Interoperable concepts for dialogue act annotation

Harry Bunt and Amanda Schiffrin
Dept. of Communication and Information Sciences, Tilburg University
harry.bunt@uvt.nl, a.schiffrin@uvt.nl

Abstract

This paper describes a range of issues that were encountered when attempting to define a widely usable set of dialogue act concepts. This encompasses theoretical as well as experimental investigations into the applicability of the proposed concepts and recent comparative research for the development of semantic annotation concepts. We aim to show that there are a number of important considerations that are not currently dealt with in most dialogue act annotation schemes, and to give concrete suggestions for how such considerations might be integrated into a set of non-scheme-specific dialogue act concepts.

1 Introduction

Language resources with semantic annotation are increasingly important both for the development of the next generation of language-based applications and services, and for the development of theoretical and computational models of language and information processing that have a strong empirical basis. Semantic annotation is not yet a well-established technique, however, and suffers from a lack of methodological maturity as well as well-established sets of concepts for semantic annotation. For this reason, department TC 37, *Terminology and other Language Resources Management*, of the International Standards Organisation (ISO) has initiated an effort to develop interoperable concepts for semantic annotation (i.e. concepts that should not just be limited to a particular theoretical or computational framework, but useful across a wide range of approaches and applications). This effort is carried out by an expert group[1] that has identified five areas in which initially to develop such concepts in the form of certified entries in an on-line registry; such entries (designed according to ISO standard 12620) are called

[1] Thematic Domain Group TC 37/SC 4/TDG 3 (http://let.uvt.nl/research/ti/iso-tdg3).

data categories. To support this activity further, the eContent project LIRICS[2] was started in 2005 with the same goal (see Bunt & Schiffrin, 2006), and works in concert with the ISO expert group. One of the focal areas of these two related research activities is that of designing a set of interoperable data categories for dialogue acts. We report here on a number of issues of a rather fundamental nature that came up during this work, and which we believe should be addressed in any attempt to design a comprehensive set of such categories. These issues can be divided into two types:

Foundational concepts: General questions concerning the segmentation of dialogue into functional units; methodological and representational problems.

Annotation issues: Specific phenomena that raise problems during multidimensional dialogue act annotation.

In the following sections we deal with groups of each kind of issue in turn, suggesting in all cases a possible approach to resolving them. Our focus is initially on concepts for human annotation, but also with a view to supporting automatic annotation by the introduction of communicative functions with different degrees of specificity, typically corresponding to differences in how close these functions are to surface phenomena (see below, Section 2.2).

2 Foundational concepts

2.1 Dialogue acts, utterances and other units

The term 'dialogue act' is sometimes used rather informally, in the sense of 'speech act used in dialogue'. Accordingly, a dialogue act has a certain function or purpose, corresponding to the 'illocutionary force' and 'propositional content' of speech act theory. In order to be optimally useful for dialogue analysis purposes however, a more precise and self-contained notion of dialogue act is preferable, focusing on its role in assigning meanings to utterances in dialogue. In line with the 'information-state update' or 'context-change' approach to dialogue which has been advocated for the ISO and LIRICS work (see Bunt & Romary, 2002), we define a dialogue act as a semantic unit in the description of utterance meaning, which has two main components: a communicative function and a semantic content. The semantic content comprises the information that the sender of the dialogue act is bringing to the addressee's attention, and the communicative function

[2] Linguistic InfRastructure for Interoperable ResourCes and Systems (http://lirics.loria.fr).

specifies what the addressee should do with the semantic content, i.e., in what way the addressee should use it to update his information state (or context model) upon understanding the utterance. This approach has been taken in recent dialogue studies and projects (see e.g. Traum & Larsson, 2003), including the DIT approach to dialogue act definition (Bunt, 2000a; 2006).

The notion of 'utterance' as discussed here for the purpose of assigning dialogue acts, should not be taken for granted. The process of segmenting dialogue behaviour into markable units ('utterances') to which communicative functions can be assigned is itself a notorious issue in dialogue act annotation. These units can often be identified on syntactic and prosodic grounds, but not always and not reliably so; this is known as the segmentation problem for dialogue act assignment (see e.g. Larsson, 1998). Sometimes, for instance, a functional unit is discontinuous, encompassing a part that has a different communicative function. For instance, in (1) we see a WH-Answer interrupted by a Stalling act:

(1) The next train is at... *let me see...* 3.46.

A consequence of the multifunctionality of communicative behaviour is that the behaviour may have to be segmented in different ways for different dimensions. A functional unit may even consist of parts that belong to different turns; an example is (3), below in Section 3.3, where a WH-Answer is split up into parts, interrupted by turn management acts. From these observations, it would follow that these units can and should be defined from a functional point of view rather than in terms of syntactic and prosodic properties alone.

Figure 1 shows a representation of the fundamental upper-level concepts that support dialogue act annotation. Dialogue acts are tied to utterances, which are parts of (one or more) turns, which are themselves parts of a dialogue. A defining feature of a turn is that it is performed by only one sender, regardless of whether the addressee interjects with feedback or backchannels (this is seen as a separate though concurrent turn). In the case of multiparty dialogue, there may be multiple addressees, and not all the utterances in a turn necessarily have the same addressee(s). Also, there may be additional participants who witness the dialogue without belonging to the intended audience (though both sender and addressee(s) might be aware of their presence, and take that into account); their role may be called that of 'overhearer'.

The metamodel in Figure 1 captures the considerations we have been discussing in this section, and additionally includes the potential for showing functional dependencies between dialogue acts. This would encompass such phenomena as indicating to which question a response is intended to be an answer, etc. Note also, and most importantly, that an utterance may correspond to multiple dialogue acts, due to the multidimensional nature of communication and the multifunctionality of natural language utterances (see Allwood, 2000; Bunt, 2006).

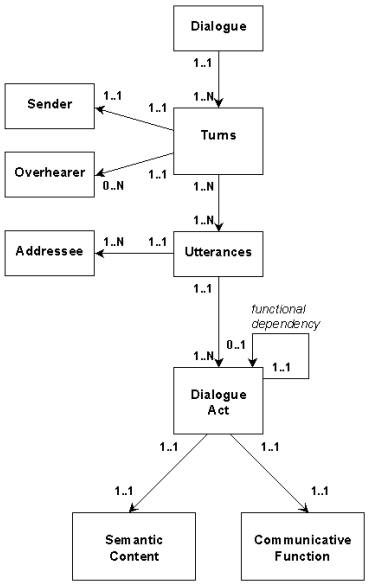

Figure 1: Dialogue act metamodel

2.2 Deep and surface functions

A distinction is sometimes made between 'surface speech acts' and 'deep speech acts'. What is meant by this distinction is that the interpretation of an utterance as an illocutionary act can be made either in a way that stays close to surface forms, or in a way that digs deeply into the speaker's underlying intentions and beliefs, without relating directly to surface forms. A 'surface speech act' would as such be determined strictly by the syntactic and lexical properties of an utterance, and would correspond to what the speaker is 'literally saying', or what he appears or purports to be doing, rather than what he is really trying to achieve (see e.g. Appelt, 1982; Allen & Perrault, 1979; Traum, 1999). This discussion relates closely to that on indirect speech acts and literal meaning – see Section 3.2 below on indirect dialogue acts.

Many approaches to dialogue act annotation do not take a clear position in this matter, defining some communicative functions in terms of beliefs and intentions and others in terms of surface form. This occurs for instance in DAMSL, in which most of the time a 'deep' approach is taken, but not consistently so. The function ASSERT is for example defined by the following two 'deep' conditions:

1. Speaker makes a claim about the world
2. Speaker is trying to change the belief of the addressee

By contrast, the function OPENING, meaning 'a phrase conventionally used to summon the addressee and/or to start the interaction', e.g. *Hi* (Core & Allen, 1997), is defined partly in terms of surface form.

We propose to take a strictly 'deep' view, while insisting that every communicative function must also be empirically justifiable, in the sense that there must be ways in which a speaker can indicate that particular function through the surface form of his behaviour. Successful communication depends on both participants understanding the communicative functions of each others' utterances, which they deduce from the utterance surface forms plus general background knowledge and their context models, as built up through the dialogue. The deep/surface distinction is particularly relevant in connection with the differences between human and automatic annotation. Human annotators are better at understanding and annotating dialogue utterances deeply, because they have richer background knowledge of frames of intentional behaviour and context models. As a dialogue annotation scheme 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, the ideal scheme should also contain concepts that match a more superficial form of annotation that stays relatively close to the surface.

2.3 Intentionality and perspective

Another fundamental issue in the definition of communicative functions is that of perspective. Allwood (1997) has noted that actions in general and communicative acts in particular, can be identified in four ways:

1. from the agent's point of view – depending on his intentions;
2. from an interpreter's point of view – how is the addressee affected;
3. from an observational point of view – what observable characteristics does the action have;
4. from the point of view of its effects – what happened as the immediate result of the action.

(See also Traum, 2000 for further discussion.) Applied to communicative action, the third option corresponds to a pure 'surface' approach to dialogue acts, which we have just argued would not be appropriate. The second and fourth possibilities come down to roughly the same thing, for in a

context-change approach, the immediate effects of a dialogue act constitute the way it affects the addressee's context, and that is precisely what comprises the addressee's interpretation of the speaker's behaviour. Identifying the dialogue acts in a dialogue from the addressee's point of view might be interesting in the case of some misunderstanding between speaker and addressee. However, the very fact that we can speak of a 'misunderstanding' strongly suggests that, if the addressee interprets the speaker's behaviour as expressing a (set of) dialogue act(s) which is different from what the speaker intended, then the addressee is *wrong*; it is the speaker who is always the best arbiter of what he meant, and so the best way of identifying the dialogue acts is from his point of view. We propose for these reasons to take a strictly speaker perspective, looking at what his goal or purpose is and which context changes he wishes to achieve. A consequence of this approach is that, whereas many kinds of action may in general be performed unintentionally, dialogue acts cannot be.

3 Annotation schema issues

3.1 Negated and modified actions

Promises, offers, instructions, suggestions, and other commissive and directive acts are usually defined in terms of commitments and pressures on the speaker or addressee to perform certain actions. This approach is too simple, however, as the examples in (2) illustrate:

(2) a. I promise *not* to raise the issue.
b. I offer to do it *to the best of my abilities*.
c. I suggest that you take this medication *three times every day*.

Apparently, the various commissive and directive acts can be about a certain way or a certain frequency (including the frequency zero) of performing an action. This can be handled fairly elegantly by taking the semantic content of these functions to consist of two parts: (1) an action; (2) predicates describing a manner and/or frequency of performing an action. A bonus of this solution is that we can do without such functions as 'Disallow', 'Dissuade', and other 'negative-polarity' dialogue acts.

3.2 Indirect dialogue acts

Indirect dialogue acts have been till now largely ignored in any dialogue act taxonomy. This may be for a number of reasons. Indirect dialogue

acts are the subject of some controversy linked to the related topic of literal meaning, about which there is little agreement and much heated and polarised argument within the linguistic and psycholinguistic community (Gibbs, 2002; see also Schiffrin, 2005 for an overview). We propose that these so-called indirect dialogue acts should be included, because their performance will have different effects on the update of the dialogue context; they are arguably functionally different from their 'direct' counterparts. For example, the difference between an indirect and direct WH-QUESTION is that in the indirect version the speaker is not expressing any expectation that the addressee knows the answer to the question, whereas in the direct version, there is an assumption that he does (cf. Bunt, 2000a). The beliefs that the speaker is subscribing to in uttering such questions are not the same in each case; therefore this provides a basis for their differentiation regardless of competing theories of meaning ascription. No scheme except for DIT currently makes this distinction.

3.3 Partial acts

There are at least two different types of partial act that can take place in a dialogue. The first type corresponds to what is in essence an interrupted, incomplete or punctuated act. Typically this occurs when the information carried by a dialogue act is too lengthy or complex to be transmitted in one go and is therefore split into more manageable parts. Some general-purpose communicative functions appear to have such partial versions, for example, partial answers:

(3)
1. C: I would like to know what time there's a flight to Frankfurt
2. I: At 10.15 there's KLM, flight 251,...
3. C: Yes,...
4. I: then at 11.20 Lufthansa,...

Utterances 2 and 4 in example (3) are both parts of a WH-ANSWER.

One way to deal with this phenomenon would be to include communicative functions such as WH-ANSWER-PART in the annotation scheme. This would only be acceptable if it were clear for which other types of dialogue act such parts occur, and if these were to be added systematically to the annotation scheme. A simple alternative is to allow an utterance to be discontinuous and spread over several turns. This avoids the introduction of act-parts.

The second type of partial act is partial with reference to the semantic content. These partial acts only occur in responsive acts, as they agree with

or accept a part of a previous proposition, while simultaneously disagreeing or rejecting another part (sometimes even the same proposition when the truth of a stated fact is subject to a certain condition). For example, one can partly agree with the statement "The sky is blue" by saying, "Yes, but not always". This second type of partial dialogue act corresponds to the 'Accept-Part' and 'Reject-Part' labels in the DAMSL annotation scheme (Allen & Core, 1997).

From a logical point of view, one can fully accept a proposition, fully reject it, or partly accept (or reject) it. So full acceptance and full rejection are really two extremes on a scale. It seems methodologically rather strange to have values for the two extremes and only one value for anything in between. Moreover, we have seen in the case of commissive and directive functions that it makes sense to include in the semantic content not just an action, as standard speech act theory has it, but also a manner and frequency of performing an action. Something similar can be done for partial acts by taking the semantic content of a responsive dialogue act to include not only the proposition that is responded to, but optionally also a condition that restricts that proposition, such as a condition on its application (*but never on Sunday*), or a condition that restricts the proposition to a part of it. This makes it possible to have one communicative function ACCEPTANCE, defined essentially by the speaker S having the goal to make the addressee know that S accepts the proposition under consideration with the condition that is part of the semantic content. One extreme case is that the condition is necessarily true (unconditional acceptance); another is that the condition is necessarily false, which amounts to rejection.

3.4 Dimension-specific issues

In this section we discuss three dialogue act dimensions in which we encountered something of special note: (1) Turn Management, where utterances seem to have two functions rather than one; (2) Dialogue Structuring, where the communicative functions seem to be characterised by the structural position of utterances in a dialogue rather than by a speaker's intentions; and (3) Social Obligation Management, where the speaker seems to be driven by social conventions rather than by intentions.

3.4.1 Turn Management

According to the standard theory of turn taking (Sacks et al., 1974), a speaker can decide at the end of each turn construction unit whether he

wants to continue, wants the dialogue partner to continue, or is indifferent about who should continue. The speaker can thus choose between three types of dialogue act: for the first case a TURN KEEPING act; for the second a TURN GIVING act, and for the third a TURN RELEASE act.

Besides making decisions about turn allocation at the end of a turn construction unit, dialogue participants also make decisions at the beginning of such units. A participant who is given (or 'assigned') the turn has to decide whether to accept it; and when the last speaker has released the turn, every participant has to decide whether to take the turn. For the dialogue acts performed in these cases the taxonomy should include TURN ACCEPT and TURN TAKE, respectively. There is one more case to consider: an addressee wants to seize the turn from the current speaker without waiting for the speaker to release the turn or give it to him. This is when interruptions happen; the corresponding turn management act is called TURN GRABBING.

Inherent to the notion of 'dimension' as defined within DIT for instance, is that an utterance should have at most one function per dimension. However, for the Turn Management dimension an utterance in general has *two* functions: one at the beginning and one at the end. We propose to resolve this by dividing the Turn management functions into two subsets, one for utterance-initial and one for utterance-final turn management functions. An utterance then generally has a *pair of turn-management functions*, one from the utterance-initial and one from the utterance-final functions.[3]

3.4.2 Dialogue Structuring

Several dialogue act schemes, including those of DAMSL and DIT, include the functions OPENING and CLOSING. As noted earlier, the DAMSL definitions seem rather unsatisfactory as they are given in terms of surface forms rather than speaker intentions.

OPENING and CLOSING acts tend to be characterised as 'the things that dialogue participants do to open and close a dialogue'. This sounds very much like saying that the first and last utterances of a dialogue always have the function of OPENING and CLOSING respectively, which seems a rather trivial and redundant observation. It is also unclear whether the opening should apply to more than the first utterance, and if so, to how many and which utterances.

[3] When a turn consists of several utterances, and one utterance u_i has a Turn Keeping utterance-final function, then the next utterance u_{i+1} does not have an utterance-initial turn management function.

From a 'deep' point of view, an OPENING act signals the speaker's willingness and readiness to engage in a dialogue with the addressee. Since both participants must agree on this, one would often expect to find a pair of utterances in the beginning of a dialogue, serving to establish that both participants agree to begin the interaction. Initial and response greetings can be seen as playing that role. Similarly for saying goodbye at the end of a dialogue

3.4.3 Social Obligation Management

Dialogue acts for dealing with social obligations (SOM acts) such as greeting, thanking and apologizing, to the extent that they have been included in annotation schemes, usually have a superficial characterisation in terms of their social function. This may be in part because it seems intuitively obvious what we mean by a greeting, an apology, or thanking.

There is evidence, however, that SOM acts have a useful function in structuring the interaction, in addition to their social function. A greeting, for instance, signals the speaker's presence and his awareness of the presence of the addressee(s), and thanking in a dialogue is used to indicate the approaching end of the dialogue (see Bunt, 2000b). Also, it is probably not a coincidence that in many languages the same verbal form is used for greeting in a face-to-face situation as in a telephone call or an email chat; this suggests that the notion of 'presence' that is relevant for greetings is not that of physical (co-)presence, but rather something like *perceptual* or *'communicative presence'*, i.e. the satisfaction of conditions for sending and receiving messages. This analysis, in turn, suggests that greetings have the same or nearly the same function as OPENING, and similarly, valedictions have the same function as CLOSING. Greetings in dialogue indicate the speaker's willingness and readiness to engage in a dialogue with the addressee, and put pressure on the addressee to indicate his willingness and readiness in return. As such, greetings seem to function not only as OPENINGS but also as CONTACT INDICATION or CONTACT CHECK acts.

4 Conclusions

In this paper we have attempted to isolate some of the issues that should be carefully considered before a suitable and comprehensive set of dialogue act descriptors could be proposed for inclusion in an online registry of interoperable data categories. We concentrated on those issues that have hitherto

been unaccountably overlooked or not thoroughly explored in previous research. We believe that many of these issues are brought up due to the conflation of form and function in the design of dialogue act annotation schemes. Among the most pressing in need of attention are:

- The role of intentionality and perspective in dialogue act assignment and recognition.
- The distinction between annotating with deep and surface functions.
- The multiple segmentation of dialogue into functional units.
- Negated and modified actions.
- The treatment of indirect dialogue acts.
- Dealing with partial dialogue acts.

For all of these issues we have provided a brief discussion of the problem and suggested an approach that is consistently functional in nature.

Future work will be to test the applicability of these concepts. This will be implemented by extending the initial annotation experiments reported in Geertzsen & Bunt (2006) and by the development of test suites in several languages (Dutch, English, Spanish and Italian) with the interoperable dialogue act descriptors developed in the LIRICS project. This will not only be the basis of providing a guide for researchers wishing to use dialogue act concepts, but will aid in refining the current data categories, in highlighting where there may be flaws or omissions, and in checking the general viability of the set of data categories for use in NLP implementations and systems. The test suites will thus provide a feedback mechanism for establishing the consistency, reliability and comprehensiveness of the data categories for dialogue act annotation.

5 References

Allen, J. and M. Core (1997) Draft of DAMSL: Dialog Act Markup in Several Layers.

Allen, J. and R. Perrault (1979) Plans, Inferences, and Indirect Speech Acts. In *Proceedings of the 17th Annual Meeting of the ACL.*

Allwood J. (1997) 'A critical look at speech act theory'. In: Osten Dahl (ed.) *Logic, Pragmatics, and Grammar.*

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

Appelt, D. (1982) Planning Natural Language Utterances to Satisfy Multiple Goals. SRI International AI Center Technical Note 259.

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

Bunt, H. (2000b) Non-problems and social obligations in human-computer conversation. In Proceedings 3rd International Workshop on Human-Computer Conversation, Bellagio, Italy.

Bunt, H. (2006). 'Dimensions in Dialogue Act Annotation'. In Proceedings LREC 2006 Workshop.

Bunt, H. and L. Romary (2002). 'Towards multimodal content representation'. In Proceedings LREC 2002 Workshop.

Bunt, H. & A. Schiffrin (2006) LIRICS project deliverable D4.2, Preliminary set of semantic data categories.

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

Geertzen, J. & H. Bunt (2006) 'Measuring annotator agreement in a complex, hierarchical dialogue act scheme'. In *Proc. SIGDIAL 2006*, Sydney.

Gibbs, R. W. (2002) 'A new look at literal meaning in understanding what is said and implicated', *Journal of Pragmatics*, 34: 457–486.

Sacks, H., E. Schegloff & G. Jefferson (1974) A simplest systematics for the organization of turn-taking for conversation. *Language* 53:696–735.

Larsson, S. (1998) 'Coding schemes for dialog moves'. Unpublished paper, available at http://www.ling.gu.se/sl

Schiffrin, A. (2005) *Modelling Speech Acts in Conversational Discourse*, Ph.D. Thesis: University of Leeds.

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

Traum, D. (2000) 20 Questions on Dialogue Act Taxonomies. *Journal of Semantics* 17(1): 7–30.

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.