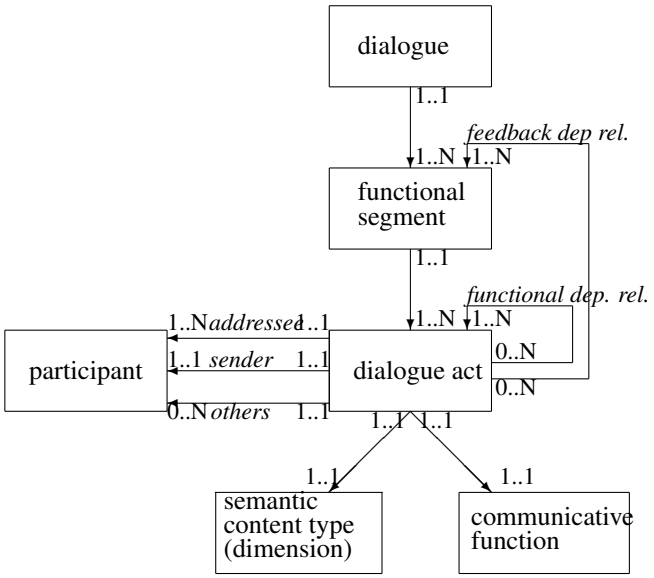


Figure 1: Metamodel for dialogue act annotation.

## 4. Communicative functions

### 4.1. Approaches to communicative function definition

Existing dialogue act annotation schemas use one or both of the following two approaches to the definition of communicative functions: (1) in terms of the intended effects on addressees; (2) in terms of properties of the signals that are used. For example, questions, invitations, confirmations, and promises are nearly always defined in terms of speaker intentions, while repetitions, hesitations, and dialogue openings and closing are defined by their form.

Defining a communicative function by its linguistic form has the advantage of making its recognition easy, but runs into the problem that the same linguistic form can often be used to express different communicative functions. For example, the utterance *Why don't you start?* has the form of a question, and can be used as such, but it can also be used to invite somebody to start. Similarly for 'declarative questions' like *You're going home tomorrow*, which look like statements but are intended as check questions.

Form-based definitions of communicative function also run the risk of being purely descriptive, rather than semantic. For example, when a speaker repeats something that was said before, this behaviour may be characterized as a 'repetition act'; however, that would only say something about the *form* of the behaviour, nothing about its communicative function. A repetition is for instance often used to check correct understanding, but it can also have different functions, as in the following example where it is used as a confirmation in response to a check question:

(3) 1. S: There are evening flights at 7.15 and 8.30.
2. C: And that's on Sunday too.
3. S: And that's on Sunday too.

We take a strictly semantic approach to the definition of communicative functions. But while we do not take linguistic form to be part of the definition of a communicative function, we do insist that for every communicative function there are ways in which a speaker can indicate that his behaviour should be understood as having that particular function, shaping his (linguistic and/or nonverbal) behaviour so as to have certain observable features which are indicative for that function in the context in which the behaviour occurs. This requirement puts all communicative functions on an empirical basis.

A particular case where form and function are not related in a straightforward way is that of indirect speech acts, which is the phenomenon that a speaker uses a linguistic form that is standardly used to express one type of dialogue act, but in context means something else. Questions of the form *Do you know [X]* illustrate this: while an utterance of this form standardly would seem to ask whether the addressee possess the knowledge $[X]$, it is often used to request the addressee to provide the information $[X]$, if he is able to do so. This makes such a question a conditional request.

The DIT$^{++}$ taxonomy of communicative functions (Bunt, 2005; Bunt & Girard, 2005) views indirect requests as having a communicative function which is slightly different from that of the corresponding direct form, because their performance is thought to have slightly different effects on information states. For example, the difference between *Where is Lee's office?* (SetQuestion) and *Do you know where Lee's office is?* (IndirectSetQuestion) would be that in the indirect version the speaker does not express an expectation that the addressee knows the answer to his question, whereas in the direct version he does. The full complexity of the phenomenon of indirect speech acts is beyond the scope of this ISO standard, but an important class of indirect speech acts can be covered as being conditional - see Section 6.4. and Petukhova & Bunt (2010).

### 4.2. Communicative function recognition

Successful communication depends on addressees understanding the communicative functions of the speaker's utterances. These functions are inferred from the utterance surface characteristics in combination with a model of the dialogue context. Such a model includes assumptions about each other's beliefs and goals, as well as knowledge of the dialogue history and of the activity which motivates the dialogue. It is in general not possible to recognize the communicative functions of utterances from their surface form only, since virtually every utterance form can be used with different functions; only in a particular dialogue context can the utterance features be interpreted unambiguously, and sometimes not even then.

The distinction between intention-based and form-based approaches to dialogue acts is relevant to consider in connection with the differences between human and automatic annotation. Human annotators are better at recognizing the intentions behind dialogue utterances, since they are experienced in interpreting intentional behaviour and they have

more comprehensive context models. Since a general dialogue annotation schema should support human annotation, it should contain concepts with a depth and granularity that matches human understanding of the functions of dialogue utterances. In order to support automatic annotation, on the other hand, the schema should also contain concepts that are suitable for a more surface-oriented form of automatic annotation that relies less on deep semantic knowledge. In order to accommodate both requirements, we propose the use of (1) hierarchies of communicative functions, and (2) function qualifiers, which make a base communicative function more specific (see below); where functions deeper down in a hierarchy or carrying a qualifier correspond to more detailed specification of intentions or assumptions on the part of the speaker than functions higher in the hierarchy and functions without qualifiers. An example is the recognition of whether an inform act should be interpreted as a justification, an explanation, or an elaboration of something that was said before.

Some dialogue act annotation schemas employ highly specialized communicative functions such as "accept_date" and "suggest_exclude_location", which mix semantic content type into the communicative function definition. Such functions clearly do not belong in a general-purpose dialogue act annotation schema, and are not part of the set of core communicative functions defined in this ISO standard; they may however be added as optional functions for use in relation to particular domains.

## 5. Dimensions

As mentioned above, the observation that dialogue utterances often have multiple communicative functions has led to the development of multidimensional annotation schemas. Until very recently these schemas, such as DAMSL (Allen & Core, 1997); COCONUT (Di Eugenio et al., 1998), and MRDA (Dhillon et al., 2004); do not base their multidimensionality on a clear notion of dimension; rather they use 'dimension' as a label for a cluster of mutually exclusive tags.

For example, the DAMSL schema distinguishes five clusters of 'forward-looking functions', including the classes of commissive and directive functions, and four clusters of 'backward-looking functions': Agreement, Understanding, Answer, and Information Relation. Core & Allen (1997) refer to these nine subclasses as 'dimensions'.

Popescu-Belis (2005) mentions the following five aspects of utterance function that could be relevant for choosing dimensions in a multidimensional schema: (1) the traditional clustering of illocutionary forces in speech act theory into five classes: Representatives, Commissives, Directives, Expressives and Declarations; (2) turn management; (3) adjacency pairs; (4) topical organization in conversation; (5) politeness functions; and (6) rhetorical roles.

Bunt (2005; 2006) proposes to base the design of a multidimensional tag set on a well-founded notion of *dimension*, inspired by the observation that participation in a dialogue involves a range of communicative activities beyond those strictly related to performing the task that underlies the dialogue. Allwood (2000) notes that in natural conversation, among other things, dialogue participants constantly *"evaluate whether and how they can (and/or wish to) continue, perceive, understand and react to each other's intentions"*. Communication is thus a complex, multi-faceted activity, which is enabled by the *multifunctionality* that dialogue utterances often display (Bunt, 2009; 2010). Dialogue participants share information not only about the task or activity that they pursue with the help of the dialogue, but also about the processing of each other's messages, about the allocation of turns, about contact and attention, about the use of time, and about various other aspects of the interaction. They thus perform various types of communicative activity, such as giving and eliciting feedback, taking turns, stalling for time, establishing contact, and showing attention. Each of these types of activity is concerned with a different kind of information which they can have as semantic content.

We use the term 'dimension' to refer to these various types of semantic content or, equivalently, to the types of communicative activity concerned with these types of information. This leads to considering dimensions such as *feedback*, *turn management*, and *time management*, besides the dimension formed by the activity that motivates the dialogue. In the next section we will discuss the specific set of dimensions proposed for the ISO dialogue annotation standard, and their empirical, theoretical, and practical justification.

## 6. Core Concepts

### 6.1. Dimensions

'Core dimensions' are those dimensions whose relevance does not depend on the domain of application. In order to identify such dimensions, Petukhova & Bunt (2009a) formulate and test a number of criteria that a core dimension should satisfy, which are listed in (4). The first four of these criteria apply to the identification of dimensions more generally; the fourth criterion applies to the choice of a coherent *set* of dimensions, and the final fifth criterion applies specifically to 'core' dimensions.

(4)
1. Eachdimension has a clear empirical basis, corresponding to observed forms of behaviour in dialogue.
2. Each dimension is theoretically justified, corresponding to a well-established class of communicative activities, such as taking turns or giving feedback.
3. Each dimension is recognizable with acceptable precision by human analysts, in particular by annotators, as well as by dialogue understanding and dialogue annotation systems.
4. Each dimension in a multidimensional system can be addressed by dialogue acts independent from addressing other dimensions (the dimensions are *independent* or *orthogonal*).
5. Each *core* dimension is present in many existing dialogue act annotation schemes.

In their study, Petukhova and Bunt (2009a) survey the literature and analyse the contents of 18 existing annotation schemes in order to verify these criteria for a range of proposed dimensions. To examine the recognizability and

orthogonality criteria, they present the results of annotation experiments, empirical data on co-occurrence relations among dialogue acts and dimensions, tests of independent addressability, measures of semantic relatedness, and data on human and machine recognition of dimensions. Their study confirms that the following nine dimensions qualify as core dimensions.

1. Task (or Activity): dialogue acts dealing with the task or activity that motivates the dialogue;

2-3. Auto-and Allo-Feedback; dialogue acts providing or eliciting information about the processing of previous utterances by the current speaker (*auto*) or the current addressee (*allo*);

4. Turn Management: activities for obtaining, keeping, releasing, or assigning the right to speak;

5. Time Management: acts for managing the use of time in the interaction;

6. Discourse Structuring: dialogue acts dealing with topic management, opening and closing (sub-)dialogues, or otherwise structuring the dialogue;

7-8. Own and Partner Communication Management: actions by the speaker for editing his current contribution, or for editing (e.g. completing) the current contribution of another current speaker;

9. Social obligations Management: dialogue acts for dealing with social conventions such as greeting, introducing oneself, apologizing, and thanking, and responses to these acts, such as accepting an apology

### 6.2. Communicative Functions

The choice of communicative functions to populate a multidimensional schema can be based on similar criteria as the choice of core dimensions. The following six criteria have been identified, of which the last one is perhaps best viewed as a desirable property rather than a strict requirement:

(5)
1. *Empirical validity:* for every communicative function there exist linguistic or nonverbal means which can be used by speakers to indicate that their behaviour has that function.
2. *Theoretical validity:* every communicative function has a precise definition which distinguishes it semantically from other functions.
3. The set of communicative functions applicable in a certain dimension provides a good coverage of the phenomena in that dimension.
4. Each communicative function can be recognized with acceptable precision by humans and by machines.
5. Each *core* communicative function occurs in many existing annotation schemas.
6. Any two communicative functions that can be used in a given dimension are either mutually exclusive, i.e. if one of them applies to a given functional segment then the other one does not, or one function is a specialization of the other.

The latter property has the effect that an annotator, when deciding that a functional segment addresses a given dimension $D_i$, can choose from the set of communicative functions available for $D_i$ that unique function which, among the functions that might be applicable, is the most specific one for which there is sufficient evidence. For example, in (6) B's utterance provides information in response to A's Check Question (addressing the task dimension).

(6) A: And that's the first flight tomorrow, right?
B: That's correct.

An annotator will consider assigning an information-providing function to B's utterance such as *Inform, Answer, Agreement, or Confirm,* Of these candidates, *Inform* and *Agreement* are not optimally specific, since they miss the fact that B is responding to a question. *Confirm* is a specialization of *Answer*, and since the use of the word *"correct"* is a sign of confirmation, the most appropriate tag is *Confirm*. Note also that the information-providing functions *Disagreement, Correction* and *Disconfirm*, which do not apply here since there is nothing adversary in what B says, are mutually exclusive with *Agreement* and *Confirm*. The property of mutual exclusivity or specialization (5.6), together with the orthogonality of dimensions (4.4), has the consequence that every functional segment can be interpreted as having at most as many communicative functions as there are dimensions.

### 6.3. Dimension-specific and general-purpose functions

Some communicative functions are specific for a particular dimension; for instance *Turn Accept* and *Turn Release* are specific for turn management; and *Stalling* and *Pausing* are specific for time management. Other functions can be applied in any dimension; for instance a *Check Question* can be used with task-related semantic content in the Task dimension, but can also be used for checking correct understanding (feedback). In general, all types of question, statement, and answer can be used in any dimension, and the same is true for commissive and directive functions, such as *Offer, Suggest,* and *Request*. These communicative functions are therefore called *general-purpose* functions, as opposed to *dimension-specific* functions. The use of *Inform* acts in different dimensions is illustrated in (7)

(7)
1. We will be open this Sunday. [*Task*]
2. I didn't hear what you said. [*Auto-Feedback*]
3. You misunderstood me. [*Allo-Feedback.*]
4. I have something more to add. [*Discourse Structuring*]
5. I need a moment to check this. [*Time Management*]
6. I think Peter should continue. [*Turn management*]
7. I'm very grateful for you help. [*Social Obligations Man.*]

Some examples of dialogue acts with dimension-specific functions are shown in (8).

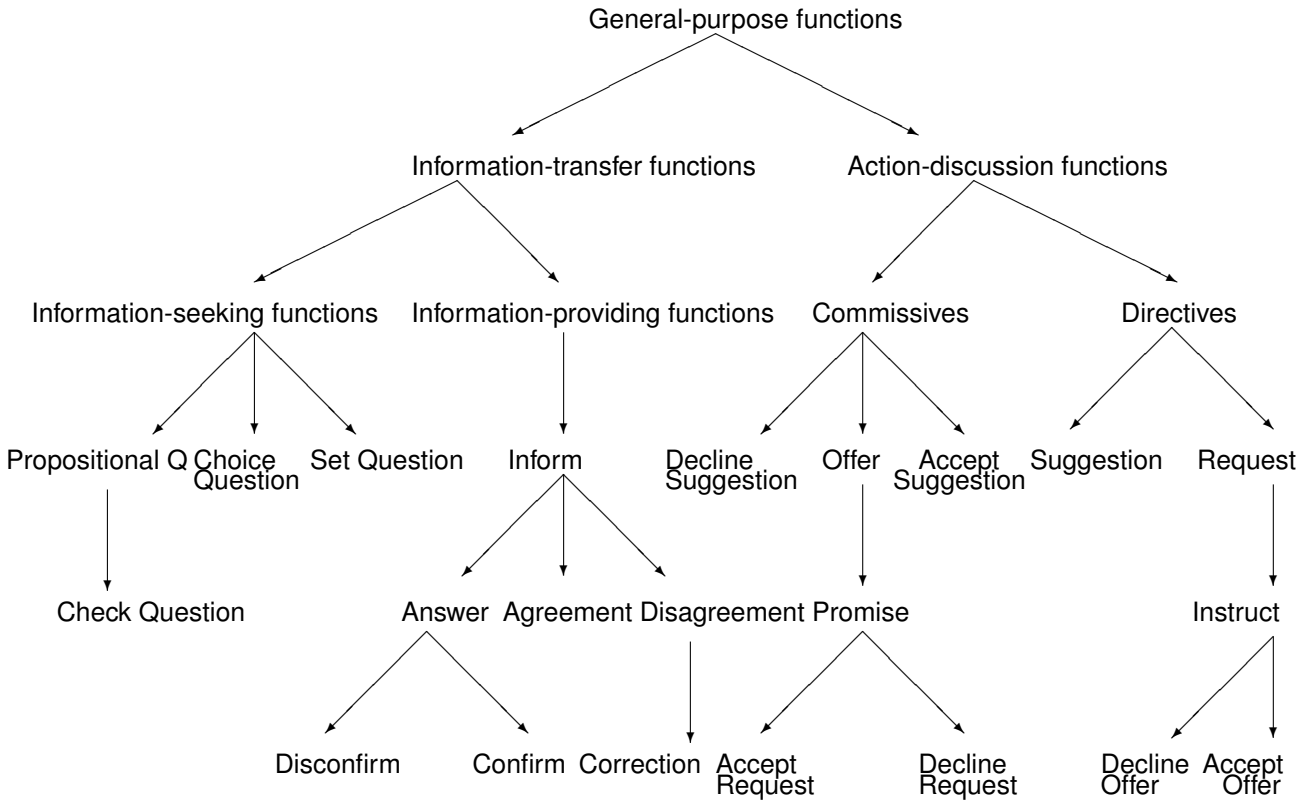


Figure 2: Taxonomy of general-purpose functions

(8) 1. Okay. [*Positive Auto-Feedback*]
2. I beg your pardon? [*Negative Auto-Feedback*]
3. *slow, short nods* [*Positive Auto-Feedback*]
4. *hold gesture with hand* [*Turn Keeping*]
5. Ehm,,, [*Stalling*]
6. I'm sorry. [*Apology*]
7. No problem. [*Accept Apology*]

The specification of communicative functions in the ISO standard should be seen in relation to the data categories containing the definitions of these concepts in the ISOcat Data Category Registry (`http://www.isocat.org`). As long as the ISOCat registry is not yet fully operational, the data categories are temporarily made available at the website `http://let.uvt.nl/research/ti/semantic-annotation/iso/diaml/cf-datcats.pdf`. This includes the following categories of core communicative functions:

- for general-purpose functions:
  - 4 information-seeking functions,
  - 7 information-providing functions,
  - 6 commissive functions,
  - 5 directive functions;
- for dimension-specific functions:
  - 2 auto-feedback functions;
  - 3 allo-feedback functions;
  - 2 time management functions;
  - 6 turn management functions;
  - 3 discourse structuring functions;
  - 2 own communication management functions;
  - 2 partner communication management functions;
  - 10 social obligation management functions.

Due to the different levels of specificity that they are associated with, the set of general-purpose functions has a hierarchical structure which is shown in Figure 2.

### 6.4. Function Qualifiers

A limitation of virtually every dialogue act taxonomy is that it fails to capture certain subtleties in the performance of communicative actions, relating to modality, conditionality, partiality, and accompanying emotions and attitudes. For instance, it is common to distinguish only two possible responses to an offer: accepting it and refusing it. However an offer may be responded to in less clear-cut ways, and can be accepted conditionally (as in (9.3), or partly (as in (9.2), or with a certain emotion (as in (9.4):

(9) 1. A: How about a cup of coffee and a cookie?
2. B: Just coffee please.
3. B: Coffee only if you have it ready please.
4. B: Ah, lovely!

Information-providing dialogue acts may also express the speaker's uncertainty about the information that he provides, as in (10):

(10) A: Do you know who's coming tonight?
B: I have a hunch that Alice won't come.

Studies of these phenomena (see Petukhova & Bunt (2010) indicate that (un)certainty, partiality, and conditionality can be captured in most cases by means of a binary distinction. For representing a speaker's emotional stance with respect to the semantic content of the act, or his attitude towards the addressee, a wide variety of descriptors has been proposed in the literature, ranging from Kendon's basic 6 emotions to classifications of several hundred possible values. In view of this, the ISO proposal contains three binary attributes, for representing conditional, partial, and uncertain variants of dialogue acts, and an attribute with an open class of values for dealing with emotions and attitudes. This is summarized in Table 1.

| qualifier attribute | qualifier values | CF category |
|---|---|---|
| modality | uncertain, certain | info-providing functions |
| mode | angry, happy, surprised, ... | info-providing functions; feedback functions |
| conditionality | conditional, unconditional | action-discussion functions |
| partiality | partial, complete | responsive functions; feedback functions |

Table 1: *Qualifier attributes, values, and function categories*

## 7. DiAML: Dialogue Act Markup Language

The Dialogue Act markup Language (DiAML complies with the ISO Linguistic Annotation Framework (LAF; Ide & Romary, 2003) in making a distinction between *annotations* and *representations*. The term 'annotation' refers to the linguistic information that is added to segments of language data and/or nonverbal communicative behaviour; this notion is independent of the format in which the information is represented. The term 'representation' is used to refer to the format in which an annotation is rendered, independent of its content. According to LAF, annotations rather than representations are the proper level of standardization.

This distinction is reflected in the definition of DiAML, which consists of an *abstract syntax* with a semantics, and a *concrete syntax*. The abstract syntax specifies the elements making up the information in an annotation and how these elements may be combined; these combinations are defined as set-theoretical structures. The concrete syntax specifies a way to represent annotation structures in XML.

Following the requirement that semantic markups should have a well-defined semantics (Bunt & Romary, 2002), the DiAML language has a formal semantics which rests on the interpretation of communicative functions as information-state update schemas. Communicative function qualifiers are interpreted as operations which make these update schemas more specific. The schemas can be instantiated with a given semantic content to define a specific update operation.

Annotations may be attached directly to primary data, such as stretches of speech defined by begin-and end points, but more often they will be attached to structures at other levels of analysis, such as the output of a tokenizer. In (11-12) we give an example of a dialogue act annotation represented in DiAML, using the joint TEI-ISO standard ISO 24610-1 for attaching information to digital texts. In the example, we assume that the dialogue participants are identified in the metadata of the primary data as "p1" and 'p2", and that their utterances are identified as the functional segments "fs1", "fs2", and "fs3". In the example, p1 asks p2 a question in an indirect way, which in terms of the DiAML concepts is described as a conditional request (*"Please tell me... if you know"*). Utterance P2 is segmented into two overlapping functional segments; one in the Auto-Feedback dimension and one in the Task dimension, with value 'answer' qualified as 'uncertain'.

(11)

| | |
|---|---|
| P1: | *Do you know what time the next train to Utrecht leaves?* |
| P2: | *The next train to Utrecht leaves I think at 8:32.* |
| AuFB | The next train to Utrecht [*positiveAutoFeedback*] |
| TA | The next train to Utrecht leaves at 8:32. [*answer, uncertain*] |

(12)
```
<diaml
xmlns:"http://www.iso.org/diaml/">
<dialogueAct xml:id="da1" sender="#p1"
addressee="#p2" target="#fs1"
communicativeFunction="request"
dimension="task"
conditionality="conditional"/>
<dialogueAct xml:id="da2" sender="#p2"
addressee="#p1" target="#fs2"
communicativeFunction="overallPositive"
dimension="autoFeedback"
feedbackDependenceTo="fs1"/>
<dialogueAct xml:id="da3" sender="#p2"
addressee="#p1" target="#fs2"
communicativeFunction="answer"
dimension="task"
functionalDependenceTo="da1"/>
</diaml>
```

## 8. Concluding Remarks

In addition to a structured collection of domain-independent, theoretically and empirically grounded communicative functions with precise definitions, this ISO standard will also provides guidelines for how it may be extended with additional dimensions or functions, as may be required for particular domains or annotation purposes, as well as for how to restrict the schema to a coherent subschema that could be adequate for a particular application.

## 9. References

Allen, J., L. Schubert, G. Ferguson, P. Heeman, C.H. Hwang, T. Kato, M. Light, N. Martin, B. Miller, M. Poesio, D. Traum. and the TRAINS Team (1994) *The TRAINS project: A case study in defining a conversational planning agent*. Technical Report 532, Computer Science Department, University of Rochester.

Allen, J. & M. Core (1997) DAMSL: Dialogue Act Markup in Several Layers (Draft 2.1). Available at http://www.cs.rochester.edu/research/cisd/resources/damsl/RevisedManual/

Alexandersson, J., B. Buschbeck-Wolf, T. Fujinami, M. Kipp, S. Koch, E. Maier, N. Reithinger, B. Schmitz & M. Siegel (1998). *Dialogue acts in VERBMOBIL-2*. Verbmobil Report 226. Saarbrücken: DFKI.

Allwood, J. (1992). *On dialogue cohesion.* Gothenburg Papers in Theoretical Linguistics 65. Gothenburg University, Department of Linguistics.

Allwood, J. (2000) An activity-based approach to pragmatics. In: Harry Bunt and William Black, eds. *Abduction, Belief and Context in Dialogue.* Benjamins, Amsterdam, p. 47–80.

Bunt, H. (1994) Context and Dialogue Control. *Think Quarterly* 3, 19–34.

Bunt, H. (2000a) Dialogue Pragmatics and Context Specification. In: Harry Bunt & William Black (eds.) *Abduction, Belief and Context in Dialogue*. Benjamins, Amsterdam, p. 81–150.

Bunt, H. (2005) A Framework for Dialogue Act Specification. *4th Joint ISO-SIGSEM Workshop on the Representation of Multimodal Semantic Information,* Tilburg, January 2005. http://sigsem.uvt.nl/wg/meeting4.fdas.pdf

Bunt, H. (2006). Dimensions in Dialogue Act Annotation. In *Proceedings Fifth International Conference on Language Resources and Systems (LREC 2006)*, Genova.

Bunt, H. (2009a) The DIT$^{++}$ taxonomy for functional dialogue markup. In *Proceedings AMAAS 2009 Workshop "Towards a Standard Markup Language for Embodied Dialogue Acts'*, Budapest.

Bunt, H. (2009b) Multifunctionality and multidimensional dialogue semantics. In *Proceedings of DiaHolmia Workshop on the Semantics and Pragmatics of Dialogue*, Stockholm.

Bunt, H. (2010). Multifunctionality in Dialogue. Forthcoming in *Computers, Speech, and Language*, Special Issue on Dialogue Modelling.

Bunt, H. and L. Romary (2002). Towards multimodal content representation. In *Proceedings of LREC Workshop on International Standards of Terminology and Language Resources*, Las Palmas, ELRA, Paris, p. 54-60..

Bunt, H. & Y. Girard (2005) Designing an open, multidimensional dialogue act taxonomy. In C. Gardent & B. Gaiffe (eds.) *DIALOR'05, Proceedings 9th Workshop on the Semantics and Pragmatics of Dialogue*, Nancy, 37-44.

Carletta, J., A. Isard, S. Isard, J.Kowtko & G. Doherty-Sneddon (1996) HCRC dialogue structure coding manual. Technical Report HCRC/TR-82.

Clark, H.H. (1996) *Using language.* Cambridge University Press, Cambridge, UK.

Core, M., M. Ishizaki, J. Moore, C. Nakatani, N. Reithinger, D. Traum & S. Tutiya (1998) Report of The Third DRI Workshop, Chiba University, Chiba, Japan.

Core & Allen (1997) Coding Dialogs with the DAMSL Annotation schema. *AAAI Fall Symposium on Communicative Action in Humans and Machines*, Boston, MA.

Dhillon, R., S. Bhagat, H. Carvey and E. Shriberg (2004) Meeting Recorder Project: Dialog Act Labeling Guide. Technical Report TR-04-002, ICSI, Berkeley.

Di Eugenio, B., P.W. Jordan and L.Pylkkänen (1998) The COCONUT project: dialogue annotation manual. ISP Technical Report 98-1, University of Pittsburgh.

Geertzen, J. (2009) *Dialogue act recognition and prediction.* PhD Thesis, Tilburg University.

Ide, N. and L. Romary (2003). International Standard for a Linguistic Annotation Framework. *Natural Language Engineering* 10: 211–225.

ISO DIS 24617-2 (2010) Language resource management – Semantic annotation framework (SemAF), Part 2: Dialogue acts. ISO, Geneva, January 2010.

Jurafsky, D., E. Schriberg, and D. Biasca (1997) Switchboard SWBD-DAMSL Shall-Discourse-Function Annotation Coders Manual.
Available at http://stripe.colorado.edu/$\sim$jurafsky/manual.august.html

LIRICS (2006a) D4.2, Preliminary set of semantic data categories. Available at http://lirics.loria.fr.

LIRICS (2006b) D4.3, Documented compilation of semantic data categories. Available at http://lirics.loria.fr.

LIRICS (2007) D4.4, Multilingual test suites for semantically annotated data, Available at http://lirics.loria.fr.

Petukhova, V. and H. Bunt (2009a) Dimensions of communication. Technical Report TR 2009-002, Tilburg center for Creative Computing, Tilburg University.

Petukhova, V. and H. Bunt (2009b) The independence of dimensions in multidimensional dialogue act annotation. In *Proceedings 2009 NAACL-HLT Conference*, Boulder, Co.

Petukhova, V. and H. Bunt (2010) Introducing communicative function qualifiers. In *Proceedings ICGL 2010, Second 2nd International Conference on Global Interoperability for Language Resources*, Hong Kong, January 2010.

Popescu-Belis, A. (2005) Dialogue Acts: One or More Dimensions? ISSCO Working Paper 62, ISSCO, Geneva. Available at http://www.issco.unige.ch/publications/working-papers/papers/apb-issco-wp62b.pdf

Popescu-Belis, A. (2008) Dimensionality of Dialogue Act Tagsets: An Empirical Analysis of Large Corpora. *Language Resources and Evaluation* 42(1): 99–107.

Traum, D. (1999) Speech Acts for Dialogue Agents. In: M. Wooldridge & A. Rao (eds.) *Foundations of rational agency*, Kluwer, Dordrecht, 169–201.

Traum, D. & E. Hinkelman (1992) Conversation acts in task-oriented spoken dialogue. *Computational Intelligence* 3 (8): 575-599.

Traum, D. & S. Larsson (2003) The Information State Approach to Dialogue Act Management. In J. van Kuppevelt & R. Smith (edr.) *Current and new Directions in Discourse and Dialogue*, Kluwer, Dordrecht.