Conversational Games Within Dialogue*

Jacqueline C. Kowtko†and Stephen D. Isard
Human Communication Research Centre, Edinburgh‡
Gwyneth M. Doherty
Human Communication Research Centre, Glasgow§

November 12, 1993

Abstract

An analysis of task oriented dialogue has been developed around goal-directed exchanges labelled *conversational games*. Games account for that aspect of discourse coherence that is manifested in initiation–response–feedback patterns, and they do so by relating the form of dialogue to underlying non-linguistic goals. The system applies to real, spontaneous dialogue and may prove useful in the design of computer dialogue systems. A study was conducted to determine agreement among analysts, upon coding two sets of dialogues. One expert and three novice analysts agreed on an average of 78% of the moves assigned. When certain consistent differences are taken into account, however, the accuracy increases to 88%.

1 Introduction

We have developed a system of discourse analysis which structures conversations on two functional levels. The level of *conversational games* is associated with mutually understood conversational goals such as obtaining information or getting one's conversational partner to perform an action. Games are analyzed as sequences of conventional *moves*—opening, closing, etc.—including the possibility of nested subgames. One ultimate purpose in performing this sort of analysis is to test a theory of intonational pragmatics in which the role of an intonational tune is to signal the sort of move being made in a conversational game. Another is to provide an objective description of the verbal channel in order to study non-verbal communication. A prerequisite to carrying through such a programme is to establish the game and move structure on a firm footing, which is the subject of the present paper.

Although we designed the analysis on the basis of dialogues arising from one particular task—the map task as described below—it is intended to apply to other types of task-oriented dialogue, and we report an experiment in which the system is applied to dialogues

*An earlier version of this paper appears in the Proceedings of the ESPRIT Workshop on Discourse Coherence, University of Edinburgh, 4-6 April 1991.

†The writing of this paper was partially supported by a UK Overseas Research Student Award.

‡HCRC, University of Edinburgh, 2 Buccleuch Place, Edinburgh EH8 9LW SCOTLAND

§HCRC, University of Glasgow, 62 Hillhead Street, Glasgow G12 8RT SCOTLAND

about a different task. The games and moves into which we analyze map task dialogues presumably do not exhaust the speaker's repertoires. Our observations suggest that they are sufficient in this particular setting, but a different task might well call different games into play. For instance, none of our participants felt the need for the game very common in classroom situations where a question is asked not in order to learn the answer, but to find out whether the other person knows it. Our supposition is that an individual's game repertoire—like their vocabulary—is finite at any given moment, but that new games can be learned in order to cope with new social situations. Two members of the same speech community can be expected to have a substantial overlap in their repertoires of games.

Moves in our analysis have functional definitions. To some extent function is signalled by form, but in our analysis we are not making assumptions about the form to function correspondence. It is, however, possible to examine the correspondence between form and function, which is what we intend to do later, with intonation.

2 Task-Oriented Dialogue for Analysis

Task-oriented dialogue has two particular advantages for our purposes. Firstly, turns are fairly short, as each party checks that the communication has been understood and actions properly performed. Secondly, when conversants are working to solve a problem, their motivations in conversing become clearer to an outside observer. Since our analysis involves imputing intentions to speakers, it is easier to apply the analysis to dialogue where motives for speaking are more salient to the analyst.

The task used to elicit dialogue is one in which two participants are given similar, though slightly different maps, one with a path marked on it and one without (see Boyle, 1990, and Anderson *et al.*, 1991, for a description of the HCRC map task corpus). The person with the path is designated information giver, and the other, information follower. The participants sit on opposite sides of a screen and cannot see each other's maps. The information giver tells the follower how to draw the path on their map. Since the maps differ in terms of location and presence of some landmarks, the conversants sometimes encounter communication obstacles and must discuss the problem to solve it. In fostering misunderstanding, this task produces interaction similar to that in real life conversation as well as in a potential "conversation" with a machine, in which problems might occur.

This type of task invites conversants to produce spontaneous, natural sounding speech. Since the follower normally draws a reasonably accurate path, we also know that communication has been effective.

The system of game analysis was largely developed on map task dialogues, but consideration was given to another type of dialogue: maze game dialogues (e.g. Garrod and Anderson, 1987). For the maze task, the two participants sit in adjoining rooms, each in front of a computer screen displaying a diagram of a maze, and they communicate with each other via microphones and headphones. The game requires each participant to move a symbol through the maze from some start position to some goal position. If a player becomes blocked by barriers, s/he must enlist the help of the other player to break free again. An interesting difference between the two tasks is that while in the map task one person instructs another, the players in the maze task do not have specific roles imposed upon them.

3 The Dialogue Analysis

A game consists of the turns which are necessary to accomplish one conversational sub-goal. For example, an *Instructing* game begins when an instruction is given and ends when the follower indicates that the instruction is either finished or well on its way to being completed, without any other communication being necessary to do so. A game typically proceeds smoothly from initial move to final move; however, two types of variation may occur: nesting or a break within a game. Nesting is considered to occur when a sub-game is initiated whose purpose can clearly be seen as contributing to the goal of the current game, e.g. when a follower asks for extra information in order to carry out an instruction. A break, such as the announcement of a misunderstanding, puts the status of the current game in doubt; the game might continue if the misunderstanding is cleared up, or it might be abandoned.

A move is an utterance, a partial utterance or a group of utterances which convey the same, specific intent. For instance, a complicated description of where a landmark lies on a map could require several grammatically complete sentences. On the other hand a move such as an acknowledgement may consist of the brief utterance "Right." We define moves by their function and do not have specific requirements about their form.

In describing utterances in terms of moves and games, we ascribe intentions with respect to local exchanges of information and accomplishment of general goals within the overall task. Since we cannot read the conversants' minds, we can have no guarantee that our description is correct. The best we can hope for is agreement on the part of several observers as to their reading of what a given speaker was trying to do.

4 Historical Background

Our game and move framework is based on Power (1979) and Houghton (1986; Houghton and Isard, 1987). The purpose of this earlier work was to produce a theory of how non-linguistic goals give rise to conversation. In both cases, the theory was presented in the form of an AI model, in which a pair of "robots" conversed in order to achieve simple goals which neither was capable of achieving on its own. In Houghton's model, conversation was integrated with the rest of the robots' capacities for interacting with their world. The robots knew that successful conversational games could result in the transfer of information or in their partner performing some non-linguistic act. So a robot wanting to know whether a door was bolted could consider having a look, giving a push, asking Fred, or even asking Fred to give a push as alternative means of finding out. Here "asking Fred" and "asking Fred to give a push" are used as shorthand for "engaging Fred in the FIND_OUT conversational game" and "engaging Fred in the GET_DONE conversational game." The actual asking is only the first move in the relevant game. The significance of seeing it as part of a game lies in the askee knowing what sort of response is required, in the asker knowing what to make of the response, and in both players knowing that getting the response is the point of starting the game in the first place.

The robot conversations generated in this way were sensible and coherent, if somewhat stilted. The question we are addressing is whether such a game analysis—possibly with an extended set of games—can be imposed on spontaneous conversations.

Sinclair and Coulthard (1975) represents a previous attempt at applying a broadly similar analysis scheme to classroom dialogues. Our games are approximately equivalent to their exchanges. Two important differences are that Sinclair and Coulthard do not characterize their exchanges in terms of goals and intentions, and they have a theoretical bias against nesting one exchange inside another. We view such nesting as a natural reflection of a recursive planning structure with goals and subgoals.

5 The System of Dialogue Analysis

We have based our analysis upon a set of several map task dialogues, averaging 10 minutes each in length. They were selected at random from the dialogues (in the HCRC database) involving pairs of participants who had not previously met. We avoided dialogues between conversants familiar with each other because they tend to interrupt each other more often—making transcription and analysis more difficult—and sometimes stray from the task topic (Anderson, personal communication).

We began with Houghton's (1986) repertoire of four games and added further ones to account for frequent patterns of exchange that did not fit the original four.

5.1 Initiating Moves of Games

Based upon data from the dialogues we have coded, our analysis differentiates between six types of games and twelve different moves. Six of the classes of moves are the initiating moves of games. They are as follows:

INSTRUCT Communicates a direct or indirect request or instruction, to be done immediately or shortly. e.g. "You then go down two inches."

CHECK Checks self-understanding of a previous message or instruction by requesting confirmation directly or indirectly; makes sure that a complicated instruction is understood. e.g. "So you want me to go down two inches?"

QUERY-YN, QUERY-W Yes-No question (QUERY-YN) or open-answer Wh-question (QUERY-W); asks for new or unknown detail about some part of task; does not request clarification about instructions (that would be a CHECK); e.g. "Do you have a rockfall?"

EXPLAIN Describes status quo or position in task with respect to the goal; freely offered, not elicited; provides new information. e.g. "I've got a cairn."

ALIGN Checks the other participant's understanding or accomplishment of a goal; elicits a positive response which closes a larger game; checks alignment of both participants' plans or position in task with respect to goal; checks attention, agreement, or readiness. e.g. "Ok?", meaning, "Are you with me?"

Most games consist of two or more moves, which may in turn initiate sub-games. Some games contain additional feedback and response moves which loop until both participants agree that the goal has been satisfied and the game is thus completed. Sometimes whole games nest within other games. Often such nesting occurs when one of the participants decides that more information needs to be exchanged before the initial move can receive a fair response. This may take the form of a CHECK to see if an instruction was indeed understood, a QUERY if more information is needed, or an EXPLAIN to communicate the status quo.

Since the map task involves giving a series of instructions, the resulting dialogue contains a series of *Instructing* games, with an occasional *Querying*, *Checking*, or *Explaining* game in between.

5.2 Other Moves

The following six moves serve as response and feedback moves within a game:

CLARIFY Clarifies or rephrases what has previously been said; usually repeats given or known information; elicited by other person. e.g. "South, two inches."

REPLY-Y, REPLY-N Affirmative (REPLY-Y) or negative (REPLY-N), elicited response to QUERY-YN, CHECK, or ALIGN; also indicates agreement, disagreement, or denial; e.g. "Yes, I have."

REPLY-W An elicited reply to QUERY-W or CHECK; can be a response to QUERY-YN that is not easily categorizeable as positive or negative (REPLY-Y/N). e.g. "Down."

ACKNOWLEDGE Vocal acknowledgement of having heard and understood; not specifically elicited but often expected before the other speaker will continue; announces readiness to hear next move—in essence a request to 'please continue'; may close a game. e.g. "All right" or "Oh right, I see what you mean."

READY Indicates intention to begin a new game and focuses attention on oneself, in preparation for the new move; an acknowledgement that the previous game has just been completed, or leaving the previous level or game; consists of a cue-word. e.g. "Now" or "Right."

In conjunction with move classes, we include in our analysis a set of features which may be attached to the labels for each conversational move. They serve not as an integral part of the analysis, but to provide additional information to aid different types of analyses. The features specify whether a move is continued, abandoned, mumbled, repeated, interjected, or at a meta-conversational level.

In our coding we have assigned some moves dual function—such as answering a question and also providing unelicited information.

6 Examples of Games Structure in Conversation

Map task dialogues typically consist of a series of *Instructing* games, although a small percentage begin with *Explaining* games in which each participant compares landmarks on the maps. The following excerpt from dialogue NAQ3C8 shows the structure of moves and games. It consists of one large *Instructing* game with three nested games: *Explaining*, *Querying-yn*, and *Checking*. (Angled brackets below indicate overlapped speech; double vertical lines indicate the boundary of a move; *g and *f are for instruction giver and instruction follower, respectively.)

I-begin	*g	Right, em, go to your right towards the carpenter's house.
		INSTRUCT
	*f	Alright ‖ well I'll need to go below. I've got a blacksmith marked.
		ACKNOWLEDGE
E-begin		EXPLAIN
	*g	Right, well you do that.
E-end		ACKNOWLEDGE
Q-begin	*f	Do you want it to go below the carpenter?
		QUERY-YN
	*g	< No, I want you to go up the left hand side of it towards /
		REPLY-N
	*f	Ok.
		ACKNOWLEDGE
	*g	Green Bay and make it a slightly diagonal line, towards, em sloping to the right.>
Q-end		REPLY-N–continued
C-begin	*f	So you want me to go above the carpenter?
		CHECK
	*g	Uh-huh.
C-end		REPLY-Y
	*f	Right.
I-end		ACKNOWLEDGE

The next example, shows an *Aligning* game following the point where the information giver thinks the requested action has been completed and it is safe to ask if the hearer's path and location in the dialogue are up to date, expecting an affirmative response.

Q-begin	*f	How f-...How far up? QUERY-W
	*g	< Ehm, about six centimetres till you're about, ah, till you're about ehm, about the same distance away from... you're about five centimetres...six centimetres from the top of the page / REPLY-W
	*f	Right. O.K. ACKNOWLEDGE
	*g	now.> REPLY-W–continued
I-end	*f	O.K. ACKNOWLEDGE
A-begin	*g	Right,‖ you're to the left hand side of the page. READY, ALIGN
A-end	*f	Yes. REPLY-Y

7 Assessment of the Analysis System: Experiment

We conducted a study to determine the ease and accuracy of of applying our coding system to labelling conversational moves in two different types of task-oriented dialogue.

7.1 Subjects

The four subjects of the experiment were psychology undergraduates, two males and two females, who had no experience of coding dialogues or of the two types of task-oriented dialogue which were used. The expert coders had been using the Game Analysis for about eleven months to code dialogues.

7.2 Materials

The materials used were five task-oriented dialogues. Two of these dialogues were taken from the HCRC map task corpus and three were from a corpus of maze game dialogues (see descriptions in Section 2).

Here is a brief excerpt from a coded maze dialogue:

E-begin	*A	i'm in E one moving to E two and my goal's in D four EXPLAIN
	*B	ok
E-end		ACKNOWLEDGE
E-begin		and my goal's in E one and i'm in C four at the moment i'm going into B four because the other two have got barricades on them EXPLAIN
	*A	right ok
E-end		ACKNOWLEDGE

7.3 Procedure

Since the subjects had no knowledge of discourse analysis, they were given literature to read, including an explanation of the game coding system. The subjects also had several tutorials on the conversational game analysis with the expert coder and were supplied with written instructions on how to classify conversational moves.

The subjects were each given copies of the dialogue transcripts with accompanying audio tapes. To avoid confusions over move boundaries, the expert coder segmented the map dialogues into moves on the transcripts before the subjects coded. The maze dialogues have much shorter turns, and subjects were left to decide on their own where to place the few within-turn boundaries in each maze dialogue. Subjects were asked to assign move boundaries to another, unmarked, map task dialogue so that we could measure agreement between both move classification and decisions of boundary placement within the map dialogues. Subjects were instructed to assign one label to each designated move. No mention was made that some moves may have a dual function.

Although the subjects were coding the same dialogues they did so independently of one another. While they were coding, the subjects were given the opportunity to question the expert coder if they felt they did not understand anything about the coding system. The subjects spent several hours coding over a period of about four weeks, coming to the expert coder with any non-specific problems that they had, usually concerning their understanding of the coding instructions they had been given.

Once they had completed this task, the subjects were given copies of the transcripts with the expert's coding. They then compared their own coding with the expert's coding and represented this on confusion matrices. The measure of consistency used was the percentage of moves in a dialogue on which the subject and expert agreed on classification.

7.4 Results

One of the male subjects made changes to some of his classifications having seen the expert coding and was therefore not included in these results.

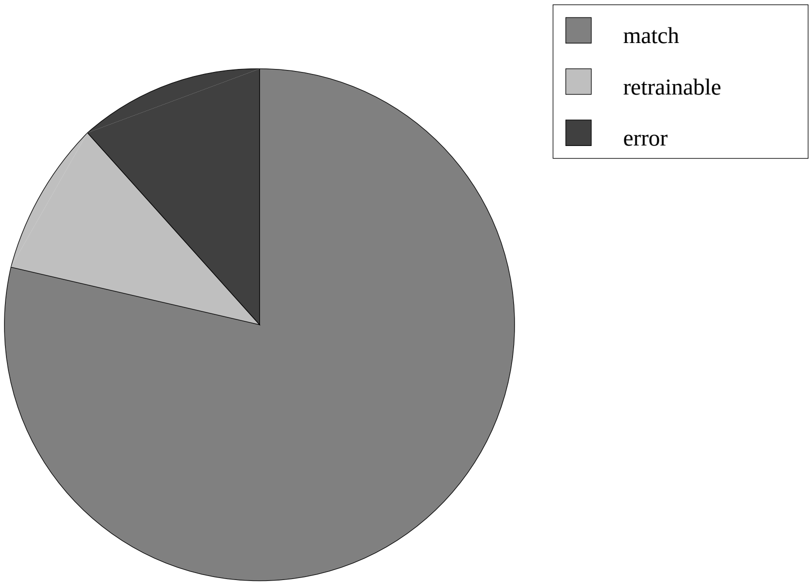

Figure 1: Distribution of Classification Error

The percentage agreement on moves for the three remaining subjects for both tasks is shown in Table 1. The mean consistency score for the three subjects on all the dialogues was 78%.

Table 1: Percentage of Move Classification Agreement

Dialogue:	Map 1	Map 2	Maze 1	Maze 2	Maze 3
Subject					
1	84	86	77	70	75
2	86	81	72	73	85
3	81	77	72	69	77

All instances of erroneous coding on the part of the novices (defined as mismatch with expert coding) were extracted and examined. These errors were classified in terms of which types of moves the novices confused with which type of expert coding. For example an expert's REPLY-Y move was confused by a novice as an ACKNOWLEDGE move. Twenty different types of mismatch were found to occur; however, some were very infrequent overall: for example, expert CLARIFY confused with novice REPLY-W accounted for only 0.2% of the errors. Figure 1 illustrates the distribution of match and mismatch found.

Some of the mismatch types (those in Figure 1) were classified as "retrainable". These accounted for 44% of the total error. Retrainable mismatches were most likely due to the inexperience of the novice coders with the coding system and their misinterpretation of some of the coding instructions they were given. This view is supported by the self-reports of the novice coders after the exercise was completed and various mistakes were pointed out. Also, it is easy to see possible causes of the errors in the retrainable cases. See Table 2 for a description of these possible causes for each retrainable mismatch type.

Table 2: Plausible Cause for Each Retrainable Mismatch Type

Mismatch Type		Plausible Cause
Expert	*Novice*	
REPLY-Y	ACKNOWLEDGE	Novice misinterpreted an elicited response as an unelicited response.
REPLY-W	EXPLAIN	Novice misinterpreted an elicited response as an unelicited information-giving move.
CHECK	QUERY-YN	Novice has not realised that an interrogative was about known information or does not realise that this is the distinction between QUERY-YN and CHECK.
ACKNOWLEDGE	CLARIFY	Novice has misunderstood the definition of CLARIFY. The function of clarifying or rephrasing has been confused with an acknowledgement of information in which the phrase is repeated.
CHECK	CLARIFY	Novice has misinterpreted a question about information from another speaker as a clarification of that information. A conversant only clarifies information they have previously given.
QUERY-YN	QUERY-W	Novice has not realised that a question requires only affirmation or negation, or does not appreciate this distinction between this and a more open content ('wh') question. Often the syntactic form of a question can cause this confusion.
INSTRUCT	EXPLAIN	Novice has not realised that the function of a move was to instruct a co-participant in an action. Again, the form of an utterance is confused with its function; it looks like an offering of information rather than an instruction.
REPLY-Y	CLARIFY	Novice misinterpreted an affirmative response to a question as a clarification of previous information. It would only be a clarification if the affirmative response was a response to a question about a message given by the clarifier.
ALIGN	QUERY-YN	Novice was again misled by form. ALIGN describes moves which have the function of obtaining feedback about a listener's knowledge or accomplishment of some task. A QUERY-YN has the function of eliciting new information. The novice has not appreciated this distinction.
INSTRUCT	QUERY-YN	Novice has overlooked the INSTRUCT function of a move because its form was interrogative.

Each of these mismatches can be associated with a specific misinterpretation of instructions. For example, the mismatch which accounted for the greatest proportion of error was when a move was coded as a REPLY-W by the expert and by the novice as EXPLAIN. In such a case, the novice had overlooked the fact that the move had been elicited by the other dialogue partner and was not a spontaneous information-giving move.

Therefore, it can be seen that many of the mistakes made by novice coders could have been avoided given two conditions. Firstly, a greater appreciation of some of the finer distinctions between move categories would be required. This would be possible with clarification of coding instructions. Secondly, practice at considering move content, form and context simultaneously would also improve the coders' categorisation of move function.

Results of move boundary assignment for one map task dialogue appear in Table 3. In this dialogue (NAQ2C2), the subjects both inserted more move boundaries and disagreed with the expert's boundary placement. The mean agreement between novice and expert is 88%.

Table 3: Percentage of Move Boundary Agreement

Subject	Map 3
1	88
2	89
3	86

8 Discussion

There are ten types of disagreement in move classification which it is predicted will be eliminated by clarifying instructions and allowing subjects more practice in coding (the items in Table 2). The exclusion of this "retrainable" error brings overall coding agreement from 78% to 88%. The percentage of error will decrease, as the formerly novice coder practices and becomes an expert. Some of the remaining error arises from mistakes, some may be due to weakness in the coding system itself, and some derives from the ambiguity inherent in assessing a speaker's intent.

The relatively high percentage of agreement on move boundaries, averaging 88%, shows that subjects understand the concept of a move in addition to performing well on move classification.

9 Conclusion

We have provided a framework in which the function of an utterance can be referred to as move X within game Y. This facilitates a systematic comparison of function to form in a particular move when occurring at a certain place in a specific game, for example the first move in game X or the last Y move in game Z. In particular, we have begun investigations into the relationship between function described in this way and intonational form (see Kowtko, 1992).

10 References

Anderson, A. H. (pers. comm.) Human Communication Research Centre and Department of Psychology, University of Glasgow.

Anderson, A. H., M. Bader, E. G. Bard, E. Boyle, G. Doherty, S. Garrod, S. Isard, J. Kowtko, J. McAllister, J. Miller, C. Sotillo, H. Thompson, and R. Weinert (1991) "The HCRC Map Task Corpus." *Language and Speech*, **34**(4): 351-366.

Boyle, E. A. (1990) "User's Guide to the HCRC Dialogue Database," Technical Report HCRC/TR-11, September, 1990, Human Communication Research Centre, University of Edinburgh.

Garrod, S. & A. Anderson (1987) "Saying what you mean in dialogue: A study in conceptual and semantic coordination," *Cognition*, **27**, 181-218.

Houghton, G. (1986) *The Production of Language in Dialogue: A Computational Model.* PhD thesis, University of Sussex.

Houghton, G. & S. D. Isard (1987) "Why to speak, what to say, and how to say it," in P. Morris (ed.), *Modelling Cognition*, Wiley, 249-267.

Kowtko, J. C. (1992) "On the intonation of mono- and di-syllabic words within the discourse framework of conversational games," in *Proceedings of the 30th Annual Meeting of the Association for Computational Linguistics*, Delaware.

Power, R. (1979) "The organisation of purposeful dialogues," *Linguistics*, **17**, 107-152.

Sinclair, J. & M. Coulthard (1975) *Towards an Analysis of Discourse: The English used by teachers and pupils.* Oxford University Press.