Toward a Semantics for a Speech Act Based Agent Communications Language*

Ira A. Smith
Center for Human-Computer Communication
Department of Computer Science and Engineering
Oregon Graduate Institute

Philip R. Cohen
Center for Human-Computer Communication
Department of Computer Science and Engineering
Oregon Graduate Institute

September 18, 1995

Abstract

Implementations of systems based on distributed agent architectures require an agent communications language. The language must have a clearly defined semantics; without such a semantics, no agent can be sure what another agent's commitment to perform a task means. In this context, agent to agent communications can be viewed as a task oriented dialogue for building, maintaining and disbanding teams. A speech act model based on joint intentions can provide the means for agents to build maintain and disband teams. It can also not only describe the outward appearance of a task oriented dialogue, but can also provide motivation for the step by step structure of the dialogue. In this paper we begin to develop a semantics for an agent communications language based on the interactive abilities of agents and use these semantics to analyze the Winograd & Flores *basic conversation for action.*

1 Introduction

In this paper we begin to develop a semantics for an agent communications language based on the interactive abilities of agents and use these semantics to analyze the Winograd & Flores *basic conversation for action.* Since the language must provide operators corresponding to agent capabilities, its semantics must be based on an underlying theory of agent interaction. We use the Cohen – Levesque theory of Joint Intentions as the underlieing theory for the description of agent behavior. This work is relevant to ongoing agent communication language development activities [FFMM94, LF94, BDBW95].

*This research was supported by a grant from the New Energy Development Organization (Japan) and by the Advanced Research Projects Agency (contract number DABT63-95-C-0007) of the Department of Defense. The results presented here do not reflect the position or policy of the either the Japanese Government or the US Government.

2 The Cohen – Levesque Theory of Joint Intentions

This theory describes agent behavior in terms of characterizations of an agent's internal state. Cohen and Levesque present the theory in a series of papers written between 1990 and 1995 [CL90b, CL90c, CL90a, CL91a, Coh94]. A fundamental point of Cohen and Levesque's theory, with respect to teamwork, is that joint action is more than just a coordination of the actions of the individual team members. In addition to coordination of actions, there must also be a coordination of the agents' mental states. The team is formed by each agent having a joint intention with the other agents forming the team with respect to a specific goal. This concept of a joint intention, the heart of a team, is defined in terms of the individual intentions of the team members. Individual intentions are formed from the knowledge, beliefs, and goals of the individual agent. In this paper we are only concerned with *achievement goals* — goals yet to be achieved.

Agents can hold goals with varing degrees of commitment. One degree of commitment is a *persistent goal.* An agent with a **PGOAL** of p will not give up p until either the goal has been achieved (and become a maintenance goal), or until the agent knows the goal will never be true.

Definition 1 *Persistent Goal*
$(\textbf{PGOAL}\ x\ p) \triangleq (\textbf{AGOAL}\ x\ p) \wedge (\textbf{BEFORE}\ ((\textbf{BEL}\ x\ p) \vee (\textbf{BEL}\ x\ \Box\neg p))\ \neg(\textbf{AGOAL}\ x\ p))$

An *Intention* is a particular kind of **PGOAL** .

Definition 2 $(\textbf{INTEND}\ x\ a) \triangleq (\textbf{PGOAL}\ x\ (\textbf{DONE}\ x\ (\textbf{BEL}\ x\ (\textbf{HAPPENS}\ a))?;a))$

The agent has a **PGOAL** not only to have **DONE** a, but also to believe he is about to do a, and then do it[1]. The additional requirement, to believe he is about to do a, forces x to pursue an active strategy to achieve a. An agent cannot be committed simply to wait for a to be true.

PGOAL , and hence **INTEND**, represent a fanatical commitment to a goal. It is beneficial to allow agents to have a commitment to a goal less than fanatical than a **PGOAL** but more binding than a simple **AGOAL** .

Definition 3 *Relativized Persistent Goal*

$$(\textbf{RGOAL}\ \mathrm{x\ p\ q}) \triangleq (\textbf{AGOAL}\ \mathrm{x\ p}) \wedge (\textbf{BEFORE}\ [\ (\textbf{BEL}\ \mathrm{x\ p}) \vee (\textbf{BEL}\ \mathrm{x}\ \Box\neg \mathrm{p}) \vee (\textbf{BEL}\ \mathrm{x}\ \neg \mathrm{q})]\ \neg(\textbf{GOAL}\ \mathrm{x}\ (\textbf{LATER}\ \mathrm{p})))$$

A *relativized persistent goal*, **RGOAL** , is a **PGOAL** held relative to a qualifying proposition q. If the agent discovers q is no longer true he will drop the goal. The theory deliberately places no restrictions on the nature of q. A new version of *Intend* is defined in terms of **RGOAL** .

Definition 4 *Relativized Intention*
$(\textbf{INT}\ x\ a\ p) \triangleq (\textbf{RGOAL}\ x\ (\textbf{DONE}\ x\ (\textbf{BEL}\ x\ (\textbf{HAPPENS}\ x\ a))?;a)\ p)$

The revised definition allows an agent to have intentions that are contingent upon other goals, environmental factors or the goals of other agents. Should the relativizing condition of one of his intentions embedded **RGOAL** become impossible, the agent will be able to drop that intention.

[1] In [CL90a] two types of **INTEND** are defined. The other intend takes a proposition as its second argument. In this case the agent is committed to performing a sequence of events, after which the proposition will be true.

3 TEAMS

A team is a set of agents having a shared goal and a shared mental state. Both the goal and the mental state are required for a team to exist. Without both, there is no unified activity and hence no team. Cohen-Levesque [LCN90, CL91b] characterizes the shared mental states of the team members through the definitions of a *mutual goal, weak achievement goal, weak mutual goal, joint persistent goal* and *joint-intention.* These mental states are the multi-individual equivalents of individual goals, persistent goals, etc.

A *mutual goal* is a generalization of an individual goal; that is, each agent has the goal and there is the mutual belief (**MB**) that each has the goal.

Definition 5 *Mutual Goal*
$(\boldsymbol{MG}\ x\ y\ p) \triangleq (\boldsymbol{MB}\ x\ y\ (\boldsymbol{GOAL}\ x \Diamond p) \wedge (\boldsymbol{GOAL}\ y \Diamond p)$

While mutual goals are necessary to have a team, they are not quite enough. A problem arises if one team member decides to drop the goal, he may have found out the goal has achieved or impossible. If the team has a mutual goal, the member is under no obligation to inform other team members of his change in mental state, thereby leaving the remainder of the team to continue in a now futile effort. We need a mental state that will require the team member to communicate with the rest of the team, allowing them to drop the goal as well.

Definition 6 *Weak Achievement Goal*

$$\begin{aligned}(\mathbf{WAG}\ \mathrm{x\ y\ p\ q}) \triangleq & [\neg(\mathbf{BEL}\ \mathrm{x\ p}) \wedge (\mathbf{GOAL}\ \mathrm{x} \Diamond \mathrm{p})] \vee \\ & [(\mathbf{BEL}\ \mathrm{x\ p}) \wedge (\mathbf{GOAL}\ \mathrm{x} \Diamond (\mathbf{MB}\ \mathrm{x\ y\ p}))] \vee \\ & [(\mathbf{BEL}\ \mathrm{x} \Box \neg \mathrm{p}) \wedge (\mathbf{GOAL}\ \mathrm{x} \Diamond (\mathbf{MB}\ \mathrm{x\ y} \Box \neg \mathrm{p}))] \vee \\ & [(\mathbf{BEL}\ \mathrm{x} \neg \mathrm{q}) \wedge (\mathbf{GOAL}\ \mathrm{x} \Diamond (\mathbf{MB}\ \mathrm{x\ y} \neg \mathrm{q}))]\end{aligned}$$

A weak achievement goal (**WAG**) is a mental state of each member of a team. The team member either believes p doesn't currently hold, and has as a goal that p be true in the future; or upon coming to believe that either p currently holds or that p will never be true or that q is not true, is left with the residual goal of establishing the same belief in the other members of the team. A *weak mutual goal* is a generalization of a mutual goal, substituting a **WAG** for the **GOAL** in the original definition.

Definition 7 *Weak Mutual Goal:* $(\boldsymbol{WMG}\ x\ y\ p) \triangleq (\boldsymbol{MB}\ x\ y\ (\boldsymbol{WAG}\ x\ y\ p) \wedge (\boldsymbol{WAG}\ y\ x\ p))$

We can now define a *joint persistent goal.*

Definition 8 *Joint Persistent Goal*

$$\begin{aligned}(\mathbf{JPG}\ \mathrm{x\ y\ p\ q}) \triangleq & (\mathbf{MB}\ \mathrm{x\ y} \neg \mathrm{p}) \wedge (\mathbf{MG}\ \mathrm{x\ y\ p}) \wedge \\ & (\mathbf{BEFORE}\ [(\mathbf{MB}\ \mathrm{x\ y\ p}) \vee (\mathbf{MB}\ \mathrm{x\ y} \Box \neg \mathrm{p}) \vee (\mathbf{MB}\ \mathrm{x\ y} \neg \mathrm{q})]\ (\mathbf{WMG}\ \mathrm{x\ y\ p}))\end{aligned}$$

The characteristics embedded in the **WAG** and **JPG** are necessary to allow teams to have a coherence. The last three clauses in the **WAG** definition provide a way to balance the responsibilities an individual has towards the team with the requirement that an individual team member be allowed to drop a goal under certain reasonable conditions [CL91b]. If an agent discovers the goal has been accomplished or is impossible, or if he discovers the relativizing condition is no longer true, he is allowed to drop the goal. The agent is left with a goal to make his discovery mutually believed by the rest of the team.

4 Speech Acts

Speech acts will serve as the means agents use to communicate their mental states and to ask for the assistance of other agents in achieving their goals. By performing a particular speech act, an agent is deliberately attempting to alter the state of the world, therefore there is intention associated with the performance of the act. In general there is a specific result the act is intended to accomplish — its perlocutionary effect. This result is related to one or more of the agent's goals. Based on an assumption of reliable communication channels, the communications act will succeed in the sense that the addressee will receive it. However, since the speaking agent's goal is to alter the addressee agent's mental state in a particular way, there is no guarantee the act will achieve this ultimate goal. Therefore, in addition to the agent's actual goal, we specify a minimum achievement the act will accomplish — its illocutionary effect. A Speech act is defined as an attempt. The attempt operator requires four arguments: the agent, the act the agent will perform, the goal of the attempt, and the result the agent has committed to. The two final arguments are the agents ultimate goal and the minimum acceptable achievement. The following formalization of an attempt is from [CL90c].

Definition 9 *Attempt:*
$(\boldsymbol{ATT}\ x\ e\ p\ q) \stackrel{\triangle}{=} (\boldsymbol{BEL}\ x\ \neg p) \wedge (\boldsymbol{GOAL}\ x\ (\boldsymbol{HAPPENS}\ x\ e;p?)) \wedge (\boldsymbol{INTEND}\ x\ e;q?)$

The agent believes some condition p to not be true at the present time, and is going to do an action (e). The ultimate goal of the agent's act is to bring about some condition p, with the intention of achieving at least q. The agent has only a limited commitment to the ultimate goal (p). In contrast, the agent has an intention to achieve q. If she were to come to the conclusion that the attempt failed to achieve even this, we could predict the agent would reattempt; that is she would either perform e again or perform a similar action to achieve the same result.

4.1 Request

An agent will use the request speech act to attempt to recruit another agent to perform a task. Often this will be a task that fits as a subtask in the overall goal of the requesting agent.

Definition 10 *Request*
$(\boldsymbol{REQ}\ x\ y\ e\ a\ p) \stackrel{\triangle}{=} (\boldsymbol{ATT}\ x\ e\ \phi\ \psi)$
where ϕ *is:*
$\Diamond(\boldsymbol{DONE}\ y\ a) \wedge (\boldsymbol{INT}\ y\ a\ [\boldsymbol{WAG}\ x\ y\ (\boldsymbol{DONE}\ y\ a)\ p])$

and ψ *is :*
*(**BMB** y x (**WAG** x y [(**DONE** y a)* $\wedge$ *(**WAG** y x (**DONE** y a) (**WAG** x y a p))]))*

The goal (ϕ) of a request consists of two parts, the first is the straight-forward requirement that the addressee perform the requested act. The second part of the goal places a requirement on the addressee's mental state. The requested action is that y not only perform a, but perform it with respect to x's **WAG** that y do a (relative to p). The requesting agent is not only specifying that the addressee perform an act but that he perform it in a particular mental state. It is not enough that y do a, y must also intend to do a relative to x's goal that she do it. If y were to do a accidentally the act would not meet the goal of x's request because the requisite mental state would be absent.

The request may not fully succeed; the attempt may not achieve ϕ for the requester. The addressee may refuse the request for any number of reasons. The requested action may conflict (either directly or indirectly) with other of addressee's goals, the addressee may not have the resources to adopt the request due to prior commitments or the addressee may just arbitrarily refuse the request. Under any of these circumstances, the speaker is not obligated to perform the request again.

The minimum result x is committed to is ψ; that y believe x has a weak achievement goal that both she eventually do a, and she have a weak achievement goal to eventually do a. Should x come to believe that even this result has not been achieved, by the definition of an attempt we would expect x to redo the request.

Our definition of a team is based on the notion of a **JPG** (joint persistent goal); the **JPG** is defined in terms of mutual belief in the existence of each individual team member's **WAG**. Having publicly committed to the **WAG** the requester has informed the requestee that she has the **WAG**. Thus the requester has already made the individual commitments required for the formation of a team, this agent is already treating the requestee as a team member. By accepting this persistent goal, the requester is committing to acting as though there is a team to achieve the goal, and the requestee is a team member. This requirement forces the requesting agent to commit resources to team obligations, the agent receiving the request is under no such obligation. The requesting agent is expending resources to form a team, and is committed to a future expenditure of resources, however minimal, to maintain the team, this commitment is practical because the requesting agent is able to assume the requesting agent will notify him with either a confirm or a refuse. From the definition of a request we can prove the requester has a persistent goal to achieve a. Our chain of reasoning will be based on a sincerity assumption and on the definition of a weak achievement goal.

Proposition 1 *(**DONE** (**REQ** x y e a p))* $\Rightarrow$ *(**WAG** x y (**DONE** y a) p)*

Proof: By assumption x is sincere. Since x is committed to establish a belief by y that x has a **WAG** for y to do a, x must have the **WAG** that y do a.

Proposition 2 *(**DONE** (**REQ** x y e a p))* $\Rightarrow$ *(**PGOAL** x (**DONE** y a) p)*

Proof: By Proposition 1 x has a **WAG** of (**DONE** y a). The definition of a **WAG** requires x to either have a **PGOAL** of (**DONE** y a), to believe it to be already true or impossible or to believe the relativizing condition p to be false (in the latter three cases x has a **PGOAL** to notify y). Since a **REQ** is an *Attempt*, and x is a sincere agent, he cannot believe (**DONE** y a) to be either true or impossible. For the same reason he cannot believe p to be false.

4.2 Refuse

Receipt of a request does not commit the receiving agent to accept the request; it is possible to refuse — We make no assumption of helpfulness. A refusal is a notification to the requesting agent that this agent will not commit to the goal of the request.

Definition 11 Refuse

$(\boldsymbol{REFUSE}\ y\ x\ e\ a\ \psi) \stackrel{\triangle}{=} (\boldsymbol{ATT}\ y\ e\ \phi\ \phi)$
where ψ *is*
$\quad (\boldsymbol{WAG}\ x\ y\ (\boldsymbol{DONE}\ y\ a)\ p)$
and ϕ *is*
$\quad (\boldsymbol{BMB}\ x\ y\ \neg(\boldsymbol{WAG}\ y\ x\ (\boldsymbol{DONE}\ y\ a)\ \psi))$

In the definition of a refuse, the goal and the minimal result are the same. Unlike the request, where the requester is attempting to have the addressee take on a particular mental state, that of commitment to a future action that will have an associated cost, the refuse is simply an attempt by the speaker to make known to the addressee that she will not commit to any future action. Our assumption of reliable communication channels and the sincerity axiom imply the attempt will succeed, therefore the acts goal and minimal result are the same.

The definition of a refuse speech act does not imply y cannot undertake a; but only that y will not commit to a with respect to x's **WAG** . If y were to achieve a by accident or as a by product of some other goal it is pursuing (either independently or in conjunction with other agents); her actions would be consistent with her refusal. This direct consequence of the requester's goal may seem a meaningless distinction has practical consequences. If y were to achieve a after refusing the **WAG** , she would be under no requirement to notify x, hence x might never find out. For x to gain what he wants, a team to form as a result of the request, y would have to accept the **WAG** .

The main effect of a refuse is to notify the requesting agent that the requested agent can not be counted upon as a means to achieve the requested action. A result of a refuse is the requesting agent is freed of the team obligations with respect to the requestee that were incurred by making the original request. These obligations take the form of the **PGOAL** that was embedded in the requester's **WAG** .

Theorem 1

$$\begin{aligned}\models \textbf{HAPPENED}\ [&(\textbf{REQ}\ \mathrm{x\ y\ e\ a\ p}));\\ &(\textbf{REFUSE}\ \mathrm{y\ x\ e\ a}\ (\textbf{WAG}\ \mathrm{x\ y}\ (\textbf{DONE}\ \mathrm{y\ a})\ \mathrm{p}))] \Rightarrow\\ \neg(&\textbf{PGOAL}\ \mathrm{x}\ (\textbf{DONE}\ \mathrm{y\ a})\ \mathrm{p})\end{aligned}$$

y's refusal tells x there is no team, and frees him from any commitments toward y with respect to the original request. The refusal would have no effect on x's **PGOAL** of a, if x has such a **PGOAL** that is independent of his goal of (**DONE** y a). If this is the case, we would expect x to continue to pursue the achievement of a by some other means than that of y's cooperation.

4.3 Confirm

The confirm speech act is used by an agent to notify a requesting agent that she is accepting the weak achievement goal in the request. By accepting the requesters **WAG** , the speaker is committed to do a and is also committed to the other obligations of team membership.

Definition 12 *Confirm*

(**CNFM** y x e a (**WAG** x y (**DONE** y a) p)) $\triangleq$
[**ATT** y e (**BMB** x y (**WAG** y x (**DONE** y a) (**WAG** x y (**DONE** y a) p)))
(**BMB** x y (**WAG** y x (**DONE** y a) (**WAG** x y (**DONE** y a) p)))]

In our definition of the Confirm speech act, the ultimate goal and the minimum acceptable goal of the underlieing attempt are the same. As was the case with Refuse we can define an attempt in this manner because of the assumptions we made about the agents and their environments. The existence of reliable communications channels means the message will get to the listener and be understood. The speaker is sincere, the addressee knows the speaker is sincere, hence the addressee will believe the speaker.

A set of propositions that are analogous to those of the request speech act hold for confirm.

Proposition 3

(**DONE** (**CNFM** y x e a (**WAG** x y (**DONE** y a) p))) $\Rightarrow$
(**WAG** y x (**DONE** y a) p)

Proposition 4

(**DONE** (**CNFM** y x e a (**WAG** x y (**DONE** y a) p))) $\Rightarrow$
(**PGOAL** y (**DONE** y a) p)

I have omitted proofs for these propositions as they are analogous to those of propositions 1 and 2.

4.4 Assert

An assert speech act is used by an agent to inform the addressee that she should believe a particular state of affairs to be true.

Definition 13 *Assert*
ASSERT *(y x e q)* $\triangleq$*[**ATT** y e (**BMB** x y q) (**BMB** x y (**BEL** y q))]*

The speaker's goal is for the addressee to believe that q is true. Performing the attempt does not guarantee this goal will be achieved; for example the addressee could have access to information (that the speaker doesn't have) indicating the speaker is wrong, in this case the addressee will continue to believe that q does not hold. The minimum acceptable result for the speaker is that the addressee believe the speaker believes that q is true. The speaker is committed to making the addressee believe a fact about his mental state. Our assumption of reliable communication channels guarantees the achievement of this result.

By performing an Assert, the speaker is claiming to believe the truth of q. The definition and the sincerity axiom are enough to prove the following proposition.

Proposition 5 $\models$*(**HAPPENED** (**ASSERT** y x e q)) $\Rightarrow$ [(**BEL** y q)?]*

5 Building and Disbanding Teams

Now that we have defined a semantics for the speech acts, we will show how these acts are used to create and dissolve teams. Under normal circumstances, a request followed by a confirm will establish a joint persistent goal between x and y, relative to p, to achieve a. The under normal circumstances caveat is necessary because we make an implicit assumption that no event occurs between the the **REQ** and the **CNFM** that either makes the goal of the **JPG** true or impossible, or makes the relativizing condition false.

Theorem 2 $\models$ *[**HAPPENED** (**REQ** x y e a p);*

*(**CNFM** y x e1 a (**WAG** y x (**DONE** y a) p))] $\Rightarrow$*
*(**JPG** x y (**DONE** y a) (**WAG** y x (**DONE** y a) p))*

The combination of the **REQ** and the **CNFM** have created the precise set of **WAG**s needed to form a **JPG** and therefore have built a team. Having created a team, we must supply a method to dissolve the team once the **JPG**s goal has been accomplished. Just as team creation is a process that builds interlocking **WAG**s into a **JPG**, dissolving the team is a process that unwinds the **WAG**s and the **JPG**. This process is accomplished with a series of speech acts.

The Assert act will be used is to inform an agent that a previously requested goal has been accomplished. A series of assertions by the team members will allow a team to be disbanded. If the Assert succeeds in achieving its goal, that is if the addressee comes to believe the goal has been achieved, the **PGOAL** associated with the assertion can be dropped. For instance the assertion that a goal has been achieved, followed by the addressee believing the asserted fact is a sufficient condition for the team members to each drop their **PGOAL** .

Theorem 3 $\models$ *[**HAPPENED** (**REQ** x y e0 a p);*

*(**CNFM** y x e1 a (**WAG** x y (**DONE** y a) p));*
*e3;(**ASSERT** y x e2 (**DONE** y a));(**BEL** x (**DONE** y a))?] $\Rightarrow$*
*[$\neg$(**PGOAL** x y a p) $\wedge$ $\neg$(**PGOAL** y x a p)]*

The event sequence *e3* in the statement of this theorem represent a sequence of unspecified actions that y has performed or caused to be performed (possibly by recruiting yet other agents) that have achieved the goal a.

It is important to note that although the team members can drop their individual persistent goals, the necessary conditions that will allow the team to disband have still not yet been achieved. Each of the team members are left with the requirement to establish mutual belief that the goal has been accomplished. Until this occurs, the team members still have obligations from their **WAG**s and from the **WMG** from the **JPG** . One way to establish mutual belief will be for x to assert that she also believes y has achieved the goal.

Theorem 4

$$\begin{aligned}\models [&\textbf{HAPPENED}\ (\textbf{REQ}\ \text{x y e0 a p});\ (\textbf{CNFM}\ \text{y x e1 a}\ (\textbf{WAG}\ \text{y x}\ (\textbf{DONE}\ \text{y a})\ \text{p}));\\ &\quad \text{e4};(\textbf{ASSERT}\ \text{y x e2}\ (\textbf{DONE}\ \text{y a}));(\textbf{BEL}\ \text{x}\ (\textbf{DONE}\ \text{y a}))?;\\ &\quad (\textbf{ASSERT}\ \text{x y e3}\ (\textbf{BEL}\ \text{x}\ (\textbf{DONE}\ \text{y a})))] \Rightarrow\\ &(\textbf{MB}\ \text{x y}\ (\textbf{DONE}\ \text{y a}))\end{aligned}$$

x's assertion that he agrees that y has **DONE** a is enough to establish mutual belief that the teams goal has been accomplished. Having discharged the **PGOAL** (theorem 3) and established mutual belief, the agents have satisfied the **WAG**. Mutual belief also satisfies the **JPG** , allowing the agents to drop the associated **WMG**. With this last requirement discharged, the team is dissolved.

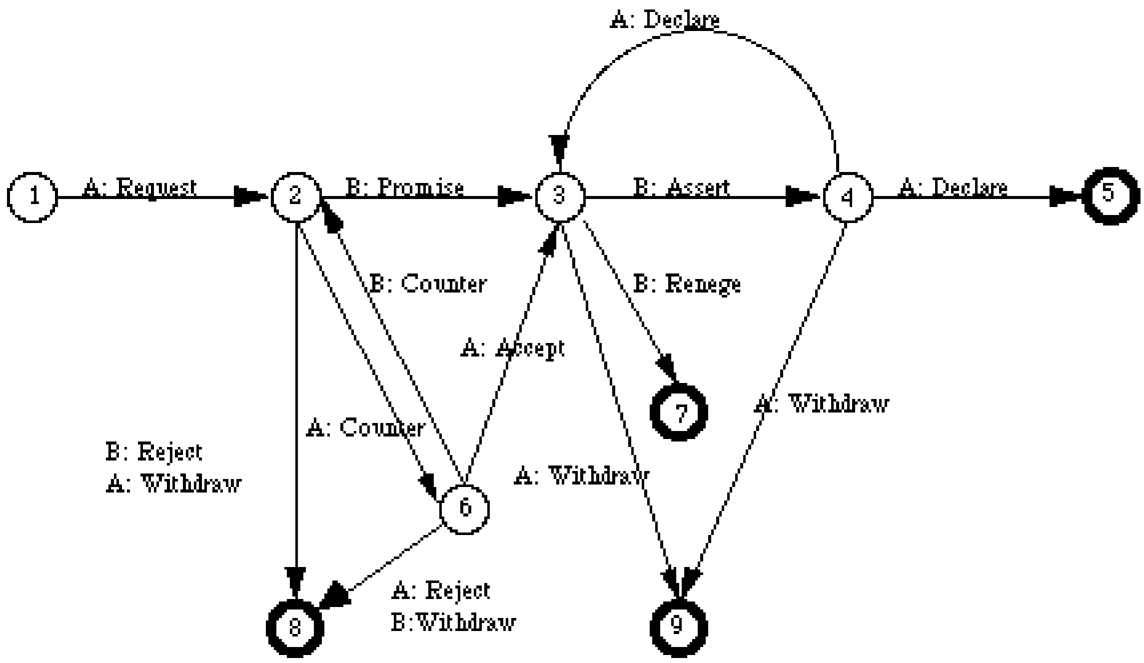
6 Application of the Theory

In this section we will examine the Winograd and Flores' [WF88] model of basic conversation for action, with the intent of showing that the behavior of the transition graph of this model of dialogue is an emergent behavior of agents behaving according to the principles in our model of interagent communication.

The diagram, reproduced as Figure 1, is easy to follow and may indeed describe the basic flow of a task oriented dialogue; however it offers little or no explanation of why the conversation is structured in this way.

In the figure the nodes (circles) represent the states of the conversation, and the arcs (lines) represent speech acts that cause transitions from state to state in the conversation. Winograd and Flores assert that states 5, 7, 8 and 9 represent final states of the conversation, state 1 initiates the conversation and the other states (2, 3, 4 and 6) represent an intermediate stage of the dialogue. We are not claiming these assertions are in error, in fact our claim is using the formalization of speech acts based on the Joint Intention Theory we can explain the reasons for this classification, and show the transitions in the figure are correct.

Winograd and Flores state the initial request from state 1 to state 2 represents a request stating some *conditions of satisfaction* (their emphasis [WF88, pg. 65]). In our model this is a request of the form (**REQ** A B e p q). A is requesting that B perform some action p with respect to a relativizing condition q. This is a task oriented dialogue, the purpose is to build a team to



The basic conversation for action (Winograd & Flores, 1986, p. 65)

Figure 1:

accomplish p. A's **REQ** speech act starts that process by informing B that A has a task to be done, and that A wants B to do the task. State 1 serves as the initial state of the dialogue. In this model it is the only initial state. By performing the REQUEST, A has committed resources towards team formation. From 1 and 2 A has outstanding goals. The dialogue must continue until A can discharge those goals. States in which A is able to accomplish this will be the final states of this dialogue.

The remainder of this section will examine the paths through the diagram that start at the initial state and end in some final state. Each of these paths will be analyzed with regard to our inner level of analysis. In addition we will be able to characterize the type of outer level completion (successful or unsuccessful with regard to the completion of the original task) a conversation following that path achieved.

Starting from the initial state Winograd and Flores claim there are exactly five alternative sets of paths for the conversation to take.

> ... the hearer can accept the conditions ... , can reject them, or can ask to negotiate a change in the conditions of satisfaction (counteroffer). The original speaker can also withdraw the request before a response, or can modify its condition. [WF88, pg. 65]

The paths representing the hearer rejecting the task are the paths leading to state 8, namely $1-2-8$ and $1-(2-6)^{*}-2^{*}-8$. The first of these is clearly the exact sequence of acts that allow the application of Theorem 1. The result of this theorem is the requester no longer has any obligations toward potential team members, as is signified by his withdrawal; with the hearer's immediate refusal, he never assumed any obligations. An analysis of the other paths leading to state 8 requires a definition of a counteroffer. A counter offer is a series of events that express two speech acts. The original hearer is saying that he refuses the original request, but if the original requester would make a (slightly) different request the hearer will commit to that request.

Definition 14 *Counteroffer*

*(**CO** x y a a1 (**WAG** y x (**DONE** x a) q))* $\triangleq$
*(**REF** x y e a (**WAG** y x (**DONE** x a) q));(**ASSERT** x y e1 ϕ)*
where ϕ is:
*IF (**ASSERT** y x (**WAG** y x (**DONE** x a1) q))*
*THEN (**INT** y a1 (**WAG** y x (**DONE** x a1) q))*

The context for a counteroffer is a **WAG** that was communicated to the party making the counteroffer. In this dialogue, the **WAG** is supplied by the original request or by a previous counteroffer in the case of a counter-counteroffer. In all cases a series of counteroffers followed by a rejection and withdrawal allow the application of Theorem 1. The counter offer illustrates an important point in the design of an agent communication language. There is no need to have a large extendible set of primitive communication acts in the language. When new capabilities are required, they can be built (by composition) from the existing set of communication acts.

Since all the paths that lead to state 8, leave all parties free of team obligations, 8 is a final state. The dialogue has been brought to a successful conclusion. However, on the outer level of dialogue analysis, state 8 represents a failure; the team was never fully formed, and the task was left unaccomplished.

All other paths to final states lead through state 3. The simplest path to 3 is $1-2-3$. This path is a **REQ** followed by a **CNFM** . As we have shown in theorem 2, this establishes a **JPG** for A and B. The other path to 3 is $1-(2-6)^{*}-3$. We have already analyzed the first two sets of links in this path, the arc from $6-3$ is labeled with A:Accept. A is accepting B's counteroffer, that is A is performing (**ASSERT** A B (**WAG** A B (**DONE** B a1) q)). *a1* and *q* are bound by B's counteroffer. B's counteroffer followed by A's accept have created a set of interlocking **WAG**s, which establish a **JPG** . All paths leading to 3 are paths where a **JPG** exists as the state is entered. This state represents a point where theorem 2 applies. A team has been formed, both A and B have undischarged goals as the result of the **JPG** . Any path leading to a final state from state 3 must provide a way to discharge both sets of goals.

The shortest paths out of state 3 are those leading directly to states 7 and 9. One is labeled B:Renege, the other A:Withdraw. In the case of the Renege, B is performing (**ASSERT** B A $\Box\neg$(**DONE** B a)). From this we know (**BEL** B $\Box\neg$(**DONE** B a)), and with reliable communication channels A will know this as well. This allows the two partys to discharge the **JPG** and disband the team. Actually there is a requirement that A communicate his acceptance to B before B can completely discharge the **WAG** — in a real world application of the theory, this final communication might take place between the two party's attorneys. In a similar manner the arc labeled A:Withdraw represents (**ASSERT** A B $\neg$(**WAG** A B (**DONE** A a) q)). Receipt of this communication allows B to discharge his **WAG**. Again, B will have to communicate back to fully dissolve the **JPG**.

The remaining paths from state 3 require the analysis of the $(3-4)^{*}$ segment. In this segment, B is performing (**ASSERT** B A (**DONE** B a))—that is, B is claiming to have finished the task. The arc from state 3 to 4 is A disagreeing—(**ASSERT** A B $\neg$(**DONE** y a)). The paths $3-(4-3)^{*}-7$ and $3-(4-3)^{*}-9$ have the same meaning as the corresponding paths without the $(4-3)^{*}$ segment.

We have finished examining all the paths into states 7 and 9, in all cases the goals that were active in state 3 have been discharged upon entering these states. These final states represent conditions under which a team was formed, and then was later disbanded without accomplishing its task.

The last path to be analyzed is the one leading from state 3 to 5. The only segment that remains to be discussed is $4-5$. On this arc A is communicating (**ASSERT** A B (**DONE** B a))), that is A is agreeing with B's prior assertion. This represents the situation described in theorem 4. All goals have been discharged, and the team is disbanded with its task accomplished.

7 Conclusions and Future Work

This paper has attempted to show that we can define communications acts in terms of the requisite mental states of the agents performing the act, and that these acts are effective ways to form, regulate and disband teams of agents. The mental states of the agents represent the commitments the agents in a team have toward each other and toward the team's task. These communications acts and the mental states they represent can be used as the basis for an agent communication language's semantics. We have applied the theory to a model of task oriented dialogue, and shown the theory successfully explains the structure of the dialogue.

References

[BDBW95] Jeffery Bradshaw, Stewart Dutfield, Pete Benoit, and John D. Wooley. Kaos: Toward an industrial-strength open agent architecture. In J. M. Bradshaw, editor, *Software Agents*. AAAI/MIT Press, 1995.

[CL90a] Philip R. Cohen and Hector J. Levesque. Intention is choice with commitment. *Artificial Intelligence*, 42:213–261, 1990.

[CL90b] Philip R. Cohen and Hector J. Levesque. Persistence, intention and commitment. In Cohen et al. [CMP90], chapter 2, pages 33–71.

[CL90c] Philip R. Cohen and Hector J. Levesque. Rational interaction as the basis for communication. In Cohen et al. [CMP90], chapter 12, pages 221–256.

[CL91a] Philip R. Cohen and Hector J. Levesque. Confirmations and joint action. pages 951–957. IJCAII, Morgan Kaufmann, August 1991.

[CL91b] Philip R. Cohen and Hector J. Levesque. Teamwork. *NOÛS*, 21:487–512, 1991.

[CMP90] Philip R. Cohen, Jerry Morgan, and Martha E. Pollack, editors. *Intentions in Communication*. System Development Foundation Benchmark Series. Bradford Books, MIT Press, 1990.

[Coh94] Philip R. Cohen. Models of dialogue. In T. Ishiguro, editor, *Cognitive Processing For Voice and Vision*. Society of Industrial and Applied Mathematics, 1994.

[FFMM94] Tim Finin, Richard Fritzson, Don McKay, and Robin McEntire. Kqml as an agent communication language. In *Proceedings of the Third International Conference on Information and Knowledge Management*. ACM Press, 1994.

[LCN90] Hector J. Levesque, Philip R. Cohen, and José H. T. Nunes. On acting together. In *Proceedings of the Eighth National Conference on Artificial Intelligence*, pages 94–99. AAAI, AAAI Press, 1990.

[LF94] Yannis Labrou and Tim Finin. A semantics approach for kqml – a general purpose communication language for software agents. In *Proceedings of the Third International Conference on Information and Knowledge Management*. ACM Press, 1994.

[WF88] Terry Winograd and Fernando Flores. *Understanding Computers and Cognition*. Addison-Wesley, 1988.