Journal of Semantics 17: 7–30

20 Questions on Dialogue Act Taxonomies

DAVID R. TRAUM
University of Southern California

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 must be addressed in order to facilitate the shared understanding and use of taxonomies. The discussion is framed in terms of 20 questions, the answers to which will help make the meanings of taxonomy elements more clear to different communities of users.

1 INTRODUCTION

When engaging in a study related to dialogue pragmatics, a researcher is confronted with a bewildering range of theories and taxonomies of dialogue acts[1] to choose from. Moreover, specific deficits in any given theory often 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 new formal languages (e.g. programming languages like Java) and (at a slower pace) natural languages continue to be created. 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, various conditions are used to characterize a dialogue act labeled as *inform*, including those listed in (1).

[1] By 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 & Hinkelman 1992), *conversational moves* (Carletta *et al.* 1997), and *dialogue moves* (Cooper *et al.* 1999). 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 forum on dialogue.

[2] Washington DC Public Advocate David Howard, in February 1999.

When one encounters such a label, it is often not clear which subset of the constraints in (1) (or perhaps none of them, when an entirely different formulation is used to define an *inform*) are meant by the labeler to characterize the labeled utterance. 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 & 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 concepts as 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 the claim that dialogue acts are a useful concept, given the wide variety of uses to which they are put. Some of these uses include:[3] representations of the pragmatic meaning of utterances in dialogue theories (Vanderveken 1991; Bunt 1996; Poesio & Traum 1997, 1998), building blocks for grammars of dialogue (Winograd & Flores 1986; Bilange 1991), labels for corpus annotation (Carletta *et al.* 1997; Alexandersson *et al.* 1998), 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 & 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 researchers other than the taxonomy designers. Here I will briefly raise some of the issues that have often caused confusion when interpreting one taxonomy of dialogue acts from within the viewpoint of another. These issues must be addressed in order to have a clearer idea of what one means by saying that 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, I focus here on 20, formulated as questions, in homage to the

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

'dialogue game' named in the title. For convenience, these questions are grouped into sections of related questions.

2 DEFINING DIALOGUE ACTS

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

This question has implications beyond just dialogue act definitions, and applies to 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 all of the properties of the intuitive concept (e.g. normal modal logics are very popular as models of *belief* (c.f. Hintikka 1962), yet they have the property of logical omniscience, which is certainly undesirable as a model of human belief)? Or should one 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 well-defined concept within a formal system may not be close enough to the underlying concept to be useful, but an underspecified concept without some of these formal properties may suffice for the task at hand (corpus labeling 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 for 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, but 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.

Question 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 use in analyzing 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 naïve coders without instructions to the contrary, may tend to focus on the former enterprise, which may have undesirable consequences depending on how 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 containing a speech act verb to describe an occurrence. For Lexical Semantics, these linguistic intuitions (or similar examinations of actual usage) are paramount (barring issues of polysemy). On the other hand, the intuitions might not be useful when devising an ontology of action—such an ontology might, for independently motivated reasons, diverge from the classifications made in natural languages.

Question 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. On the other hand, none of these conditions is necessary for action ascription.

(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 generalizations over acts that consumers of a dialogue act taxonomy would like to express. For example, one may be interested strictly in the result, intention, or context, or perhaps in 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 using only one of the other criteria, as this would lead to an unjustified claim of the results holding. Using different criteria (e.g. results only vs. intention only) 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 comes to believe 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 context conditions hold, or in which the speaker has a different intention.

Which kinds of conditions and whether they are necessary will also depend on the task being attempted. Compare, for example, the tasks of discovering lexical semantics compared with the task of constructing an action ontology, as discussed in the previous question. Also, it makes a difference 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. an offline annotator of a pre-collected dialogue corpus (see also question 6).

Question 4: What is the role of speaker intention?

Intention is usually given a somewhat privileged position with respect to determining what dialogue acts (or actions, in general) have been performed, viz. 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 with this approach 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, at times, 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 act or 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, **p**, 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 intends to cause the hearer to believe 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 the issues of intention recognition and dialogue act attribution, even though they are related.

Question 5: What is the role of addressee uptake?

Regardless of speaker intention, many dialogue act definitions 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 addressee 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 & Schaefer 1989; Traum & 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 & Hirst 1995), severing completely the link between the dialogue effects and original speaker intentions or addressee uptake.

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

There are several points of view that 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–addressee team, which may 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, in order to determine what acts have been performed (e.g. whether a speaker has committed herself). There is also the issue of time with respect to coding or ascription of acts: is

it an on-line decision made at the time of performance (or using only information available at the time of performance), or is one allowed to consider subsequent utterances/action, as well, before deciding what happened?

Point of view is relatively straightforward from the internal perspective of a dialogue system (although a system might still need to reason about the interlocutor's point of view, including information discovered at subsequent time points, in diagnosing misunderstanding (McRoy & Hirst 1995) or constructing a negotiated view). It is, however, far from clear what point of view should be taken by coders (and how they should estimate the speaker's or addressee's point of view without mind-reading). Likewise, in defining the acts or giving them a logical semantics, it may be necessary to take point of view 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 *acknowledgment*, *repair*, *request for repair*, or *request for confirmation* (Traum & Allen 1992). The latter two could perhaps be distinguished from the former by prosody: questioning intonation could indicate lack of certainty as well as desire for further feedback, while declarative intonation could indicate one of the former functions. One could distinguish acknowledgment from repair by deciding whether the second utterance repeats (or paraphrases) the information in the former (*acknowledgment*), or changes some part of it (*repair*). However, this decision requires a point of view, indicating who believes it to be the same or different. Especially with current technology speech recognition systems, there is a significant likelihood that a system may 'repeat' what it thought it heard, while producing something different 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.

3 DIALOGUE ACT COMPONENTS

Question 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). Theories of action proposed by Artificial Intelligence (AI) researchers 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 is characterized by 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 performance of the main action. In addition, something like the schema in (5) is used (although usually only in an abductive or circumscriptive sense, rather than as a sound axiom describing all circumstances), for reasoning from subaction to intention ascription (plan recognition). These axioms can also be used to 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. $Pre(Y, prev) \wedge decomp(Y, \{X_1, \ldots, X_n, \}) \wedge$
$\forall X_{i:1 \leq i \leq n} : Done(X_i, now) \rightarrow Done(Y, now)$

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

Question 8: What is context?

Given the general framework for actions in the discussion of the previous question, a large question remains as to which aspects of the situation are relevant as potential conditions for defining types of dialogue act performance, and which aspects 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

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

[5] Details 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.

(Winograd & Flores 1986; Traum & Allen 1992; Lewin 1998), or some other structural representation of context (e.g. Ginzburg 1998). Using the dialogue grammar approach, certain acts may be defined with a pre-condition that the dialogue be in a particular state in the transition network, while the effects will include a transition to a new state in the network. Or, using a different notion of information state, one could stipulate a pre-condition that it is only possible to perform an *answer* if there is a relevant question under discussion (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 & Perrault 1979; Allen & 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 kind of information is in terms of the social obligations and commitments undertaken by the dialogue participants (Allwood 1994; Poesio & 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 be formulated for dialogue acts such *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.

Question 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. If the enabling conditions do not hold, a more complex plan is formed to achieve the conditions so that the action under consideration may be attempted; if the applicability constraints do not hold, the action will be dropped from consideration, and some other action (or set of actions) will be considered instead. 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 cannot 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, but this flexibility comes at the price of having to determine whether an action is felicitous and also needing to characterize *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 *dialogue act recognition*. For the former (e.g. using axiom (4a)), one might care more about sufficient rather than necessary conditions. For the latter (e.g. using axiom (4b)), however, one might be more interested in necessary conditions (to use this as an axiom rather than as a default rule).

Question 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 agent A to agent B, for B to do some action x, schematically: $Request(A, B, Do(B, x))$. One must now determine which of the conditions in (6) to associate with an attempt vs. success vs. satisfaction. Condition (6f) seems sufficient to describe an attempted request, while (6a) is necessary for a fully satisfied request. 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) (in the latter case, requiring an additional assumption of cooperativity to lead to (6b) and then (6a) (Cohen & Perrault 1979)). The social commitments approach would favor (6c) (Allwood 1994), or (6d) (Traum 1994) (in the latter theory, (6c) would come about only as a result of acceptance of the request).

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

Another issue concerns the kinds of actions 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 & Schaefer 1989; Traum 1994) needed? It is certainly most likely that an addressee will not perform a requested act (6a) (or intend to perform it (6b)) if she does not hear or understand the request. Likewise, it is debatable whether one would even have the obligation to perform the act (6c) under such conditions. Should one then say that the act was not performed (equating success with achieved result, as in (3c)), or that the results do not necessarily hold when an act has been performed?

4 RELATIONSHIPS AND COMPLEX ACTS

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

One of the main intuitions behind speech act theory (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 to account for both. Litman & Allen (1987) distinguished dialogue acts as being *meta-acts*, defining discourse plans as having other plans (domain or discourse) as parameters. Lambert & 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. Another difference between dialogue acts and many sorts of physical action is that dialogue acts involve multiple agents, since there is at least a speaker and addressee involved. See also question 13.

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

Dialogue structure is used for a variety of purposes, e.g. for calculating referential accessibility, topic and focus, and global coherence of utterances. There are several options as to how to view the relationship between dialogue structure and the dialogue acts that have been performed. Some conceive dialogue structure as being wholly dependent on the structure of

performance of dialogue acts (e.g. grammar-based approaches such as Sinclair & Coulthard 1975). Others use a different sort of structure, not directly composed of the performance of dialogue acts, which is sensitive to other aspects of the utterances, or is primarily constructed from the activity that the participants are engaged in (Allwood 1995; Grosz & 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, serving as the source of pre-conditions for act definitions, and as input for a process of action recognition. For example, one might want to say that an *answer* act is only possible given some configuration of dialogue structure. One might frame this either in terms of previous acts (e.g. an *information request* act had just been performed), or in terms of other sorts of structure (e.g. there is a *Question Under Discussion* for which this act provides the answer, regardless of whether any particular act happened to bring the question under discussion).

Question 13: Are there multi-agent dialogue acts?

As mentioned in relation to question 5, some researchers view the performance of most illocutionary acts as a 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 & Coulthard 1975), or collaborative completions where one speaker finishes another's sentence. There are several difficulties with these kinds of acts, however. The first is related to reliable tagging and deciding what aspects of a dialogue are 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 & Traum 1999). This difficulty is compounded when there are multiple acts with different boundaries (e.g. the multi-agent act and the single-agent component of a multi-agent act performed by a speaker within an utterance).

Another issue is that one will need a more complex logic to represent multi-agent action than is needed for representing single agent action. For example, if one needs to reason about the single-agent components as well as the multi-agent act, then one needs a logic allowing simultaneous action and a method for relating the two actions (e.g. using something like the proposal in Goldman (1970).

Question 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 four tiers in the system of Sinclair & Coulthard (1975) is conceived of as ranks within a general 'discourse' level. Also, the *check game* in the Maptask coding scheme (Carletta *et al.* 1997) is composed of an initiating *check* move, along with other moves that accomplish the purpose of the check. On the other hand, the multi-tiered system in Traum & Hinkelman (1992) is organized in levels (at least for the lower three strata), 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 that are realized.

Within the plan ontology described in the discussion of question 7, this amounts to a question of whether the decomposition of a dialogue act contains only other dialogue acts, or involves some other sort of realization.

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

Since most utterances have multiple functions, the answer will be 'yes', given most definitions formulated in terms of conditions and effects. There are, however, 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 & 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 (McRoy & Hirst 1995) or statistical methods (Reithinger & Klesen 1997). Such methods generally are used to decide on a particular label while excluding others, e.g. deciding 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 all functions, and so designers of coding manuals often instruct annotators to label only the most significant function (in the opinion of the coding task designer), e.g. the *code high* principle in Condon & 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: assumed occurrence (perhaps on the basis of some other tag), assumed 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 the occurrence of a 'lower' act.

5 TAXONOMIC CONSIDERATIONS

Question 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, which is the concern of question 20. Relevant types of meta-activities include logical reasoning, system participation in a dialogue, and corpus analysis. For the meta-activity of attributing dialogue acts to utterances, there is also the issue of whether this is an on-line or off-line attribution, and the amount of lookahead allowed (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 have been designing taxonomies of dialogue acts for. Some examples include casual conversation (Jurafsky *et al.* 1997), classroom discourse (Sinclair & Coulthard 1975), and various flavors of task-oriented dialogue, such as information seeking (van Vark *et al.* 1996), collaborative scheduling (Alexandersson *et al.* 1998), 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. This can be seen in Table 1, which compares eight coding efforts, showing for each the percentage of utterances that were labeled with various tags. Most cells actually show the percentage of utterances labeled with one of a set of tags rather than an individual tag,

using the inter-scheme equivalences proposed in (Klein *et al.* 1999).[6] The first two columns use the DRI coding scheme (Discourse Resource Initiative 1997) and the manual in Allen & Core (1997). Two different tasks were coded in these efforts, however: in the first, dialogues were about planning the movements of trains and commodities (Heeman & Allen 1994),[7] while in the second, the dialogues involved more complex disaster management planning (Stent 2000). The third column shows coding of the Switchboard Corpus (Jurafsky *et al.* 1998) using a variant of the DAMSL tags (Jurafsky *et al.* 1997). The next two columns depict coding efforts using the HCRC coding scheme (Carletta *et al.* 1996), on the same task, Maptask. The fourth column shows results for the HCRC corpus (Carletta *et al.* 1997),[8] involving Scottish students, while the fifth column shows the DCIEM corpus, involving Canadian military personnel (Taylor *et al.* 1998). The last three columns involve variants of the Verbmobil Tasks. The last one uses the first Verbmobil coding scheme (Jekat *et al.* 1995), on a German corpus of scheduling dialogues (Kipp 1998). The sixth and seventh use the revised scheme (Alexandersson *et al.* 1998) on a wider variety of tasks, the sixth column showing English-speaking subjects, the seventh showing German-speaking subjects.[9]

As can be seen, there are some striking differences in distributions of act types across the various domains, schemes, and corpora. For example, roughly 50% of utterances are *statements* in the Switchboard corpus, which is concerned with casual conversation, while the Maptask efforts, concerned with instruction giving/following, have only 8% of utterances labeled with the equivalent tag, *explain*. Conversely, the Maptask dialogues have over 15% of utterances marked as *instruct*, while Switchboard has less than 1% of utterances labeled as *action-directive*. 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. While it is hard to see from Table 1, since individual tags are clustered for comparison purposes, there are also large differences

[6] The comparisons here are very rough, since the proposed equivalences might not hold in all cases. For one thing, the first scheme (in the first two columns) allows an utterance to be labeled with multiple tags, while the latter do not. The numbers for these columns thus do not equal 100%, since the percentages are based on utterances rather than tags. Other columns fail to reach 100% due to some codes in incomparable categories. Also, while there is no corresponding category for *questions* in the Verbmobil scheme, it is likely that the subjects did ask questions, though these probably were coded as requests. Likewise, many of the feedback codes are probably also acknowledgments.

[7] TRAINS Statistics from Mark Core, personal communication. See Core (1998) for details of the annotation.

[8] HCRC Maptask Statistics, personal communication from Amy Isard.

[9] Verbmobil II statistics, personal communication from Michael Kipp.

Table 1 Percentage distributions of dialogue acts in corpus coding

Damsl TRAINS	Damsl Monroe	SWBD-Damsl Switchboard	HCRC HCRC Maptask	HCRC DCIEM Maptask	Verbmobil II Verbmobil English	Verbmobil II Verbmobil German	Verbmobil I Verbmobil I German
statement 45.9	51.4	49	explain 7.9	7.9	Inform, . . . 22.8	21.2	12.2
info-request 15.2	9.9	questions 4.9	query, check, align 23.5	20.3			
action-dir, oo 12.2	12.9	0.7	instruct 15.6	15.2	request, suggest 26.0	27.0	32
commit, offer 23.8	16.8	0.1			commit 0.5	0.8	
conventional 2.5	0.6	1.4			13.4	15.6	16.5
answer 14.7	8.4	3	reply, clarify 22.8	20	feedback 15.2	9.8	0.6
accept 30.0	23.0	5			accept, confirm 10.3	12.3	13.5
reject 2.2	0.5	0.2			reject, explained 3.3	4.4	8.2
other agree 3.6	1.8	0.3			clarify 2.3	1.9	8.9
Understanding 30.2	28.5	23	acknowledge 20.5	28.1	backchannel 3.6	3.3	
non-understand 1.2	0.5	0.1					

between the number of tags in these coding schemes, e.g. 12 tags in the HCRC scheme vs. 34 tags in Verbmobil II.

Some researchers 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 use act distributions to identify activities or genres of activities as well as episodes within an activity. The DRI group has been working toward the goal of a general purpose scheme 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).

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

As well as considering different communicative activities, we may also consider 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 possibly even animals with machines). 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 for any of these sets of agents. Some (e.g. Jönsson 1995), however, have pointed to the differences in communication styles between human–human and human–machine communication as a motivation for different taxonomies, and not carrying over too many insights from one setting to the other.

Even when only humans are communicating, there is still an important issue of the medium, e.g. face to face, spoken language only, or multi-modal computer mediated communication of various flavors. These issues will certainly have a bearing on the distribution of act types. For example, there is much more explicit grounding in spoken dialogue ($> 95\%$ (Traum & Heeman 1997)) than computer chat ($\sim 40\%$ (Dillenbourg *et al.* 1997)),[10] and more explicit verifications from computer systems with relatively poor speech recognition than between fluent humans.

We can see from Table 1 that even within the same task group and using the same medium of spoken language, we can see significant differences in some of the act distributions, e.g. the different amount of acknowledgments performed during Maptask by Canadian military and Scottish students, or

[10] Note, however, that these studies concerned different tasks.

for the Verbmobil II participants, contrasting English and German speakers. While using the same coding scheme for different corpora involving different participant groups may allow investigation into social and/or stylistic differences between speaker groups and between individuals, it may not be ideal for e.g. purposes of statistical training of computer systems within the style of a single group (in which case, one would want a scheme with maximal discrimination of the coding decisions).

Question 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*, *confess*, *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. Also, there is often a trade-off between proposing many acts, to precisely capture subtle differences in conditions and effects and the reliability that can be attached to a coding effort using these tags, given inevitable ambiguity in particular situations (not to mention the potential for coders not sharing an understanding of the intended distinctions).

If possible it may be best to arrange these fine distinctions within a hierarchical or lattice structure (as is done by e.g. Allen & Core 1997; Alexandersson *et al.* 1998), so that a degree of specificity may be chosen that is 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.

Question 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, which requires multiple labels for each utterance, one for each function. The advantage is an ability to use fairly simple act definitions, each with fairly clear semantics and ascription conditions. The disadvantage is that there are a large number of tagging decisions—one for each functional dimension, which, if coded by human annotators, leads to a fairly onerous tagging task and lower reliability on some dimensions

depending on annotator attention and atunement to each phenomenon. This approach is taken by Discourse Resource Initiative (1997) and Allen & Core (1997).

The other extreme is to combine sets of coherent bundles of dialogue functions into complex labels and use these labels for coding dialogues. 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 different collections of functions co-occur in the corpus. If only some of these function-bundles 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 is taken by the first Verbmobil coding scheme (Jekat *et al.* 1995).

It is also possible to find taxonomies that take a more intermediate position than either extreme, attempting to capture some of the advantages of each. For example, the Switchboard DAMSL scheme uses many ideas from Discourse Resource Initative (1997) and Allen & 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 still retain the theoretical connection to the multi-layer DRI scheme, and so it should still be relatively straightforward to determine individual functions. Likewise, it should be possible to define optional rather than mandatory *macros* which combine convenient bundles of features together, simplifying the coding tasks while still maintaining the full flexibility of coding multiple functions. This is the method advocated in Poesio *et al.* (1999) and Cooper *et al.* (1999).

Question 20: Can a taxonomy used for tagging dialogue corpora be given a formal semantics and/or be 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 (see Poesio & Traum 1998 for an attempt to formalize the acts in Discourse Resource Initiative 1997; Allen & Core 1997). There are some difficulties, however. One is the issue of

different resources—one may require not just the act category, but also details of the content of an act in order to use in a dialogue system or provide an appropriate semantic interpretation, yet providing this information 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 (however, these intuitions may perhaps be learned from a corpus, using machine learning techniques (Reithinger & Klesen 1997; Samuel 1998; Poesio & Mikheev 1998; Wright *et al* 1999)). These different skill sets may tend to make taxonomies designed for different purposes diverge.

6 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, hopefully leading to wider understanding and applicability of the taxonomies that are used.

Acknowledgements

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, because there are too many to mention without fear of forgetting someone important) thank the many colleagues who helped me formulate my ideas on the topics discussed in this paper. Thanks also to Jan Alexandersson, Mark Core, Amy Isard, Michael Kipp, and Amanda Stent for providing information on the various coding efforts reported in Table 1. Finally, I would like to thank Jens Allwood, William Mann, Henk Zeevat, and an anonymous reviewer for helpful comments on previous versions of this paper.

DAVID R TRAUM
USC Institute for Creative Technologies
13274 Fiji Way
Marina del Rey, CA 90292 USA
traum@ict.usc.edu

Received: 01.09.1999
Final version received: 21.07.2000

REFERENCES

Airenti, G., Bara, B. G., & Colombetti, M. (1993), 'Conversation and behavior games in the pragmatics of dialogue', *Cognitive Science*, 17, 197–256.

Alexandersson, J., Buschbeck-Wolf, B., Fujinami, T., Kipp, M., Koch, S., Maier, E., Reithinger, N., Schmitz, B., & Siegel, M. (1998), 'Dialogue Acts in VERBMOBIL-2', second edition, Verbmobil-Report 226, DFKI Saarbrücken, Universität Stuttgart, Technische Universität Berlin, Universität des Saarlandes.

Allen, J. & Core, M. (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., & 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)*, 62–70.

Allen, J. F. & Perrault, C. R. (1980), 'Analyzing intention in utterances'. *Artificial Intelligence*, 15, 3, 143–78.

Allwood, J. (1976), 'Linguistic communication as action and cooperation', Ph.D. thesis, Göteborg University, Department of Linguistics.

Allwood, J. (1977), 'A critical look at speech act theory', in Ö. Dahl (ed.), *Logic, Pragmatics and Grammar*, Studentlitteratur, Lund.

Allwood, J. (1980), 'On the analysis of communicative action', in M. Brenner, (ed.), *The Structure of Action*, Basil Blackwell, Oxford. 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 R. Andersson, Å. Abelin, J. Allwood & P. Lindblad (eds), *Fonetik 99: Proceedings from the Twelfth Swedish Phonetics Conference, Gothenburg Papers in Theoretical Linguistics 81*, 5–9. Department of Linguistics, Göteborg University.

Allwood, J., Nivre, J., & 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, Cambridge, MA.

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*, 83–8.

Bretier, P. & Sadek, M. D. (1996), 'A rational agent as the kernel of a cooperative spoken dialogue system, implementing a logical theory of interaction', in J. P. Müller, M. J. Wooldridge & N. R. Jennings (eds), *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)*, 187–98.

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

Carletta, J., Isard, A., Isard, S., Kowtko, J. C., Doherty-Sneddon, G., & 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, Chicago.

Clark, H. H. & Schaefer, E. F. (1989), 'Contributing to discourse', *Cognitive Science*, 13, 259–94. Also appears as Chapter 5 in Clark (1992).

Cohen, P. R. & Levesque, H. J. (1990), 'Rational interaction as the basis for communication', in P. R. Cohen, J. Morgan & M. E. Pollack (eds), *Intentions in Communication*, MIT Press, Cambridge, MA.

Cohen, P. R. & Perrault, C. R. (1979), 'Elements of a plan-based theory of speech acts', *Cognitive Science*, 3, 3, 177–212.

Condon, S. & 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., & Traum, D. (1999), 'Coding instructional dialogue for information states. Deliverable D1.1, Trindi Project.

Core, M. (1998), 'Analyzing and predicting patterns of DAMSL utterance tags', in *Working Notes AAAI Spring Symposium on Applying Machine Learning to Discourse Processing*, 18–24.

Core, M., Ishizaki, M., Moore, J., Nakatani, C., Reithinger, N., Traum, D., & Tutiya, S. (1999), 'The report of the third workshop of the Discourse Resource Initiative', Chiba University and Kazusa Academia Hall. Technical Report No.3 CC-TR-99-1, Chiba Corpus Project.

Dillenbourg, P., Jermann, P., Schneider, D., Traum, D., & Buiu, 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)*, 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.cselt.stet.it/fipa/spec/fipa97/f8a21.zip.

Ginzburg, J. (1998), 'Clarifying utterances', in J. Hulstijn & A. Niholt (eds), *Proc. of the Twente Workshop on the Formal Semantics and Pragmatics of Dialogues*, 11–30, Enschede, Universiteit Twente, Faculteit Informatica.

Goldman, A. I. (1970), *A Theory of Human Action*, Prentice Hall Inc., New York.

Grice, H. P. (1957), 'Meaning', *Philosophical Review*, 66, 377–88.

Grosz, B. J. & 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, Heidelberg.

Heeman, P. A. & Allen, J. (1994), 'The TRAINS 93 dialogues', TRAINS Technical Note 94-2, Department of Computer Science, University of Rochester.

Hintikka, J. (1962), *Knowledge and Belief: An Introduction to the Logic of the Two Notions*, Cornell University Press, Cornell, NY.

Jackendoff, R. (1983), *Semantics and Cognition*, MIT Press, Cambridge, MA.

Jekat, S., Klein, A., Maier, E., Maleck, I., Mast, M., & 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*, 1405–11, Montreal, Canada.

Jurafsky, D., Bates, R., Coccaro, N., Martin, R., Meteer, M., Ries, K., Shriberg, E., Stolcke, A., Taylor, P., & Ess-Dykema., C. V. (1998), 'Switchboard discourse language modeling project final report',

Research Note 30, Center for Speech and Language Processing, Johns Hopkins University.

Jurafsky, D., Shriberg, E., & 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.

Kipp, M. (1998), 'The neural path to dialogue acts', in *Proceedings of ECAI 98*, 175–9.

Klein, M., Bernsen, N. O., Davies, S., Dybkjaer, L., Garrido, J., Kasch, H., Mengel, A., Pirrelli, V., Poesio, M., Quazza, S., & Soria, C. (1999), 'Supported coding schemes', deliverable D1.1, MATE Project, available at http://www.dfki.de/mate/d11/.

Lambert, L. & Carberry, S. (1991), 'A tripartite plan-based model of discourse', in *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics*, 47–544.

Lewin, I. (1998), 'The autoroute dialogue demonstrator', Technical Report CRC-073, SRI Cambridge Computer Science Research Centre.

Litman, D. J. & Allen, J. F. (1987), 'A plan recognition model for subdialogues in conversation', *Cognitive Science*, 11, 163–200.

McCarthy, J. & Hayes, P. (1969), 'Some philosophical problems from the standpoint of artificial intelligence', in B. Meltzer & D. Michie (eds), *Machine Intelligence 4*, Edinburgh University Press, Edinburgh, 463–502. Also appears in N. Nilsson & B. Webber (eds), *Readings in Artificial Intelligence*, Morgan-Kaufmann. Los Altos, California.

McRoy, S. W. & Hirst, G. (1995), 'The repair of speech act misunderstandings by abductive inference', *Computational Linguistics*, 21, 4, 5–478.

Nakatani, C. H. & Traum, D. R. (1999), 'Coding discourse structure in dialogue (version 1.0)', Technical Report UMIACS-TR-99-03, University of Maryland.

Poesio, M., Cooper, R., Larsson, S., Matheson, C., & Traum., D. (1999), 'Annotating conversations for information state update', in *Proceedings of Amstelogue '99 Workshop on the Semantics and Pragmatics of Dialogue.*

Poesio, M. & Mikheev, A. (1998), 'The predictive power of game structure in dialogue act recognition, experimental results using maximum entropy estimation', in *Proceedings of ICSLP-98*, Sydney, 1998.

Poesio, M. & Traum, D. R. (1997), 'Conversational actions and discourse situations', *Computational Intelligence*, 13, 3.

Poesio, M. & 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 P. R. Cohen, J. Morgan & M. E. Pollack (eds), *Intentions in Communication*, MIT Press, Cambridge, MA.

Reithinger, N. & Klesen, M. (1997), 'Dialogue act classification using language models', in *Proc. Eurospeech '97*, 2235–8, 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)*, 1199, AAAI Press, Menlo Park.

Searle, J. R. (1969), *Speech Acts*, Cambridge University Press, New York, NY.

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 forteenth National Conference of the American Association for Artificial Intelligence (AAAI-94)*, 814–19.

Sinclair, J. M. & Coulthard, R. M. (1975), *Towards an Analysis of Discourse: The English Used by Teachers and Pupils*. Oxford University Press, Oxford.

Singh, M. P. (1998), 'Agent communication languages: rethinking the principles', *IEEE Computer*, 31, 12, 40–7.

Stent, A. J. (2000). 'The Monroe corpus', Technical Report 728, Computer Science Dept., University of Rochester.

Taylor, P. A., King, S., Isard, S. D., & Wright, H. (1998), 'Intonation and dialogue context as constraints for speech recognition', *Language and Speech*, 41, 493–512.

Traum, D. R. (1994), 'A computational theory of grounding in natural language conversation', Ph.D. 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 A. Rao & M. Wooldridge (eds), *Foundations of Rational Agency*, Kluwer, Dordrecht.

Traum, D. R. & Allen, J. F. (1992), 'A speech acts approach to grounding in conversation', in *Proceedings 2nd International Conference on Spoken Language Processing (ICSLP-92)*, 137–40.

Traum, D. R. & Heeman, P. (1997), 'Utterance units in spoken dialogue', in E. Maier, M. Mast & S. Luperfoy (eds), *Dialogue Processing in Spoken Language Systems: ECAI-96 Workshop*, Lecture Notes in Artificial Intelligence, 125–40. Springer-Verlag, Heidelberg.

Traum, D. R. & Hinkelman, E. A. (1992), 'Conversation acts in task-oriented spoken dialogue', *Computational Intelligence*, 8, 3, 575–99. Special Issue on non-literal language.

van Vark, R., de Vreught, J., & 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 P. R. Cohen, J. Morgan & M. E. Pollack (eds), *Intentions in Communication*, MIT Press, Cambridge, MA.

Vanderveken, D. (1990/1991), *Meaning and Speech Acts*, Cambridge University Press, Cambridge.

Winograd, T. & Flores, F. (1986), *Understanding Computers and Cognition*, Addison-Wesley. Redding, MA.

Wright, H., Poesio, M., & Isard, S. (1999), 'Automatic extraction of game structure for dialogue act recognition using prosodic features', in *Proc. of the ESCA Workshop on Dialogue and Prosody*, Eindhoven.