

20 Questions on Dialogue Act Taxonomies

David R. Traum

UMIACS

University of Maryland

Abstract

There is currently a broad interest in dialogue acts and dialogue act taxonomies, and new uses, taxonomies, and standardization efforts continue to be proposed. This paper presents a discussion of issues that are important to be addressed, if taxonomies are meant to be shared and understood the same way by proposers and others. The discussion is framed in terms of 20 questions, the answers to which will help make the meanings of taxonomy elements more clear to disjoint communities of users.

Introduction

There is currently a very wide range of theories and taxonomies of dialogue acts¹ available for a researcher to choose from. Moreover, specific deficits in any theory lead researchers to continue to develop new taxonomies to suit their particular purposes. To some degree, this is to be expected; dialogue act taxonomies can be seen as a kind of language for describing communicative events, and there is certainly no deficit in the continued creation and development of new formal languages (e.g., programming languages like Java) or (at a slower pace) natural languages. On the other hand, in both natural and artificial languages, the use of similar signs for different concepts can cause confusion and misunderstanding, often with serious undesirable consequences (e.g., in programming languages, the use of = as an assignment rather than equality operator in a boolean context; or the firing of an American city official for using the word *niggardly* (of independent Scandinavian origin) because it sounded too similar to an offensive racial epithet euphemistically referred to as “the N word”²). Similar confusions often occur when one researcher tries to interpret the dialogue act taxonomy of another. For example, when encountering an act labeled as *inform*: which subset of the constraints in (1) can said to be claimed (or perhaps none of them, depending on some other formulation entirely). This kind of confusion has led some (e.g., (External Interfaces Working Group, 1993; Discourse Resource Initiative, 1997; FIPA, 1997)) to propose standard theories that could be well-defined and understood and used across groups while others (e.g. (Allwood, 1977; Cohen and Levesque, 1990)) prefer to treat dialogue act (i.e., illocutionary force) identification as of only secondary importance, as a derived concept within a more general theory of rational interaction, using other primitives.

- (1) a. declarative mood was used
- b. propositional information was expressed
- c. new information was expressed
- d. the addressee came to believe what was expressed
- e. what was expressed is actually believed by the speaker
- f. what was expressed is actually true

It is hard to dispute that dialogue acts are a useful concept, given the wide variety of uses to which they are put. Some of these uses include³: representations of the pragmatic meaning of utterances in dialogue theories (Vanderveken, 1991; Bunt, 1996; Poesio and Traum, 1997; Poesio and Traum, 1998), building blocks for grammars of dialogue (Winograd and Flores, 1986; Bilange, 1991), labels for corpus annotation (Carletta et al., 1997; Alexandersson et al., 1997), agent communication languages (External Interfaces Working Group, 1993; Sidner, 1994; FIPA, 1997; Singh, 1998), object of analysis in dialogue systems (Allen et al., 1996; Bretier and Sadek, 1996), and element of a logical theory of rational interaction (Sadek, 1991). Despite this popularity of the concept, there are still a number of issues that present significant challenges for creating a taxonomy of dialogue acts that can be understood and used by other groups. Here, I will briefly raise some of the issues that have often caused confusion when interpreting one taxonomy of dialogue acts within the viewpoint of another. These issues must be addressed in order to have a clearer idea of what one means by saying a dialogue act occurred, whether the dialogue act taxonomy is meant for labeling a naturally occurring corpus, as part of a formal theory of action, or as a system-internal representation of the dialogue. Although there are many such issues and variations on the ones listed below, I focus here on 20, in the form of questions, in homage to the “dialogue game” named in the title. These questions are grouped, for convenience into sections of related questions.

Defining Dialogue Acts

1: Which is most important: fit to intuitions or formal rigor?

This question has implications beyond just dialogue act definitions, but is applicable for any attempt to provide a formal theory of commonsense notions. Very often it is difficult to precisely formulate complex intuitions using available formal techniques. The question then arises as to which goal to sacrifice for the time being: should one formalize a simpler notion that does not have the same properties of the intuitive concept (e.g., for belief, an undesirable property of logical omniscience, in a normal modal logic), or sacrifice some desirable formal properties, such as a model-theoretic semantics, necessary and sufficient conditions for categories, or soundness and completeness of an inference system. The answer will depend on the purposes to which the concept is to be put: if the primary goal is to discover and prove properties of the system, formal

properties are not easily sacrificed. On the other hand, if the goals are more empirically motivated, a formal system with undesirable properties may not be close enough to be useful, while a concept without some of these properties may suffice for the task at hand (corpus labelling or use in a computer program). There should also be a place for intermediate points, that make some sacrifices at each side, while striving for maximum utility in a given purpose.

In particular, with respect to dialogue acts, it can be relatively easy to state precise definitional conditions of occurrence within a formal logic of action, however a problem may arise when these conditions diverge from a more intuitive (and intuitively useful) notion of action that empirical analysts and dialogue system designers would actually like to use.

2: Is the definition of a dialogue act an issue of Lexical Semantics or Ontology of Action?

There are different tasks one might be attempting when defining the meaning of a dialogue act. Is it to provide an account of when someone might be justified in describing an occurrence using a sentence headed with a particular verb (e.g., *inform*, *request*), or to provide a technical vocabulary to compactly describe various types of occurrences in convenient ways for analyzing other aspects of interaction. As (Allwood, 1977) warns, these endeavors should be clearly separated, even if one might want to use similar categories to describe each (as is done in (Allwood, 1980)), or maintain a position of identity of semantic and conceptual structure (Jackendoff, 1983). Intuitions, or annotation by naive coders without instructions to the contrary may tend to focus on the former enterprise, which may have undesirable consequences for the way in which the taxonomy is to be used. The key question is how much weight, if any, should be given to linguistic intuitions about when it is true or appropriate to use a particular sentence to describe an occurrence. For lexical semantics, this is the paramount question (barring issues of polysemy), while it might not really be a factor for an ontology which might diverge from language classifications for independently motivated reasons.

3: Under what conditions may an action be said to have occurred?

There are a number of different criteria that are being used to decide whether or not an action occurs in a given situation. (Allwood, 1980) uses four criteria, shown in (2), each of which can be a sufficient condition for ascribing that an action has occurred, while none is necessarily present.

- (2) a. intention of performer
- b. form of the behavior (e.g., linguistic form)
- c. achieved result
- d. context in which the behavior occurs

While it is certainly coherent to define actions in terms of meeting minimal conditions along any of these dimensions, it is less clear that this is the most useful way of capturing the generalities over acts that consumers of a dialogue act taxonomy would like to express. E.g., one may be interested strictly in the result, intention, or context, or perhaps between the relationship between

form and result. In the most central case, all four kinds of conditions will hold, however one must know what to do when only some but not others hold. One should especially take care to avoid defining dialogue acts according to, say, a certain set of results holding, and then identify instances of these acts occurring strictly by one of the other criteria, leading to an incorrect claim of the results holding. Using different criteria (e.g., just results vs. intention, or vs. any of the four) can also lead to misunderstandings between theorists (or coders) as to whether a particular act has been performed, and whether the performance of an act implies a particular result holding.

As an example, consider a characterization of an *inform* act, given in (3).

- (3) a. intention of performer: that receiver believes proposition **p**.
- b. form of the behavior: speaker utters a declarative sentence with propositional content **p**.
- c. achieved result: speaker and hearer mutually believe **p**.
- d. context in which the behavior occurs: Speaker and hearer in contact, speaker believes **p**, hearer does not believe **p**.

One could, of course, quibble with any of these characterizations in terms of being too strong or too weak to capture the meaning of “inform”, or perhaps decide that they are more appropriate for some other act (e.g., *statement*, *assertion*). For example, one might produce an utterance of the same form, when not all of the conditions hold, or in which the speaker has a different intention.

Which kinds of conditions and whether only some or all of them might be necessary will also depend on the task being attempted. E.g., lexical semantics or action ontology, as in the previous question. Also, whether this ascription is made from the point of view of an online dialogue participant (such as a dialogue system) or an external observer, e.g., in labelling a corpus (see also question 6).

4: What is the role of speaker intention?

Intention is usually given a somewhat privileged position with respect to constituting dialogue acts (or action in general), e.g., the first criterion in (2). Some would define dialogue acts on the basis of the intention behind them, while others would equate illocutionary acts with recognition of this intention (based on the notion of meaning in (Grice, 1957)). A problem is that definitively interpreting the intention of the speaker requires mind-reading on the part of the hearer. Another problem is that some dialogue acts (like other acts) can be performed unintentionally or with an only ex post facto commitment. Finally, as with other acts, one may perform them with various goals in mind – it may be unnecessary to discover the actual intention in order to recognize an actor its effects in context. For example, a declarative utterance might be performed with the intention to cause the hearer to adopt a belief in the stated proposition, as in (3a). However, the same utterance might very well be performed if the speaker intends instead to cause the hearer

to believe that the speaker believes *p*. Or that the speaker wants the hearer to believe *p*. Or the conjunction of some set of these (or other similar conditions).

For these reasons, some prefer to keep distinct (though related) the issues of intention recognition and dialogue act attribution.

5: What is the role of addressee uptake?

Regardless of speaker intention, many dialogue acts require for even the most limited notion of success some changes to the addressee based on understanding of the utterance in a particular way. Noticing whether the hearer has actually understood in a particular way can often require just as much mind-reading on the part of the speaker as intention recognition requires on the part of the hearer. Later utterances in a dialogue often provide more clues, and thus some, e.g., (Clark and Schaefer, 1989; Traum and Hinkelman, 1992) require a *grounding* process (in the later case by performing other kinds of dialogue acts) before considering some dialogue acts, such as *inform*, *request* to have been successfully performed. This involves the giving of positive and negative feedback (Allwood et al., 1992) about how utterances were perceived and understood.

A negotiation of meaning can also occur (McRoy and Hirst, 1995), severing completely the link between the dialogue effects and original speaker intentions or addressee uptake.

6: What point of view should be taken regarding performance of acts?

There are several points of view which may be taken when regarding the performance of dialogue acts. Relating to the previous two questions are the speaker's and hearer's point of view, respectively. Also, there is a *negotiated* collaborative point of view of the speaker-hearer team, which might differ from the private views of each of the participants. There is also a normative-conventional point of view, which can make reference to social institutions beyond just the speaker hearer pair. There is also the issue of time with respect to coding or ascription of acts: is it an on-line decision just at the time of performance, or is one allowed to view subsequent utterances/action, as well, before deciding what happened?

Point of view is relatively straightforward from a dialogue system internal perspective (though a system might still need to reason about the interlocutor's point of view And subsequent time points in diagnosing misunderstanding (McRoy and Hirst, 1995) or constructing a negotiated view), however it is far from clear what point of view should be taken by coders (and how they should estimate speaker or hearer point of view without mind-reading). Likewise, in defining the acts or giving them a logical semantics, point of view may be necessary to take into account.

As an example, consider the case of a feedback reply of a word or phrase following a declarative utterance by the other speaker. There are several different grounding functions that could be performed by this second utterance, such as *acknowledgement*, *repair*, *request for repair*, or *request for confirmation*. The latter two could perhaps be distinguished from the former by prosody: questioning intonation indicating lack of certainty and desire for further feedback, while declarative intonation indicating one of the former functions. One could distinguish acknowledgement from repair by deciding whether the second utterance repeats (or paraphrases) the former (acknowledgement) or replaces some part of it (repair). However, this decision re-

quires a point of view for who believes it to be the same. Especially for current technology speech recognition systems, there is a significant likelihood that a system may repeat what it understood, which differs from what was actually said. It is also possible (though perhaps less likely) that a system intends to correct but ends up repeating what was really said. The same issues come up (though with less frequency) in human-human conversation.

Dialogue Act components

7: How are actions used in a logic?

In formal theories, actions are usually seen as transitions from states to states (or worlds to worlds), while dialogue acts are seen as special cases of actions (though see question 11). AI theories of action generally associate several sets with actions: a set of effects (constraints on the resulting state), a set of pre-conditions (constraints on the initial state), and decompositions (subactions that, performed together constitute the action).⁴

In terms of the categories given in (2), the effects corresponds to the achieved result, aspects of context and intention may be related to the pre-conditions, and the form of the behavior amounts to the decompositions. The AI theories of action generally include requirements on each of these aspects, so that the axioms in (4) hold (where X is an action type, Pre and $Effects$ are the preconditions and effects of this action type, and $prev\ now$ and $next$ are “consecutive” time points).⁵ (4a) involves reasoning from felicitous performance to effects, (4b) involves reasoning from performance to preconditions having held, and (4c) involves reasoning from performance of subactions to main action. In addition, something like the schema in (5) is used (though in only an abductive or circumscriptive sense, rather than a sound axiom), for reasoning from subaction to intention ascription (plan recognition). Use of these kinds of axioms can also help determine inconsistency of a (default) interpretation, which may then be a cue of an indirect speech act, or a misunderstanding.

- (4) a. $Pre(X, now) \wedge Try(X, now) \rightarrow Effects(X, next)$
 b. $Done(X, now) \rightarrow Pre(X, prev)$
 c. $Pre(Y, prev) \wedge decomp(Y, \{X_1, \dots, X_n, \}) \wedge \forall X_i : Done(X_i, now) \rightarrow Done(Y, now)$

- (5) $Do(A, X) \wedge decomp(Y, \{\dots, X, \dots\}) \rightarrow Intend(A, Y)$

8: What kinds of information are the conditions and effects about?

Given the general framework for actions in the discussion of the previous question, there’s still a large question of what aspects of the situation are relevant to defining conditions for dialogue act performance, and what kinds of things are (directly) affected. Some logical models might allow the truth value of any representable proposition to be a possible condition or effect. This must,

of course, be filtered through the lens of “point of view” (see question 6). Generally there are three more special sorts of information used for conditions and effects of dialogue acts.

First, there is a notion of dialogue state, as encoded as state in a dialogue grammar (Wino-grad and Flores, 1986; Traum and Allen, 1992), or other structural representation of context (Ginzburg, 1998). For example, certain acts may have as their pre-condition that the dialogue be in a particular state in the transition network, or, e.g., that an *answer* is possible only if there is a question under discussion. The effects can be a transition to a new state in the network, or other effects to the dialogue structure (see also question 12).

The second kind of information, the most popular in the planning approach, is in terms of mental states (e.g., belief, intention) of the speaker and addressee(s) (Cohen and Perrault, 1979; Allen and Perrault, 1980). For instance, pre-conditions of an *inform* act may include the latter two conditions in (3d). Effects will include newly adopted beliefs and intentions.

A third alternative is in terms of the social obligations and commitments undertaken by the dialogue participants (Poesio and Traum, 1998; Traum, 1999; Singh, 1998). Example effects include commitments to stated propositions, and commitments to do promised actions. Pre-conditions of this sort are more rare, though could include things like *excuses*, which presuppose a sort of obligation to act (which has not been or will not be performed).

Most approaches will actually combine two or three of these kinds of conditions and effects. There may also be other types of effects, not easily classifiable into these categories.

9: What kind of conditions are most appropriate ?

The notion of *pre-condition* is often criticized as meaning too many different things in relation to planning and reasoning about action (e.g., (Pollack, 1990)). First of all, there is the general issue of enabling conditions vs. applicability constraints – the former being those that can be planned to achieve, while the latter describe conditions in which this kind of action should be considered. There is also the issue of whether these conditions are necessary or sufficient for (successful) performance of the action.

Many convenient dialogue acts actually have few if any actual pre-conditions, in the sense that the action can not occur if the conditions are not met. Conditions are often formed in terms of either normal conditions or in terms of what is required for felicitous performance of the action (Searle, 1969). Formulating conditions in this way does give greater flexibility, however at the expense of having to determine whether the conditions are felicitous (in addition to determining whether the action has been performed) and having to also describe *non-felicitous* performance.

The kinds of conditions to represent in a theory will also depend on the type of cognitive tasks to be performed using the acts: dialogue act planning and performance or recognition. For the former (e.g., using axiom (4a)), one might care more about sufficient rather than necessary conditions. However, for the latter (using e.g., axiom (4b)), one might be more interested in necessary conditions (to use this as an axiom rather than default rule).

10: How should an unsuccessful act be distinguished from a failed attempt to perform an act?

This question is related to the difference between *success* and *satisfaction* of a speech act (Vanderveken, 1990). The former has to do with whether the act was actually fully performed, the latter with whether the propositional content is (or becomes) true. If one uses a social commitment approach, then one may say the act has been performed if the commitments are established, and (fully) successful if its intended perlocutionary effects (Sadek, 1991) or evocative intentions (Allwood, 1995) are achieved.

As an example, consider a request by A to B, for B to do some action x, schematically: Request(A, B, Do(B, x)). The question becomes which of the conditions in (6) do we associate with an attempt vs success vs. satisfaction? Condition (6f) seems sufficient for an attempt, while (6a) is necessary for full satisfaction. Success criteria are more difficult to agree upon, however (see also question (8)). According to the mental states approach, successful performance of the request might be (6b), or (6e), with an additional assumption of cooperativity to lead to (6b) and then (6a) (Cohen and Perrault, 1979). The social commitments approach would favor (6c) (Allwood, 1994), or (6d) (Traum, 1994), with (6c) coming only on acceptance of the request.

- (6) a. Do(B, x)
- b. Intend(B, Do(B, x))
- c. Obligated(B, Do(B, x))
- d. Obligated(B, Address(B, Act1))
- e. Believe(B, Want(A, Do(B, x)))
- f. Try(A, Request(A, B, Do(B, x)))

Another issue is what kinds of actions are involved in leading to success of the action (and the associated effects)? Is a single utterance (in the appropriate circumstances) enough, or is a grounding process (Clark and Schaefer, 1989; Traum, 1994) needed

Relationships and Complex Acts

11: What is the relationship between dialogue acts and other (e.g., physical) acts?

One of the main intuitions behind speech act theory (doing things with words (Austin, 1962)) was to connect speech acts with other actions. However different theories may maintain a crisp or more blurred distinction between dialogue acts and non-communicative acts. Some want a clear distinction, while others would want to use the same logic of action account to account for both. (Litman and Allen, 1987) distinguished dialogue acts as being *meta-acts*, defining discourse plans as having other plans (domain or discourse) as parameters. (Lambert and Carberry, 1991) also distinguish discourse, domain and problem solving plans and actions.

Depending on the answer to question 8, some may want to describe dialogue acts as having a different sort of effect on the dialogue context, mental states, or social context than can be achieved with other kinds of action.

12: What is the relationship between dialogue acts and dialogue structure?

There are several options. Some conceive dialogue structure as being wholly dependent on the structure of dialogue acts performed (e.g., grammar-based approaches like (Sinclair and Coulthard, 1975)) Others use a different sort of structure, not directly composed of dialogue act performance to represent things like accessibility, topic and focus, and global coherence, which can be sensitive to other aspects of the utterances, or be primarily constructed from the activity that the participants are engaged in (Allwood, 1995; Grosz and Sidner, 1986). In this latter case, it remains to be explicated what effect (if any) performance of different kinds of dialogue acts have on this dialogue structure. Dialogue structure is also often used as one of the aspects of context for dialogue act performance, which can serve as the source of pre-conditions and input for action recognition.

13: Are there multi-agent dialogue acts?

As mentioned related to question 5, some see the performance of most illocutionary acts as the collective performance of multiple agents, in virtue of the grounding process. Other candidates for multi-agent action include notions of higher-level activity such as *games* (Severinson Eklundh, 1983) or *exchanges* (Sinclair and Coulthard, 1975), or collaborative completions where one speaker finishes another's sentence. There are several difficulties with these kinds of acts, however. First is related to reliable tagging and computing proper inclusion/exclusion of relevant parts of the collaborative action. Finding the right "units" at which to apply the tags can be a difficult process (see, e.g., discussions in (Discourse Resource Initiative, 1997; Nakatani and Traum, 1999)). This difficulty is compounded when there are multiple acts with different boundaries (e.g., the single-agent act and multi-agent component performed by a speaker within an utterance).

Another issue is the kind of logic that will allow this kind of action will need to be more complex than that required for single agent action.

14: Can dialogue acts be "composed" of more primitive acts

If a dialogue act taxonomy has multiple strata of acts, then the question becomes whether these strata are conceived of as *levels* or *ranks*, according to the terminology of (Halliday, 1961), that is, whether there could be some grammar or recipe for performance of an act of one stratum using acts of a lower stratum, in the way that sentences can be composed of words and phrases (rank), or whether these are different kinds of phenomena, like the distinction between phonology and syntax (level). For example, the 4 tiered system in (Sinclair and Coulthard, 1975) is conceived of as ranks within a general "discourse" level, and e.g., the *check game* in the Maptask coding scheme (Carletta, 1992) is composed of an initiating check moves, along with other moves that

accomplish the purpose of the check. On the other hand, the multi-tiered system in (Traum and Hinkelman, 1992) is conceived of in terms of ranks (at least for the lower three levels), and, although core speech acts like *inform* are only successfully realized at the point of a completed structure of *grounding acts*, there is no relationship between the type or sequence of grounding acts performed and the type of core speech acts which are realized.

Within the plan ontology described related to question 7, this amounts to a question of whether the decomposition of a dialogue act can contain other dialogue acts, or some other sort of realization.

15: Can multiple dialogue acts occur at the same time (performed through the same utterance)?

Since most utterances have multiple functions, the answer, given most definitions according to conditions and effects, will be “yes.” However there are a number of complications, depending on the use to which the taxonomy is put. For logical theories, one important question is whether the logic can accommodate simultaneous action or *level-generation* (Goldman, 1970). Simple versions of, e.g. the situation calculus (McCarthy and Hayes, 1969) or dynamic logic (Harel, 1979) do not, which makes it difficult to formalize this kind of phenomenon. Likewise, within dialogue systems, reasoning about act occurrence is often made not on the basis of necessary and sufficient conditions, but on closeness of fit, using abductive or statistical methods. Such methods generally are used to decide on a particular label to the exclusion of others, e.g., that an interrogative utterance is an indirect request but not a question. Finally, in tagging a corpus, it is often tedious and unreliable to try to code all possible occurrences of a particular function, and so instructions are designed so as to only code the most significant (in the opinion of the coding task designer), e.g., the *code high* principle in (Condon and Cech, 1992). It is important to be explicit about such assumptions, and whether multiple dialogue acts are assumed to be allowed to happen at the same time, and what the meaning of something not being coded is: non-occurrence or no statement about occurrence or non-occurrence. In the Condon-Cech scheme, one could deduce that a “higher” act had not occurred, but no such deduction is warranted about a “lower” act.

Taxonomic Considerations

16: Can the same taxonomy be used for different kinds of activities?

There are two relevant notions of activity here. First is the meta-activity of recognizing or coding dialogue acts, that is the concern of question 20. Relevant types of activities include logical deduction, system participation, and corpus analysis. There is also the issue of on-line or off-line coding and amount of lookahead (see question 6). Here I will concentrate on the activities that the dialogue participants are engaged in.

There are a number of different dialogue activities that people are interested in designing taxonomies of dialogue acts for. Some examples include casual conversation (Jurafsky et al., 1997), classroom discourse (Sinclair and Coulthard, 1975), and various flavors of task-oriented

dialogue, such as information seeking (van Vark et al., 1996), collaborative scheduling (Alexandersson et al., 1997), and direction following (Carletta et al., 1997)).

Taxonomies designed for different tasks or genres of dialogue tend to be quite different (e.g., even within the general realm of task-oriented cooperative dialogue, meeting-scheduling vs. direction following). To some extent, this is to be expected, since different genres will have different frequencies of acts. For example, roughly 50% of utterances are *statements* in the Switchboard corpus, which is concerned with causal conversation (Jurafsky et al., 1998), while the HCRC Maptask corpus, concerned with instruction giving/following (Carletta et al., 1997), has only 8% of utterances labelled with the equivalent tag, *explain*.⁶ Conversely, Maptask has 15.6% of utterances marked as *instruct*, while Switchboard has less than 1% of utterances labelled as *action-directive*. Interestingly, though, both have around 20% of utterances marked as *acknowledgement*. Different tasks and coding purposes may also place different demands on specificity of a taxonomy (see question 18), e.g., to have an appropriate reliability and perplexity for a given coding purpose.

Some hope that these different task specific “sub-taxonomies” might be fit together within a coherent general taxonomy of acts in dialogue. A general theory might also better allow one to identify activities as well as episodes within an activity and genres of activities. The DRI group has been working toward schemes that might have more general applicability (at least within the general category of task-oriented dialogue) (Discourse Resource Initiative, 1997; Core et al., 1999)). The SLSA project at Gothenburg University is investigating more generally the issue of corpus collection and dialogue coding of spoken language activities (Allwood, 1999).

17: Can the same taxonomy be used for different kinds of agents?

Related to the above question, is one of whether the same taxonomy could cover situations of humans communicating with humans, humans with machines, and machines with machines. Other possibilities could also include humans with animals or animals with animals (or animals with machines?). Even when only humans are communicating, there is still an important issue of the medium, e.g., face to face, spoken language only, multi-modal computer mediated communication of various flavors. These issues will certainly have a bearing on the distribution of act types. E.g., much more explicit grounding in spoken dialogue (> 95% (Traum and Heeman, 1997)) than computer chat (~ 40% (Dillenbourg et al., 1997))⁷, and more explicit verifications from computer systems with relatively poor speech recognition than between fluent humans.

Again, the hope of many researchers is that the same taxonomies (at a suitably abstract level, concerning some of the lack of subtlety of machine communication) could be used concerning any of these sets of agents. Some, however, have pointed to the differences in communication styles between human-human and human-machine communication as a reason for anticipating different taxonomies, and not carrying over the insights from one to the other (Jönsson, 1995).

18: How detailed should a dialogue act taxonomy be?

There are many subtle gradations in speech act verbs, often relating to different facets of the participants or normative attitudes towards the content of the act (e.g., state, assert, inform, con-

fess, concede, maintain, . . .). The question arises as to how many of these distinctions should be captured within a dialogue act taxonomy. One key issue is whether one wants to capture generalizations or distinctions. There is also often a trade-off between precisely capturing differences in conditions and effects and confidence in a label.

If possible it may be best to arrange these fine distinctions within a hierarchical or lattice structure (as is done by, e.g., (Allen and Core, 1997; Alexandersson et al., 1997)), so that a degree of specificity may be chosen appropriate to the particular task. One issue is whether theorists and coders can agree on the hierarchical structure of related acts, which, in some cases, may be more controversial than the base labels themselves.

19: Where should complexity be realized in a coding taxonomy?

Given that utterances in dialogue are generally multi-functional, the question arises as to how best to capture this multiplicity of functions in a taxonomy. There are two extremes: one is to separate out each function and code it separately, requiring multiple labels for each utterance, one for each function. The advantage is fairly simple act definitions, each with fairly clear semantics and ascription conditions. The disadvantage is a large number of tagging decisions — one for each functional dimension, leading to a fairly onerous tagging task, and lower reliability on some dimensions depending on coder attention and attunement to each phenomenon. This approach is taken by (Discourse Resource Initiative, 1997).

The other extreme is to combine sets of coherent bundles into complex labels and code with these. The advantage is a potentially easier and more reliable coding task, especially if the same bundles appear repeatedly within a given coding effort. The disadvantage is that there might be many possible acts, if many gradations appear. If only some of these are assigned labels, then it may be difficult to decide how to code an utterance that shares some (but not all) of the features of one label, while having some features from another. This approach can also lead to missing connections between different acts that share some of the features, making it hard to analyze existence of these features from the coded data. This approach may be typified by the first Verbmobil coding scheme (Jekat et al., 1995).

It is possible both to find taxonomies that take a more middle position than either extreme, and that capture some of the advantages of the other scheme (while lessening the disadvantages of their own). For example, the Switchboard DAMSL scheme uses many ideas from (Discourse Resource Initiative, 1997; Allen and Core, 1997), while moving toward the other extreme of coding in discrete, mutually exclusive bundles rather than multiples dimensions. There are also proposals to do this for the main DRI scheme as well (Core et al., 1999). These schemes will still retain the theoretical connection to the multi-layer scheme, and so will be more easy to determine individual functions. Likewise, it should be possible to define optional rather than mandatory *macros* which combine convenient bundles of features into simplified coding tasks, while still maintaining the full flexibility of the multiple layer approach. This is the method advocated in (Cooper et al., 1999).

20: Can a taxonomy used for tagging dialogue corpora be given a formal semantics and/or used in a dialogue system?

The hope of many researchers is definitely a “yes” answer to this question: the purpose of tagging or formal semantics is often for use within a dialogue system. Moreover a clear semantics may help one to formulate sharper principles for a tagging exercise. There are some difficulties, however. One is the issue of different resources - one may require details of the content of an act in order to use in a system or provide semantics, yet this may be too onerous for a tagging exercise. Likewise, formal representations of context built from incorporation of previous acts may not be available during a coding task. On the other hand, human coders may be able to use complex intuitions in their coding which are difficult to incorporate in a formal description or implementation (though these may perhaps be learned from a corpus, using machine learning techniques (Reithinger and Klesen, 1997; Samuel, 1998)). These different skill sets may tend to make taxonomies designed for different purposes diverge.

Discussion

Given that the above questions are not exhaustive or binary, and have remained mostly at the meta-level, we can certainly see that formulating the ultimate dialogue act taxonomy is a much harder problem than the game of 20-questions. The discussion above is also far from the last word on any of these topics. The hope is that further research may yield some more definitive answers or at least better understanding of the issues involved. Meanwhile, the above discussion may help dialogue act theorists be clearer about some of the meanings of their taxonomy, in the hopes of wider understanding and applicability of the taxonomies that are used.

Acknowledgments

The author was supported during the writing of this paper by the TRINDI (Task Oriented Instructional Dialogue) project, EU TELEMATICS APPLICATIONS Programme, Language Engineering Project LE4-8314. I would also like to (anonymously) thank the many colleagues who helped me formulate my ideas on the topics discussed in this paper, and Jens Allwood and William Mann for helpful discussions on previous versions of this paper.

DAVID R TRAUM
UMIACS
A. V. Williams Building
University of Maryland
College Park, MD 20742 USA

Notes

¹By the term *dialogue acts*, I don't mean to limit discussion to those theories and taxonomies that explicitly use this term. Other terms used for the same general concept include *locutionary*, *illocutionary*, and *perlocutionary acts* (Austin, 1962), *speech acts* (Searle, 1969), *communicative acts* (Allwood, 1976; Sadek, 1991; Airenti et al., 1993),

conversation acts (Traum and Hinkelman, 1992) and *conversational moves* (Carletta et al., 1997). My remarks here are intended to apply to the general phenomenon described by this range of terms. *Dialogue acts* can perhaps be seen as most generic, at least in the context of a workshop on dialogue.

²Washington D.C. Public Advocate David Howard, in February 1999.

³Here and elsewhere in the papers, examples are meant to be representative rather than exhaustive; there is a large amount of work in some of these areas.

⁴(Pollack, 1990) focuses instead on *enabling conditions* rather than pre-conditions, and *generation conditions* rather than decomposition, (following (Goldman, 1970).

⁵Details of axioms of this sort obviously vary quite a bit depending on the syntax and semantics of the logic used, e.g., whether *Done* means “happened in the immediately prior state transition” or some looser sense of happened recently.

⁶Maptask Statistics from personal communication from Amy Isard.

⁷Although these studies concerned different tasks.

References

- Airenti, G., Bara, B. G., and Colombetti, M. (1993). Conversation and behavior games in the pragmatics of dialogue. *Cognitive Science*, 17:197–256.
- Alexandersson, J., Buschbeck-Wolf, B., Fujinami, T., Maier, E., Reithinger, N., Schmitz, B., and Siegel, M. (1997). Dialogue acts in VERBMOBIL-2. Verbmobil Report 204, DFKI, University of Saarbruecken.
- Allen, J. and Core, M. (Draft, 1997). Draft of damsl: Dialog act markup in several layers. available through the WWW at: <http://www.cs.rochester.edu/research/trains/annotation>.
- Allen, J. F., Miller, B. W., Ringger, E. K., and Sikorski, T. (1996). A robust system for natural spoken dialogue. In *Proceedings of the 1996 Annual Meeting of the Association for Computational Linguistics (ACL-96)*, pages 62–70.
- Allen, J. F. and Perrault, C. R. (1980). Analyzing intention in utterances. *Artificial Intelligence*, 15(3):143–178.
- Allwood, J. (1976). *Linguistic Communication as Action and Cooperation*. PhD thesis, Göteborg University, Department of Linguistics.
- Allwood, J. (1977). A critical look at speech act theory. In Dahl, Ö., editor, *Logic, Pragmatics and Grammar*. Studentlitteratur.
- Allwood, J. (1980). On the analysis of communicative action. In Brenner, M., editor, *The Structure of Action*. Basil Blackwell. Also appears as Gothenburg Papers in Theoretical Linguistics 38, Dept of Linguistics, Göteborg University.
- Allwood, J. (1994). Obligations and options in dialogue. *Think Quarterly*, 3:9–18.
- Allwood, J. (1995). An activity based approach to pragmatics. Technical Report (GPTL) 75, Gothenburg Papers in Theoretical Linguistics, University of Göteborg.
- Allwood, J. (1999). The Swedish Spoken Language Corpus at Göteborg University. In Andersson, R., Abelin, Å., and an d Per Lindblad, J. A., editors, *Fonetik 99: Proceedings from the Twelfth Swedish Phonetics Conference*, number 81 in Gothenburg Papers in Theoretical Linguistics, pages 5–9. Department of Linguistics, Göteborg University.

- Allwood, J., Nivre, J., and Ahlsen, E. (1992). On the semantics and pragmatics of linguistic feedback. *Journal of Semantics*, 9.
- Austin, J. A. (1962). *How to Do Things with Words*. Harvard University Press.
- Bilange, E. (1991). A task independent oral dialogue model. In *Proceedings of the Fifth Conference of the European Chapter of the Association for Computational Linguistics*, pages 83–88.
- Bretier, P. and Sadek, M. D. (1996). A rational agent as the kernel of a cooperative spoken dialogue system: Implementing a logical theory of interaction. In Müller, J. P., Wooldridge, M. J., and Jennings, N. R., editors, *Intelligent Agents III — Proceedings of the Third International Workshop on Agent Theories, Architectures, and Languages (ATAL-96)*, Lecture Notes in Artificial Intelligence. Springer-Verlag, Heidelberg.
- Bunt, H. (1996). Interaction management functions and context representation requirements. In *Proceedings of the Twente Workshop on Language Technology: Dialogue Management in Natural Language Systems (TWLT 11)*, pages 187–198.
- Carletta, J. (1992). *Risk-taking and Recovery in Task-Oriented Dialogue*. PhD thesis, University of Edinburgh.
- Carletta, J., Isard, A., Isard, S., Kowtko, J. C., Doherty-Sneddon, G., and Anderson, A. H. (1997). The reliability of a dialogue structure coding scheme. *Computational Linguistics*, 23(1):13–31.
- Clark, H. H. (1992). *Arenas of Language Use*. University of Chicago Press.
- Clark, H. H. and Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science*, 13:259–294. Also appears as Chapter 5 in (Clark, 1992).
- Cohen, P. R. and Levesque, H. J. (1990). Rational interaction as the basis for communication. In Cohen, P. R., Morgan, J., and Pollack, M. E., editors, *Intentions in Communication*. MIT Press.
- Cohen, P. R. and Perrault, C. R. (1979). Elements of a plan-based theory of speech acts. *Cognitive Science*, 3(3):177–212.
- Condon, S. and Cech, C. (1992). Manual for coding decision-making interactions. unpublished manuscript, updated May 1995, available at: ftp://sls-ftp.lcs.mit.edu/pub/multiparty/coding_schemes/condon.
- Cooper, R., Larsson, S., Matheson, C., Poesio, M., and Traum, D. (1999). Coding instructional dialogue for information states. Deliverable D1.1, Trindi Project.
- Core, M., Ishizaki, M., Moore, J., Nakatani, C., Reithinger, N., Traum, D., and Tutiya, S. (1999). The report of the third workshop of the discourse resource initiative, chiba univeristy and kazusa academia hall. Technical Report No.3 CC-TR-99-1, Chiba Corpus Project.
- Dillenbourg, P., Jermann, P., Schneider, D., Traum, D., and Buii, C. (1997). The design of moo agents: Implications from a study on multi-modal collaborative problem solving. In *Proceedings of the 8th World Conference on Artificial Intelligence in Education (AI-ED 97)*, pages 15–22.
- Discourse Resource Initiative (1997). Standards for dialogue coding in natural language processing. Report no. 167, Dagstuhl-Seminar.
- External Interfaces Working Group (1993). Draft specification of the kqml agent-communication language. available through the WWW at: <http://www.cs.umbc.edu/kqml/papers/>.
- FIPA (1997). Fipa 97 specification part 2: Agent communication language. working paper available at <http://drogo.cse.stet.it/fipa/spec/fipa97/f8a21.zip>.
- Ginzburg, J. (1998). Clarifying utterances. In Hulstijn, J. and Niholt, A., editors, *Proc. of the Twente Workshop on the Formal Semantics and Pragmatics of Dialogues*, pages 11–30, Enschede. Universiteit Twente, Faculteit Informatica.
- Goldman, A. I. (1970). *A Theory of Human Action*. Prentice Hall Inc.

- Grice, H. P. (1957). Meaning. *Philosophical Review*, 66:377–88.
- Grosz, B. J. and Sidner, C. L. (1986). Attention, intention, and the structure of discourse. *Computational Linguistics*, 12(3):175–204.
- Halliday, M. A. K. (1961). Categories of the theory of grammar. *Word*, 17:241–92.
- Harel, D. (1979). *First Order Dynamic Logic*. Springer-Verlag.
- Jackendoff, R. (1983). *Semantics and Cognition*. The MIT Press.
- Jekat, S., Klein, A., Maier, E., Maleck, I., Mast, M., and Quantz, J. (1995). Dialogue Acts in VERBMOBIL. Technical Report 65, BMBF Verbmobil Report.
- Jönsson, A. (1995). Dialogue actions for natural language interfaces. In *Proc. of the 14th IJCAI*, pages 1405–1411, Montreal, Canada.
- Jurafsky, D., Shriberg, E., and Biasca, D. (1997). Switchboard swbd-damsl shallow-discourse-function annotation coders manual. Technical Report 97-02, University of Colorado Institute of Cognitive Science. Draft 13.
- Jurafsky, D., Shriberg, E., Fox, B., and Curl, T. (1998). Lexical, prosodic, and syntactic cues for dialog acts. In *Proceedings of ACL/COLING-98 Workshop on Discourse Relations and Discourse Markers*, pages 114–120, Montreal, Canada.
- Lambert, L. and Carberry, S. (1991). A tripartite plan-based model of discourse. In *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics*, pages 47–544.
- Litman, D. J. and Allen, J. F. (1987). A plan recognition model for subdialogues in conversation. *Cognitive Science*, 11:163–200.
- McCarthy, J. and Hayes, P. (1969). Some philosophical problems from the standpoint of artificial intelligence. In Meltzer, B. and Michie, D., editors, *Machine Intelligence 4*, pages 463–502. Edinburgh University Press. Also appears in N. Nilsson and B. Webber (editors), *Readings in Artificial Intelligence*, Morgan-Kaufmann.
- McRoy, S. W. and Hirst, G. (1995). The repair of speech act misunderstandings by abductive inference. *Computational Linguistics*, 21(4):5–478.
- Nakatani, C. H. and Traum, D. R. (1999). Coding discourse structure in dialogue (version 1.0). Technical Report UMIACS-TR-99-03, University of Maryland.
- Poesio, M. and Traum, D. R. (1997). Conversational actions and discourse situations. *Computational Intelligence*, 13(3).
- Poesio, M. and Traum, D. R. (1998). Towards an axiomatization of dialogue acts. In *Proceedings of Twendial'98, 13th Twente Workshop on Language Technology: Formal Semantics and Pragmatics of Dialogue*.
- Pollack, M. E. (1990). Plans as complex mental attitudes. In Cohen, P. R., Morgan, J., and Pollack, M. E., editors, *Intentions in Communication*. MIT Press.
- Reithinger, N. and Klesen, M. (1997). Dialogue act classification using language models. In *Proc. Eurospeech '97*, pages 2235–2238, Rhodes, Greece.
- Sadek, M. D. (1991). Dialogue acts are rational plans. In *Proceedings of the ESCA/ETR workshop on multi-modal dialogue*.
- Samuel, K. (1998). DISCOURSE LEARNING: Dialogue act tagging with transformation-based learning. In *Proceedings of the 15th National Conference on Artificial Intelligence (AAAI-98) and of the 10th Conference on Innovative Applications of Artificial Intelligence (IAAI-98)*, pages 1199–1199, Menlo Park. AAAI Press.
- Searle, J. R. (1969). *Speech Acts*. Cambridge University Press, New York.
- Severinson Eklundh, K. (1983). The notion of language game – a natural unit of dialogue and discourse. Technical Report SIC 5, University of Linköping, Studies in Communication.

- Sidner, C. L. (1994). An artificial discourse language for collaborative negotiation. In *Proceedings of the fourteenth National Conference of the American Association for Artificial Intelligence (AAAI-94)*, pages 814–819.
- Sinclair, J. M. and Coulthard, R. M. (1975). *Towards an analysis of Discourse: The English used by teachers and pupils*. Oxford University Press.
- Singh, M. P. (1998). Agent communication languages: Rethinking the principles. *IEEE Computer*, 31(12):40–47.
- Traum, D. R. (1994). *A Computational Theory of Grounding in Natural Language Conversation*. PhD thesis, Department of Computer Science, University of Rochester. Also available as TR 545, Department of Computer Science, University of Rochester.
- Traum, D. R. (1999). Speech acts for dialogue agents. In Rao, A. and Wooldridge, M., editors, *Foundations of Rational Agency*. Kluwer.
- Traum, D. R. and Allen, J. F. (1992). A speech acts approach to grounding in conversation. In *Proceedings 2nd International Conference on Spoken Language Processing (ICSLP-92)*, pages 137–40.
- Traum, D. R. and Heeman, P. (1997). Utterance units in spoken dialogue. In Maier, E., Mast, M., and Luperfoy, S., editors, *Dialogue Processing in Spoken Language Systems — ECAI-96 Workshop*, Lecture Notes in Artificial Intelligence, pages 125–140. Springer-Verlag, Heidelberg.
- Traum, D. R. and Hinkelman, E. A. (1992). Conversation acts in task-oriented spoken dialogue. *Computational Intelligence*, 8(3):575–599. Special Issue on Non-literal language.
- van Vark, R., de Vreught, J., and Rothkrantz, L. (1996). Analysing ovr dialogue coding scheme 1.0. Technical Report 96-137, TU Delft Faculty of Technical Mathematics and Informatics.
- Vanderveken, D. (1990). On the unification of speech act theory and formal semantics. In Cohen, P. R., Morgan, J., and Pollack, M. E., editors, *Intentions in Communication*. MIT Press.
- Vanderveken, D. (1990-1991). *Meaning and Speech Acts*. Cambridge University Press.
- Winograd, T. and Flores, F. (1986). *Understanding Computers and Cognition*. Addison-Wesley Publishing Company, Inc.