

Linguaggio e realtà sociale nei sistemi di agenti artificiali

Marco Colombetti^{1,2}, Nicoletta Fornara² e Mario Verdicchio¹

¹ marco.colombetti@polimi.it <http://www.elet.polimi.it/>
mario.verdicchio@elet.polimi.it <http://www.elet.polimi.it/>
Politecnico di Milano, Dipartimento di elettronica e informazione,
piazza Leonardo da Vinci 32 20133 Milano, Italia

² marco.colombetti@lu.unisi.ch <http://www.lu.unisi.ch/>
nicoletta.fornara@lu.unisi.ch <http://www.lu.unisi.ch/>
Università della Svizzera italiana, Facoltà di Scienze della comunicazione,
via Giuseppe Buffi 13 6900 Lugano, Svizzera

Abstract. The development of open multiagent systems requires the definition of a standard communication language, in which agents can carry out their interactions. In turn, specifying a communication language presupposes the definition of a system of concepts constituting the core of an ontology of a society of agents. In this paper we show how the semantics of a communication language can be defined within a language institution centred on the concept of commitment. Our proposal allows for a simple and uniform treatment of all types of communicative acts that appear to be interesting for artificial agent interaction.

Sommario. Lo sviluppo di sistemi multiagente aperti richiede la definizione di un linguaggio di comunicazione standard in cui gli agenti possano condurre le loro interazioni. A sua volta, la specifica di un linguaggio di comunicazione presuppone la definizione di un sistema di concetti che costituiscono il nucleo di un'ontologia di una società di agenti. In questo articolo mostriamo come la semantica di un linguaggio di comunicazione possa essere definita all'interno di un istituto del linguaggio centrato sul concetto di impegno. La nostra proposta consente di trattare in modo semplice e uniforme tutti i tipi di atti comunicativi necessari per l'interazione fra agenti artificiali.

1 Introduzione

Dall'inizio degli anni novanta si è diffuso nell'informatica l'uso del termine 'agente artificiale', o più semplicemente 'agente', per designare un componente software dotato di certe caratteristiche. Il concetto di agente ha preso rapidamente piede, tanto che oggi esistono testi, riviste scientifiche e congressi periodici interamente dedicati all'argomento. Come spesso succede tutto ciò è avvenuto, e sta avvenendo, in assenza di una definizione chiara, univoca e universalmente accettata di che cosa sia un agente: abbiamo infatti a che fare con una metafora di notevole potere euristico, che non è semplice tradurre in un concetto tecnico preciso.

Il concetto di agente non è poi così recente, dato che possiamo farlo risalire al modello computazionale degli actor proposto per la prima volta da Hewitt e colleghi nel 1977; gli agenti, tuttavia, hanno assunto un ruolo di primo piano solo con l'avvento dei sistemi distribuiti su Internet. In questo articolo non tenteremo di analizzare tutti gli aspetti significativi del concetto di agente: il lettore interessato può consultare testi recenti (Ferber 1999; Weiss 1999; Wooldridge 2000, 2002; d'Inverno e Luck 2001) e accedere al materiale sull'argomento reperibile sul Web. Ci baseremo invece su una definizione parziale, ma sufficiente per i nostri scopi: vedremo un agente come componente di un *sistema software distribuito e aperto*. Chiariamo i termini: un sistema software si dice *distribuito* quando il suo funzionamento si basa sull'interazione di programmi che operano (di fatto o in linea di principio) su diversi calcolatori; un sistema distribuito si dice poi *aperto* se i suoi componenti non sono rigidamente prestabiliti, ma possono entrare o uscire dal sistema dando luogo a gruppi d'interazione sempre diversi. Un sistema di questo genere lo chiameremo, più brevemente, *sistema multiagente*.

Un esempio di sistema multiagente è fornito dai cosiddetti *mercati elettronici* (*electronic marketplaces*): si tratta di sistemi che consentono scambi commerciali automatizzati fra agenti che ricoprono diversi ruoli, quali venditore, cliente, intermediario. Attualmente i sistemi di questo genere sono più che altro un oggetto di ricerca e non hanno ancora dato luogo a prodotti commerciali di ampia diffusione; alcuni prototipi sperimentali sono però già operativi; fra questi vale la pena di citare il sistema utilizzato per il *TAC Game*, una competizione internazionale che si svolge ogni anno dal 2000 e che simula l'esecuzione di una complessa asta elettronica (<http://www.sics.se/tac/>) e la rete sperimentale di agenti in corso di sviluppo nell'ambito di *Agentcities*, un consorzio di centri di ricerca finanziato dall'Unione Europea (<http://www.agentcities.org/>).

È facile rendersi conto che uno degli aspetti più importanti di un sistema multiagente è il processo di comunicazione fra gli agenti che lo compongono: infatti un sistema distribuito veramente aperto deve consentire l'interazione fra agenti realizzati in tempi e luoghi diversi da progettisti indipendenti; in parole povere, gli agenti devono poter interagire anche se non si sono mai "incontrati" prima d'ora. Alcuni dei prototipi citati precedentemente (ad esempio il *TAC Game*) raggiungono l'obiettivo rendendo pubbliche le *specifiche d'interfaccia*, ovvero i formati di scambio dei dati e le procedure che gli agenti devono seguire per partecipare a un'interazione; si tratta però di una soluzione impraticabile su larga scala, perché non consente agli agenti di interagire con sistemi diversi. Affinché i sistemi multiagente siano veramente aperti è necessario che le interazioni si svolgano secondo standard universalmente accettati – un'esigenza, questa, che porta in primo piano l'attività delle organizzazioni dedite alla definizione di standard, quali la FIPA (*Foundation for Intelligent Physical Agents*, www.fipa.org).

L'introduzione di standard per l'interazione nei sistemi multiagente è tutt'altro che semplice; si tratta infatti di definire e fissare un insieme di "strutture sociali": come minimo un linguaggio di comunicazione e un repertorio di concetti e di regole condivisi. Per rendersi conto della portata dell'impresa basta chiedersi che cosa occorra per condurre, ad esempio, una transazione commerciale completamente automatizzata: oltre a un linguaggio comune, gli agenti devono operare all'interno di un sistema che definisca in modo adeguato gli istituti della proprietà, del denaro e dello scambio. Come già accaduto in passato, il mondo dell'informatica si trova quindi ad affrontare un'impresa dalle evidenti implicazioni filosofiche. Nel resto

dell'articolo cercheremo di mostrare come sia possibile costruire un "quadro sociale" all'interno del quale degli agenti artificiali possano interagire, concentrandoci soprattutto sull'approccio che stiamo sviluppando presso il Politecnico di Milano e l'Università della Svizzera italiana (Colombetti 1999, 2000a, 2000b, 2000c; Colombetti et al. 2002; Colombetti e Verdicchio 2002; Fornara e Colombetti 2002, 2003; Verdicchio e Colombetti 2002, 2003). Più precisamente, nel paragrafo 2 analizzeremo il ruolo degli istituti nel mondo dell'artificiale; nel paragrafo 3 presenteremo il nostro punto di vista sull'istituto del linguaggio; nel paragrafo 4 descriveremo il ruolo degli istituti extralinguistici nell'interazione fra agenti artificiali; nel paragrafo 5, infine, avizzeremo alcune considerazioni conclusive.

2 Gli istituti nel mondo dell'artificiale

Come abbiamo visto nell'introduzione ci troviamo di fronte al problema di definire un concetto di struttura sociale per i sistemi multiagente, allo scopo di consentire agli agenti di compiere azioni che presuppongono opportuni istituti, come nel caso delle transazioni commerciali.

2.1 Il concetto di istituto

Nella nostra analisi del concetto di istituto facciamo riferimento soprattutto all'opera di John Searle (1995). Secondo questo autore la realtà sociale è costruita dagli esseri umani grazie a meccanismi che presuppongono specifiche facoltà mentali, come l'intenzionalità (e in particolare l'intenzionalità collettiva) e la capacità di attribuire funzioni (con particolare riguardo all'attribuzione di *status*). Una società in senso proprio presuppone dunque che gli individui che la compongono possiedano queste capacità specificamente umane. Per sgomberare il campo da possibili equivoci, vogliamo precisare subito che non riteniamo plausibile l'esistenza di una società di sistemi computazionali autonoma rispetto alla società umana. In ultima analisi gli agenti artificiali agiscono per conto di persone umane, fisiche o giuridiche che siano: il denaro ceduto da un agente artificiale in cambio di un bene è di fatto scambiato fra esseri umani, e il bene acquistato passa da mano umana a un'altra mano umana. Se vogliamo, possiamo vedere i sistemi multiagente come un prolungamento tecnologico che consente agli esseri umani di interagire fra loro in modo nuovo; proprio per questo, la "società degli agenti" dovrà almeno in parte replicare alcuni aspetti fondamentali della società umana.

Per lo più, le azioni compiute dagli agenti sono azioni istituzionali (ovvero, rese possibili da opportuni istituti), come l'esecuzione di atti comunicativi e le transazioni commerciali. Sempre seguendo Searle, ogni atto istituzionale è definito da *regole costitutive* della forma "*X* vale come (*counts as*) *Y* nel contesto *C*"; ad esempio, la firma in calce a un contratto vale come accettazione delle clausole del contratto, qualora la firma sia eseguita in condizioni normali (ad es., in assenza di minacce personali) da una persona legalmente autorizzata.

Sulla struttura delle regole costitutive vale la pena di fare alcuni commenti. In primo luogo il nesso fra *X* e *Y* definito da una regola costitutiva è puramente convenzionale: così come nella nostra società si assumono certi tipi di impegni firmando contratti, è pur vero che in altri luoghi o in altri tempi si assumevano

impegni analoghi mettendo a contatto il sangue dei contraenti; la firma è una pratica più igienica, ma altrettanto convenzionale. In secondo luogo il contesto gioca, nelle regole costitutive, un ruolo essenziale. In realtà, con il termine 'contesto' si fa riferimento a condizioni di natura disparata. Consideriamo ad esempio l'apertura dell'assemblea annuale di un'associazione: per aprire la seduta, il presidente proferisce l'enunciato 'l'assemblea è aperta' o 'dichiaro aperta l'assemblea'. Come è noto dalla teoria degli atti linguistici (Searle 1969, 1975; Searle e Vanderveken 1985), il proferimento dell'enunciato vale come un atto illocutivo di dichiarazione e ha l'effetto istituzionale di aprire effettivamente l'assemblea; tuttavia, l'enunciato vale come dichiarazione solo se sussistono certe condizioni contestuali. Un primo tipo di condizioni riguarda la possibilità logica di aprire l'assemblea: ad esempio, l'enunciato non ha alcun effetto se viene proferito per errore quando l'assemblea è già aperta; in altre parole, aprire un'assemblea presuppone che l'assemblea non sia ancora aperta. Questo tipo di condizioni non è esclusivo delle azioni istituzionali: allo stesso modo, aprire una bottiglia *presuppone* che la bottiglia non sia ancora aperta. Un secondo tipo di condizioni riguarda invece i poteri del parlante: l'enunciato ha effetto soltanto se viene proferito da chi possiede l'autorizzazione necessaria (nel nostro caso, il presidente dell'associazione); in altre parole, l'apertura dell'assemblea *richiede* che l'atto sia eseguito dal presidente. Al contrario del primo, il secondo tipo di condizioni è caratteristico delle azioni istituzionali; qualora sia fisicamente possibile, infatti, l'azione di aprire una bottiglia può essere compiuta anche da chi non ne ha l'autorizzazione.

Le regole costitutive vanno considerate in opposizione alle *regole regolative*, che non definiscono un nuovo tipo di azione istituzionale ma si limitano a regolare un comportamento preesistente, naturale o istituzionale. Di questo genere è ad esempio la regola 'non bisogna parlare ad alta voce nella sala di lettura', che non introduce una convenzione per l'esecuzione di azioni istituzionali, ma piuttosto vieta, sotto determinate condizioni, un comportamento che esiste indipendentemente dalla regola (parlare ad alta voce). Su questi concetti avremo modo di ritornare in seguito.

2.2 La comunicazione fra agenti

Alcuni anni or sono abbiamo iniziato a occuparci della semantica dei linguaggi per la comunicazione fra agenti (ACL, acronimo di *Agent Communication Language*), spinti dalla convinzione che le proposte avanzate fino a quel momento presentassero notevoli problemi. Negli anni novanta la maggior parte degli studiosi riteneva che fosse possibile definire la semantica di un ACL esplicitando le condizioni per l'esecuzione razionale di un atto comunicativo in termini di stati mentali individuali, quali credenze, desideri e intenzioni (Cohen e Levesque, 1990; Wooldridge, 2000); l'aspetto istituzionale della comunicazione linguistica non era invece preso in considerazione. Verso la fine degli anni novanta alcuni ricercatori si sono convinti che l'approccio alla comunicazione centrato sulla razionalità individuale è insufficiente; le alternative proposte sono essenzialmente due: trattare la comunicazione fra agenti formalizzando nel modo più fedele possibile la teoria searlina degli atti illocutivi (Chaib-draa e Vanderveken 1998), o introdurre un modello comunque derivato dalla teoria degli atti illocutivi ma più semplice, basato sul concetto di impegno (Singh 1998; Colombetti 2000a, 2000b, 2000c). Tutti e due gli approcci hanno serie motivazioni, anche se noi, per le ragioni che diremo in seguito, preferiamo il secondo; in ambedue i casi, comunque, si recupera la dimensione sociale e

Linguaggio e realtà sociale nei sistemi di agenti artificiali

istituzionale della comunicazione. Una volta imboccata questa strada, tuttavia, ci si accorge che le strutture necessarie per l'interazione comunicativa vanno molto al di là di ciò che è richiesto per dotare di semantica il nucleo minimale di un ACL. Certo, un agente deve essere in grado di informare un altro agente, di chiedere o proporre qualcosa, di accettare o respingere una richiesta o una proposta, e queste azioni possono essere viste come semplici atti illocutivi, completamente definiti all'interno dell'istituto del linguaggio. Ma ciò non basta: un agente per il commercio elettronico, ad esempio, deve anche vendere e comprare, fissare il prezzo di una merce e negoziare i prezzi fissati da altri, aprire e chiudere le fasi di una negoziazione; in altre parole, deve sapersi muovere all'interno degli istituti della proprietà, del denaro, dello scambio commerciale.

Fra gli istituti, il linguaggio occupa un ruolo fondamentale perché ogni altro istituto presuppone qualche forma di espressione linguistica: il linguaggio può esistere senza il denaro, non il denaro senza linguaggio. Il primo ACL che sia stato proposto e utilizzato porta il nome di KQML (*Knowledge Query and Manipulation Language*; vedi Finin et al. 1995). L'aspetto più interessante di questo linguaggio è il riconoscimento esplicito del ruolo degli atti linguistici nella comunicazione: in KQML l'unità di analisi della comunicazione, il messaggio, è interpretato come strumento convenzionale per l'esecuzione di atti illocutivi e la forza illocutiva di ogni atto è rappresentata esplicitamente da un simbolo, detto *performativo*, che consente di evitare le difficoltà legate all'interpretazione degli atti illocutivi umani (si pensi ad esempio al problema di riconoscere la forza degli atti illocutivi indiretti: Searle 1975b, Perrault e Allen 1980). Più recentemente la già citata FIPA ha proposto un altro linguaggio, noto come FIPA ACL, la cui struttura non si discosta molto da KQML; anche qui l'unità di analisi della comunicazione è l'atto illocutivo, ribattezzato per l'occasione *atto comunicativo*. Sia in KQML che in FIPA ACL, la tipologia degli atti comunicativi che costituiscono il nucleo del linguaggio è dettata da un'analisi dei processi comunicativi di maggior interesse per il mondo degli agenti artificiali. Se ci rifacciamo alla nota tassonomia searliana (Searle, 1975a), (Searle e Vanderveken, 1985) è facile riconoscere un ruolo importante agli atti di tipo assertivo, direttivo e commissivo; FIPA ACL, ad esempio, definisce atti assertivi come *inform* e *confirm*, atti direttivi come *request* e *query*, atti commissivi come *agree* e *accept*, e atti ibridi direttivo-commissivi come *propose*. Restano giustamente esclusi gli atti di tipo espressivo (come ringraziare, congratularsi e così via), troppo legati alla specificità delle interazioni sociali umane.

Il quinto tipo di atti della tassonomia searliana, le dichiarazioni, è invece sorprendentemente trascurato sia da KQML sia da FIPA ACL. Che cosa succede, ad esempio, quando in un'asta elettronica un cliente viene designato come vincitore? Se seguiamo lo standard FIPA, il battitore dell'asta *informa* il cliente che è risultato vincitore; in realtà non ci troviamo di fronte a un atto assertivo, bensì a una dichiarazione: comunicando a un cliente che ha vinto, il battitore esegue una dichiarazione il cui effetto è di stabilire il vincitore dell'asta. Perché allora le dichiarazioni sono state trascurate nelle prime proposte di ACL? La risposta è semplice. Per trattare le dichiarazioni non è sufficiente l'istituto linguistico: in generale le dichiarazioni presuppongono altri istituti, che chiameremo *istituti extralinguistici* per distinguerli dall'istituto fondamentale del linguaggio. Gli istituti extralinguistici svolgono una doppia funzione: da un lato *definiscono un tipo di realtà istituzionale* che senza l'istituto non esisterebbe neppure (nel caso dell'esempio: vincere un'asta); da un altro lato, stabiliscono chi è *autorizzato a*

eseguire un'azione istituzionale (sempre riferendoci all'esempio: il vincitore è designato dal battitore, non dai clienti). Come si vede, non è possibile dotare un linguaggio di dichiarazioni senza sviluppare nel contempo un modello sufficientemente dettagliato degli istituti extralinguistici. Nel paragrafo 3 ci occuperemo dell'istituto del linguaggio, per passare poi agli istituti extralinguistici nei paragrafi successivi.

2.3 Come definire un istituto

In generale la definizione di un istituto deve specificare:

- l'*ontologia* delle entità che costituiscono il frammento di realtà sociale creata dall'istituto;
- le *autorizzazioni* per l'esecuzione degli atti istituzionali;
- le *convenzioni* per l'esecuzione degli atti istituzionali; e infine
- le *norme* per l'esecuzione degli atti istituzionali.

Ci soffermeremo ora separatamente sui quattro punti.

L'ontologia della realtà sociale

L'aspetto più importante di un istituto è la creazione di nuovi componenti della realtà sociale: ad esempio, in assenza di un istituto del denaro i concetti di denaro, prezzo e pagamento semplicemente non esistono;¹ e in assenza di un istituto del linguaggio non si può informare, richiedere, proporre.

L'ontologia di un istituto può definire entità di diversa natura, quali oggetti, attributi ed eventi; consideriamo ad esempio l'istituto della proprietà: essere proprietario di un bene è un attributo istituzionale, mentre un passaggio di proprietà è un evento istituzionale. Ci si può chiedere come sia possibile che un nuovo aspetto della realtà nasca, quasi magicamente, da un istituto. Come abbiamo già accennato, nel caso della società umana la cosa sembra riconducibile all'intenzionalità della mente (in particolare all'intenzionalità collettiva) e alla capacità umana di assegnare funzioni alle entità del mondo (Searle, 1995). Tutto ciò è molto importante per lo studio della società umana, che in ultima analisi esiste soltanto grazie alle capacità mentali degli esseri umani; ma le cose stanno diversamente per le "società" di agenti artificiali: in questo caso, come abbiamo già affermato, non si può parlare di una struttura sociale autonoma e indipendente dalla società umana. Una società di agenti artificiali risulta dalla proiezione della società umana sul mondo dell'artificiale. Negli istituti artificiali non è dunque in gioco alcuna "intenzionalità collettiva" o "capacità di attribuzione funzionale" degli agenti; al contrario un istituto artificiale è creato dagli esseri umani per attribuire certe conseguenze istituzionali ai processi di computazione eseguiti dagli agenti (ad esempio: la trasmissione di una firma elettronica da parte di un agente vale come accettazione delle clausole di un contratto): la società degli agenti non è una società autonoma, bensì un prolungamento tecnologico della società umana.

Anche ridimensionando il concetto di società di agenti, tuttavia, resta il problema di come si possa definire in pratica un'ontologia istituzionale, in modo tale da

¹ Possono naturalmente esistere dei precursori di tali concetti: il denaro è stato inventato quando il baratto era già diffuso e assestato nelle società umane.

Linguaggio e realtà sociale nei sistemi di agenti artificiali

consentire l'esecuzione di azioni istituzionali da parte di sistemi artificiali. Questo problema verrà affrontato nei paragrafi successivi, in cui mostreremo come definire l'istituto del linguaggio per i sistemi multiagente.

Di particolare interesse è la definizione delle azioni istituzionali: si può infatti pensare che la funzione essenziale degli istituti sia di creare nuove possibilità d'azione. Nell'informatica, le azioni sono solitamente definite in base alle loro condizioni di eseguibilità e ai loro effetti; se si vuole seguire questo schema si dovrà dunque chiarire quali siano, di volta in volta, le condizioni e gli effetti delle azioni istituzionali – condizioni ed effetti che, a loro volta, coinvolgeranno fatti istituzionali.

Un esempio di azione istituzionale può essere la cessione della proprietà di un oggetto, come avviene quando un bene viene donato o comunque scambiato; questo tipo di azione è definito all'interno dell'istituto della proprietà.

Autorizzazioni e ruoli

Come abbiamo già rilevato, l'esecuzione di un atto convenzionale comporta conseguenze istituzionali solo se viene eseguito in condizioni opportune e, in particolare, da un agente autorizzato. Ad esempio, potremmo stabilire che l'unico agente autorizzato a trasferire la proprietà di un bene è il proprietario del bene. Tuttavia è possibile concepire un istituto più vasto in cui un primo agente, in virtù del ruolo che ricopre, è autorizzato a trasferire la proprietà di un bene da un secondo a un terzo agente; una simile azione istituzionale sarebbe un caso di esproprio. Come vedremo più avanti, casi analoghi sono presenti anche nell'istituto del linguaggio.

In generale, un agente gode di determinate autorizzazioni in virtù del *ruolo* che ricopre: se Tizio è il presidente di un'associazione, ad esempio, è corretto dire che Tizio è autorizzato ad aprire l'assemblea sociale, ma di questa autorizzazione Tizio gode appunto in quanto ricopre il ruolo di presidente. Ci possono essere eccezioni: in certi casi Tizio potrebbe godere di autorizzazioni che sono state concesse soltanto a lui, ad esempio grazie a una delega *ad personam*. Ci sembra peraltro che l'"autorizzazione personale" sia un fenomeno strettamente legato alla natura delle relazioni umane; come tale potrebbe non avere nessuna utilità nei sistemi artificiali.

Le convenzioni

Un'azione istituzionale è un atto astratto; ciò significa che l'azione è definita dalle condizioni di eseguibilità e dagli effetti, non dal modo in cui è eseguita. Per consentire l'effettiva esecuzione delle azioni, un istituto deve però specificare anche le modalità per l'esecuzione concreta delle azioni. Un aspetto importante è che tali modalità sono *convenzionali*, nel senso che soltanto una convenzione arbitraria lega un'azione istituzionale alla sua modalità di esecuzione concreta.

Ritorniamo alla forma generale delle regole costitutive suggerita da Searle (par 2.1). Nell'espressione "X vale come Y in C" possiamo ora interpretare Y come un'azione istituzionale, C come le sue condizioni di eseguibilità e X come *l'atto strumentale* la cui esecuzione, per convenzione, vale come esecuzione di Y. Nella società umana, le convenzioni di un istituto possono essere specificate con estrema nettezza e precisione oppure ricondursi a un insieme di pratiche alquanto sfumate. Un esempio di sistema convenzionale netto è costituito dalle procedure legali; per contrasto si pensi a quanto siano flessibili le modalità con cui si può regalare un oggetto a un amico. Nel caso degli agenti artificiali, riteniamo, i sistemi di convenzioni saranno sempre del primo tipo.

Le norme

In generale un istituto non si limita a specificare chi è autorizzato a eseguire le azioni istituzionali, ma fissa anche un insieme di obblighi e permessi. Ad esempio, il presidente di un'associazione non solo è autorizzato a convocare l'assemblea dei soci: tipicamente è anche obbligato a farlo almeno una volta all'anno, e ha il permesso di farlo più spesso se lo considera opportuno.

Nel prossimo paragrafo seguiremo lo schema qui delineato per definire l'istituto del linguaggio.

3 Comunicazione e linguaggio

Comunicare è un'attività: l'unità di analisi della comunicazione è quindi *l'atto comunicativo*, inteso come atto istituzionale eseguito mediante l'invio di un *messaggio* da parte di un *mittente* a uno o più *destinatari*. Più precisamente, possiamo dire che *l'invio di un messaggio*, sotto opportune condizioni contestuali, vale come esecuzione di un atto comunicativo, i cui attributi sono funzione delle caratteristiche del messaggio e di certi elementi del contesto.

Ma che cos'è esattamente un atto comunicativo? Rispondere a questa domanda significa:

- determinare il rapporto fra la forma dei messaggi e gli atti comunicativi che vengono eseguiti inviandoli;
- specificare le condizioni sotto cui l'invio di un messaggio realizza felicemente l'atto comunicativo corrispondente;
- definire gli effetti dell'esecuzione di un atto comunicativo.

Nel seguito vedremo due possibili approcci coerenti con l'assunzione che comunicare sia un'attività istituzionale.

3.1 Gli atti comunicativi come atti illocutivi

Come abbiamo già visto, un modello di riferimento naturale per la definizione degli atti comunicativi è la teoria degli atti illocutivi umani; questa strada è effettivamente suggerita da alcuni studiosi (Chaib-draa e Vanderveken 1998). La prima, ovvia, difficoltà che questo approccio incontra è la notevole complessità del sistema illocutivo umano. Se siamo restii a seguire questa strada, tuttavia, è soprattutto per un altro motivo: la teoria dell'illocuzione descrive il linguaggio come istituto autonomo della società umana, mentre la teoria degli atti comunicativi artificiali può dare per scontato il fatto che la società degli agenti non è autonoma, ma esiste in quanto prolungamento della società umana; le conseguenze di questo cambiamento di prospettiva non sono indifferenti, e per rendersene conto è utile analizzare un esempio. Seguendo Searle e Vanderveken (1985), l'atto assertivo di *informare* può essere definito nel modo seguente. In primo luogo abbiamo le condizioni costitutive comuni a tutti gli atti assertivi:

- il *punto illocutivo* è presentare il contenuto dell'atto come rappresentazione di uno stato di cose che sussiste effettivamente (p. 37);

Linguaggio e realtà sociale nei sistemi di agenti artificiali

- le *condizioni di contenuto proposizionale* sono vuote (p. 43);
- le *condizioni preparatorie* richiedono che il mittente sia in possesso di evidenza a favore della verità del contenuto dell'atto (p. 44);
- le *condizioni di sincerità* richiedono che il mittente creda effettivamente al contenuto dell'atto (p. 45).

In secondo luogo abbiamo le condizioni aggiuntive dell'atto di informare:

- le *condizioni preparatorie specifiche* dell'atto di informare richiedono che il destinatario non sia già a conoscenza del contenuto (p. 185).

Come si vede, le condizioni costitutive dell'atto di informare sono piuttosto complesse. Esse fanno riferimento in modo essenziale al sussistere di uno stato di cose, all'evidenza a favore di una proposizione e agli stati mentali del mittente e dell'ascoltatore (stati epistemici, nel caso in esame). Queste categorie concettuali, e altre ancora, sono necessarie per rendere conto di come un atto eseguito da un essere umano possa valere come un atto di informare; ma non è ovvio che lo siano se lo scopo è rendere conto degli effetti di uno scambio comunicativo fra agenti artificiali. Consideriamo ad esempio il ruolo degli stati epistemici nella condizione di sincerità (il mittente crede a ciò che dice) e nella condizione preparatoria specifica (il destinatario non conosce ancora ciò che gli viene detto): come possono essere utilizzate condizioni simili nell'interazione fra due agenti artificiali?

Una soluzione spesso proposta nella letteratura suggerisce di considerare gli agenti artificiali come sistemi effettivamente dotati di stati mentali; in questa linea sono stati sviluppati i modelli logico-formali noti come *modelli BDI (Belief, Desire, Intention)*; vedi ad es. Wooldridge 2000). Al di là del fatto che questa posizione è assai discutibile, non è per nulla chiaro in che modo queste condizioni possano essere utilizzate nei concreti processi di comunicazione fra agenti²: il problema più vistoso è che, secondo la teoria searliana degli atti linguistici, alcune condizioni costitutive degli atti illocutivi sono inosservabili. Nel caso del linguaggio umano, che certe condizioni siano inosservabili non può essere considerato un difetto: la possibilità di mentire è una caratteristica specifica ed essenziale della comunicazione umana. Naturalmente, per il destinatario di un messaggio è spesso importante stabilire se il mittente è o non è sincero; in questo caso gli esseri umani cercano di sfruttare tutte le informazioni di cui dispongono sull'interlocutore, sul contesto della comunicazione, e così via; insomma, la capacità di muoversi in un mondo in cui è possibile la menzogna è una caratteristica specificamente umana. Non è invece plausibile, crediamo, pretendere tanto nel mondo dell'artificiale: gli agenti si troveranno in generale a condurre interazioni brevi, con interlocutori di cui sanno poco o nulla e in assenza delle informazioni contestuali (come l'espressione facciale, la postura, il tono di voce) che sono così preziose nel campo dell'interazione umana.

C'è però una soluzione alternativa, a nostro avviso più adeguata alla comunicazione artificiale, che ora descriveremo.

² Un sintomo di questa incertezza è la frequente lettura regolativa delle condizioni di sincerità, che nel caso degli assertivi vengono interpretate come una sorta di nono (od ottavo) comandamento ('non mentire'). In realtà secondo Searle la condizione di sincerità è costitutiva e non regolativa; nel caso degli assertivi comporta che se il mittente non crede alla verità del contenuto di un atto illocutivo l'atto eseguito è oggettivamente "difettoso", indipendentemente da quanto possano ritenere il destinatario o un osservatore indipendente.

3.2 Comunicazione e impegno

Come abbiamo già sottolineato, la definizione di un atto comunicativo richiede di determinare il rapporto fra la forma dei messaggi e gli atti comunicativi eseguiti, specificare le condizioni sotto cui l'invio di un messaggio realizza l'atto comunicativo corrispondente, e definire gli effetti dell'esecuzione di quest'ultimo. Nel caso della comunicazione fra agenti artificiali, è sensato aggiungere che la determinazione dell'atto comunicativo eseguito dev'essere oggettiva e non ambigua. Questa richiesta, in verità, ha conseguenze assai rilevanti: essa comporta che nel definire un atto comunicativo si possa fare riferimento unicamente a fatti ed eventi *pubblici*, cioè rilevabili oggettivamente dagli agenti che partecipano allo scambio comunicativo. Le prime vittime di questo approccio sono le condizioni espresse in termini di stati mentali, poiché questi sono privati e inosservabili; in particolare cadono tutte le condizioni di sincerità, nonché alcune condizioni preparatorie, come ad esempio la condizione preparatoria specifica dell'atto di informare; ma neppure la definizione del punto illocutivo resta indenne, almeno per alcune categorie di atti comunicativi. Consideriamo ad esempio gli atti direttivi: secondo Searle e Vanderveken (1985), il punto illocutivo di un direttivo è costituito dal *tentativo*, da parte del mittente, di far sì che il destinatario realizzi l'azione descritta dal contenuto dell'atto linguistico (p. 39); ma il concetto di tentativo ha una componente intrinsecamente mentale: tentare di fare Y significa infatti eseguire X con l'*intenzione* che ciò comporti Y.

Se vogliamo seguire l'approccio testé delineato, la prima cosa da fare è identificare uno o più concetti che ci permettano di esprimere l'effetto dell'esecuzione di un atto comunicativo; in ogni caso, gli effetti dovranno essere eventi istituzionali oggettivi, completamente determinabili in base a informazioni pubbliche. In verità, la soluzione del problema è immediata per la categoria degli atti commissivi, come ad esempio le promesse: infatti l'effetto di un commissivo è un impegno (*commitment*); più precisamente, l'esecuzione di un atto commissivo impegna il mittente, nei confronti del destinatario, a eseguire l'azione descritta dal contenuto. Il concetto di impegno soddisfa tutti i criteri precedentemente enunciati; infatti:

- un impegno è un fatto istituzionale, ovvero un frammento di realtà sociale;
- che un impegno sussista o non sussista può essere stabilito in base a informazioni oggettive e pubbliche, quali ad esempio la firma su un contratto o l'esecuzione di un atto di promessa.

Naturalmente non è sorprendente che un atto commissivo possa essere definito ricorrendo al concetto di impegno; il problema è piuttosto se questo modo di procedere possa essere esteso anche ad altre categorie di atti comunicativi. Singh 1998 è stato il primo a suggerire che tutti gli atti comunicativi possano essere definiti in termini di impegni; da questo punto di vista, l'esecuzione di un atto comunicativo è vista come la manipolazione della rete di impegni che legano un agente agli altri membri di un gruppo. Tuttavia, né il lavoro citato né i successivi articoli di Singh (Singh, 2000) presentano un quadro sistematico di definizioni in grado di coprire tutti gli aspetti della comunicazione fra agenti; nel seguito presenteremo quindi soprattutto l'approccio che stiamo sviluppando e che generalizza significativamente la proposta originale di Singh.

3.3 L'istituto del linguaggio

Mostreremo ora come l'istituto del linguaggio possa essere definito seguendo lo schema proposto nel paragrafo 2.

L'ontologia istituzionale

Come abbiamo già suggerito nel paragrafo 3.2, l'entità concettuale su cui fondiamo la nostra definizione del linguaggio è l'*impegno*. Non è nostra intenzione ricostruire i processi mentali e sociali che stanno alla base di questo concetto; al contrario lo assumeremo come entità primitiva e lo descriveremo in modo assiomatico.

Dal punto di vista della *struttura*, un impegno è un'entità costituita da tre componenti:

- l'agente che ha l'impegno, che chiameremo *debitore*;
- l'agente verso cui il debitore è impegnato, che chiameremo *creditore*;
- la proposizione alla cui verità il debitore è impegnato, che chiameremo *contenuto*.

Per la trattazione che segue è molto utile che gli impegni siano trattati come oggetti astratti, la cui esistenza è garantita dall'istituto pertinente; ciò consente in particolare di fare riferimento a impegni specifici, trattandoli come individui. Utilizzeremo quindi la notazione simbolica

$C(c,a,b,p)$

per asserire che c è un impegno (*commitment*) con l'agente a come debitore, l'agente b come creditore e la proposizione p come contenuto.

Nel nostro modello introduciamo anche il concetto di *preimpegno* (*precommitment*), che va inteso come un impegno non ancora fissato, e soggetto all'accettazione o al rifiuto da parte del debitore. Per asserire che c è un preimpegno con l'agente a come debitore, l'agente b come creditore e la proposizione p come contenuto scriveremo

$P(c,a,b,p)$.

La prima proprietà che vogliamo esprimere è che un impegno può essere *pendente*, *soddisfatto* o *violato*. Intuitivamente, un impegno è pendente se il suo contenuto non è (ancora) né vero né falso; soddisfatto se il suo contenuto è vero; e violato se il suo contenuto è falso. Supponiamo ad esempio che alle ore 10 del mattino di un giorno qualunque l'agente a dica a b : 'pioverà entro oggi'; questa *previsione* (un atto assertivo) impegna a , nei riguardi di b , al fatto che piova entro le ore 24 del giorno corrente. Supponiamo poi che non piova dal momento in cui la previsione è stata fatta fino alle ore 11 comprese: in queste condizioni l'impegno è pendente, perché il suo contenuto non può ancora essere considerato né vero né falso. Supponiamo infine che alle ore 12 inizi a piovere: il contenuto dell'impegno diventa allora vero e l'impegno è soddisfatto; se al contrario non piove per tutto il giorno, non appena vengono superate le ore 24 il contenuto diventa falso e l'impegno è violato. In conclusione abbiamo che un impegno c è:

- pendente se $C(c,a,b,p)$ e p è indefinito,
- soddisfatto se $C(c,a,b,p)$ e p è vero,
- violato se $C(c,a,b,p)$ e p è falso.

Queste definizioni sembrano presupporre una logica a tre valori, ma in effetti possono essere tradotte in una logica temporale ramificata con due valori di verità³ (Verdicchio e Colombetti, 2003): in uno stato del mondo una proposizione è allora indefinita se sono possibili stati futuri in cui il suo valore di verità è 1 e stati futuri in cui il suo valore di verità è 0, è vera se in ogni stato futuro il suo valore è 1, ed è falsa se in ogni stato futuro il suo valore di verità è 0.

Il secondo problema da affrontare è la *dinamica* dell'impegno: gli impegni infatti possono essere creati, negoziati, accettati e cancellati dagli agenti. Ammettiamo quindi che esistano un certo numero di possibili operazioni che hanno questi effetti; più precisamente:

- l'operazione $mc(a,b,p)$ (*make commitment*) crea un nuovo impegno c tale che $C(c,a,b,p)$;
- l'operazione $mp(a,b,p)$ (*make precommitment*) crea un nuovo preimpegno c tale che $P(c,a,b,p)$;
- l'operazione $cc(c,a,b,p)$ (*cancel commitment*) cancella un impegno preesistente c della forma $C(c,a,b,p)$;
- l'operazione $cp(c,a,b,p)$ (*cancel precommitment*) cancella un preimpegno preesistente c della forma $P(c,a,b,p)$; e infine
- l'operazione $ap(c,a,b,p)$ (*accept precommitment*) trasforma un preimpegno preesistente c della forma $P(c,a,b,p)$ in un impegno a pieno titolo della forma $C(c,a,b,p)$.

Queste operazioni (insieme con le definizioni d'impegno pendente, soddisfatto e violato) definiscono completamente l'ontologia dell'impegno. Va notato che l'ontologia dell'impegno si limita ad assumere l'esistenza delle operazioni descritte, mentre non dice nulla su chi sia autorizzato a eseguirle; dal punto di vista dell'ontologia dell'impegno, insomma, è indifferente che a creare un impegno sia il debitore, il creditore o un terzo agente: le autorizzazioni sono infatti definite da un'altra componente dell'istituto del linguaggio, che ora descriveremo.

Ruoli e autorizzazioni

Con l'eccezione delle dichiarazioni, che tratteremo più avanti, le azioni eseguibili all'interno dell'istituto del linguaggio consistono essenzialmente nella manipolazione di impegni. Per l'esecuzione di queste azioni valgono le autorizzazioni seguenti: un agente può

- creare un impegno arbitrario di cui è debitore;
- creare un preimpegno arbitrario di cui è creditore;
- cancellare un impegno o preimpegno preesistente di cui è creditore;
- cancellare o accettare un preimpegno preesistente di cui è debitore.

Come vedremo in seguito, ulteriori autorizzazioni possono essere definite da istituti extralinguistici.

Le azioni eseguibili all'interno dell'istituto del linguaggio sono gli atti comunicativi; nel seguito presentiamo alcune definizioni seguendo da vicino la

³ Il fatto che una proposizione possa essere "ancora" indefinita, ma diventare vera o falsa nel futuro, è noto come problema dei *contingenti futuri*. Come rilevato nel testo, questo problema trova un'elegante soluzione formale nell'ambito della logica temporale ramificata a due valori di verità.

Linguaggio e realtà sociale nei sistemi di agenti artificiali

tassonomia searliana degli atti illocutivi (Searle, 1975). È importante ricordare che le definizioni delle azioni istituzionali devono rispettare le autorizzazioni previste dall'istituto.

Per quanto riguarda gli atti di tipo assertivo, facciamo l'ipotesi che asserendo qualcosa il mittente si impegni, nei confronti del destinatario, alla verità di quanto afferma:

assertivo (*a*: mittente; *b*: destinatario; *p*: contenuto) =
condizioni – ;
effetto *a* è impegnato con *b* al fatto che *p* sussista.

Analogamente, con un atto commissivo il mittente si impegna, nei confronti del destinatario, a eseguire un'azione del tipo specificato:

commissivo (*a*: mittente; *b*: destinatario; *p*: contenuto) =
condizioni – ;
effetto *a* è impegnato con *b* ad eseguire un esempio di azione del tipo descritto da *p*.

Passiamo ora agli atti direttivi. Va detto subito che questa categoria di atti comunicativi sembra rappresentare un problema per l'approccio basato sul concetto di impegno; tuttavia, un suggerimento proviene dall'analisi di alcuni direttivi particolari, come gli atti di *ordinare* e di *richiedere*. Un ordine, emesso da un mittente che ne abbia l'autorità, crea un impegno del *destinatario* nei confronti del mittente (o dell'organizzazione che il mittente rappresenta). In effetti la stessa cosa non si può dire dei direttivi in generale: ad esempio una richiesta non crea immediatamente un impegno del destinatario; tuttavia, anche una richiesta crea un impegno del destinatario *quando viene accettata*. Un atto direttivo pone quindi il destinatario in uno stato istituzionale di preimpegno, che il destinatario può rifiutare o trasformare in impegno compiuto con un atto di accettazione; in certi casi, la situazione istituzionale complessiva in cui sono immersi il mittente e il destinatario comporta la trasformazione per così dire automatica del preimpegno in impegno compiuto (e ciò spiega il funzionamento di un ordine). In conclusione:

direttivo (*a*: mittente; *b*: destinatario; *p*: contenuto) =
condizioni *p* è un'azione futura di *b*;
effetto *b* è preimpegnato con *a* a eseguire un esempio di azione del tipo descritto da *p*.

A sua volta, un preimpegno può essere trasformato in impegno eseguendo un atto di accettazione:

accettazione (*a*: mittente; *b*: destinatario; *p*: contenuto) =
condizioni *p* è un'azione futura di *a*;
a è preimpegnato con *b* a eseguire *p*;
effetto *a* è impegnato con *b* a eseguire un esempio di azione del tipo descritto da *p*.

Alternativamente, un preimpegno può essere rifiutato dal debitore:

rifiuto (a : mittente; b : destinatario; p : contenuto) =

<i>condizioni</i>	p è un'azione futura di a ; a è preimpegnato con b a eseguire p ;
<i>effetto</i>	a non è più preimpegnato né impegnato con b a eseguire un esempio di azione del tipo descritto da p .

Come si vede confrontando le due definizioni, l'atto di accettazione è un particolare atto commissivo. A questo punto, se seguiamo la tassonomia di Searle 1975a dobbiamo ancora occuparci di due categorie di atti comunicativi: gli espressivi e le dichiarazioni. Per quanto riguarda i primi proponiamo una soluzione piuttosto drastica, che consiste nell'eliminazione pura e semplice degli espressivi dal sistema comunicativo degli agenti artificiali. Non ci sembra una grave perdita: per quanto ci consta, nessuno ha ancora trovato una funzione per scuse, saluti e congratulazioni nell'interazione fra agenti: gli espressivi sembrano indissolubilmente legati alla natura specifica dei rapporti umani ed è quindi possibile che restino esclusi anche in futuro dalla comunicazione artificiale. Diverso è il discorso per quanto riguarda le dichiarazioni: come abbiamo già rilevato nel paragrafo 2.2, infatti, si tratta di una categoria di atti comunicativi essenziale per l'esecuzione di molte azioni istituzionali. Per analizzare le dichiarazioni, tuttavia, il concetto di impegno non è sufficiente perché le dichiarazioni, in generale, presuppongono frammenti di realtà sociale creati da istituti extralinguistici (vedi il par. 2.1); rimandiamo pertanto la definizione degli atti di dichiarazione al paragrafo 4, dedicato all'analisi degli istituti extralinguistici.

Le convenzioni

Le azioni istituzionali sono convenzionali, nel senso che per eseguire un'azione istituzionale è necessario compiere un'azione strumentale che, per convenzione, vale come esecuzione dell'azione istituzionale. Nel caso dei linguaggi di comunicazione per agenti, l'azione strumentale che vale per l'esecuzione di un atto comunicativo è l'invio di un messaggio di formato prestabilito. Ad esempio, FIPA ACL prevede tre diversi formati: uno pensato per la facilitare la lettura umana, uno espresso in XML, e uno progettato per minimizzare i caratteri trasmessi. Per la trattazione presente non è importante stabilire una convenzione specifica: ci basta infatti riconoscere che è possibile fissare una convenzione qualsiasi.

Le norme

Nella comunicazione umana è facile identificare norme che regolano l'uso del linguaggio. Una norma molto generale, ad esempio, stabilisce che si dovrebbe sempre rispondere a una richiesta, in senso positivo o negativo; percepiamo infatti la mancata risposta come una violazione di una regola dell'interazione sociale. A nostro avviso, tuttavia, una regola del genere non appartiene direttamente all'istituto del linguaggio, bensì all'istituto delle conversazioni, che preferiamo vedere come un primo esempio di istituto extralinguistico: qualificando un istituto come "extralinguistico", infatti, non vogliamo suggerire che il linguaggio non vi abbia un ruolo (non c'è istituto in cui il linguaggio non giochi un ruolo fondamentale), bensì rimarcare il fatto che l'istituto va al di là della pura semantica degli atti comunicativi. Per questo motivo preferiamo assumere che l'istituto del linguaggio vada considerato come privo di norme. Le norme assumono invece un ruolo importante negli istituti extralinguistici, come vedremo nel prossimo paragrafo.

4 Istituti extralinguistici e organizzazioni

Può sembrare, a questo punto, che l'obiettivo che ci eravamo dati all'inizio sia raggiunto: l'istituto del linguaggio definisce i tipi di atti comunicativi eseguibili, sancendo contestualmente le autorizzazioni a compierli, precisandone gli effetti istituzionali (in termini di impegni e preimpegni) e fissando le convenzioni per eseguirli. Tuttavia la nostra trattazione non è ancora completa, e di ciò è facile rendersi conto prendendo nuovamente in considerazione l'atto di ordinare. Come abbiamo già suggerito, non è possibile distinguere una richiesta da un ordine senza chiamare in causa qualche istituto extralinguistico che autorizzi un agente, in qualità di creditore, a impegnare direttamente un altro agente come debitore. Nella società umana, gli esempi di istituti di questo genere abbondano, dato che è possibile emanare ordini in qualunque organizzazione fortemente gerarchica; è chiaro però che per affrontare questa problematica la nostra analisi deve uscire dall'ambito linguistico.

4.1 Le norme

Un esempio interessante, nell'ambito della applicazioni dei sistemi multiagente al commercio elettronico, è dato dal concetto di asta. Come tutti sappiamo, le aste sono schemi d'interazione, strettamente codificati, per l'esecuzione di scambi commerciali; si tratta in effetti di un frammento di realtà sociale costruito grazie a un insieme di regole costitutive e quindi, secondo il nostro modo di vedere, di un istituto⁴.

Nelle applicazioni informatiche è usuale vedere un'asta come un "protocollo d'interazione" (Fornara e Colombetti 2003). Questo punto di vista ci sembra sostanzialmente corretto; va però rilevato che ogni protocollo, stabilendo quali azioni possano essere eseguite in ciascun momento dell'interazione da ciascun agente, ha l'effetto di sancire implicitamente un insieme di autorizzazioni, stabilendo nel contempo un insieme di obblighi e di permessi relativi all'esecuzione delle azioni autorizzate.

Consideriamo ad esempio un momento particolare di un'asta inglese: il battitore ha appena raccolto le offerte relative al bene attualmente battuto e, in attesa di eventuali offerte superiori, ha temporaneamente assegnato il bene al miglior offerente. Il battitore è in effetti l'unico agente *autorizzato* ad eseguire l'azione di assegnare il bene al miglior offerente; inoltre, però, il battitore è *obbligato* a eseguire l'azione quando se ne verificano le condizioni. Simmetricamente, al momento opportuno ogni cliente dell'asta può lanciare un'offerta, purché superiore all'offerta correntemente accettata: il cliente è *autorizzato* a eseguire l'azione, ma non ne ha l'*obbligo* bensì il *permesso*. In effetti, come abbiamo già sottolineato, in ogni istituto le autorizzazioni sono accompagnate da obblighi o da permessi, ed è importante non confondere il concetto di autorizzazione con il concetto di permesso: quando diciamo che un agente è autorizzato a eseguire un'azione, diciamo che gli è stato accordato un certo *potere istituzionale* (Jones e Sergot 1996); a sua volta, l'esercizio di questo potere può essere soggetto a una regolamentazione, descrivibile in termini di obblighi e di permessi.

⁴ Ci si può chiedere se esista un unico istituto delle aste o una pluralità di istituti (asta inglese, asta olandese e così via). Il problema può essere risolto adottando per la definizione degli istituti un approccio gerarchico, come nei sistemi informatici orientati agli oggetti.

Come le autorizzazioni, anche gli obblighi e i permessi non sono solitamente legati ad agenti specifici, ma al ruolo che gli agenti ricoprono. Certi ruoli possono essere svolti da un solo agente per volta (il battitore di un'asta), altri da più agenti (i clienti dell'asta); inoltre può capitare che un singolo agente svolga più ruoli, alternativamente o contemporaneamente. Intuitivamente, i ruoli previsti da un istituto possono essere visti come "caselle" che devono essere riempite da specifici agenti affinché l'interazione abbia luogo; come vedremo fra poco, tuttavia, perché ciò sia possibile è necessario che l'istituto sia realizzato concretamente da un'organizzazione.

4.2 Le organizzazioni

Mentre l'istituto del linguaggio è, per così dire, un istituto diffuso (ovvero presente in modo pervasivo in tutte le interazioni umane) gli istituti extralinguistici sono generalmente reificati da *organizzazioni*⁵ specifiche. È facile rendersi conto della differenza fra istituti e organizzazioni: ad esempio l'asta è un istituto, mentre una casa d'aste è un'organizzazione che reifica una particolare versione dell'istituto delle aste. Va detto, tuttavia, che i due concetti sono stati spesso confusi nell'ambito della ricerca sui sistemi multiagente (Jones e Sergot, 1996), con conseguenze negative sul piano teorico.

Per i nostri interessi immediati la differenza principale fra un istituto e un'organizzazione è che un'organizzazione può essere vista come un agente, mentre lo stesso non vale per un istituto; ad esempio, un'azienda esegue azioni come comprare, produrre e vendere; quando viene eseguito un atto comunicativo, invece, ad agire non è certo l'istituto del linguaggio. Nelle organizzazioni umane, gli atti vengono eseguiti da persone che le rappresentano: non c'è azione appartenente alla sfera umana, infatti, che non venga eseguita da un essere umano. Nel mondo dell'artificiale, invece, almeno in linea di principio è possibile concepire un'intera organizzazione come un singolo agente: qui possiamo ritagliare gli individui come ci conviene, perché non ci sono vincoli biologici da soddisfare. Nei casi di complessità elevata, tuttavia, possiamo attenderci che un'organizzazione artificiale sarà realizzata come un'*agenzia*, ovvero come un macro-agente costituito al suo interno da una collezione di agenti coordinati e cooperanti.

Fra le organizzazioni e gli istituti intercorre una precisa relazione: un'organizzazione, infatti, reifica uno o più istituti extralinguistici, nel senso che realizza concretamente qualche forma specifica di "modello istituzionale". Ad esempio, una casa d'aste reifica una forma specifica di asta, un esercito realizza una forma specifica di gerarchia, un'agenzia turistica realizza una forma specifica d'intermediazione commerciale e così via. I ruoli previsti dall'istituto reificato sono allora occupati da agenti specifici: si costituisce in questo modo un *gruppo* organizzato di agenti fra cui può avere luogo un'interazione comunicativa.

⁵ Anziché di 'organizzazioni' si potrebbe parlare di 'istituzioni'; il secondo termine ci sembra però potenzialmente fuorviante, sia perché in diverse lingue si tende a utilizzare lo stesso termine per indicare istituti e istituzioni, sia perché il termine 'istituzione' non è comunemente applicato a certi tipi di organizzazioni, quali le imprese.

4.3 Le dichiarazioni

Possiamo ora ritornare all'istituto del linguaggio per definire l'ultimo tipo di atti comunicativi: le *dichiarazioni*. Presentiamo subito la nostra definizione, cui seguiranno alcuni commenti:

dichiarazione (*a*: mittente; *p*: contenuto; *i*: istituto) =
condizioni *p* è un fatto istituzionale definito in *i*;
p non sussiste nello stato corrente;
a è autorizzato in *i* a dichiarare *p*;
effetto *p* sussiste.

La prima osservazione è che fra gli argomenti dell'azione manca un destinatario. Le dichiarazioni, infatti, non sembrano avere un destinatario ben identificato; ad esempio, quando un giudice dichiara assolto un imputato, la dichiarazione è fatta *nel* tribunale, ma non al tribunale, né tanto meno all'imputato. Del resto le dichiarazioni, al contrario degli atti comunicativi visti precedentemente, non creano in generale un oggetto istituzionale che leghi il mittente a uno specifico destinatario. La seconda osservazione è che una dichiarazione, in generale, presuppone un istituto extralinguistico, tipicamente reificato da un'organizzazione specifica. Nell'esempio precedente viene presupposto l'istituto della giustizia, reificato dall'organizzazione giudiziaria; nella dichiarazione del vincitore di un'asta viene presupposto l'istituto delle aste, reificato da una specifica casa d'aste; e così via.

Abbiamo detto che "in generale" un istituto è reificato da un'organizzazione. Ciò non è però vero in senso assoluto: certi istituti, infatti, sono attivi nella società umana senza bisogno di alcuna organizzazione esplicita che ne delimiti il funzionamento; l'esempio principe di questi *istituti impliciti* è, naturalmente, il linguaggio umano. È interessante notare che anche gli istituti impliciti possono sostenere un sistema di dichiarazioni; infatti, le ben note *esecuzioni performative* di atti illocutivi non sono altro che dichiarazioni, ma non presuppongono in generale alcun istituto extralinguistico (Searle 1975a). Nel mondo degli agenti artificiali, tuttavia, ci sembra che le cose vadano diversamente: qui anche l'istituto del linguaggio sarà presumibilmente reificato da un'organizzazione esplicita, come ad esempio il ben noto W3C (World Wide Web Consortium). Resta il fatto, forse importante anche dal punto di vista applicativo, che tutti gli atti comunicativi definiti nel paragrafo 3.3 possono essere realizzati da un sistema di dichiarazioni. Vediamo come ciò sia possibile.

Gli atti comunicativi come dichiarazioni

È possibile definire un linguaggio per agenti artificiali in modo tale che gli atti comunicativi siano eseguiti in modo performativo, ovvero come dichiarazioni. Il primo passo è assumere come predefinito l'istituto del linguaggio, con particolare riguardo alle operazioni per la manipolazione dell'impegno e alle relative autorizzazioni. A questo punto, gli atti comunicativi possono essere definiti nel modo seguente:

assertivo (*a*: mittente; *b*: destinatario; *p*: contenuto) =
dichiarazione (*a*: mittente; *mc(a,b,p)*: contenuto; *linguaggio*: istituto);

Colombetti, Fornara e Verdicchio

commissivo (a : mittente; b : destinatario; p : contenuto) =
dichiarazione (a : mittente; $mc(a,b,p)$: contenuto; *linguaggio*: istituto),
dove p descrive un'azione futura del mittente;

direttivo (a : mittente; b : destinatario; p : contenuto) =
dichiarazione (a : mittente; $mp(a,b,p)$: contenuto; *linguaggio*: istituto),
dove p descrive un'azione futura del destinatario.

Il vantaggio di questo approccio è che diviene necessario un solo tipo di atto comunicativo primitivo, la dichiarazione. L'estensione del linguaggio con nuovi atti comunicativi diventa allora molto semplice: è sufficiente introdurre nuovi tipi di dichiarazioni, nel rispetto delle autorizzazioni stabilite dall'istituto del linguaggio.

5 Conclusioni

Nei paragrafi precedenti abbiamo introdotto un insieme di nozioni che riteniamo essenziali per lo sviluppo di sistemi multiagente; ci sembra quindi interessante chiedersi come i concetti sin qui delineati si possano tradurre in un sistema informatico funzionante. Il primo passo, crediamo, è produrre un *modello logico-formale* che definisca in modo rigoroso i concetti di impegno e preimpegno, di messaggio e di atto comunicativo, di istituto e di organizzazione. Allo stadio di sviluppo attuale abbiamo proposto un modello logico dell'impegno basato una logica temporale ramificata (Colombetti 2000c; Verdicchio e Colombetti 2003); in particolare questo modello formalizza le operazioni elementari sugli impegni e mostra come una libreria di atti comunicativi possa essere definita in termini di queste operazioni.

Nella comunità che si occupa degli agenti artificiali è diffusa l'opinione che i modelli logico-formali si giustifichino soprattutto come strumenti per la dimostrazione di proprietà generali dei sistemi multiagente; noi preferiamo vedere tali modelli anche come un primo passo verso una metodologia di progetto e sviluppo di prototipi concreti. Per raggiungere questo scopo, tuttavia, non ci si può fermare al livello della formalizzazione logica: la distanza fra un modello logico e il progetto di un sistema software è troppo grande. Per accorciare questa distanza è opportuno sviluppare un *modello operativo*, rispetto al quale il modello logico si comporta come una specifica. Rispetto a un modello logico un modello operativo, espresso nei termini di un'opportuna macchina astratta, ha il vantaggio di mostrare, almeno in termini generali, come i concetti formalizzati possano essere realizzati da un sistema computazionale. I primi passi in questa direzione sono descritti in Fornara e Colombetti (2002, 2003).

Dedica e ringraziamenti

Questo articolo è dedicato con affetto a Marco Somalvico, pioniere dell'intelligenza artificiale in Italia e fondatore del Progetto di intelligenza artificiale

Linguaggio e realtà sociale nei sistemi di agenti artificiali

del Politecnico di Milano, la cui recente scomparsa ha lasciato un grande vuoto fra i suoi colleghi e amici.

Numerosi colleghi ci hanno aiutato, con commenti e critiche a precedenti lavori, a sviluppare il punto di vista esposto in questo articolo. In particolare vorremmo ringraziare Antonella Carassa, Cristiano Castelfranchi, Andrew Jones, Alessio Lomuscio, Eddo Rigotti, Andrea Rocci, Peter Schulz e Marek Sergot. Francesco Viganò sta curando una prima implementazione del modello come estensione dell'ambiente JADE (Java Agent Development Environment).

Riferimenti Bibliografici

- Chaib-draa, B. e Vanderveken, D. (1998), "Agent Communication Language: Toward a semantics based on success, satisfaction, and recursion" , in *Proc. ATAL 1998*, pp. 363–379.
- Cohen, P.R. e Levesque, H.J. (1990), "Rational interaction as the basis for communication", in Cohen, P.R., Morgan, J. e Pollack, M.E. (eds.), *Intentions in communication*, MIT Press, Cambridge, MA, pp. 221–256.
- Colombetti, M. (1999), "Semantic, normative and practical aspects of agent communication", in *Proc. IJCAI 1999 Workshop on Agent Communication Languages*, Stockholm, pp. 51–62.
- Colombetti, M. (2000a), "A commitment-based approach to agent speech acts and conversations", in *Proc. Autonomous Agents 2000 Workshop on Agent Languages and Conversation Policies*, Barcelona.
- Colombetti, M. (2000b), "A language for artificial agents", in *Studies in Communication Sciences* 1, (1–32) (scaricabile da www.scoms.ch).
- Colombetti, M. (2000c), "Commitment-based semantics for Agent Communication Languages", in *Proc. 1st International Workshop on the History and Philosophy of Logic, Mathematics and Computation (HPLMC-00)*, San Sebastian, E.
- Colombetti M., Fornara, N. e Verdicchio, M. (2002), "The role of institutions in multiagent systems", in *Atti del VII convegno dell'Associazione italiana per l'intelligenza artificiale (AI*IA 02)*, Siena.
- Colombetti, M. e Verdicchio, M. (2002), "An analysis of agent speech acts as institutional actions" , in *Proc. AAMAS 2002*, Bologna, pp. 535–542.
- d'Inverno, M. e Luck, M. (2001), *Understanding agent systems*, Springer, Berlin.
- Ferber, J. (1999), *Multi-Agent Systems: An introduction to Distributed Artificial Intelligence*, Addison-Wesley, Harlow, UK.
- Finin, T., Labrou, Y. e Mayfield, J. (1995), "KQML as an agent communication language", in Bradshaw, V. (ed.), *Software agents*, MIT Press, Cambridge, MA.
- Fornara, N., e Colombetti, M. (2002), "Operational specification of a commitment-based communication language", in *Proc. of AAMAS 2002*, Bologna, pp. 1157–1164.
- Fornara, N. e Colombetti, M. (2003), "Defining interaction protocols using a commitment-based Agent Communication Language", in *Proc. of AAMAS 2003*, Melbourne (accettato per la pubblicazione).
- Hewitt, C., Bishop, P. e Steiger, R. (1973), "A Universal Modular ACTOR Formalism for Artificial Intelligence", in *Proc. IJCAI 1973*, pp. 235–245.
- Jones, A. e Sergot, M.J. (1996), "A formal characterisation institutionalised power", in *Journal of the IGPL*, 4, pp. 429–445.

Colombetti, Fornara e Verdicchio

- Perrault, C.R. e Allen, J.F. (1980), "A plan-based analysis of indirect speech acts" in *American Journal of Computational Linguistics*, 6, pp. 167–182.
- Searle, J.R. (1969), *Speech Acts*, Cambridge University Press, Cambridge, UK.
- Searle, J.R. (1975a), "A taxonomy of illocutionary acts" in Gunderson, K. (ed.), *Language, mind, and knowledge (Minnesota Studies in the Philosophy of Science VII)*, University of Minnesota Press, pp. 344–369.
- Reprinted in Searle, J.R. (1979), *Expression and meaning*, Cambridge University Press, Cambridge, UK, pp. 1–29.
- Searle, J.R. (1975b), "Indirect speech acts", in Cole, P. e Morgan, J. (eds.), *Syntax and semantics 3: Speech acts*, Academic Press.
- Reprinted in Searle, J.R. (1979), *Expression and meaning*, Cambridge University Press, Cambridge, UK, pp. 30–57.
- Searle, J.R. (1995), *The construction of social reality*, Simon & Schuster.
- Searle, J.R. e Vanderveken, D. (1985), *Foundations of illocutionary logic*, Cambridge University Press, Cambridge, UK.
- Singh, M.P. (1998), "Agent communication languages: Rethinking the principles", in *IEEE Computer* 31, pp. 40–47.
- Singh, M.P. (2000), "A social semantics for Agent Communication Languages", in *Issues in Agent Communication*, pp. 31–45.
- Verdicchio, M. e Colombetti, M. (2002), "Commitments for agent-based supply chain management", in *ACM SIGecom Exchanges*, 3-1, pp. 13–23.
- Verdicchio, M. e Colombetti, M. (2003), "A logical model of social commitment for agent communication", in *Proc. AAMAS 2003*, Melbourne (accettato per la pubblicazione).
- Weiss, G. (1999), *Multiagent systems: A modern approach to distributed artificial intelligence*, MIT Press, Cambridge, MA.
- Wooldridge, M. (2000), *Reasoning about rational agents*, MIT Press, Cambridge, MA.
- Wooldridge, M. (2002), *An introduction to multiagent systems*, John Wiley, Chichester, UK.