

UNIVERSIDAD NACIONAL DEL SUR



TESIS DE MAGISTER EN  
CIENCIAS DE LA COMPUTACIÓN

**Especificación de los actos del habla  
en sistemas multi-agente**

Guillermo Carlos Aguirre

BAHÍA BLANCA

ARGENTINA

2018



# Prefacio

Esta Tesis es presentada como parte de los requisitos para optar al grado académico de Magister en Ciencias de la Computación, de la Universidad Nacional del Sur, y no ha sido presentada previamente para la obtención de otro título en esta Universidad u otras. La misma contiene los resultados obtenidos en investigaciones llevadas a cabo en el Departamento de Ciencias de la Computación de la UNSL, bajo la dirección del Dr. Guillermo R. Simari, Profesor Titular del Departamento de Ciencias de la Computación en la UNS y del Dr. Marcelo Luis Errecalde, profesor Asociado del Departamento de Informática de la UNSL.

Guillermo Carlos Aguirre

`gaguirre@unsl.edu.ar`

DEPARTAMENTO DE INFORMÁTICA  
UNIVERSIDAD NACIONAL DE SAN LUIS

San Luis, Octubre del 2017.



# Agradecimientos

Agradezco a la UNSL y a la UNS el haberme abierto sus puertas de manera tan generosa, brindando tanto, conocimiento por parte de su gente, como cobijo en sus aulas. De igual manera lo hago extensivo a todas las instituciones educativas públicas en las que estudié. Formar parte del LIDIC representa integrar un grupo que transita por ámbitos del conocimiento desafiantes, que nos estimulan a superarnos y es un privilegio estar ahí. Mi más profundo agradecimiento a los miembros de mi familia, por el imprescindible apoyo afectivo que significan.

Son muchas las personas que me han estimulado a realizar este trabajo, principalmente agradezco al Dr Guillermo Simari, por haberme dado la oportunidad y por establecer el nivel de exigencia del trabajo. Invalorable ha sido contar con la colaboración del Dr Marcelo Errecalde, ya que supo transmitirme su entusiasmo y orientación en la investigación. Destaco el valioso aporte que tuvieron para mí los debates sobre temas de argumentación con el Dr Edgardo Ferretti, su aporte ha sido irremplazable. También agradezco al Dr Alejandro Grosso por servirme de modelo como profesor y apego al trabajo. Finalmente, agradezco a alguien que con su legado me transmitió el amor por investigar y enseñar, el Dr Raúl Gallard.



# Resumen

En el vasto campo de estudio sobre la comunicación entre humanos, un aspecto destacado es el rol que desempeña la pragmática en el uso del lenguaje; más específicamente la teoría de los actos del habla. Una parte importante de la interpretación pragmática de un enunciado en lenguaje natural, es la elección del acto del habla correcto para el contexto en que se usa. Adaptar las técnicas de comunicación, tan naturales para los humanos, al ámbito de la Inteligencia Artificial no es algo sencillo. Tal es así que un pionero como Alan Turing estableció un fuerte vínculo entre comunicación e inteligencia. De ahí que para lograr que los agentes computacionales se comuniquen, es necesaria una adecuada integración de varios procesos. Para clarificar algunas de estas ideas, mostrar distintos tipos de ejemplos y para motivar la lectura, se presenta una situación ficticia adecuada para evaluar los intercambios de actos del habla y los conflictos que pueden surgir cuando los interlocutores son agentes. En este marco ficticio, los agentes se emplean para analizar los diálogos entre soldados que operan en una coalición internacional en un país ficticio de oriente medio.

Más allá de la flexibilidad que caracteriza a los agentes, existe mucha investigación sobre el uso de recetas o planes para guiarlos hacia determinados objetivos o metas. Los pasos o progresos que realiza el agente muchas veces involucran el intercambio de actos del habla con otros pares. El progreso en ese intercambio se consigue empleando “operadores” o mensajes que se pueden aplicar siempre que se disponga de un determinado conocimiento. Eso es algo propio de los lenguajes de comunicación de agentes conocidos, como KQML y ACL, por esa razón en su especificación de los mensajes, establecen cual es el estado mental requerido para emplear un mensaje determinado y que cambios se producirán al aplicarlo.

El ser social es una cualidad requerida en la definición de agente, asociado con esto está la idea de compromiso con los demás miembros del sistema multi-agente. Cada afir-

mación transmitida, cada requerimiento hecho, cada promesa realizada, se mantienen en un almacén de compromisos, gracias al cual los demás pueden formar un perfil o conjunto de antecedentes sobre el cual es posible razonar. Pero también es necesario considerar los problemas que introduce la transmisión, como demoras y extravíos de los mensajes, por los que seguramente surgen diferencias en las percepciones de los interlocutores. Se deben cumplir ciertas condiciones que garanticen que los agentes tengan las mismas observaciones para exigir que, si alguien es acreedor de un compromiso, el correspondiente deudor se asuma como tal. Esta situación, en la que los compromisos están alineados, es deseable; resulta conveniente, en consecuencia, establecer los requisitos para que la evolución del sistema vaya en esa dirección.

Para lograr que los agentes realmente se puedan comunicar, es necesario que empleen razonamiento. Si en la búsqueda de cumplir con sus metas, los agentes deben superar situaciones de conflicto, como ser diferencias en los puntos de vista individual, la comunicación es un medio que ayuda. Es conocida la tradición del diálogo para lograr acuerdos, por ejemplo, entre un feudal y su vasallo o en los ámbitos legales. Se requieren reglas claras que motiven el avance hacia una resolución que sea aceptada por las partes, reglas con las cuales se pueda razonar, hacer afirmaciones, requerir fundamentos o argumentos y plantear contrargumentos. Estos entornos son conocidos como juegos de diálogo, en los cuales se usan distintos tipos de lógicas para soportar las conclusiones que son obtenidas a medida que progresa el intercambio de mensajes. En particular la persuasión resulta de interés entre agentes independientes, aquellos que mantienen una interacción ocasional; como puede ser el caso de los soldados en el contexto ficticio introducido para recrear ejemplos.



# Abstract

Communication among humans beings is a widespread topic. In this context, Pragmatics plays a very important role in establishing how language is used, in which Speech Acts Theory stands out. An important part of pragmatic interpretation in natural language utterances is the choice of the appropriate speech act for each context. Adapting human communication techniques, which are so natural for humans, is not simple in the area of Artificial Intelligence. Alan Turing, a pioneer in this field, established a strong link between communication and intelligence. Thus, the adequate integration of several processes is necessary in order to have computer agents communicate. To clarify some of these ideas, show diferent types of examples, and motivate reading, an appropriate fictitious situation is presented. This is to evaluate the exchange of speech acts and problems that may arise when the interlocutors are agents. In this fictitious framework, agents are used to analyze the dialogues between soldiers who are part of an international coalition in a fictitious Middle Eastern country.

Beyond the flexibility that characterizes the agents, there are plenty of research works on the use of recipes or plans to guide agents towards certain objectives or goals. The exchange of speech acts with other peers is often involved in the steps or progress made by the agents. Progress in this exchange is achieved by using “operators” or messages that can be send as long as certain knowledge is available. This characteristic is typical in classical agent communication languages, such as KQML and ACL. For this reason, the specification of messages in these languages establish both; the mental state required in order to employ a certain message, and the changes produced when applying the message.

A very important characteristic required in the definition of an agent is to be a social being accompanied by the idea of commitment with other members of the multi-agent system. Every statement transmitted, every request made, and every promise done are kept in a commitment store, through which, others agents can create a profile or set of

backgrounds from which it is possible to reason. Besides it is necessary to consider the problems introduced by the transmission, such as delays and misled messages, which are likely to foster differences among the interlocutors perceptions. It is essential the existence of specific conditions to ensure that all agents have the same observations. That is to say; if someone is a creditor of a commitment, the corresponding debtor assumes himself as such. The situation in which the commitments are aligned, is desirable. Therefore, it is convenient to establish the conditions for the system to evolve in the direction of the alignment.

In order to achieve that the agents can successfully communicate, it is necessary that they reason. In the pursuit of fulfilling their goals, agents must overcome situations of conflict, such as differences in individual points of view in which communication is a means that helps. The dialogue has a very well known tradition to reach agreements, for example, between a feudal and his vassal or in legal matters. Clear rules are required to motivate the progress in the potential resolution accepted by the parties, rules by which it is possible to reason, make statements, demand arguments or propose counterarguments. These ruling environments are known as dialogue games in which different types of logics are used to support the inferences obtained, while the exchange of messages progresses. In this context, persuasion is of a particular interest among independent agents who maintain occasional interactions. This may be the case of soldiers in the fictional context introduced to recreate the examples

# Índice general

<b>1. Introducción</b>	<b>1</b>
1.1. Organización de la tesis . . . . .	3
<b>2. Consideraciones sobre comunicación</b>	<b>5</b>
2.1. Comunicación entre seres humanos . . . . .	5
2.2. Comunicación en Inteligencia Artificial . . . . .	12
2.3. Comunicación en un escenario ficticio . . . . .	16
<b>3. La perspectiva mental</b>	<b>19</b>
3.1. Los actos del habla basados en planes . . . . .	19
3.1.1. El operador REQUERIR . . . . .	20
3.1.2. El operador INFORMAR . . . . .	23
3.1.3. Plan para INFORMAR un DESEO . . . . .	25
3.1.4. REQUERIR vs INFORMAR un DESEO . . . . .	27
3.2. Lenguaje KQML . . . . .	28
3.3. Lenguaje ACL. Actos del habla según FIPA . . . . .	32
<b>4. Perspectiva Social</b>	<b>39</b>
4.1. Comunicación desde la perspectiva social . . . . .	39
4.1.1. Uso de los compromisos en la comunicación . . . . .	40
4.2. La alineación de los compromisos . . . . .	41

4.2.1.	Primera Interpretación del efecto de los mensajes . . . . .	42
4.2.2.	El origen de los problemas . . . . .	44
4.2.3.	Relación entre compromisos . . . . .	44
4.2.4.	Transacción independiente . . . . .	45
4.2.5.	Predicados en las transacciones . . . . .	45
4.3.	Postulados para razonar sobre los compromisos . . . . .	47
4.3.1.	Cuando es posible dar de baja un compromiso . . . . .	48
4.3.2.	Desconectar una parte del antecedente . . . . .	48
4.3.3.	Aumento de la fuerza del antecedente . . . . .	48
4.3.4.	Unificar dos antecedentes mediante disjunción . . . . .	49
4.3.5.	Unificar dos consecuentes mediante conjunción . . . . .	49
4.3.6.	No existe un compromiso cuyo consecuente sea falso . . . . .	50
4.3.7.	El consecuente no será inferencia directa del antecedente . . . . .	50
4.3.8.	Relajación en el consecuente . . . . .	51
4.4.	Características técnicas del enfoque social . . . . .	51
4.4.1.	Mensajes y observaciones . . . . .	52
4.4.2.	Instantánea de los agentes y mantenimiento local . . . . .	53
4.5.	Estados del sistema relevantes para la alineación . . . . .	55
4.6.	Progreso hacia estados de alineación . . . . .	56
<b>5.</b>	<b>Perspectiva Argumentativa</b>	<b>63</b>
5.1.	Los sistemas de diálogo . . . . .	67
5.1.1.	Sistema de persuasión . . . . .	71
5.1.2.	Una instancia de sistema de persuasión . . . . .	74
<b>6.</b>	<b>Conclusiones</b>	<b>79</b>

<b>A. El sistema argumentativo PraTor</b>	<b>85</b>
A.1. Introducción . . . . .	85
A.2. El sistema PraTor . . . . .	88
A.3. Formas dialécticas de argumentación . . . . .	92



# Capítulo 1

## Introducción

La comunicación de agentes es el estudio de como dos o más entidades de software pueden comunicarse entre sí. Los aspectos de investigación en este dominio son profundos y antiguos. Uno de los desafíos es la dificultad de asignar significado a las frases, debido que el significado preciso de una frase depende de varios factores: el contexto en el cual es emitida, la posición en la secuencia de sentencias previas, la naturaleza de la sentencia (por ejemplo una proposición, un compromiso para tomar alguna acción, un requerimiento, etc), los objetos mencionados en la sentencia (tales como un objeto del mundo real, un estado mental, un estado del mundo futuro, etc) y las identidades del orador y del oyente. Otro desafío, quizás insuperable, es la verificación semántica: cómo verificar el significado de lo que dice un agente en una frase. En un sistema de agentes heterogéneos, un agente normalmente no es capaz de ver el código interno de otro agente a fin de verificar sus afirmaciones; aun si fuese posible, un agente suficientemente inteligente podría siempre simular cualquier estado mental deseado, al ser inspeccionado por otro agente.

Resulta clave para este área, la necesidad de estudiar las teorías relevantes sobre la temática y desarrollar un marco unificado para ellas. En particular, es necesaria una teoría formal del lenguaje de agentes y protocolos, capaz de estudiar exhaustivamente las propiedades que éstos poseen y que permita comparar rigurosamente un lenguaje o un protocolo con otro. En resumen, se necesita avanzar hacia un mejor entendimiento de la correcta integración de diferentes lenguajes de comunicación de agentes, lenguajes de contenido y protocolos en los distintos posibles dominios de aplicación, para que los resultados de la investigación sean adoptados más ampliamente.

El estudio de los lenguajes y protocolos se puede comenzar a partir de un componente elemental como son las frases o sentencias. La naturaleza de una sentencia puede ser caracterizada de distintas maneras: una primer aproximación es usar la comunicación para cuestiones básicas como transmitir un conocimiento o deseo interno del agente; también para requerir o preguntar algo a otro agente. Un segundo enfoque es comunicarse para coordinar acciones entre los agentes, lo cual requiere un mínimo entendimiento mutuo y confianza. Este requerimiento representa un desafío importante y difícil de superar, ya que no es una tarea sencilla llegar a verificar la veracidad de las afirmaciones que se realizan durante la comunicación. En ese sentido, el uso de convenciones como los compromisos asumidos de manera social representan una de las herramientas más usadas para cimentar confianza. El tercer enfoque es poder razonar sobre las implicancias que tienen las frases hechas durante un diálogo. Para eso es necesario establecer una definición precisa de diálogo, ¿cuáles son las restricciones impuestas al intercambio de frases?, ¿cómo se pueden emplear distintos tipos de formalismos para reducir al mínimo posible las diferencias de interpretación?. Una posibilidad para enfrentar este desafío es usar diálogos de persuasión, donde un agente tiene la posibilidad de persuadir a otro agente sobre la honestidad de su afirmación.

En líneas generales, el párrafo anterior presenta el contexto de investigación de nuestro trabajo. Sobre este contexto existe mucha bibliografía en la que se discuten las tres perspectivas de comunicación introducidas: mental, social y argumentativa [Woo09, Kib06, CL85, Pra09]. Por encima de nuestra área de principal interés está la Inteligencia Artificial, donde también se aborda la comunicación desde una perspectiva más amplia, como ser el uso de la comunicación en ausencia de convenciones conocidas por todos los interlocutores. En ocasiones no es necesario recurrir a mensajes o frases para lograr que otro agente adopte determinados planes ya que lo mismo se puede lograr mediante la aplicación de leyes sociales [RN03]. Existen numerosas formas de razonar frente a los hechos, no sólo usando alguna variante de las lógicas conocidas, sino también es posible hacerlo recurriendo a la teoría de probabilidad.

La motivación principal del trabajo es averiguar si existe alguna forma de unir las tres perspectivas de comunicación. Esa “forma” consiste en encontrar algún elemento que posibilite la integración de esos tres enfoques. La búsqueda pretende determinar cual es el recurso sobre el cual estructurar una visión amplia que abarque a las otras.

Para llegar a un entendimiento más profundo del problema y al mismo tiempo cubrir



algunos de los temas de investigación más relevantes, es que se aborda cada perspectiva por separado. En ese camino se pretende identificar las características distintivas de cada aproximación y resaltar los recursos principales usados en cada una de ellas.

Como hipótesis que guía y motiva nuestro trabajo, pensamos que una única perspectiva no es suficiente para ayudar a que un agente, en distintas circunstancias, decida cual es el acto del habla adecuado para usar. Entendemos que los agentes deben emplear los estados mentales individuales y los elementos conocidos por toda la sociedad de agentes como recursos con los cuales razonar y poder elegir las opciones de comunicación más convenientes.

## 1.1. Organización de la tesis

El resto de la tesis se organiza de la siguiente manera.

**Capítulo 2.** Consideraciones sobre comunicación, en tres ámbitos diferentes:

\**La comunicación entre los seres humanos*, brinda una rápida descripción de dos enfoques sobre el tema: la teoría de los actos del habla y la acción comunicativa. En el primero se describen los actos y las condiciones que deben estar presentes en la comunicación. En el segundo se consideran los compromisos sociales y los requisitos para cuestionar una frase.

\**La comunicación en inteligencia artificial*: se discuten algunas características generales que deben estar presentes en la comunicación entre los agentes computacionales. Estas características se sintetizan en siete procesos, algunos realizados por el emisor y otros por el oyente. Se menciona un esfuerzo científico para estandarizar la comunidad de agentes.

\**La comunicación en un escenario ficticio*: Holistan es el nombre elegido para denominar una región de oriente medio donde distintas fuerzas internacionales intervienen para mantener el orden. Los distintos actores se encuentran en diferentes situaciones de colaboración, intercambio de información y de mutua dependencia.

**Capítulo 3.** La perspectiva mental, discute cómo el conocimiento incorporado en los agentes se emplea para definir operadores usados en planes. Los lenguajes de comunicación de agentes más conocidos también recurren a la “mente” de los agentes para definir los distintos mensajes. En el caso de ACL, esta definición emplea una lógica multimodal.

**Capítulo 4.** La perspectiva social, procura llegar a una definición para caracterizar el momento en el cual los compromisos de los agentes estén *alineados*, una situación que llevada al ámbito de los sistemas distribuidos, resulta igualmente requerida. También se introduce el *almacén de compromisos* como un recurso que le permite a los agentes razonar sobre el diálogo actual.

**Capítulo 5.** La perspectiva argumentativa, es el capítulo donde se introducen los juegos de diálogo, se menciona como evolucionaron hasta convertirse en una serie de reglas o procedimientos formales y de que manera se usan para regular los diálogos entre agentes. Aquellos juegos de diálogo cuya formalización recurre a una lógica formal deductiva, presentan ciertas limitaciones sobre el tipo de inferencias que pueden hacer y tienen el inconveniente de no permitir plantear contraargumentos. Por su parte, aquellos que usan razonamiento rebatible, pueden emplear la argumentación para atacar las frases del interlocutor en más de un modo. Para esta última clase, se desarrolla un formalismo en un protocolo particular y se dan ejemplos de uso.

**Capítulo 6.** Conclusiones, sintetiza las principales ideas presentadas en cada uno de los tres capítulos dedicados a las perspectivas de comunicación consideradas. Como resultado de la elaboración realizada sobre cada perspectiva, se desprende la contribución principal: la comunicación efectiva requiere razonar con sentido común. La veracidad de esta afirmación se sustenta en el hecho que este tipo de razonamiento se puede integrar en los distintos puntos de vista considerados y permite adecuar las características de los actos del habla al ámbito de los agentes computacionales.

**Apéndice A.** El sistema argumentativo PraTor, describe un recurso pensado para usar en ámbitos legales, tiene la particularidad de que permite razonar sobre la prioridad que existe entre las reglas. PraTor tiene particular importancia por ser empleado como contexto de argumentación en uno de los juegos de diálogo del Capítulo 5.

# Capítulo 2

## Consideraciones sobre comunicación

En este capítulo se eligieron tres ámbitos para comenzar con la discusión general del tema. Uno más amplio y de conocimiento general, descrito en la primer sección, otro más específico y un poco más técnico, desarrollado en la segunda sección, donde se establece el contexto general que sirve para ayudar a interpretar el resto de los capítulos. El tercer ámbito es donde se pondrán a prueba algunas de las consideraciones más importantes para evaluar el uso de los actos del habla en la comunicación de los agentes computacionales.

### 2.1. Comunicación entre seres humanos

El estudio del lenguaje natural usualmente asume una división del campo en tres áreas: *sintaxis*, *semántica* y *pragmática*. La sintaxis es el estudio de las expresiones visibles del lenguaje, las palabras, y las reglas que establecen cómo estas palabras se emplean para crear expresiones bien formadas. Por su parte la semántica y la pragmática se combinan para determinar cual es el significado otorgado a las expresiones del lenguaje. La semántica frecuentemente es considerada como el verdadero núcleo de condiciones de la sentencia: condiciones bajo las cuales la sentencia, o la proposición que expresa, es verdadera. La semántica de una expresión, por lo tanto, define su significado literal - el aspecto del significado de la expresión que es común a través de todo contexto de uso. La pragmática por otro lado, está relacionada con aquel aspecto del significado que surge desde contextos específicos de uso, y cómo este contexto contribuye tanto al significado total como a los efectos de una frase. Aquí, “contexto” se refiere a las circunstancias relevantes en las

cuales un acto comunicativo específico ocurre, tales como el discurso previo, el entorno físico, el entorno social, propiedades del orador y del oyente, suposiciones compartidas, objetivos retóricos y comunicativos de los participantes, etc. Así, la pragmática considera factores más allá de las condiciones estáticas de verdad especificadas por una semántica particular - los factores que determinan aquellos aspectos de la interpretación del lenguaje natural que dependen de un entorno comunicativo particular.

Un subcampo destacado dentro de la pragmática es la teoría de los *actos del habla* (en inglés *speech acts*). Una parte importante de la interpretación pragmática de un enunciado en lenguaje natural es la determinación de cual es el acto del habla (o acto ilocucionario, como se explica más adelante) usado en una expresión determinada. Por ejemplo la frase: “Hace frío en esta habitación”, tiene un análisis sintáctico y un significado literal, o semántico, que se mantienen a través de todos los usos posibles. Digamos que la temperatura de la habitación es baja de acuerdo al criterio del orador. Sin embargo el acto del habla que el orador intenta realizar al usar esta frase depende del contexto en el que fue enunciada. La sentencia pudo ser usada para establecer un hecho, para requerirle al oyente que cierre la ventana, para que el oyente no entre en la habitación, o para otra clase de propósito. En efecto las frases en lenguaje natural son usadas frecuentemente y de manera consciente, para varios propósitos al mismo tiempo. Cada tipo de meta que el orador persigue con un enunciado particular se corresponde, de alguna manera, con el *tipo* de acto del habla elegido para ese enunciado. Por este motivo, hay actos del habla para informar, para hacer un requerimiento, etc.

Los actos comunicativos tienen su origen en la *teoría de los actos del habla* [Aus62] a comienzo de los años 60. A partir de allí se produce un cambio en la manera de analizar los enunciados empleados en la comunicación. Hasta entonces, solamente se rescató la valoración lógica (como *correctas* o *incorrectas*) que pudiera hacerse sobre las frases, reflejando con este enfoque el pensamiento positivista propio de la época. Austin en cambio, resalta la importancia de ampliar el análisis del lenguaje hacia un universo mucho más amplio: el lenguaje cotidiano, es decir, rescatar el valor pragmático de las frases que la gente común usa a diario. Una pregunta, una promesa o una orden tienen sentido más allá de clasificarlas como verdaderas o falsas; son *exitosas* en la medida que consiguen las intenciones del orador.

Austin es ampliamente reconocido como el inventor del concepto *Teoría de los Actos del Habla*. Él examinó el uso de los *enunciados ejecutivos* en el lenguaje, investigó las

diferencias entre sentencias declarativas (aquellas que se pueden constatar) y las sentencias realizables: “Las sentencias comprobables tienen la propiedad de ser verdaderas o falsas. Las sentencias realizables, por su parte, carecen de esas dos propiedades: tienen su propia tarea especial: se usan para realizar una acción. Emitir esta clase de sentencia es realizar la acción - una acción que difícilmente se podría hacer de otra manera.” Uno de los ejemplos clásicos de sentencia realizable es: “Llamo a este barco el Reina Elizabeth” (sentenciado mientras se rompe una botella de champán contra el casco). “Parece claro que pronunciar esa sentencia (desde luego, bajo las circunstancias apropiadas), no es describir lo que se está haciendo... es hacerlo. ...[Una sentencia ejecutiva] indica que la declaración de la frase, es la realización de una acción - en general no es considerado simplemente como decir algo”.

La teoría de los actos del habla se desarrolló como una serie de seminarios dictados por Austin en Oxford entre 1952 y 1954 bajo la denominación de “Palabras y Hechos” (en inglés Words and Deeds) y las conferencias William James en la Universidad de Harvard durante el año 1955. Luego de su muerte estos fueron publicadas en un libro: “Cómo hacer cosas con palabras”, el cual es considerado como la referencia estándar del trabajo de Austin sobre los actos del habla.

Según esta teoría, el oyente podrá interpretar adecuadamente una frase en la medida que entienda que pretende lograr el orador con esa frase, la cual forma parte de un plan que tiene quien la dijo. Es decir, conviene concentrarse en entender “porque” se usó una frase, más que analizar la secuencia de palabras que la conforman ya que las palabras sueltas no pueden asociarse con un plan [BN78].

Surge así una nueva percepción del *lenguaje como acción*. Cuando un juez dicta sentencia en la corte y dice: “Culpable”, se produce un cambio en el mundo real tan concreto como el producido por cualquier acción física.

Se pueden reconocer en la propuesta austiniana dos ejes fundamentales: a) la consideración del lenguaje operando sobre lo real, lo concreto -no lo simbólico- y b) una nueva percepción del éxito de un enunciado en función de llegar a satisfacer las intenciones que le dieron origen. En la teoría de los actos del habla se identifican tres actos o dimensiones para cada enunciado:

- el acto de emitir los sonidos correspondientes al vocabulario usado, el acto *locutorio*.

- el acto *illocutorio* que se lleva a cabo **cuando** se dice algo: prometer, afirmar, amenazar, y
- el acto *perlocutorio* que se produce **como consecuencia** de decir algo: intimidar, asombrar, convencer, ofender.

La realización exitosa de un enunciado ejecutivo (en inglés *performative*) requiere satisfacer ciertas condiciones relativas a las circunstancias y personas involucradas, las que deben participar del procedimiento completo y de una manera sincera (sin actuar). Estas condiciones fueron refinadas por John Rogers Searle a fin de establecer el contexto más propicio para lograr los resultados esperados. Searle profundizó el trabajo sobre la teoría de los actos del habla, pero también estableció algunas diferencias. Él plantea que no existe una frase con fuerza illocucionaria neutral. Es decir, cada sentencia tiene una potencial fuerza illocucionaria concebida con su significado. Searle establece que si bien los conceptos illocucionario y acto illocucionario son distintos, la diferencia conceptual no es suficiente para establecer una diferencia entre clases distintas de actos, y realmente cada acto locucionario es un acto illocucionario. El principal aporte de Searle fueron las denominadas *condiciones de los actos del habla*, que consistió en un conjunto de reglas extraídas de las condiciones necesarias y suficientes para que un acto del habla fuese exitoso y no presentara defectos al incorporarse como parte de una frase.

Tomando como punto de partida la teoría de los actos del habla, el filósofo alemán Jürgen Habermas propuso extensiones al trabajo de Austin y Searle en una teoría alternativa denominada *Teoría de acción comunicativa*[Ver97], la cual se aboca a la coordinación de acciones en sistemas sociales por medio del lenguaje. Habermas critica el trabajo de Searle en tres ejes principales: 1) diferencias entre lo que llama una *acción estratégica* y una *acción comunicativa*, 2) la consideración de los denominados *reclamos de validez* para las frases que se intercambian y 3) una taxonomía para los actos comunicativos propuestos diferente a la empleada para los actos del habla.

La distinción entre acción estratégica y acción comunicativa es ejemplificada por la distinción entre un *reclamo de poder* y un *reclamo de validez*. Por ejemplo el éxito de una frase imperativa (“Trae tu informe”) está basado en un reclamo de poder (por miedo a sanciones), y el éxito de una directiva (“¿Puedes cerrar la ventana?”) está basado en un reclamo de validez (donde prevalece la regulación social). De esta forma, sólo los actos comunicativos que permiten críticas a los reclamos de validez, motivan al oyente a aceptar

la acción requerida sin problema y de esta manera se consigue un mecanismo efectivo para la coordinación de las acciones.

Junto con los actos comunicativos que se intercambian, tanto en las acciones estratégicas como en las comunicativas, el orador siempre introduce algún reclamo de validez; este elemento es el que permite que se inicie la coordinación de las acciones. Dependiendo del tipo de acción iniciada por el orador, el oyente puede continuar con alguna de las alternativas posibles de comportamiento. Considerando que el reclamo dominante en las acciones estratégicas es el reclamo de poder, en ese caso el oyente sólo tiene dos alternativas para responder: SI o NO. Pero si la acción iniciada es comunicativa, quiere decir que el orador incluyó en la frase tres reclamos de validez: 1) establece que lo que dice es cierto (reclamo de verdad); 2) que la frase es adecuada en el contexto actual (reclamo de justicia) y 3) que la intención expresada es auténtica (reclamo de sinceridad).

Los reclamos de validez están inevitablemente ligados con las prácticas cotidianas de quienes dialogan y actúan. La posibilidad de criticar los reclamos de validez constituye el mecanismo adecuado para lograr la coordinación de las acciones de los interlocutores.

Un acuerdo mutuo se puede lograr a través de la comunicación, pero solamente si se han logrado fusionar las perspectivas de los individuos que intervienen, lo cual requiere llegar a un acuerdo sobre la validez de los actos comunicativos que se comparten. La teoría de acción comunicativa afirma que los actos comunicativos compartidos entre los individuos durante una comunicación conllevan tres tipos de pretensiones o reclamos de validez: verdad, justicia y sinceridad. Los tres, silenciosamente pero con insistencia, demandan ser justificados con fundamentos. El simple proceso de llegar a un entendimiento con otros individuos conlleva la necesidad de ser responsables de lo que cada uno dice y de ser capaces de justificar los reclamos o pretensiones de validez que se plantean en relación con las cuestiones objetivas, normativas y verificables.

Habermas identifica dos mecanismos excluyentes para coordinar acciones sociales: consenso e influencia. Su teoría es que el consenso constituye el mecanismo fundamental de coordinación social. El consenso está relacionado con la idea de *entendimiento* (en alemán *verständigung*), un término que puede ser empleado para indicar tanto comprensión como acuerdo.

La distinción entre consenso e influencia en la teoría de Habermas, se corresponde con la distinción entre *acción comunicativa* y *acción estratégica*.

En la *acción estratégica* cada agente tiene sus propios objetivos y planes. Cada individuo

evalúa la reacción que manifiestan los otros agentes y en consecuencia intentará convencerlos o derrotarlos. Los agentes tienen una motivación concreta: procurar maximizar sus ganancias y minimizar sus pérdidas. La coordinación está basada en contingencias experimentales especialmente en lo relativo al poder, tanto en el reclamo como en el uso del mismo.

En la *acción comunicativa* la coordinación se consigue por un entendimiento mutuo tanto de la situación como de los objetivos perseguidos. La acción comunicativa debiera ser interpretada como una “interacción entre sujetos que se involucran en una relación social”, con la idea de alcanzar un mutuo entendimiento de la situación, a fin de ser capaces de coordinar sus acciones. Su motivación es racional: la gente responde a los requerimientos porque presupone que esos requerimientos podrán ser justificados cuando sean cuestionados.

Los reclamos de validez definen las condiciones para la creación de compromisos. En principio cualquiera de los reclamos mencionados más arriba pueden ser cuestionados. Cuando el oyente realiza un planteo porque no se cumple con algún reclamo, el orador deberá brindar las razones que le sirven como soporte. Por ejemplo, un oyente puede desafiar la veracidad de los dichos del interlocutor, diciendo: “¿Qué razones tienes para decir eso?”. El oyente también puede desafiar el derecho del orador para decir lo que dice cuestionando la legitimidad del contexto normativo del enunciado. (reclamo de justicia). Por ejemplo, diciendo: “¿Qué razones tienes para decirme eso ahora?”. Otra posibilidad es que el oyente cuestione la sinceridad del orador, por ejemplo “¿Qué razones tienes para esperar que crea eso?”.

En los reclamos se distinguen tres mundos: objetivo, social y subjetivo. El reclamo de veracidad se refiere al mundo objetivo (como son las cosas), el reclamo de justicia se refiere al mundo social (como los participantes se posicionan en relación con los demás) y el reclamo de sinceridad se refiere al mundo subjetivo (como el orador percibe el mundo).

Las críticas de Habermas se refieren a los primeros trabajos de Searle, mejorados luego en publicaciones posteriores, donde *las condiciones de los actos del habla* proporcionan una forma para tratar con *los reclamos de validez*. Sin embargo las críticas básicas se mantuvieron, ya que para Searle un acto del habla puede llegar a ser exitoso considerando solamente el rol del orador, y en realidad de esa manera se oculta la negociación sobre los reclamos de validez. Por esa razón, Habermas entiende que Searle desconoce un principio que subyace en toda comunicación y que es lo que explica la esencia del éxito: los parti-



cipantes tienden a llegar a un acuerdo mutuo.

Otro punto de crítica es acerca de la clasificación de los actos del habla. Habermas critica a Searle por no considerar la distinción entre actos del habla basados en reclamos de poder y aquellos basados en reclamos de validez. También por no distinguir entre aquellos actos del habla que expresan un reclamo de justicia (tales como una promesa) y aquellos que expresan un reclamo de sinceridad (tales como una intención).

Habermas presenta una clasificación basada en una única dimensión, el reclamo dominante presentado por el orador. La distinción de Habermas entre la orientación práctica y la orientación racional de la comunicación se observa por ejemplo en que las órdenes directas, que no se pueden cuestionar, se diferencian de los requerimientos políticos, ya que estos últimos quedan abiertos a discusión.

Searle \ Habermas	Asertivas	Directivas	Compromisos	Expresivas	Declarativas	Clase de Reclamo
Imperativas		Futuro Simple (En inglés WILL)				Poder
Constatables						Verdad
Regulativas		Requerimiento Comando	Promesa			Justicia
Expresivas			Intención			Sinceridad

Figura 2.1: Taxonomía de Searle vs taxonomía de Habermas

En la figura 2.1 se comparan las dos clasificaciones de los actos del habla: las columnas corresponden a la clasificación de Searle y las filas a la clasificación de Habermas. Al final de cada fila aparece el reclamo de validez impuesto por el orador. La línea horizontal más gruesa separa la fila correspondiente a las frases imperativas (acciones estratégicas), de los verdaderos actos comunicativos (acciones comunicativas) que ocupan las otras tres filas. Los cuadros grises representan las similitudes entre los dos enfoques, por ejemplo las frases Asertivas de Searle son similares a las frases Constatables de Habermas. Las Directivas

de Searle se corresponden en parte con las Imperativas y en parte con las Regulativas de Habermas.

En esta sección se ha presentado una descripción general de la comunicación desde la óptica de los seres humanos, la siguiente sección presenta una perspectiva más técnica, desde ciencias de la computación, concretamente, desde el campo de la Inteligencia Artificial.

## 2.2. Comunicación en Inteligencia Artificial

Fue necesario recorrer un largo camino para que los humanos llegaran a producir y entender frases de manera natural. Esta destreza es muy difícil de replicar mediante un procesamiento automático, por lo que se hace necesario acotar y regular la manera en que se deben comportar cada una de las partes involucradas. Al reducir la consideración a este contexto, los sujetos que intervienen son denominados *agentes computacionales*, que implementan determinados algoritmos de razonamiento y están limitados a una escasa cantidad de recursos. Además es necesario dar más precisión a los términos empleados y un mayor detalle en la especificación de las tareas involucradas.

La comunicación es el intercambio de información realizado de manera intencional, por medio de la producción y la percepción de los signos derivados de un sistema convencional común. Muchos animales usan signos para representar mensajes importantes: aquí hay comida, hay un predador cerca, pueden aproximarse, deben retirarse, vamos, en marcha. En ocasiones la comunicación puede ayudar a los agentes a tener éxito ya que pueden conseguir información que es observada o inferida por otros agentes.

Entre otras cosas, lo que diferencia a los humanos de otros animales es el sistema complejo de mensajes estructurados llamado lenguaje, que permite comunicar mucho del conocimiento acerca del mundo. Si bien, los chimpancés, los delfines y otros mamíferos muestran un vocabulario de cientos de signos y cierta destreza para combinarlos, sólo los humanos pueden comunicar de manera confiable un número ilimitado de mensajes realmente distintos.

Tal es así, que Alan Turing basó su famosa prueba -el test de Turing- en el uso del lenguaje, ya que el lenguaje está íntimamente relacionado con el pensamiento. Esta prueba fue diseñada para dar una definición operacional de inteligencia. En lugar de

proponer una larga y quizás controversial lista de calificativos requeridos para establecer lo que es inteligencia, él propuso como base para su prueba, la imposibilidad de encontrar diferencias entre aquello que se desea calificar como inteligente y una entidad -como los seres humanos- que indudablemente es inteligente. Por ejemplo una computadora supera la prueba si un usuario luego de realizar algunas preguntas, no puede decidir si las respuestas fueron elaboradas por la computadora o por un ser humano. Claramente el desafío de la inteligencia artificial no es crear artefactos que superen la prueba, sino estudiar los principios subyacentes como ser las características que deben estar presentes en una comunicación efectiva. La consideración de esas características es el campo en el cual se desarrolla esta tesis.

A continuación se mencionan los pasos requeridos para la comunicación. Un episodio de comunicación típico, en el cual un orador  $O$  desea informar a un oyente  $H$  sobre una proposición  $P$  usando las palabras  $W$ , está compuesto de siete procesos:

- 1) Intención. Por algún motivo el orador  $O$  decide que hay alguna proposición  $P$  que es conveniente decir al oyente  $H$ .
- 2) Generación. El orador planifica convertir la proposición  $P$  en un enunciado tal que al ser recibido por el oyente en una situación particular pueda inferir el significado de  $P$  (o algo similar).
- 3) Síntesis. El orador produce la realización física  $W'$  de las palabras  $W$ . Esto puede ser por medio de tinta sobre papel, vibraciones en el aire, o algún otro medio. En la figura 2.2 se muestra al agente sintetizando una cadena de sonidos  $W'$ .

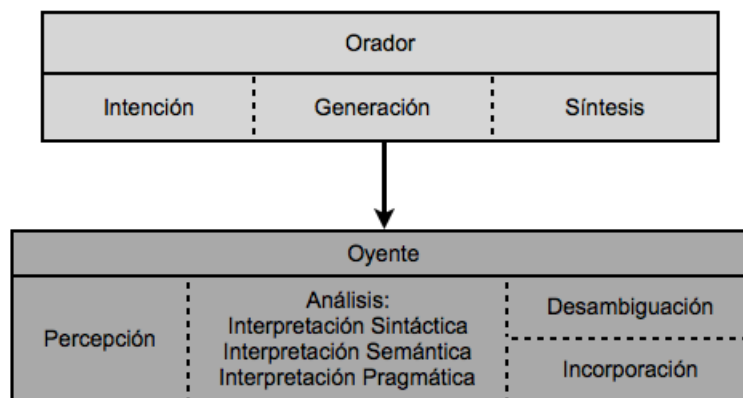


Figura 2.2: Procesos involucrados en la comunicación

4) Percepción.  $H$  percibe el fenómeno físico  $W'$  como  $W'_2$  y lo decodifica como las palabras  $W_2$ . Cuando el medio es la palabra, el paso de percepción es llamado reconocimiento del discurso; cuando el medio es escrito se denomina reconocimiento óptico de caracteres.

5) Análisis.  $H$  infiere que  $W_2$  tiene como posibles significados  $P_1, \dots, P_n$ . Es posible dividir el análisis en tres partes principales: interpretación sintáctica, interpretación semántica e interpretación pragmática. La interpretación sintáctica evalúa si cumple con los requerimientos para continuar con los otros procesos. La interpretación semántica extrae el significado como una expresión en algún lenguaje de representación; un enunciado se dice que es ambiguo cuando es posible conseguir más de una interpretación para el mismo. La interpretación pragmática tiene en cuenta que la misma palabra puede tener distintos significados en diferentes situaciones. Mientras que la interpretación sintáctica es una función de un único argumento (la secuencia de caracteres) la interpretación pragmática es función del enunciado y el contexto o situación del enunciado. Generalmente la interpretación pragmática ayuda mucho a la interpretación final del enunciado.

6) Desambiguación.  $H$  infiere que  $O$  intenta comunicar  $P_i$  (donde idealmente  $P_i = P$ ). Muchos oradores no son intencionalmente ambiguos, pero muchos enunciados tienen varias interpretaciones posibles. La comunicación se logra porque el oyente consigue entender qué interpretación es la que el orador probablemente quiso transmitir. Hay que destacar la calificación de “probable”, ya que esa es la razón por la cual en este proceso - desambiguación- se emplea razonamiento incierto. El análisis genera interpretaciones posibles, si hay más de una, la desambiguación elige la mejor.

7) Incorporación.  $H$  decide creer (o no)  $P_i$ . Un agente totalmente ingenuo puede creer todo lo que oye, pero un agente sofisticado considera el acto del habla como evidencia para  $P_i$ , no su confirmación.

Los actos del habla son denominados *actos comunicativos* por Sadek [Sad91] y la misma denominación fue adoptada por FIPA<sup>1</sup>. Nosotros usaremos ambas denominaciones indistintamente. En todos estos contextos un acto comunicativo es toda acción realizada por un agente con la intención de que sea percibida por (al menos) otro agente.

Se espera que los actos comunicativos sean *planeados* y *realizados* por un lado y

---

<sup>1</sup>En FIPA Communicative Act Library Specification. [www.fipa.org/specs/fipa00037](http://www.fipa.org/specs/fipa00037)

*reconocidos y consumidos* por el otro. Los elementos que componen este modelo de acto comunicativo se determinan en base a estas funcionalidades, las cuales corresponden al *emisor* y al *receptor* del acto. Como es de esperar en toda acción racional, debe haber un motivo por el cual se selecciona un acto comunicativo particular y también determinadas condiciones requeridas para poder planificar el acto. El primero es denominado *efecto perlocucionario* (en inglés *Perlocutionary Effect* o *PE*) y las segundas *precondiciones de factibilidad* (en inglés *Feasibility Preconditions* o *FPs*). En la sección 3.3 se verá un ejemplo donde se especifican estos términos.

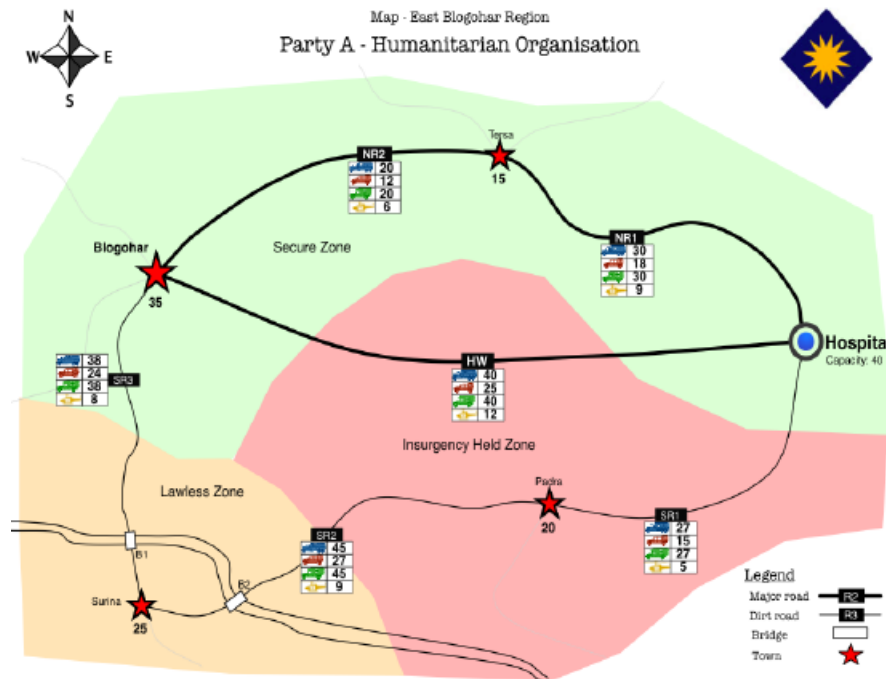
En 1996 se promovió la creación de FIPA (Foundation for Intelligent Physical Agents), una asociación internacional sin fines de lucro que agrupa compañías y organizaciones que aúnan esfuerzos para producir especificaciones estándar para la tecnología de agentes. Cuenta con un consejo principal encargado de aprobar las reglamentaciones y comités técnicos que realizan propuestas y verifican la compatibilidad de las distintas reglamentaciones. El primer conjunto de especificaciones fue publicado en 1997 y a fines de 2002 FIPA presentó su propuesta definitiva. El principal objetivo es lograr una adecuada interoperabilidad, por esa razón se concentra en el comportamiento externo de los componentes de los sistemas, dejando abierto los detalles de implementación y la arquitectura interna de los mismos. El estándar de FIPA adopta el paradigma de agente y define el modelo de referencia de las plataformas de agentes y los servicios que éstas deben proveer. La colección de estos servicios y sus interfaces estándar, representan las reglas normativas que permiten a toda sociedad de agentes existente, operar y poder ser administrada.

Si la comunicación se desarrolla bajo ciertas condiciones, allana el camino para que los agentes puedan lograr sus objetivos. Se puede decir que cuando hay una efectiva comunicación se alcanza una forma simple de colaboración. Para conseguir esto es necesario que los interlocutores compartan un mismo significado de los mensajes, lo cual ha sido un reclamo frecuente en los trabajos sobre lenguaje de comunicación de agentes. En este sentido en el presente documento se consideran dos enfoques mencionados anteriormente (en la sección 2.1); el que será denominado *mentalista*, que toma como punto de partida la descripción del aspecto *perlocutorio* que ha sido establecida por Austin en su *Teoría de los actos del habla*; y otro que propone Habermas, que será llamado *social*, en el cual además de los actos del habla se considera el contexto social del agente. Adicionalmente, un aspecto requerido para conseguir que la comunicación sea efectiva, es contar con herramientas mediante las cuales cada agente pueda controlar que no hayan contradicciones

entre sus creencias privadas y las afirmaciones que hacen públicas los otros agentes.

### 2.3. Comunicación en un escenario ficticio

En esta sección se describe una situación posible en la cual resulta de interés evaluar la utilización de los actos del habla en la comunicación. Será considerado el despliegue de una coalición internacional para cumplir tareas militares y humanitarias en una región remota, que podría ser un país ficticio de oriente medio denominado *Holistan* [RLV07]. La coalición está conformada por dos fuerzas principales, una destinada a las tareas militares y la otra encargada de las tareas humanitarias. Cada fuerza mantiene su propia información que en algunos casos puede ser compartida con miembros de la otra fuerza. En la figura 2.3 se observa el mapa de la fuerza humanitaria, con información sobre la cantidad de soldados hospitalizados en las principales ciudades de la región.



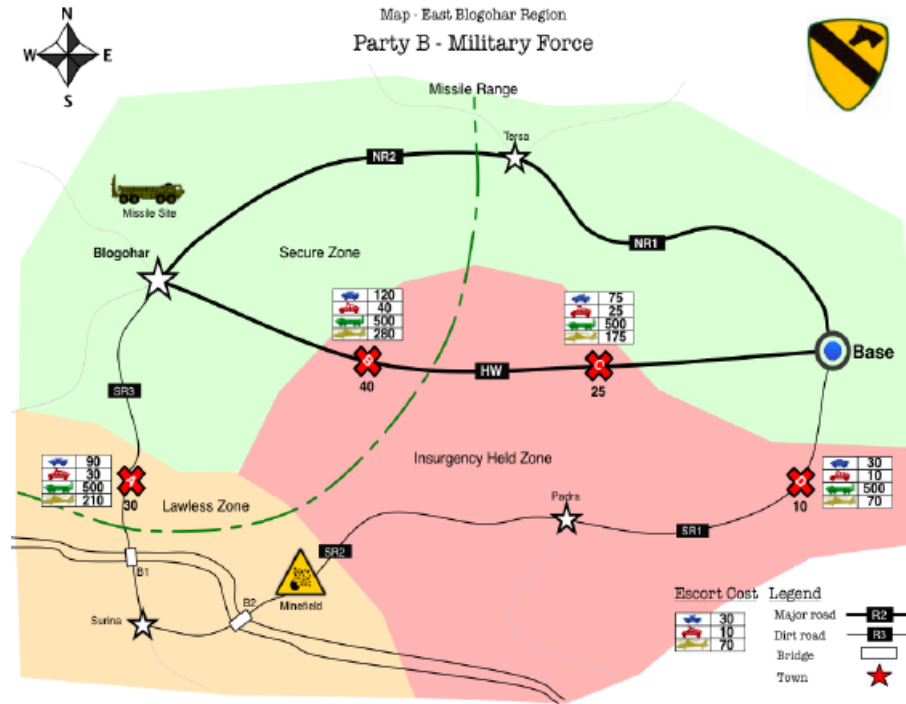


Figura 2.4: Mapa de la fuerza militar en Holistan

Con la finalidad de evaluar el uso de los actos del habla durante la interacción entre los miembros de la coalición, se propone la incorporación de agentes computacionales con las capacidades que serán descritas en los siguientes capítulos. El escenario internacional planteado puede servir para evaluar las hipótesis experimentales propuestas sobre la utilización de los actos del habla. Tanto el escenario como las situaciones de comunicación que ocurren en él, ilustran cómo se pueden aplicar algunos de los enfoques más conocidos sobre el tema. En los distintos contextos considerados se involucran dos agentes (en algunos casos tres); cada agente computacional interactúa con un representante humano de alguna de las dos fuerzas que integran la coalición.

El contexto en su conjunto está pensado para identificar los posibles problemas de comunicación entre los miembros de la coalición; las operaciones suceden en el escenario ficticio, haciendo uso de una interfaz de usuario colaborativa que permite la inserción de un agente para observar e interactuar con los participantes humanos durante las actividades de comunicación en las que estos últimos se involucran. Se necesita una interfaz que ofrezca a los participantes una forma sencilla de comunicarse con los otros integrantes de su fuerza y también con aquellos de la otra fuerza de coalición. La interfaz permite a cada uno de los agentes interceptar los diálogos e interacciones que mantienen los representantes de

las dos fuerzas. Cada agente intentará brindarle al representante humano la suficiente información de realimentación que permita mejorar la comunicación.

La interfaz debiera estar basada en la web y usar como soporte una base de datos, llamada *escenario*, en la que se mantenga información de los bienes asociados con cada representante humano. Ver Figura 2.5.

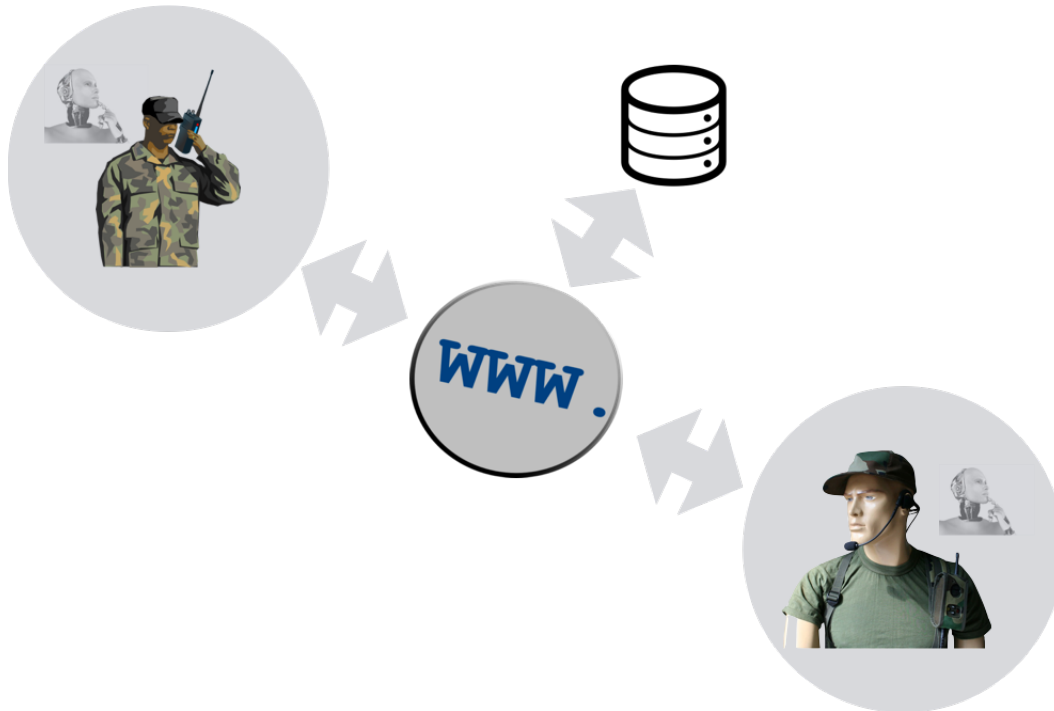


Figura 2.5: Agentes de comunicación colaborando con humanos



# Capítulo 3

## La perspectiva mental

En este capítulo se describe la propuesta de Cohen y Perrault para integrar los actos comunicativos en planes, luego se presentan los lenguajes de comunicación de agentes más conocidos (ACL y KQML), para terminar con la aproximación propuesta por FIPA.

### 3.1. Los actos del habla basados en planes

Uno de los aspectos que ha sido cuestionado de la teoría de los actos del habla es el hecho que Searle no provee demasiada justificación sobre la idoneidad de los distintos actos del habla. Muchos actos no cuentan con una definición precisa que brinde sustento a la existencia de esos actos del habla. En esta sección se considera una propuesta para definir los actos del habla mediante la cual queda en evidencia que una incorrecta especificación o cualquier cambio efectuado en su definición, afectan los planes que pueden ser realizados. Un segundo aspecto considerado, es la decisión de limitar algunas de las condiciones de Searle a ciertos actos del habla, cuando en verdad forman parte de aspectos más generales del comportamiento de los interlocutores en una comunicación.

Durante la creación de planes, a partir de los operadores que serán definidos, va surgiendo el uso de algunas de las condiciones de Searle que fueron mencionadas en el capítulo anterior. Se presentan algunos ejemplos sobre el uso de la *condición de sinceridad*, del *contenido proposicional* y de la *condición de "no-obviedad"*. A lo largo de esta sección se asumirán *condiciones de entrada/salida normales* y, antes del final de la misma, se realizará un análisis pormenorizado sobre la *condición de fuerza*. También

se explicará cómo las *condiciones preparatorias y esenciales* se corresponden con las precondiciones y los efectos de los operadores.

Tanto Searle como Austin coinciden en la necesidad de que se cumplan las condiciones preparatorias para que los actos del habla puedan ser realizados con normalidad. En particular, Austin establece diferencias entre emplear de manera inoportuna un acto del habla, de cuando se realiza de manera incorrecta. En este capítulo, las condiciones preparatorias de Searle son tomadas como las condiciones que garantizan la aplicabilidad de un operador más que su realización exitosa, y se formalizan como *precondiciones*. De esta manera, cuando se utiliza un operador sin satisfacer las precondiciones se dice que el operador fue “mal aplicado”.

Todos los esquemas de operadores tendrán dos clases de precondiciones - las de “capacidad” y las de “deseo”. Las primeras, referenciadas como CAPAZDE, indican esquemas de proposiciones que deben ser verdaderos en el modelo del mundo en el que se aplica esa instancia del operador. La precondición identificada como DESEO formaliza un principio de comportamiento intencional - el agente que hace una acción debe desear hacer esa acción.

Como regla práctica al establecer la definición de los operadores se propone aplicar el principio denominado “punto de vista”: las precondiciones comienzan con “el orador cree..” y los efectos comienzan con “el oyente cree...”.

### 3.1.1. El operador REQUERIR

A continuación se considera un primer caso: la definición del operador que corresponde a un requerimiento. Las condiciones preparatorias de Searle exigen que el oyente H sea capaz de hacer la acción ACC y que el orador S crea que H es capaz de hacer ACC. En verdad no es necesario que en el mundo real H sea capaz de hacer ACC, sólo se requiere que tanto S como H piensen que H puede hacer ACC. Para formalizar la frase “puede hacer ACC” se propone el predicado CAPAZDE(Q, ACC) que es verdadero si las precondiciones CAPAZDE de ACC son verdaderas (donde la persona Q está ligada al rol de agente que hace la acción ACC).

Por su parte la condición esencial es modelada como el efecto de un requerimiento y tiene como base la separación entre el acto ilocucionario y el efecto perlocucionario. Una

condición que ha sido establecida es que los oradores no pueden influir directamente sobre las creencias y los objetivos de los oyentes. Los efectos del operador REQUERIR son modelados como una realidad donde la finalización exitosa no requiere que el oyente desee hacer ACC. Así, el efecto se establece como la creencia del oyente en la cual el orador desea que él realice ACC. Esta formulación de la condición esencial, como se verá más adelante, es realmente una complicación. En la figura 3.1 se especifica el esquema de un operador con el cual un orador efectúa un requerimiento a un oyente para que realice la acción ACC.

REQUERIR(ORADOR, OYENTE, ACC)  
 CAPAZDE.PR:  
 ORADOR CREE OYENTE PUEDE ACC  
 ORADOR CREE OYENTE CREE OYENTE PUEDE ACC  
 DESEO.PR:  
 ORADOR CREE ORADOR DESEA instancia-requerir  
 EFECTO:  
 OYENTE CREE ORADOR CREE ORADOR DESEA ACC

Figura 3.1: Operador REQUERIR

El problema con esta definición de requerimiento es que el efecto no se ajusta a lo pretendido con el plan. Por eso hace falta un paso intermedio; es necesario inducir el deseo de hacer la acción. Para cubrir el espacio entre un requerimiento y el efecto perlocucionario por el cual se lo planifica, se postula el paso intermedio INDUCIR-DESEO (en inglés CAUSE-TO-WANT), este paso modela lo necesario para que el oyente desee hacer algo. Para conseguir el efecto requerido se establece una simplificación, mediante la cual, para que alguien desee hacer ACC sólo es necesario que esa persona sepa que su interlocutor desea que haga ACC. En la figura 3.2 se muestra la definición del paso intermedio INDUCIR-DESEO.

El plan para un requerimiento ahora es simple: el REQUERIR provee las condiciones necesarias para INDUCIR-DESEO; consiste en componer REQUERIR con INDUCIR-DESEO, esa es la manera en que se combinarán los actos del habla.

Cuando no existen creencias como para satisfacer las precondiciones CAPAZDE de una acción que el orador decide planificar para que el oyente realice, el orador necesita

INDUCIR-DESEO(AGT1, AGT, ACC)  
 CAPAZDE.PR:  
 AGT CREE AGT1 CREE AGT1 DESEA ACC  
 EFECTO:  
 AGT CREE AGT DESEA ACC

Figura 3.2: Paso intermedio INDUCIR DESEO

planificar un REQUERIR. Por ejemplo, supongamos que en Holistan se están comunicando dos miembros de la coalición; para lograr que uno de ellos, digamos Juan se mueva desde Padra hasta Surina, José, el orador, debe construir un plan para conseguir que Juan deje la ciudad de Prada y se tralade a la ciudad de Surina. El plan se muestra en la figura 3.3.

En el ejemplo del REQUERIR la precondition de deseo “JOSÉ CREE JOSÉ DESEA REQUERIR(JOSÉ, JUAN, MOVER(JUAN, PADRA, SURINA))”, es trivialmente verdadera porque es parte del plan de JOSE. En general, cuando el agente que hace una planificación determina que alguien debe hacer alguna acción, que aun no tiene pensado hacer, debe planificar un acto del tipo *directiva*.

La condición de “no-obviedad” de Searle establece que el oyente no debe tener ya previsto hacer lo que se le va a pedir. Esta condición se corresponde con un comportamiento intencional racional más amplio y no es de un acto del habla en particular. En el caso de la figura 3.3, la “no obviedad” aparece en el renglón que ha sido marcado con ++ y corresponde a la precondition de deseo de la acción que se le va a requerir a Juan. Si José ya sabe que Juan tiene planeado hacer ese viaje, entonces José no debe planificar el REQUERIR.

La condición de sinceridad establece que el orador debe desear que la acción sea realizada. En la figura esta condición ha sido marcada con +, indica la acción que se le requiere a Juan. El deseo del orador de que el oyente viaje de Prada a Surina es la razón para planificar el REQUERIR.

El contenido proposicional (la acción solicitada) se planifica previamente como la combinación de tres cosas: la precondition de deseo de la acción, el acto intermedio INDUCIR-DESEO y el efecto del REQUERIR. Esta combinación está de acuerdo con lo establecido por Searle: el contenido proposicional es una función de la condición esencial del acto, en este caso el efecto del requerimiento.

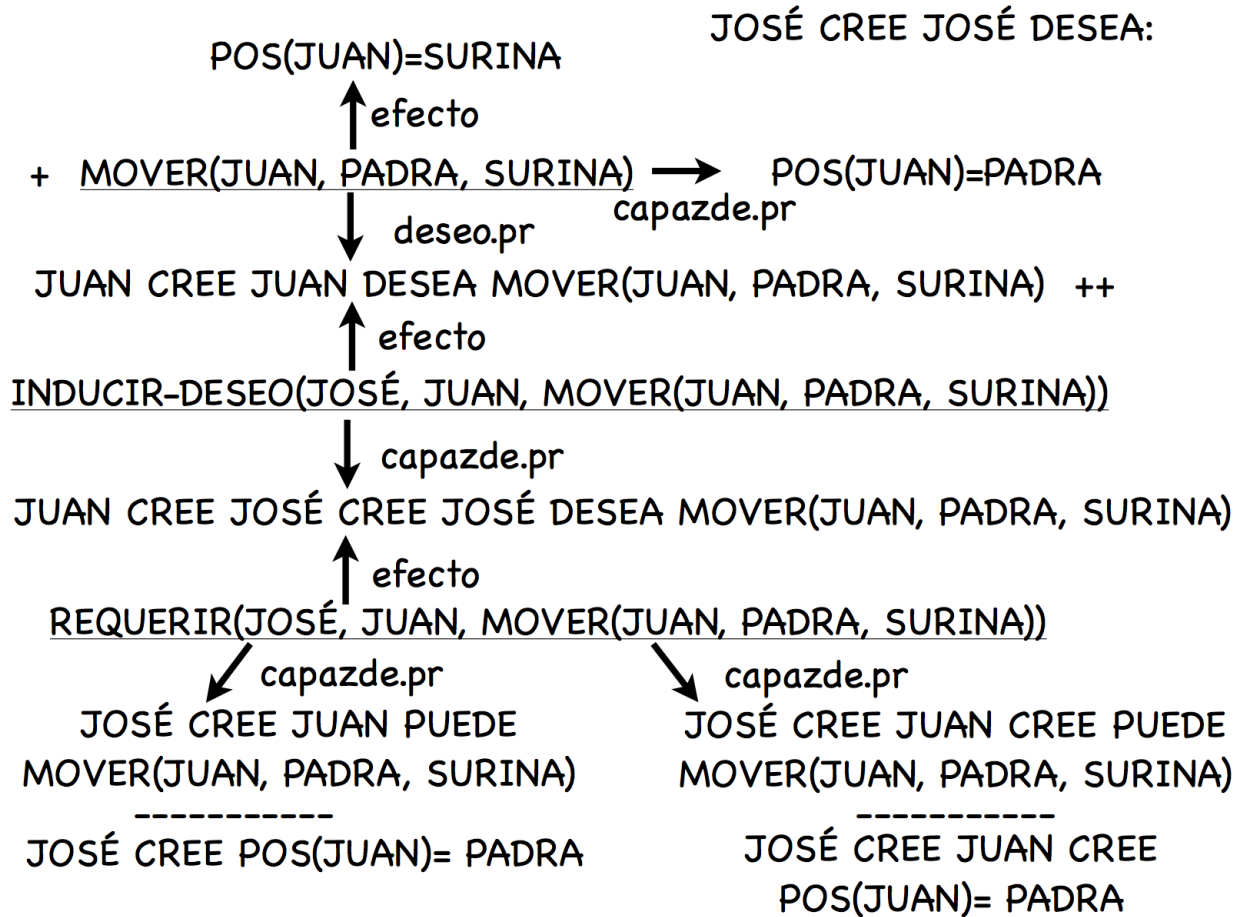


Figura 3.3: Plan para un REQUERIR

### 3.1.2. El operador INFORMAR

Se considera que un orador hace la declaración de una proposición al oyente con el objetivo de conseguir que el oyente crea que el orador cree que esa proposición es verdadera. Este acto se planifica deseando que finalmente el oyente también crea la proposición, pero eso no se consigue con la primer definición que se ve en la figura 3.4.

La precondition CAPAZDE establece que el orador debe creer la proposición. El efecto es comunicar lo que el orador cree. El oyente puede elegir rechazar la proposición sin que esto invalide la acción del operador INFORMAR que realizó el orador. Para que un agente convenza a otro de que una proposición es verdadera se define el acto intermedio CONVENCER. Este paso sirve para que AGT1 convenza a AGT sobre la verdad de la

INFORMAR(ORADOR, OYENTE, PROPOSICIÓN)  
 CAPAZDE.PR:  
 ORADOR CREE PROPOSICIÓN  
 DESEO.PR:  
 ORADOR CREE ORADOR DESEA instancia-proposición  
 EFECTO:  
 OYENTE CREE ORADOR CREE PROPOSICIÓN

Figura 3.4: operador INFORMAR

proposición y sólo es necesario que AGT crea que AGT1 cree la proposición; la definición se muestra en la figura 3.5.

CONVENCER(AGT1, AGT, PROP)  
 CAPAZDE.PR:  
 AGT CREE AGT1 CREE PROP  
 EFECTO:  
 AGT CREE PROP

Figura 3.5: Acto intermedio CONVENCER

Que el orador crea la proposición es una condición necesaria pero claramente no es suficiente para convencer a alguien para que crea algo. Un prerequisite más sofisticado de CONVENCER podría ser conocer las justificaciones que tiene AGT1 para creer la proposición, lo que lleva a que AGT tenga como creencia esas justificaciones y así siguiendo hasta establecer una cadena de razonamiento que conduce a la creencia mutua.

La planificación del acto INFORMAR es simple: para lograr que “H CREA PROP” hace falta el plan de la figura 3.6. De acuerdo al plan de la figura, no es necesario requerir que el oyente desconozca la proposición. Se puede observar que la *condición de no-obviedad* no se requiere cuando se hace un plan del acto del habla INFORMAR. También hay un cuestionamiento a la denominación de la condición de sinceridad. Lo que podría ser la condición de sinceridad del INFORMAR (S CREE PROP), en verdad, es una precondition del acto del habla más que un motivo para planificar el acto. En el caso del operador REQUERIR, la condición de sinceridad (ORADOR CREE ORADOR DESEA OYENTE haga ACC) es efectivamente un motivo para planificar el acto, ya que hay un deseo.



viaje de Padra a Surina. En la figura 3.7 se muestra un plan alternativo al REQUERIR anterior, construido usando INFORMAR un DESEO, y que tiene el mismo efecto. Se puede apreciar que a partir de la precondition del INDUCIR-DESEO todos los estados del plan alternativo son idénticos a los estados del plan REQUERIR.

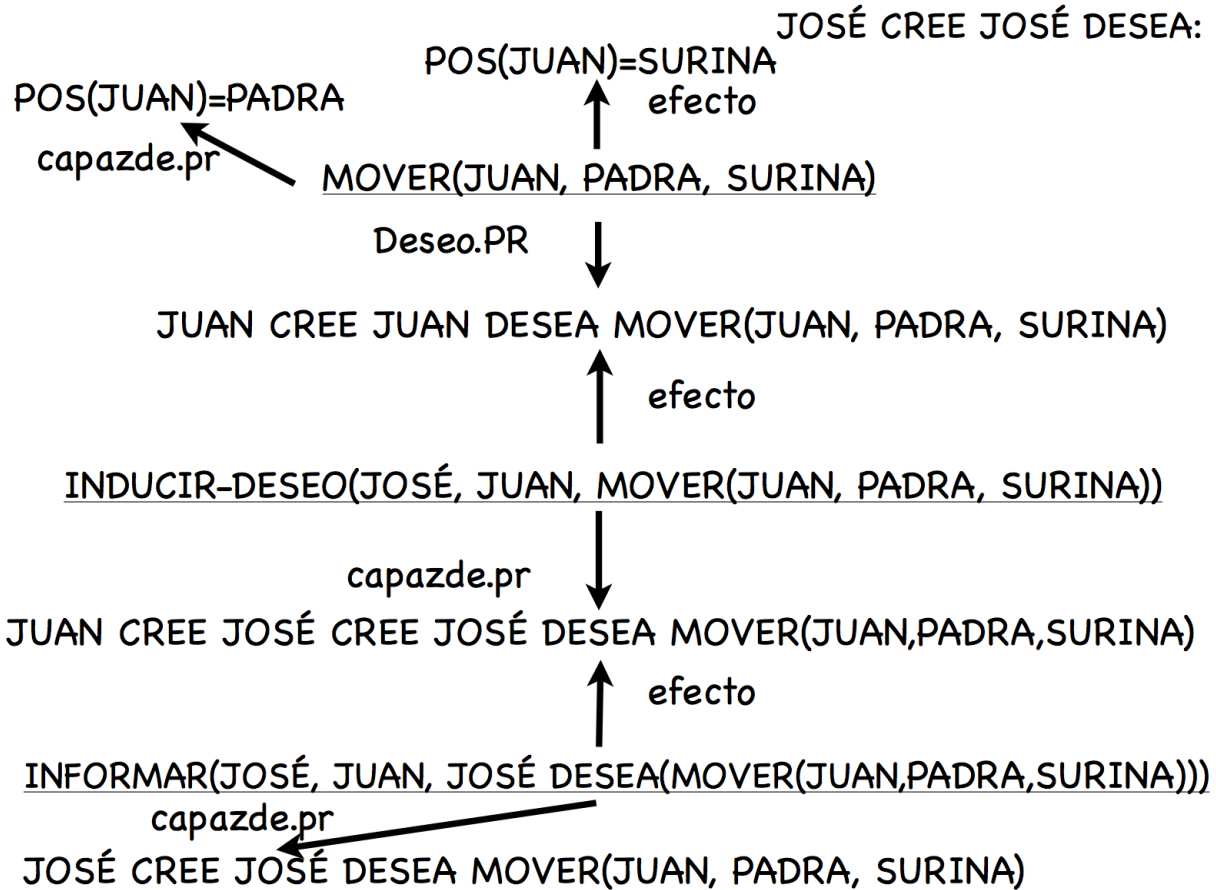


Figura 3.7: Plan para INFORMAR un DESEO

En el plan alternativo no es necesario aplicar el acto CONVENCER porque el INFORMAR un DESEO directamente produce el efecto requerido. Las dos clases de condiciones, tanto las CAPAZDE como las DESEO del operador INFORMAR se cumplen, por lo que el plan está completo.



### 3.1.4. REQUERIR vs INFORMAR un DESEO

De acuerdo a Searle, las condiciones que él propuso son necesarias y en conjunto suficientes para la correcta realización de varios actos del habla. Siguiendo esta idea, cuando se presenta un determinado comportamiento que cumple con el conjunto de condiciones de Searle, se está en presencia de un acto ilocucionario particular. En ese caso, si dos frases tienen la misma fuerza ilocucionaria, deberían ser equivalentes en términos de las condiciones de uso. Al comparar las frases “Por favor abre la puerta” y “Deseo que abras la puerta (por favor)”, se observa que ambas tienen la misma fuerza como directiva y que difieren solamente en que una es más política que la otra. Parecería que ambas se pueden planificar para conseguir el mismo resultado, pero de acuerdo a lo explicado más arriba, son distintas ya que la primera es un requerimiento y la segunda informa un deseo. La diferencia es clara porque el REQUERIR tiene como precondition “el orador cree oyente cree oyente puede hacer la acción”, mientras que el INFORMAR un DESEO no tiene esa precondition.

Para forzar la equivalencia es necesario hacer algunos cambios: quitar la precondition del REQUERIR y agregarla en el INDUCIR-DESEO. La nueva definición de INDUCIR-DESEO diría que un orador puede conseguir que un oyente desee hacer una acción con el solo hecho de creer que el orador desea que haga algo y que el oyente crea que puede hacerlo. Más formalmente, la definición de los operadores se puede ver en la figura 3.8.

Con las condiciones hasta aquí planteadas, los operadores REQUERIR e INFORMAR un DESEO tendrían el mismo efecto, pero en realidad no son intercambiables. La diferencia se observa en el supuesto que el orador quiera negar su intención de que el oyente desee hacer la acción en cuestión. Para negar que intenta conseguir el efecto del INFORMAR podría decir: “Simplemente te dije lo que yo deseaba, nada más”, mientras que después de un REQUERIR para que alguien haga X, debería decir “Simplemente te pedí que hicieras X, nada más”. Parece ser mucho más difícil negar la intención en el segundo caso. Al mantener la distinción entre el acto ilocucionario y el efecto perlocucionario, el oyente tiene la posibilidad de negarse a realizar la tarea solicitada (X); esto último es una muestra de la necesidad de mantener la distinción entre el REQUERIR y el INFORMAR un DESEO. La solución se consigue al formular el cuerpo del acto del habla como si fueran planes para alcanzar el efecto perlocutorio - aquellos planes que el oyente intenta reconocer.

REQUERIR(ORADOR, OYENTE, ACC)  
 CAPAZDE.PR:  
     ORADOR CREE OYENTE PUEDE ACC  
 DESEO.PR:  
     ORADOR CREE ORADOR DESEA instancia-requerir  
 EFECTO:  
     OYENTE CREE ORADOR CREE ORADOR DESEA ACC

INDUCIR-DESEO(AGT1, AGT, ACC)  
 CAPAZDE.PR:  
     AGT CREE AGT1 CREE AGT1 DESEA ACC  
     AGT CREE AGT PUEDE ACC  
 EFECTO:  
     AGT CREE AGT DESEA ACC

Figura 3.8: REQUERIR e INDUCIR DESEO

## 3.2. Lenguaje KQML

En el campo de ciencias de la computación, quizás el uso más difundido de la teoría de los actos del habla es para la comunicación entre aplicaciones de software. De manera creciente los sistemas de computación están siendo estructurados de tal forma que las aplicaciones individuales pueden actuar como agentes (por ejemplo con la popularización de Internet y comercio electrónico), cada uno con sus propios objetivos y mecanismos de planificación. En tales sistemas, las aplicaciones de software deben comunicarse unas con otras y con el usuario humano para conseguir la ayuda de los otros agentes para alcanzar sus metas, para comprometerse a ayudar a los otros agentes, para reportar su propio estado, para requerir un reporte del estado de otro, etc.[SLB09]

El Knowledge Query and Manipulation Language o KQML fue creado por un grupo de investigadores provenientes de la Inteligencia Artificial y de los Sistemas Distribuidos, para intercambiar conocimiento entre distintas piezas de software [FWW<sup>+</sup>93]. Este lenguaje principalmente establece el formato general que deben seguir los mensajes que se intercambian durante la comunicación. Tomando como punto de partida la teoría de los actos del habla (ver Sección 2.1) cada mensaje contiene un enunciado ejecutivo (performative) que el emisor establece como acto perlocucionario del mensaje. También se incluyen una

serie de parámetros que se especifican como pares atributo/valor para especificar una gran variedad de información complementaria como la identificación del emisor y del receptor, el contenido del mensaje y el lenguaje en el que éste último es especificado, entre otras cosas. La Tabla 3.1 muestra los parámetros habitualmente usados.

<b>Parámetro</b>	<b>Significado</b>
:content	Información sobre la cual el mensaje expresa una actitud.
:force	Si el emisor jamás negará el contenido del mensaje.
:reply-with	El emisor espera una respuesta y, de existir una, la identificación de la misma.
:in-reply-to	El rótulo esperado de una respuesta.
:language	El nombre del lenguaje usado en el parámetro :content.
:ontology	El nombre de la ontología usada en el parámetro :content.
:reciver	El receptor del mensaje.
:sender	El emisor del mensaje.

Tabla 3.1: Parámetros de los mensajes en KQML

KQML requiere que cada agente disponga de una base de conocimiento, la que se puede representar con total libertad y contiene toda la información disponible internamente para el agente. La semántica de los distintos mensajes se establece sobre el hecho que en la base de conocimiento, llamada VKB (base de conocimiento virtual), se mantienen las creencias y los objetivos del agente.

En la tabla 3.2 se describe el significado de algunos enunciados usando texto simple, claramente esta descripción de la semántica es poco precisa, pero es la única oficialmente provista.

En la figura 3.9 se observa un ejemplo de un mensaje KQML, esto puede ser parte del diálogo que mantienen dos agentes que colaboran con los representantes humanos en el contexto Holistan. Un agente consulta a otro para conseguir información sobre la seguridad para desplegarse en la región de Surina. Notar que el mensaje es un enunciado ejecutivo. El contenido del mensaje se especifica en el lenguaje KIF [GFB<sup>+</sup>92] y el significado se obtiene de la ontología Holistan.

Si como resultado del diálogo anterior, el orador determina que necesita más información sobre la región de Surina, puede solicitar al representante de la fuerza militar toda

Nombre	Significado
achieve	E desea que R haga verdadero algo de su entorno.
ask-about	E desea toda la información relevante que tenga R en su VBK.
broke-one	E desea que R encuentre ayuda para procesar un enunciado.
reply	Comunica una respuesta esperada.
sorry	E no puede enviar más información como repuesta.
stream-about	Versión de múltiple respuesta de ask-about.
tell	Una entrada en la VBK de E.

Tabla 3.2: Algunos enunciados KQML para un emisor E y un receptor R

```
(evaluate
  :sender José :receiver Juan
  :language KIF :ontology Holistan
  :reply-with q1 :content (val (seguridad en Surina)))
(reply
  :sender Juan :receiver José
  :language KIF :ontology Holistan
  :in-reply-to q1 :content (= (Surina) (minado)))
```

Figura 3.9: Diálogo entre representantes de las fuerzas en Holistan

la información que pueda suministrarle. Para eso podría usar otro enunciado ejecutivo: *stream-about* el cual sirve para solicitar al receptor toda la información que posee sobre un tema; en la figura 3.10 se muestra este diálogo. En ambos casos los campos *:language* y *:ontology* son completamente independientes de la semántica del enunciado ejecutivo empleado.

Cuando los agentes son heterogéneos, ya que han sido creados por programadores distintos, pueden haber dificultades para interpretar los mensajes de los otros agentes. Existen dos razones para estas dificultades. Primero, cuando el receptor no puede reconocer los términos que el emisor está usando en la comunicación, más allá que sean específicos de la aplicación. Segundo - y quizás más importante - cuando los componentes básicos de la comunicación no son interpretados de manera uniforme. En este sentido KQML no colabora para superar estos inconvenientes. En relación al primer tipo de dificultad, la libertad de implementación de la VKB genera situaciones difíciles de resolver. ¿Cómo

```

(stream-about
  :sender José :receiver Juan
  :language KIF :ontology Holistan
  :reply-with q1 :content Surina)
(tell
  :sender Juan :receiver José
  :in-reply-to q1 :content (= (estado Surina) (30 insurgentes)))
(tell
  :sender Juan :receiver José
  :in-reply-to q1 :content (= (estado Surina) (minado)))
(tell
  :sender Juan :receiver José
  :in-reply-to q1 :content (= (estado Surina) (alcance misil)))
(sorry
  :sender Juan :receiver José :in-reply-to q1)

```

Figura 3.10: Requerimiento de información y las respuestas correspondientes

determinar si el agente militar sabe si los misiles alcanzan al refugio A del enemigo?. Un criterio podría ser requerir que en la base de conocimiento exista una cadena de caracteres que explícitamente lo indique, pero de esa manera se excluye a otros agentes que representan esa situación de una manera distinta. Si el agente Juan conoce la ubicación del lanza misiles y de los refugios, puede calcular que A está en el radio de alcance de los misiles. Sobre la segunda dificultad, la misma ocurre debido a la gran variedad de performatives del KQML y algunas de ellas deben ser interpretadas de manera *ad hoc*. Un ejemplo es el mensaje *broker-one*, que se usa para solicitar al receptor que consiga ayuda para responder a una performative. Debido a la carencia de una semántica precisa, el receptor puede tener dificultades al interpretar este tipo de mensajes.

Para representar el contenido de los mensajes se usa KIF, un lenguaje intermedio que facilita el desarrollo independiente de programas que manipulan conocimiento. Si bien no es muy eficiente, ya que es similar al texto común, permite que un programador facilmente pueda analizar una sentencia. KIF (por sus siglas en inglés Knowledge Interchange Format) es esencialmente una versión estandarizada de una lógica de primer orden, la cual es usada para definir ontologías y el contenido de mensajes.

Hay más de 30 enunciados en KQML definidos en lo que constituye el conjunto básico, pero como este conjunto puede ser ampliado con más enunciados ejecutivos, han surgido diferentes implementaciones del lenguaje. Considerando que no existe una adecuada semántica para los enunciados, estas implementaciones muchas veces no se pueden integrar y no es fácil lograr que operen entre sí correctamente.

Se puede destacar que entre los enunciados no se incluye ninguno relacionado con los compromisos, necesarios para que un agente exprese a otro su voluntad de realizar en el futuro una tarea. Estos compromisos ayudan a coordinar el accionar entre los agentes.

### 3.3. Lenguaje ACL. Actos del habla según FIPA

Considerando que los agentes son entidades sociales que necesitan comunicarse, el Lenguaje de Comunicación de Agentes (en inglés *Agents Communication Language* o ACL) es uno de los principales aportes de FIPA para estandarizar la creación de sistemas multi-agente. El ACL de FIPA está basado en la teoría de los actos del habla como así también en ciertos principios y suposiciones sobre el paradigma de agentes que han sido ampliamente reconocidos y aceptados. El lenguaje tiene distintos tipos de mensajes, sustentados en los conceptos de creencias, deseos e intenciones de los agentes; estas son las ideas que caracterizan la arquitectura de agentes BDI, por sus siglas en inglés: Belief Desire Intention [BIP88]. Al igual que KQML, el ACL de FIPA es un lenguaje “externo” para los mensajes.

FIPA estableció una extensa biblioteca de 22 actos comunicativos que permiten materializar diferentes intenciones relacionadas con la comunicación. FIPA también define la estructura que deben respetar los mensajes, otorgando determinados lugares a los campos, como por ejemplo para la especificación del destinatario y para el emisor, o para el contenido propiamente dicho, o algunas propiedades específicas como ser el lenguaje del contenido. Otra información muy importante es la destinada a mantener conversaciones y protocolos. Por estas razones el estándar de FIPA ha sido ampliamente adoptado por la comunidad de agentes y resulta interesante analizar la perspectiva que FIPA adopta para algunos actos particulares, este análisis se hace más adelante en este capítulo.

Los mensajes transmitidos en un sistema multiagente deben respetar una sintaxis y una estructura determinadas, una parte importante del trabajo de FIPA ha sido lograr

estándares en estos sentidos. Las convenciones establecidas comúnmente especifican al menos un *orador*, que denota la identidad del agente emisor del mensaje, es decir, el agente que realiza el acto comunicativo. Generalmente, los mensajes tienen especificado este campo pero no es obligatorio ya que puede ser que el agente prefiera quedar en el anonimato. Dentro del conjunto de agentes destinatarios se incluyen todos los identificadores de agentes dentro del sistema multi-agente que aparecen en la lista de receptores del mensaje. En el mensaje también se identifica un *contenido*. El contenido del mensaje puede ser especificado a través de distintos lenguajes. Algunos de los más utilizados son FIPA-SL, FIPA-KIF, FIPA-RDF (por las siglas en inglés: Resource Description Framework<sup>1</sup>). Otro campo obligatorio es el tipo de acto ilocutorio o *enunciado ejecutivo* (en inglés *performative*) que corresponde al mensaje. Con este último campo se especifica qué acción se está realizando mediante el mensaje: preguntar, informar, requerir, etc.

**Uso de actos comunicativos según FIPA** Para describir cómo es la aproximación de los actos comunicativos en FIPA, conviene comenzar considerando un ejemplo de una situación particular en Holistan, donde un agente (de nombre José) posee algún objetivo o meta **G** y alguna intención **I**. Si como parte de la realización de **I**, el agente determina que debe enviar uno o varios mensajes a otro agente (digamos al agente Juan, el cual integra la otra fuerza de la coalición), lo primero que debe hacer José es determinar cuáles serán los mensajes correctos para poder cumplir con la intención **I**, o al menos aquellos que puedan ayudar en su realización. Si el agente José se comporta de manera razonable, no enviará un mensaje que no satisfaga su intención, porque no lo ayudaría a lograr su objetivo.

**Ejemplo 1** *Supongamos que José debe recorrer la ruta SR1 que une Prada con el Hospital para trasladar unos heridos. Esta tarea está condicionada a que no se conozcan ataques sobre esa ruta. Ver figura 2.3. ( **G** = “Ir desde Prada al Hospital”). Para eso, como primer paso, José debe consultar al agente de la fuerza militar para verificar que la ruta no es peligrosa. ( **G**' = “Saber si hay combates en SR1” ) y por eso debe enviar un mensaje a Juan consultando la situación. ( **I** = “Averiguar si hay combates en SR1”). Realmente sería un comportamiento equivocado si le pregunta “¿Cuántos rebeldes hay sobre SR1?”, ya que desde la perspectiva de José, la respuesta de Juan, no le ayudará a cumplir su objetivo **G**'. Continuando con el ejemplo, si José actúa más racionalmente, puede generar un mensaje*

<sup>1</sup><http://www.fipa.org/specs/fipa00011/XC00011B.html>

*destinado a Juan requiriendo una respuesta a la siguiente consulta: “¿Hay combates en SR1?”. De esta forma espera satisfacer su intención **I** y así cumplir su deseo **G**'.*

Para determinar cual es el mensaje correcto, José debe pensar que Juan es capaz de interpretar el mensaje y proceder racionalmente. José puede razonar que el efecto de requerir una respuesta de Juan, es que éste brinde la información que posee acerca de la consulta. Sin embargo, una vez enviado el mensaje, realmente ¿puede José asumir que en algún momento obtendrá una respuesta de Juan?. En realidad José sólo puede asumir que Juan toma conocimiento del requerimiento y que Juan sabe que José le ha pedido que confirme una cierta información. Pero por el sólo hecho de hacer el requerimiento, José no puede asumir que Juan hará lo solicitado, ya que Juan es independiente y puede estar, por ejemplo, ocupado con otras cosas.

Se presenta a continuación una descripción más detallada de dos actos comunicativos, como ejemplos de la nomenclatura que sigue FIPA.

**El acto comunicativo INFORM** Mediante este acto comunicativo el emisor informa al receptor que una proposición particular es verdadera. Usar un INFORM indica que el agente emisor:

- Sostiene que una proposición es verdadera.
- Intenta que el agente receptor también llegue a creer que la proposición es verdadera.
- Cree que el receptor no tiene conocimiento sobre la veracidad de la proposición.

Los dos primeros puntos son directos: el agente emisor es sincero y tiene la intención de que el receptor incorpore la proposición.

El último item se relaciona con la sanidad (en inglés *soundness*) semántica del acto comunicativo. Si un agente ya sabe que se observa un determinado estado del mundo (por ejemplo que el receptor ya conoce la proposición  $p$ ), no puede racionalmente adoptar la intención de alcanzar aquel estado del mundo, es decir, que el receptor llegue a conocer  $p$  como resultado del acto *inform*. En realidad esta propiedad no es tan estricta como parece. El emisor no tiene la obligación de verificar si el receptor conoce  $p$ . Sólo requiere que, en el caso que el emisor previamente conozca el estado de creencias del receptor, no debiera adoptar la intención de decirle al receptor algo que éste ya conoce.

Desde el punto de vista del receptor, recibir un mensaje *inform* lo habilita a creer que:



- El emisor cree la proposición que es el contenido del mensaje.
- El emisor desea que el receptor también crea la proposición.

Si el receptor adopta o no la proposición como una creencia, dependerá de la confianza que el receptor tenga en la sinceridad y confiabilidad del emisor.

FIPA ha resuelto usar el lenguaje formal SL (*en inglés Semantic Language*)<sup>2</sup> tanto para definir el significado de los mensajes en FIPA-ACL, como para especificar el contenido de los mensajes. Es decir que la sintáxis del contenido y la semántica asociada con el mensaje, son especificadas en FIPA-SL. Es necesario también aclarar que dentro de esta comunidad se asume que el modelo mental de un agente está basado en la representación de tres actitudes primitivas: **B**ieliefs (creencias), **U**ncertainty (incertidumbre) y **C**hoice (en algún sentido Goal (meta)). Estas actitudes son formalizadas mediante los operadores modales  $B$ ,  $U$ ,  $C$ . Las fórmulas que usan estos operadores se leen como:

- $B_i p$  quiere decir que  $i$  (implícitamente) cree  $p$ .
- $U_i p$  quiere decir que  $i$  tiene incertidumbre acerca de  $p$ , pero piensa que  $p$  es más probable que  $\neg p$ .
- $C_i p$  quiere decir que  $i$  desea que  $p$  efectivamente se cumpla.

Para simplificar la notación usaremos  $Bif_i p$  como equivalente a  $B_i p \vee B_i \neg p$ , para expresar que el agente  $i$  cree  $p$ , o bien, cree  $\neg p$ .

Formalmente el acto comunicativo *inform* puede describirse del siguiente modo<sup>3</sup>:

$$\langle i, inform(j, \phi) \rangle$$

Condiciones previas de factibilidad (en inglés Feasibility Preconditions o FP):

$$B_i \phi \wedge \neg B_i (Bif_j \phi \vee U_j \phi)$$

Efecto racional (en inglés Rational Effect o RE):  $B_j \phi$

El concepto de *Objetivo Persistente* (en inglés Persistent Goal o PG) se define a partir de la idea de creencia, elección(choice) y acción. Un agente  $i$  tiene a  $p$  como un objetivo persistente, si  $i$  tiene a  $p$  como un objetivo a ser alcanzado y está comprometido a conseguir este objetivo hasta que  $i$  crea que  $p$  ha sido alcanzado o crea que es inalcanzable. Una

<sup>2</sup>En [www.fipa.org/specs/fipa0008](http://www.fipa.org/specs/fipa0008) se especifica la gramática completa del lenguaje.

<sup>3</sup> Donde  $\phi$  denota un esquema para cualquier proposición,  $i$  es el emisor y  $j$  el receptor.

*Intención* ( $I$ ) se define como un objetivo persistente que fuerza al agente a actuar.

Las fórmulas como  $PG_i p$  e  $I_i p$  se interpretan como “ $i$  tiene a  $p$  como una meta persistente” e “ $i$  tiene la intención de conseguir  $p$ ” respectivamente.

**El acto comunicativo REQUEST** Este acto comunicativo permite que un agente requiera a otro agente la realización de alguna acción. Considerando que el pedido es iniciado por el agente  $i$ , para que el agente  $j$  realice la acción  $a$ , el modelo formal de REQUEST sería el siguiente:

$$\langle i, REQUEST(j, a) \rangle$$

Condiciones previas de factibilidad:  $B_i Agent(j, a) \wedge \neg B_i I_j Done(a)$

Efecto racional:  $Done(a)$

La expresión “ $Agent(j, a)$ ” del lenguaje SL, significa que el agente  $j$  está involucrado en la realización de la acción  $a$  y  $Done(a)$  significa que la acción  $a$  ha sido hecha. Para que este acto comunicativo pueda ser aplicado, el agente  $i$  debe creer que el agente  $j$  puede realizar la acción  $a$  (de esa manera se asegura que el mensaje está siendo enviado al tipo correcto de agente) y el agente  $i$  no debe creer que entre las intenciones del agente  $j$  está previsto realizar  $a$ .

Los actos comunicativos REQUEST son directivas que se caracterizan porque el agente emisor intenta que el receptor haga alguna acción. Es decir que son enunciados destinados a futuro. Se establecen las siguientes condiciones preparatorias para este enunciado ejecutivo.

1. El oyente es capaz de realizar el acto que se le está solicitando.
2. El orador cree el hecho del punto anterior.
3. No es obvio para ninguno de los interlocutores que el oyente vaya a realizar el acto requerido en el curso normal de eventos.

La condición esencial de una directiva como esta es intentar que el receptor del mensaje realice el acto requerido.

La condición 1 no expresa el punto de vista del agente que hace el requerimiento, en consecuencia no es una condición requerida para planificar la directiva. El mismo comentario se puede hacer acerca de la parte del emisor en la condición 3.

**Ejemplo 2** *La situación descrita en el Ejemplo 1 se muestra nuevamente, usando ahora los operadores modales recientemente introducidos. El hecho que José sepa que hay heridos en Prada es porque la proposición  $p$  = “heridos(Prada)” se ha incorporado como una creencia del agente. Para el agente José,  $p$  es el objeto al cual se aplica el operador de creencias.*

$$B_{\text{José}}p$$

*El agente tiene ahora como meta persistente averiguar si es seguro usar la ruta para trasladar esos heridos al Hospital. Verificar la veracidad de la proposición  $t$  = “No hay combates en SR1” es el objetivo persistente.*

$$PG_{\text{José}}t$$

*Adquiere una intención que le permita alcanzar su meta, o sea que la proposición  $u$  = “Averiguar si hay combates en SR1” será una tarea inmediata.*

$$I_{\text{José}}u$$

*Como consecuencia del principio de racionalidad<sup>4</sup>, el agente debiera seleccionar un acto comunicativo REQUEST para consultar a un miembro de la fuerza militar. La respuesta que reciba le confirmará si hay o no combates en SR1.*

De acuerdo a [Sad91], una intención “implica” que se genere un proceso de planificación. La acción que en definitiva se termina realizando es seleccionada de acuerdo a la circunstancia particular en que se encuentra el agente. Además de REQUEST, otros *actos comunicativos primitivos*, son: INFORM, CONFIRM y DISCONFIRM. Estos actos tienen en común el hecho que un agente puede realizarlos directamente o bien incorporarlos a sus planes ya que son operadores atómicos. - como ocurrió con el uso del REQUEST en el ejemplo anterior. Mediante un *request* el emisor solicita al receptor que realice una acción particular y no se presentan alternativas a considerar, ya que desde el punto de vista de la comunicación se lo considera realizado una vez que se emite el mensaje. Es interesante destacar que la acción solicitada en un REQUEST puede ser un nuevo un acto comunicativo como en el caso del ejemplo.

Existen también los denominados *actos comunicativos compuestos* que en su definición involucran algunas alternativas a ser resueltas por otro agente. Este tipo de actos a veces requiere el intercambio de más de un mensaje y el resultado final es uno de los varios posibles. Son los que se utilizan por ejemplo cuando un agente consulta a otro sobre la veracidad de una proposición: como lo ocurrido en el Ejemplo 2, donde las próximas

---

<sup>4</sup>“Si un agente tiene conocimiento que una de sus acciones le conduce a lograr una de sus metas, entonces el agente seleccionará dicha acción” [New82]

acciones de un agente están condicionadas a la respuesta del otro agente. En la próxima subsección se brindan más detalles del acto INFORM-IF.

**Informar si se cumple una proposición: INFORM-IF** Por ser un acto comunicativo compuesto no puede ser realizado directamente por un agente; sin embargo puede ser incorporado dentro de un plan o ser la acción objeto de un requerimiento hacia otro agente (requerir que el receptor informe si se cumple una proposición o la negación de la misma). Supongamos que en un paso del plan, es necesario que el agente informe el estado de una proposición  $\phi$ . En ese paso el agente debe informar  $\phi$ , si  $\phi$  es verdadero, pero debiera informar  $\neg\phi$  en otro caso. Lo más conveniente al hacer el plan sería usar *inform-if* ahí. Formalmente se describe:

$$\langle i, \text{inform} - \text{if}(j, \phi) \rangle \equiv \langle i, \text{inform}(j, \phi) \rangle \mid \langle i, \text{inform}(j, \neg\phi) \rangle \quad 5$$

FP:  $Bif_i\phi \wedge \neg B_i(Bif_j\phi \vee Uif_j\phi)$

RE:  $Bif_j\phi$

Los actos comunicativos compuestos se traducen en definitiva en uno de varios actos primitivos posibles. Para el caso particular del INFORM-IF, el emisor debe informar si el contenido del mensaje (que debe ser una proposición) es verdadero o si lo es su negación.

---

<sup>5</sup> La expresión de acción:  $a_1 \mid a_2$ , es una elección no determinística, en la cual, o bien aparece  $a_1$ , o aparece  $a_2$ , pero no ambas.

# Capítulo 4

## Perspectiva Social

En este capítulo se extiende la necesidad de comunicación para coordinar las acciones que realizan los agentes. Una aproximación distinta a la presentada en el capítulo anterior es considerar a los actos comunicativos como parte del desarrollo de una interacción social [CS15]. Aún si no es posible determinar si los agentes tienen un estado mental específico, se puede afirmar que los agentes de comunicación son capaces de interactuar socialmente. Esto es análogo a la distinción entre el comportamiento (externo) y el estado (interno) de un agente. Las interfaces en el diseño tradicional de programas están basadas en el comportamiento, aunque las representaciones de su estado pueden ser usadas para conseguir el comportamiento deseado [Sin98].

### 4.1. Comunicación desde la perspectiva social

En la medida que se reconoce el potencial de los sistemas multi-agente, los diseñadores de agentes no solamente deben preocuparse por el funcionamiento interno del agente sino también de la interacción con otros miembros del sistema multi-agente. El funcionamiento en conjunto representa un gran desafío: ¿Cómo lograr que los agentes heterogéneos actúen con otros agentes correctamente? [Sin00]. ¿Cómo se debe organizar un sistema multi-agente?. La organización mental enfatiza el estado mental del agente, esto quiere decir que normalmente será descripto a partir de sus intenciones y creencias. Por su parte la organización social mira a los agentes como criaturas sociales que interactúan unas

con otras. La organización mental supone que el diseño del agente incluye intenciones y creencias, pero desde el punto de vista práctico es difícil establecer si un agente particular tiene una determinada creencia, aun conociendo su diseño. Por ejemplo, ¿Cómo decidir si un agente *vendedor* conoce el libro “El factor humano”? Un criterio podría ser que solamente los agentes que tengan el string “El factor humano” en el campo “título” conozcan este libro. Pero entonces, ¿Cuál ha sido el criterio para excluir aquellos agentes que no poseen el campo “título”? Más aún, si dos agentes tuviesen el mismo campo “título”, pero métodos muy distintos entre sí en sus códigos fuente (con comportamientos marcadamente diferentes), eso se puede interpretar erróneamente como creencias distintas.

El enfoque social de la comunicación considera al sistema multi-agente como un sistema distribuido, por un lado tiene en cuenta la *condición de carrera* que se produce porque la comunicación no es inmediata y por otro lado que las acciones se coordinan sobre la idea de compromiso. Un compromiso es un predicado que se cumple cuando se hace verdadera una proposición particular.

#### 4.1.1. Uso de los compromisos en la comunicación

Los compromisos constituyen un buen punto de partida para entender la interacción entre los agentes. El éxito de esa interacción se basa en que los participantes mantengan *alineados* sus compromisos [CS09]. Es decir, no puede ocurrir que durante una interacción, un agente entienda que es acreedor de un compromiso mientras que el deudor no se reconoce como tal. En cualquier sistema distribuido, debido a la ausencia de sincronización y a las diferentes observaciones que realizan los componentes, es difícil lograr que los compromisos se mantengan alineados.

¿Cuál es rol de la comunicación en la alineación de los compromisos?. Para responder, lo primero es hacer algunas consideraciones generales sobre los compromisos. En este capítulo se sintetizan algunas de las publicaciones realizadas por Munindar P. Singh y colaboradores durante un período de más de 15 años. Se rescata el enfoque social de los sistemas multi-agente y se adopta la notación empleada en dichas publicaciones.

Un compromiso es  $C(x, y, r, u)$ . Tanto  $x$  como  $y$  son agentes, donde  $x$  es el deudor,  $y$  es el acreedor. Por su parte ambos  $r$  y  $u$  son proposiciones,  $r$  es el antecedente y  $u$  es la consecuencia. Un compromiso atraviesa por distintos estados. Si un compromiso  $C(x, y, r, u)$  está en *vigencia*, significa que  $x$  se compromete con  $y$  a que cuando  $r$  se cumpla entonces

se cumplirá  $u$ . Si  $r$  se cumple, el compromiso  $C(x, y, r, u)$  se hace *independiente* (en inglés *detached*) y entra en vigencia el compromiso *incondicional*  $C(x, y, \top, u)$ . Si  $u$  se cumple, el compromiso es *descargado* y ya no tendrá vigencia.

Para administrar los compromisos, cada agente deberá mantener información sobre los compromisos relevantes y tener la capacidad de razonar sobre los cambios en el estado de los mismos. El lugar donde se conservan es denominado *almacén de compromisos* [WK95] y allí se mantienen las afirmaciones hechas durante el diálogo. Esta información no tiene relación con el conocimiento individual de los participantes. Pero si se requiere que el conjunto de proposiciones que integran el almacén de compromisos sea consistente. Para garantizar la consistencia se utilizan una serie de postulados que son propios del tipo de diálogo y mediante ellos se determina cuando se agregan o quitan las entradas. Considerando que los agentes racionales regularmente deberán realizar la tarea de cuestionarse los fundamentos que sustentan sus compromisos, es muy importante mantener una relación de coherencia entre los compromisos dentro del almacén de compromisos.

Por ejemplo en el contexto Holistan, si la fuerza militar(FM) se compromete a brindar *escolta* a la fuerza humanitaria(FH) para rescatar los heridos de un *accidente* aéreo en Prada, se crea el compromiso  $C(FM, FH, accidente, escolta)$ . Seguramente surgirá un problema entre las fuerzas aliadas si luego que cae un helicoptero en Prada, acuden solamente los miembros de la fuerza humanitaria, confiados de que la fuerza militar brindará la escolta armada requerida, y ésta no llega.

Es decir, más allá que los agentes individuales estén bien diseñados y funcionen correctamente trabajando de manera aislada; como consecuencia de la interacción, pueden surgir situaciones de mal funcionamiento en el sistema multi-agente.

## 4.2. La alineación de los compromisos

En los siguientes ejemplos se muestran posibles secuencias de ejecución en las que se combinan las acciones de los agentes; para estos casos se usa una notación similar a los diagramas de interacción de UML (en inglés, Unified Modeling Language). Para cada agente las flechas verticales que apuntan hacia abajo representan el avance del tiempo, las flechas horizontales representan los mensajes que intercambian los agentes; siempre que un agente envía o recibe un mensaje, se indican los compromisos que se crean en ese

punto. La correspondencia que existe entre los mensajes enviados y las operaciones sobre compromisos dependen del contexto particular.

La alineación de compromisos existe cuando se puede asegurar: “siempre que un acreedor infiere un compromiso, el deudor del compromiso también infiere que existe el compromiso en cuestión”.

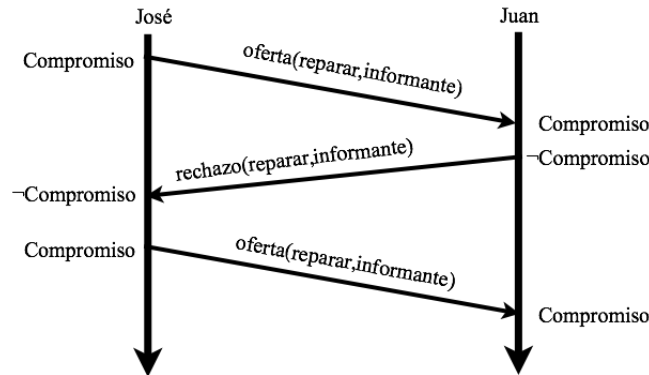


Figura 4.1: Escenario con alineación de compromisos

**Ejemplo 3** (*Alineación de Compromisos*) En la Figura 4.1 se muestra una situación en la cual el agente José de la fuerza humanitaria le ofrece al agente Juan de la fuerza militar llevar a Surina a un paciente de la insurgencia que está dispuesto a colaborar, para eso es necesario que la fuerza militar repare el puente B2 sobre la carretera SR2. Ver el mapa de la Figura 2.3. Juan rechaza la oferta, porque ahora no tiene el material necesario, pero José insiste con su propuesta inicial. En este caso los compromisos finalizan alineados. □

#### 4.2.1. Primera Interpretación del efecto de los mensajes

Los problemas que introducen los sistemas distribuidos en la alineación de los compromisos, han sido ignorados o han sido resueltos mediante soluciones hechas a medida para cada caso. Pero en la medida que los compromisos se continúen usando en los sistemas distribuidos, se hace necesario un tratamiento riguroso de los mismos.

Los sistemas distribuidos reales se comunican mediante pasaje asincrónico de mensajes, esto quiere decir que los agentes nunca se bloquean cuando deben enviar un mensaje. En esta clase de sistemas, los mensajes que un agente le envía a otro seguramente deben atravesar algún *cable* para llegar a destino. Así, en general los agentes pueden observar diferentes mensajes en diferente orden. Debido a que los mensajes se entienden en términos



de sus efectos sobre los compromisos, los agentes involucrados podrían quedar *desalineados*, es decir, llegar a conclusiones en conflicto sobre que compromisos están activos y cuales no.

Es crucial desarrollar una formalización que asegure la alineación de los compromisos, independientemente de la falta de sincronismo. Para eso, hay que tener en cuenta los siguientes puntos. Primero, la infraestructura de la computación distribuida es necesariamente asincrónica. Los sistemas grandes tienen una enorme latencia, lo que hace que las interacciones sincrónicas resulten en realidad impracticables. Segundo, cualquier formalización que trabaje de manera asincrónica, también trabajará en entornos que tienen sincronismo, es decir, funcionarán bien aún bajo aquellas restricciones adicionales impuestas al comportamiento de los agentes - como puede ser, por ejemplo, esperar turnos para enviar mensajes. Tercero, la falta de sincronismo es inherente a la autonomía de los agentes, ya que nunca serán bloqueados, por ejemplo, al enviar un mensaje o al actuar cumpliendo sus compromisos.

Considerando la ausencia de una formalización que soporte el razonamiento sobre los compromisos en entornos distribuidos, ya que toda la investigación que existe sobre aplicaciones de compromisos se limita a reportar resultados que no son suficientemente generales, o son excesivamente complejos, en este capítulo se discute el tema con bastante detalle.

Mensaje	Emisor	Receptor	Efecto Ingenuo
Crear( $x,y,r,u$ )	$x$	$y$	$C(x,y,r,u)$
Cancelar( $x,y,r,u$ )	$x$	$y$	$\neg C(x,y,r,u)$
Liberar( $y,x,r,u$ )	$y$	$x$	$\neg C(x,y,r,u)$
Declarar( $x,y,p$ )	$x$	$y$	$p$

Tabla 4.1: Efecto de los mensajes;  $x$  e  $y$  son agentes,  $r$  y  $u$  son proposiciones.

En lo que sigue  $x, y, z$  son agentes,  $p, \dots, w$  son expresiones de FNC o FND;  $\wedge, \vee, \neg$  son las conectivas usuales,  $\vdash$  el símbolo tradicional de inferencia proposicional. Para comenzar la discusión, en la tabla 4.1 se presentan los mensajes que corresponden a operaciones sobre compromisos y sus efectos tomados desde una perspectiva ingenua. En realidad esta perspectiva naïve no se puede aplicar realmente en entornos distribuidos. El mensaje *Declarar*, permite que un agente autorizado establezca un hecho de la realidad

que puede afectar directamente la evaluación de los antecedentes y de los consecuentes de los compromisos. Este mensaje se corresponde con el mensaje *Informar* presentado en las secciones 3.1.2, 3.2 y 3.3 del capítulo 3.

### 4.2.2. El origen de los problemas

Resulta claro que es necesario establecer algunos requerimientos que debe cumplir el medio en el cual se transmiten los mensajes. Por ejemplo, si como consecuencia de las demoras generadas por el uso del medio de comunicación, se produce el cruce entre los mensajes, se origina una condición de carrera en relación a cuales son los compromisos que permanecen pendientes para cada uno de los involucrados.

**Ejemplo 4** (*Falta de alineación*) El escenario de la Figura 4.2 A) muestra que la oferta de José se cruza con el rechazo de Juan.

En la Figura 4.2 B) se plantea el caso donde José presenta el compromiso a Juan, éste le envía la confirmación que el puente ha sido reparado. Antes que llegue la confirmación de Juan, José cancela la oferta. La confirmación y la cancelación se cruzan.  $\square$

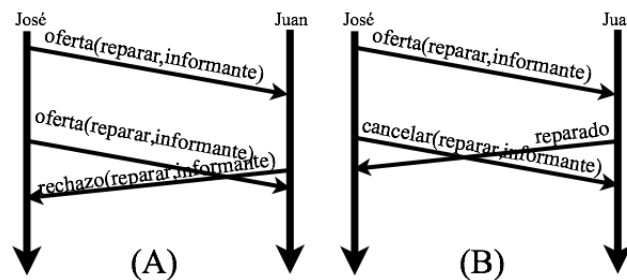


Figura 4.2: Escenarios con falta de alineación

### 4.2.3. Relación entre compromisos

La relación entre los compromisos involucra la implicación lógica entre proposiciones ( $\vdash$ ). Para reducir la notación, se omite el acreedor y el deudor de las expresiones cuando ellos se sobreentienden, por ejemplo se escribe  $C(r,u)$  en lugar de  $C(x,y,r,u)$ . Un compromiso  $C(r,u)$  es más fuerte que  $C(s,v)$  (Con el mismo deudor y acreedor), o  $C(r,u) \succeq C(s,v)$

si y sólo si  $u \vdash v$  y  $s \vdash r$ . Si  $C(r, u) \succeq C(s, v)$  pero  $C(s, v) \not\preceq C(r, u)$ , entonces  $C(r, u)$  es *estrictamente más fuerte* que  $C(s, v)$  o  $C(r, u) \succ C(s, v)$ .

**Ejemplo 5** (*Relación de fuerza entre compromisos*) En el caso del Ejemplo 3, luego de que la primer oferta de José fuese rechazada por Juan, José podría mejorar su oferta con un compromiso más fuerte. Además de llevar al informante también dejará medicamentos en Surina. Como se cumple que:

$$\text{oferta}(\text{reparar}, \text{informante} \wedge \text{remedios}) \succ \text{oferta}(\text{reparar}, \text{informante})$$

existen más posibilidades que la nueva oferta se acepte.  $\square$

Es necesario destacar la importancia que tiene la relación de fuerza entre los compromisos para la administración de los mismos; esta importancia se hace evidente más adelante, en la sección 4.3 donde se enumeran los postulados para administrar los compromisos dentro del almacén de compromisos.

#### 4.2.4. Transacción independiente

Conceptualmente, la idea detrás de una transacción es que encapsula un conjunto de eventos u operaciones relacionadas. Específicamente, los eventos en transacciones independientes no se afectarán entre sí. De esa manera, los eventos que conciernen a un compromiso en una transacción no debieran afectar los compromisos de la otra transacción. Por ejemplo, cancelar el compromiso de reparar el puente hecho en una transacción, no debiera afectar el compromiso de realizar la custodia hecho en otra transacción.

#### 4.2.5. Predicados en las transacciones

Sea  $\mathcal{P}$  un conjunto finito de predicados. Cada predicado del conjunto tiene una aridad determinada y obtiene valores desde una lista de dominios. El primer dominio es el conjunto de los números naturales ( $\mathbb{N}$ ) que es considerado el identificador del predicado. Cada dominio es un conjunto de valores, escritos como constantes en la sintaxis.

Un átomo se obtiene a partir de un predicado, aplicando el predicado a una lista de constantes, cada constante representa un valor tomado desde el correspondiente dominio (las constantes particulares elegidas dependen del razonamiento interno del agente, lo cual no será considerado aquí). Sea  $\iota$  el valor de un identificador en un átomo. Entonces se

dice que el átomo pertenece a la transacción cuyo identificador es  $\iota$ . Más simplemente, se dice que el átomo pertenece a la transacción  $\iota$ . Se requiere que cualquier expresión que involucre dos o más átomos, por ejemplo una FND o un compromiso, involucre átomos con el mismo valor para el identificador. Sea  $\iota$  el valor del identificador en una expresión de compromiso. Entonces se dice que el compromiso (la expresión) pertenece a la transacción  $\iota$ . Específicamente se escribirá  $C(\iota, r, u)$  cuando se desee hacer explícita la identificación. Todo compromiso pertenece a una única transacción pero una transacción puede incluir múltiples compromisos. Debajo, se usa  $\iota$  como el identificador genérico de transacción en las definiciones formales. Para cada transacción  $\iota$ , se define un átomo especial  $\top(\iota)$  que representa verdadero en esa transacción.

La Tabla 4.2 introduce la sintaxis que será usada. Aquí,  $\mathbb{X}$  es un conjunto de agentes. Se consideran expresiones que involucran átomos con la misma identificación. Esta es una restricción saludable porque de otro modo no habría forma de correlacionar los eventos que descargan un compromiso de aquellos que lo desconectan. Específicamente el compromiso  $C(\text{José}, \text{Juan}, \text{reparar}(20, \text{B2}), \text{trasladar}(15, \text{insurgente}))$ , está mal formado porque mezcla identificadores de distintas transacciones, en este caso 20 y 15. Pero  $C(\text{José}, \text{Juan}, \text{reparar}(20, \text{B2}), \text{trasladar}(20, \text{insurgente}))$  cumple con el requisito de integrar la misma transacción.  $C(\text{José}, \text{Juan}, \top(20), \text{trasladar}(20, \text{insurgente}))$  es la variante incondicional del compromiso y en este caso cumple con la restricción impuesta.

El hecho que el antecedente en un compromiso incondicional esté calificado con

Base	$\longrightarrow$	Comp   Átomo   Remove( $\mathbb{X}, \mathbb{X}, \mathbb{N}$ )
Comp	$\longrightarrow$	$C(\mathbb{X}, \mathbb{X}, FND, FNC)$
FND	$\longrightarrow$	And   And $\vee$ FND
FNC	$\longrightarrow$	Or   Or $\wedge$ FNC
And	$\longrightarrow$	Átomo   Átomo $\wedge$ And
Or	$\longrightarrow$	Átomo   Átomo $\vee$ Or
Mensaje	$\longrightarrow$	Declarar( $\mathbb{X}, \mathbb{X}, And$ )   Cancelar( $\mathbb{X}, \mathbb{X}, \mathbb{N}$ )   Crear( $\mathbb{X}, \mathbb{X}, FND, FNC$ )   Liberar( $\mathbb{X}, \mathbb{X}, \mathbb{N}$ )

Tabla 4.2: Sintaxis Formal

el identificador de la transacción, es clave para bloquear el razonamiento entre transacciones distintas. Por ejemplo de  $C(\text{José}, \text{Juan}, \top(20), \text{trasladar}(20, \text{insurgente}))$  y  $C(\text{José}, \text{Juan}, \top(16), \text{trasladar}(16, \text{herido}))$ , no se puede inferir un compromiso más fuer-

te por la aplicación del postulado 4.3.5, no se podría obtener  $C(\text{José}, \text{Juan}, \top(20), \text{trasladar}(20, \text{insurgente}) \wedge \text{trasladar}(16, \text{herido}))$ . Sí se podría obtener un compromiso más fuerte si los identificadores de transacción fueran el mismo en todos los casos.

Notar que en la Tabla 4.2 Cancelar y Liberar toman un identificador de transacción como argumento. Remove es un indicador que registra que un Cancelar o Liberar han sido observados.

En la siguiente sección, se establecen formalmente cuales son las consideraciones que deben realizar los agentes, cuando razonan sobre los compromisos.

### 4.3. Postulados para razonar sobre los compromisos

En esta sección se describen ocho postulados que los agentes deben emplear para razonar sobre los compromisos usando lógica proposicional [Sin08]. Los postulados establecen las reglas de razonamiento que los agentes deben aplicar sobre los compromisos para proceder de una manera consistente dentro del sistema multi-agente.

Como los compromisos pueden pertenecer a una única transacción, se garantiza que las proposiciones que se involucran durante el razonamiento de los agentes se refieren a la misma operación. Teniendo en cuenta la relación entre algunos de los compromisos y entre algunas proposiciones que cumplen el rol de antecedentes y/o consecuentes de algunos compromisos, se define cuales son las reglas lógicas establecidas por los postulados que se pueden aplicar en cada caso; el efecto de cada postulado realiza transformaciones sobre el nuevo conjunto de compromisos que será considerado por los agentes.

Algunos postulados facilitan el razonamiento de los agentes. Existen dos postulados que permiten unificar distintos compromisos en uno solo; si entre un par de agentes, por ejemplo, hay compromisos que solamente difieren en el antecedente, los agentes simplificarán su razonamiento si reemplazan a todos esos compromisos por uno sólo, en el cual el antecedente es una única expresión que combina todos los antecedentes involucrados mediante el operador de disjunción. Otra unificación similar que los agentes pueden hacer, es reunir varios consecuentes con disjunciones.

Cada postulado será descrito de tres maneras diferentes: mediante una breve explicación, un ejemplo tomado de Holistan y la sintaxis formal en lógica proposicional.

### 4.3.1. Cuando es posible dar de baja un compromiso

El hecho de que sea verdadero el consecuente de un compromiso, es razón suficiente para eliminar dicho compromiso. Por ejemplo, si la fuerza militar mantiene un compromiso incondicional con la fuerza humanitaria para quitar la mayor cantidad posible de las minas antipersonales sobre la carretera SR2, ese compromiso desaparecerá una vez que las tropas militares hayan recorrido toda la zona y verificado que no han quedado minas en ese lugar. El nombre de este postulado es *Descarga de compromiso*. Mediante este postulado se establece que no es posible que al mismo tiempo existan la consecuencia y el compromiso.  $u \rightarrow \neg C(r, u)$ .

### 4.3.2. Desconectar una parte del antecedente

Cuando el antecedente de un compromiso es una conjunción de dos átomos, se debe interpretar que se deben cumplir (hacerse verdaderos) los dos átomos del antecedente para que el compromiso se convierta en incondicional. Cuando se cumple solamente uno de ellos, el compromiso se transforma en uno que tiene como antecedente el átomo que no es verdadero. Por ejemplo, para aprovechar lo mejor posible los recursos, la fuerza humanitaria se compromete a que una vez que el puente sea reparado y lleguen los medicamentos, viajarán para iniciar la atención de los heridos en Surina. Si primero sucede que llegan los medicamentos, se deberá esperar que el puente sea reparado para recién iniciar el viaje. Cuando se verifica uno de los átomos del antecedente, se dice que una parte del antecedente se ha *desconectado* y queda lo que resta del antecedente original como antecedente de la nueva versión del compromiso. Como se trata de la misma transacción, la identificación del compromiso no cambia. Si el antecedente es la conjunción de los átomos  $r$  y  $s$ , es decir,  $r \wedge s$ , de los cuales  $r$  se verifica como verdadero, el antecedente se convierte en  $s$ .  $C(r \wedge s, u) \wedge r \rightarrow C(s, u)$ .

### 4.3.3. Aumento de la fuerza del antecedente

Si la relación entre dos átomos es tal que el agente puede inferir la verdad de uno de ellos a partir de la verdad del otro, es decir que: dados un par de átomos  $s$  y  $r$  tal que  $s \vdash r$ ; cuando  $r$  aparezca como antecedente en un compromiso  $C(r, u)$ , ese compromiso se transforma en otro compromiso  $C(s, u)$  con un antecedente  $s$  más fuerte. Como ejemplo se

puede considerar la siguiente situación: estando en la base, un chofer se ha comprometido con un miembro de la fuerza humanitaria para traer de regreso a un enfermero que está en Padra. Por otro lado como parte de sus obligaciones debe transportar dos soldados para reparar el puente B2 sobre SR2. Para reparar B2 debe pasar por Padra, por lo que el antecedente del compromiso original que fue viajar a Prada, ahora se ha transformado en reparar el puente B2. Es decir que dado  $C(r, u) \wedge s \vdash r$  el agente debiera inferir  $C(s, u)$  como un compromiso vigente. Como regla general, todo compromiso que se cumple para un antecedente más débil también se cumple para un antecedente más fuerte. Aquí,  $p \vdash q$  significa que es posible probar  $q$  a partir de  $p$ , lo cual es más fuerte que la implicación  $p \rightarrow q$ , la cual se cumple simplemente si  $p$  es falso. Claramente,  $C(r, u) \wedge \neg s \rightarrow C(s, u)$  es falso, es decir que no se puede concluir  $C(s, u)$  simplemente porque la ocurrencia de  $s$  sea falso.

#### 4.3.4. Unificar dos antecedentes mediante disjunción

Puede ocurrir que dos compromisos en una misma transacción tengan el mismo consecuente, en este caso los agentes deben considerar la existencia de un único compromiso, en el cual el antecedente es la disjunción de los antecedentes de cada uno de los compromisos originales. Como un ejemplo en el contexto de Holistan es posible que sea necesario transportar un cargamento de vacunas desde Blogohar hasta el hospital, como responsable del transporte, la fuerza militar se ha comprometido con la fuerza humanitaria a llevar las vacunas hasta el hospital. Por razones operativas una parte del cargamento irá por tierra, en camiones por la autopista HW y el resto de la mercadería llegará en helicoptero. De manera general se establece que:  $C(r, u) \wedge C(s, u) \rightarrow C(r \vee s, u)$ . La expresión a la izquierda de  $\rightarrow$  significa que  $x$  se ha comprometido a  $u$  si  $r$  o  $s$  se cumplen, que es el significado de la expresión a la derecha.

#### 4.3.5. Unificar dos consecuentes mediante conjunción

También puede ocurrir que dos compromisos que forman parte de una misma transacción, compartan un mismo antecedente y difieran en sus consecuencias. Ambos compromisos se unirán en uno solo en el cual la consecuencia es la conjunción de las dos consecuencias originales, conservando el mismo antecedente. Esta situación podría ocurrir en Holistan si la fuerza militar instalada en Prada se ha comprometido con la fuerza

humanitaria del Hospital para: reparar la ambulancia y entregar un informe del estado sanitario de los habitantes de Prada; las dos cosas, una vez que los enfermeros que vienen desde el Hospital vacunen a las tropas apostadas en esa ciudad. Con este postulado se regula la siguiente situación general: dados  $C(r, u) \wedge C(r, v) \rightarrow C(r, u \wedge v)$ . La expresión a la izquierda de  $\rightarrow$  significa que  $x$  se ha comprometido con  $u$  y  $v$  si  $r$  se cumple, que es el significado de la expresión a la derecha.

#### 4.3.6. No existe un compromiso cuyo consecuente sea falso

Este postulado se requiere para mantener consistencia en lo que un agente se puede comprometer a realizar. En el contexto de Holistan, el personal del Hospital no se pueden comprometer a recibir más de 40 pacientes, ya que esa es la capacidad máxima del Hospital.  $\neg C(r, \perp)$ . Un agente no se puede comprometer con algo que es falso.

#### 4.3.7. El consecuente no será inferencia directa del antecedente

Para los agentes, la existencia de un compromiso establece que la evolución de los hechos los llevará a un estado, en el cual será verdadera la proposición que aparece como consecuente en el compromiso. Esa evolución, entre otras cosas, deberá ser producto del accionar del deudor. Si en cambio, una vez que se cumple el antecedente, el consecuente ocurre espontáneamente, no hay necesidad de un compromiso. Los agentes deberán inferir  $\neg C(r, u)$  a partir de  $r \vdash u$ . Si cuando  $r$  se cumple también se cumple  $u$ , el compromiso  $C(r, u)$  no debiera existir. De esa manera  $C(r, u)$  será descartado por ejemplo si es el resultado de la desconexión (postulado 4.3.2) de un compromiso anterior. Esa situación llevada al ejemplo de Holistan, sería : Juan, de la fuerza militar, mantiene con José que es de la misma fuerza, un compromiso con un antecedente que es la conjunción de dos átomos: *llegar hasta el camión lanza-misiles* y *cargar el lanza-misiles*. Si el consecuente del compromiso fuese: *habrá un misil en el lanza-misiles*, una vez que Juan llegue al lugar, por la aplicación del postulado 4.3.2 el compromiso original se transformará en  $C(\text{"Cargar el lanza-misiles"}, \text{"habrá un misil en el lanza-misiles"})$ , que no es ningún compromiso.



### 4.3.8. Relajación en el consecuente

En general, si un agente crea un compromiso con un consecuente que es la conjunción de dos átomos, el agente está comprometido con cada uno de esos átomos. Sin embargo la siguiente formulación obvia: que a partir de un compromiso que tiene un consecuente que es la conjunción de dos átomos, se puede deducir la existencia de un compromiso con un consecuente más débil ( $C(r, u \wedge v) \rightarrow C(r, u)$ ), es inconsistente con el primero de los postulados enunciados; ya que si  $u$  se cumple, el postulado 4.3.1 eliminaría  $C(r, u)$ . Un médico se compromete con un soldado a que si resulta herido, el personal del Hospital se encargará de cuidarlo y alimentarlo, eso significa que mientras el soldado no esté recuperado, el médico está comprometido a curarlo. Este planteo del postulado asegura que el compromiso original no se transformará por otro que tenga un consecuente más débil. En una notación más precisa, este postulado establece lo siguiente:  $C(r, u \wedge v) \wedge \neg u \rightarrow C(r, u)$ . Cabe aclarar que debido a que el postulado 4.3.1 es fundamental para reglamentar la eliminación de un compromiso, para que este postulado de relajación del consecuente sea correcto se, incluye  $\neg$  a la izquierda de  $\rightarrow$ .

El último de los postulados enunciados captura de manera deductiva la noción de fuerza entre compromisos, presentada en la sección 4.2.3. Si el compromiso más fuerte está en vigencia, entonces por el postulado 4.3.8 se infiere que el compromiso más débil también está en vigencia.

Esta situación se puede observar en el ejemplo 5, donde se dijo que se cumplía lo siguiente:

$$oferta(reparar, informante \wedge remedios) \succ oferta(reparar, informante)$$

entonces, si existe el compromiso fuerte  $oferta(reparar, informante \wedge remedios)$  de acuerdo al postulado 4.3.8, el compromiso débil  $oferta(reparar, informante)$  es redundante por lo que se debería eliminar del almacén de compromisos del agente.

## 4.4. Características técnicas del enfoque social

Entre otras reglas establecidas por J.R. Searle [Sea69] para la comunicación, se destaca una sobre *las condiciones de entrada/salida normales*, que fue mencionada en la Sección 3.1. Esta norma establece la necesidad de que el oyente no sea sordo y que el orador no sea mudo. Es algo que puede parecer obvio y está claramente entendido en la comunicación

entre humanos, pero resulta ser algo difícil de establecer para los agentes computacionales. ¿Cómo se puede garantizar que un agente ha recibido un mensaje? ¿Qué mensajes realmente intercambian los agentes?. ¿En qué orden son percibidos esos mensajes?. ¿Cómo se percibe y registra cada mensaje individual en el estado de los agentes?. ¿Y en el sistema multi-agente en general?.

Los mensajes son eventos relevantes dentro del sistema multi-agente y cada agente involucrado debe registrar la ocurrencia de esta clase de evento. Por esa razón es necesario establecer un criterio general que los agentes deben adoptar para manejar toda la secuencia de eventos que corresponde a un diálogo y equiparlos con herramientas que permitan analizar esta secuencia completa o por secciones. De esta manera los agentes podrán hacer inferencias en base a las afirmaciones intercambiadas en cada parte del diálogo.

En el resto de esta sección se establece cómo el sistema multi-agente contabiliza los mensajes a partir de lo registrado por cada agente y cómo se pueden obtener conclusiones partiendo de un estado que garantice las condiciones adecuadas.

#### 4.4.1. Mensajes y observaciones

Dado  $\mathcal{A}$  que es un sistema de  $k$  agentes. Entonces  $O = [O_0 \dots O_i \dots O_{k-1}]$  es un *vector de observación* sobre  $\mathcal{A}$ , donde los  $O_{i,s}$  son las *secuencias de observación* de los  $k$  agentes. Cada mensaje que es enviado o recibido por un agente se registra en su secuencia de observación. Un vector de observación es así una instantánea del sistema. En lo que sigue  $O$  es un vector de observación y  $O_x$  es la secuencia de observación del agente  $x$  en el vector  $O$ . A continuación se define un concepto importante que involucra a vectores y secuencias de observación.

**Definición 1** . *El largo de una secuencia de observación es el número de observaciones que contiene. El largo de un vector de observaciones es la suma de los largos de las secuencias de observación que lo componen. Si  $O = [O_0 \dots O_i \dots O_{k-1}]$  es de largo  $n$ , entonces se usa  $O^n$  para indicar su longitud y  $O_i^m$  para denotar la secuencia de  $m$  observaciones del agente  $i$ .*

**Definición 2** . *Dado  $o'$  una secuencia de observación. Entonces  $o$  es una subsecuencia de  $o'$ , denotado por  $o \sqsubseteq o'$  si y sólo si*

- $o' = o$ , *o bien*
- *para algún mensaje  $m$ ,  $o' = o; m$ , o bien*
- *para alguna secuencia de observación  $o''$ ,  $o \sqsubseteq o'' \sqsubseteq o'$ .*

Si  $o \sqsubseteq o'$  pero  $o \neq o'$ , entonces se dice que  $o'$  extiende estrictamente a  $o$  y se escribe  $o \sqsubset o'$ .

**Definición 3** . Dado  $O = [O_0 \dots O_i \dots O_{k-1}]$  y  $O' = [O'_0 \dots O'_i \dots O'_{k-1}]$ . Se dice que  $O'$  extiende  $O$ , denotado como  $O \sqsubseteq O'$ , si y sólo si cada  $O_i \sqsubseteq O'_i$ . Si  $O \sqsubseteq O'$  pero  $O \neq O'$ , entonces se dice que  $O'$  extiende estrictamente a  $O$  y se escribe  $O \sqsubset O'$ .

Por último, la Definición 4 introduce un operador que devuelve el resto de una secuencia de observación, después de eliminar un prefijo.

**Definición 4** . Dado  $o = \langle m_0, \dots, m_l \rangle$  y  $o' = \langle m_0, \dots, m_l, m_{l+1}, \dots, m_i \rangle$ , entonces  $o' \setminus o = \langle m_{l+1}, \dots, m_i \rangle$ .

#### 4.4.2. Instantánea de los agentes y mantenimiento local

En esta sección se formaliza el tratamiento de la *instantánea* de un agente, su estado local surge de su secuencia de observación. En términos sintácticos, una instantánea es una estructura de datos finita. En términos semánticos, una instantánea produce un conjunto de proposiciones que el agente puede inferir lógicamente a partir de sus observaciones. Para permitir el progreso de la instantánea de un agente a partir de sus observaciones, se introducen dos operadores de actualización, llamados *adición*  $\oplus$  y *substracción*  $\ominus$ , los cuales respectivamente suman o restan proposiciones desde la instantánea. Estos operadores aseguran que la representación resultante es consistente con los postulados de la sección 4.3 y que solamente los cambios esenciales (es decir mínimos) son producidos por cada mensaje. A continuación se describe una forma en la cual se pueden interpretar los operadores de actualización con teoría de conjuntos.

Dado  $\mathcal{S}$ , un conjunto de proposiciones básicas, es decir, átomos, compromisos o remociones. Específicamente  $\mathcal{S}$  representa la instantánea de un agente surgida de una secuencia de observaciones  $o$ , cuando se incorpora un nuevo mensaje  $m$  a  $o$  se actualiza la instantánea

de acuerdo a  $\mathcal{S}(o; m)$ . En términos lógicos, el estado de un agente es  $\llbracket \mathcal{S} \rrbracket$ , la clausura deductiva de su instantánea. Aquí se extiende  $\vdash$  para ser aplicado a instantáneas:  $\mathcal{S} \vdash p$  significa que  $p \in \llbracket \mathcal{S} \rrbracket$ . Se considera que  $\vdash$  respeta los postulados desde 4.3.1 al 4.3.8.

Aquí se definen los operadores para modificar las instantáneas. En lo que sigue se considera que  $\beta$  es un átomo o un *remover(i)*. También que  $\mathbb{X}(\mathcal{S})$  es el conjunto de compromisos en  $\mathcal{S}$ . Es decir,  $\mathbb{X}(\mathcal{S}) = \{C(r, u) : C(r, u) \in (\mathcal{S})\}$ .

La siguiente Definición dice que al agregar un  $\beta$  a  $\mathcal{S}$  se eliminan todos los compromisos que son descargados por el operador de adición; agregar un compromiso solamente es efectivo si la transacción no ha sido removida y si el compromiso no ha sido descargado.

Se define  $\oplus$  de la siguiente manera:

- Dados  $\mathcal{S}$  y  $\beta$ ,  $\mathcal{S} \oplus \beta = (\mathcal{S} \cup \{\beta\}) \setminus \{C(r, u) : C(r, u) \in \mathbb{X}(\mathcal{S}), \mathcal{S} \cup \{\beta\} \not\vdash u\}$
- Dados  $\mathcal{S}$  y  $\beta_1 \dots \beta_k$ ,  $\mathcal{S} \oplus (\beta_1 \wedge \dots \wedge \beta_k) = (\dots (\mathcal{S}(o) \oplus \beta_1 \dots) \oplus \beta_k)$
- Dados  $\mathcal{S}$  y  $C(i, r, u)$ ,  $\mathcal{S} \oplus C(i, r, u) = \mathcal{S}$  si  $\mathcal{S} \vdash \text{remover}(i) \vee u$  sino  $\mathcal{S} \cup \{C(i, r, u)\}$

La siguiente Definición establece que *remover* una transacción equivale a eliminar todos los compromisos con el identificador de aquella transacción.

Dados la instantánea  $\mathcal{S}$  e  $i$ , se define  $\ominus$  de la siguiente manera:

$$\mathcal{S} \ominus i = \mathcal{S} \oplus \text{remover}(i) \setminus \{C(i, s, v) : C(i, s, v) \in \mathbb{X}(\mathcal{S})\}.$$

Sea  $\mathcal{S}(o)$  la instantánea correspondiente a la secuencia de observaciones  $o$ . Entonces se puede estimar  $\mathcal{S}(o; m)$  desde  $\mathcal{S}(o)$  por medio de una o más aplicaciones de los operadores de actualización de arriba. La aplicación específica de los operadores depende de la semántica del mensaje  $m$ . En los siguientes puntos  $U_1$  a  $U_5$ ,  $o$  es la secuencia de observación de  $x$  o de  $y$ .  $\mathcal{S}(\langle \rangle)$  se puede almacenar fácilmente de manera finita.

- $U_1$ .  $\mathcal{S}(\langle \rangle) = \{\top[i]\}$ , para cada posible asociación de  $i$ .
- $U_2$ .  $\mathcal{S}(o; \text{Crear}(r, u)) = \mathcal{S}(o) \oplus C(r, u)$ .
- $U_3$ .  $\mathcal{S}(o; \text{Liberar}(i)) = \mathcal{S}(o) \ominus i$ .
- $U_4$ .  $\mathcal{S}(o; \text{Cancelar}(i)) = \mathcal{S}(o) \ominus i$ .

- $U_5. \mathcal{S}(o; Declarar(\beta_1 \wedge \dots \wedge \beta_k)) = (\dots (\mathcal{S}(o) \oplus \beta_1) \dots) \oplus \beta_k$ .

La única restricción es que una vez que un agente infiere un átomo con un identificador  $i$ , ya no envía nuevos mensajes *Crear* con el identificador  $i$ . Esto refleja la idea que una vez que un agente comienza a descargar un compromiso en una transacción, no debiera incluir nuevos compromisos a la transacción.

El siguiente lema establece que si una secuencia de observaciones deriva un compromiso de identificación  $i$ , entonces no puede derivar su consecuente o *remove*( $i$ ).

**Lema 1** . Si  $\mathcal{S}(o) \vdash C(i, r, u)$ , entonces  $\mathcal{S}(o) \vdash \neg remove(i) \wedge \neg u$

## 4.5. Estados del sistema relevantes para la alineación

A continuación se establecen los requerimientos que un vector debe cumplir para ser considerado *completo* en relación a un compromiso; si de acuerdo al vector el deudor no infiere el compromiso, entonces el acreedor tiene tres alternativas: tampoco inferir el compromiso, o saber que ya fue descargado o que la transacción completa ya fue removida.

**Definición 5** .  $O$  es completo con respecto a  $C(r, u)$  si y solo si  $\mathcal{S}(O_x) \vdash \neg C(r, u) \Rightarrow \mathcal{S}(O_y) \vdash \neg C(r, u) \vee u \vee remove(i)$ .

Notar que si  $O$  es completo en relación a  $C(s, v)$  y  $C(s, v) \succ C(r, u)$ , no se puede afirmar que  $O$  sea completo en relación a  $C(r, u)$ . Para ver esto, dado  $O_x$  que sea  $\langle Crear(x, y, r, u \wedge v), Declarar(z, x, u) \rangle$  y  $O_y$  sea  $\langle Crear(x, y, r, u \wedge v) \rangle$ ; este  $O$  es completo con respecto a  $C(x, y, r, u \wedge v)$  pero no completo en relación a  $C(x, y, r, u)$ .

La Definición 6 dice que un vector está alineado en relación con un compromiso  $\iota$  si y sólo si se cumplen las siguientes condiciones: si la secuencia de observación de un acreedor en el vector deriva el compromiso  $\iota$ , entonces la secuencia de observación del deudor en el vector también deriva el compromiso  $\iota$ .

**Definición 6** .  $O$  está alineado en relación a  $C(r, u)$  si y sólo si  $\mathcal{S}(O_y) \vdash C(r, u)$  implica  $\mathcal{S}(O_x) \vdash C(r, u)$ .

Notar que la Definición 6 no se basa en ningún razonamiento interno o estado del agente. Si un compromiso tiene o no vigencia para un agente es propiedad solamente de las observaciones del agente.

Los teoremas 1 y 2 muestran que la alineación y la completitud coinciden.

**Teorema 1** . [*Alineación*] Si  $O$  es completo en relación a  $C(r, u)$ , entonces  $O$  está alineado con respecto a  $C(r, u)$ .

**Demostración.** Sea  $\iota$  el identificador de  $C(r, u)$ . Si  $O$  es completo en relación a  $C(r, u)$ , entonces por Definición 5,  $S(O_x) \vdash \neg C(r, u) \Rightarrow S(O_y) \vdash \neg C(r, u) \vee u \vee \neg \text{remover}(\iota)$ . Esto es, i)  $S(O_x) \vdash C(r, u)$  o ii)  $S(O_y) \vdash \neg C(r, u) \vee u \vee \neg \text{remover}(\iota)$ . Si i) entonces  $O$  está alineado en relación a  $C(r, u)$ . Si ii) entonces, (a)  $S(O_y) \vdash \neg C(r, u)$  entonces  $O$  está alineado en relación a  $C(r, u)$ ; (b) si  $u \vee \neg \text{remover}(\iota)$ , entonces por contraposición con el Lema 1,  $S(O_y) \vdash \neg C(r, u)$ . Por lo tanto  $O$  está alineado en relación a  $C(r, u)$ .

**Teorema 2** . [*Completitud*] Si  $O$  está alineado en relación a  $C(r, u)$ , entonces  $O$  es completo con respecto a  $C(r, u)$ .

**Demostración.** Se consideran dos casos: En el caso 1,  $S(O_y) \vdash C(r, u)$ . Entonces debido a que  $O$  está alineado en relación a  $C(r, u)$ ,  $S(O_x) \vdash C(r, u)$ . Por lo tanto por Definición 5,  $O$  es completo en relación a  $C(r, u)$ . Caso 2,  $S(O_y) \vdash \neg C(r, u)$ , por Definición 5,  $O$  es completo en relación a  $C(r, u)$ .

## 4.6. Progreso hacia estados de alineación

El efecto que la comunicación produce en la evolución del sistema lo conduce idealmente a estados en los cuales los compromisos entre los agentes están alineados. Para llegar a esa conclusión se asume que los mensajes enviados llegarán al destinatario. Un estado en el cual los compromisos están alineados (vector de observaciones alineado o completo) es requerido para un correcto funcionamiento del sistema multi-agente. Pero ¿que sucede si como producto de la interacción de los agentes se llega a un vector “no alineado” o “incompleto”? Para comenzar hace falta formalizar los elementos requeridos al momento de hacer una emisión en un entorno descentralizado.

En primer lugar es necesario asumir que se cumple el principio de *causalidad*. Un agente siempre es capaz de enviar un mensaje (autonomía), mientras que sólo será capaz de recibir un mensaje, si previamente fue enviado. También se requiere asumir el principio de *confiabilidad*. Cada mensaje que es enviado, arriba a su destino.

Los eventos relevantes para lograr alineación de compromisos son el envío y la recepción de mensajes, pero debido a la característica distribuida del sistema multi-agente, no es posible garantizar que los agentes perciban los eventos al mismo tiempo. Por eso es conveniente aclarar las condiciones requeridas para asegurar que en el sistema multi-agente se ha percibido un evento.

**Definición 7 (Percibir)** *Un sistema multi-agente  $\mathcal{A}$  de  $n$  agentes  $\{1, \dots, i, \dots, n\}$  percibe un vector de observación  $O=[O_1 \dots O_n]$  si, y sólo si, se cumple alguna de estas condiciones:*

- $O=[\langle \rangle \dots \langle \rangle]$ , un vector de  $n$  secuencias de observación vacías,  $o$ .
- $O=[O_1 \dots O_i; \text{envía}(j, m) \dots O_n]$  y  $\mathcal{A}$  percibe  $[O_1 \dots O_i \dots O_n]$ ;  $o$
- $O=[O_1 \dots O_i; \text{recibe}(j, m) \dots O_n]$ ,  $\mathcal{A}$  percibe  $[O_1 \dots O_i \dots O_n]$  y  $O_j$  contiene  $\text{envía}(i, m)$ .  $\square$

Dependiendo de los eventos y de la información que cada agente percibe, nuevos eventos de comunicación deben ser generados. Es así que bajo determinadas circunstancias se generan mensajes a otros agentes. A continuación, se presentan cuatro definiciones en las cuales se detallan situaciones de inicio de notificaciones con mensajes a otros agentes. La siguiente Definición caracteriza los mensajes que el deudor debe enviar a los acreedores de aquellos compromisos que son descargados una vez que el deudor hizo una observación, digamos la  $n + 1$ . Este conjunto de mensajes se denomina *conjunto de notificaciones por descarga para la observación  $n+1$* . La única advertencia es que se debe enviar la notificación sólo si se trata del compromiso más fuerte. La Figura 4.3 muestra como el agente  $z$  informa al agente  $y$  sobre el evento  $u'$  por el que se descarga un compromiso previo.

**Definición 8** . (*Notificaciones por descarga*).  $DA(z, n)$ , el conjunto de notificación por descarga de  $z$  en la observación  $n + 1$  es  $\{\text{Declarar}(z, y, u') : S(O_z^n) \vdash C(z, y, r, u) \text{ y } S(O_z^n) \vdash \neg C(z, y, s, v) \text{ tal que } C(z, y, s, v) \succ C(z, y, r, u) \text{ y } S(O_z^{n+1}) \vdash u' \text{ tal que } u' \vdash u\}$ .

$\square$

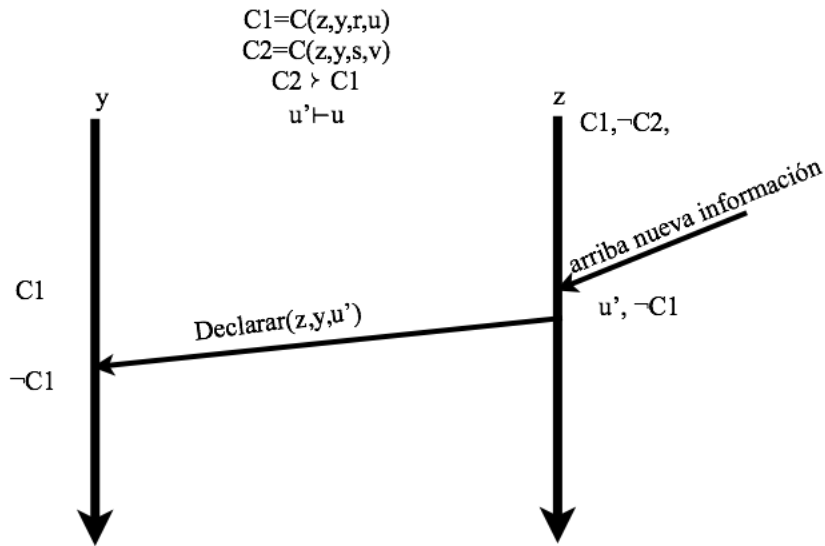


Figura 4.3: Notificación por descarga

La siguiente Definición caracteriza los mensajes que un acreedor debe enviar a sus deudores al hacer una observación que desconecta un compromiso que se encuentra activo (nuevamente, el más fuerte). La Figura 4.4 muestra cómo el agente  $z$  informa al agente  $x$  sobre la desconexión de  $s$  en un compromiso previo.

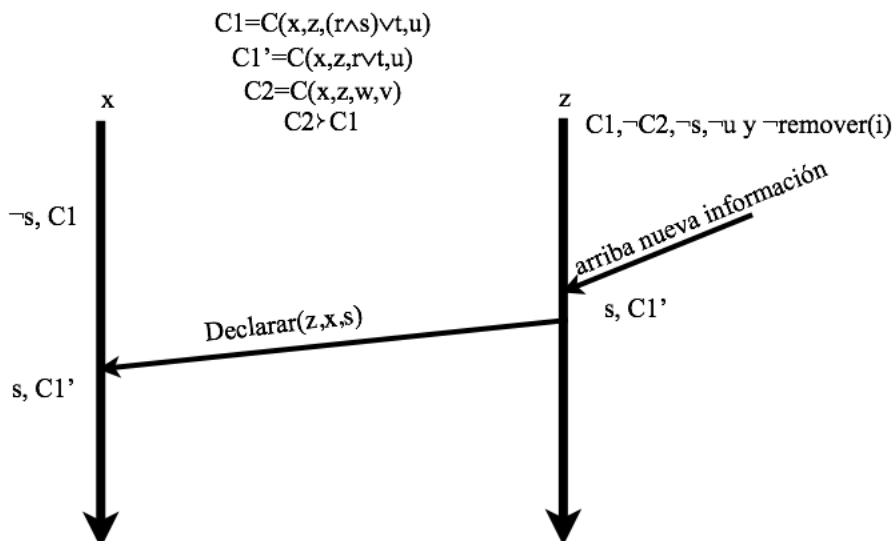


Figura 4.4: Notificación pos desconexión

**Definición 9** . (Notificaciones por desconexión posterior).  $DOPO(z, n)$ , el conjunto de notificación por desconexión posterior de  $z$  en la observación  $n + 1$  es  $\{Declarar(z, x, s) :$



$S(O_z^n) \vdash \neg C(x, z, (r \wedge s) \vee t, u) \wedge \neg C(x, z, w, v) \wedge \neg s$  tal que  $C(x, z, w, v) \succ C(x, z, (r \wedge s) \vee t, u)$  y  $S(O_z^{n+1}) \vdash s \wedge \neg u \wedge \neg \text{remover}(i)$ .  $\square$

La siguiente Definición caracteriza los mensajes que un acreedor debe enviar al deudor de un compromiso, una vez que recibe el mensaje *Crear* para un compromiso que podría ser, al menos parcialmente desconectado debido a que una parte del antecedente ya se cumple. La Figura 4.5 muestra cómo el agente  $z$  informa al agente  $x$  sobre la desconexión de  $s$  en el compromiso creado.

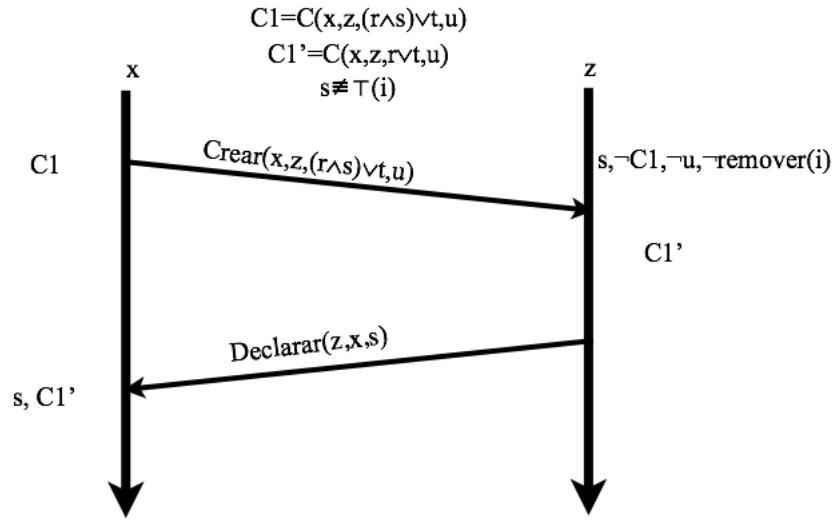


Figura 4.5: Notificación pre desconexión

**Definición 10** (Notificaciones por desconexión previa).  $DOPR(z, n)$ , el conjunto de notificación por desconexión previa de  $z$  en la observación  $n + 1$  es  $\{\text{Declarar}(z, x, s) : S(O_z^n) \vdash s \wedge \neg C(x, z, (r \wedge s) \vee t, u) \wedge \neg u \wedge \neg \text{remover}(i)$  y  $n+1$  es la observación de  $z$  recibir( $x, \text{Crear}(x, z, (r \wedge s) \vee t, u)$ ) tal que  $s \neq \top(i)\}$ .  $\square$

La siguiente Definición caracteriza el conjunto de todos los mensajes que un agente debe enviar una vez que hizo una observación. La Figura 4.6 muestra como el agente  $z$  informa al agente  $x$  sobre la descarga de  $s$  en un compromiso previo y  $z$  notifica al agente  $y$  sobre la desconexión posterior del antecedente en un compromiso que  $y$  mantiene con  $z$ .

Hay que hacer una aclaración: si la observación  $n$  es una declaración, sólo puede haber notificación por descarga o por desconexión posterior, no por desconexión previa porque el mensaje no crea ningún compromiso. Sólo si el mensaje se trata de la creación de un compromiso, puede haber una notificación por desconexión posterior.

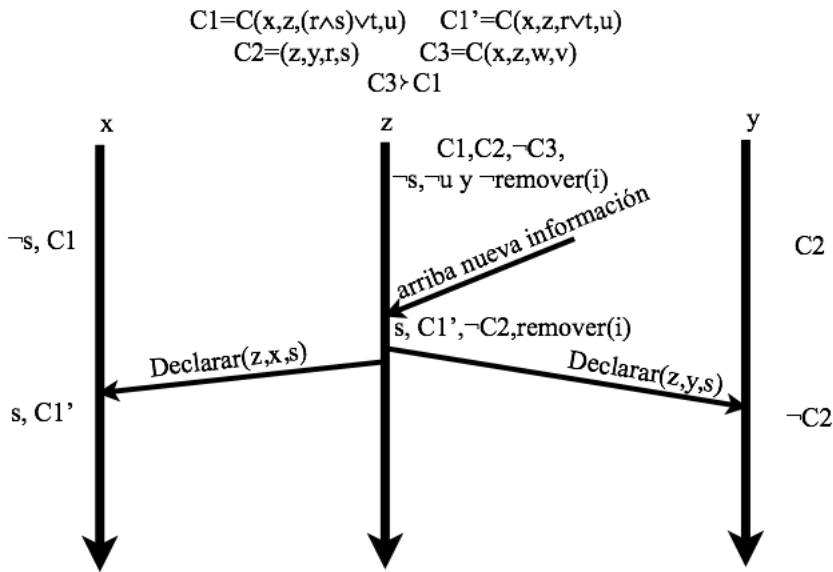


Figura 4.6: Notificación por descarga y por desconexión posterior

**Definición 11** (Notificaciones por evento).  $DS(z, n)$ , el conjunto de notificación del agente  $z$  con la observación  $n$  es  $DA(z, n) \cup DOPO(z, n) \cup DOPR(z, n)$ .  $\square$

La siguiente Definición captura el estado de un agente que ha enviado los mensajes del conjunto de notificación para una secuencia de observación.

**Definición 12** El agente  $z$  hizo la notificación de  $n$  en  $k$  si y sólo si  $O_z^n \sqsubseteq O_z^k$  y para cada mensaje  $m \in DS(z, n)$ ,  $O_z^k \setminus O_z^n$  contiene  $\text{enviar}(x, m)$ .  $\square$

Los mensajes enviados desde el conjunto de notificación pueden en sí mismos generar notificaciones. La siguiente Definición caracteriza la idea de clausura de notificaciones.

**Definición 13** El agente  $z$  ha hecho la clausura de notificación si y sólo si para cualquier  $n$  existe un  $k$  tal que  $z$  ha hecho la notificación para cada  $j$  ( $n \leq j \leq k$ ) en  $k$ .  $\square$

El siguiente Lema expresa una propiedad importante: si un agente simplemente envía mensajes en el conjunto de notificación por una observación  $n$ , sin que cualquier otra observación intervenga, entonces la secuencia resultante está garantizada que es cerrada sobre las notificaciones. Esto permite tratar el envío de notificaciones de manera atómica con la observación que las hizo necesarias.

**Lema 2** *Sea  $k$  el número más chico para el cual  $z$  ha hecho la notificación para  $n$  en  $k$ . Entonces  $z$  es cerrado sobre la notificación.*  $\square$

**Definición 14** *Un sistema  $A$  de  $n$  agentes  $\{1 \dots i \dots n\}$  aceptablemente percibe un vector de observaciones  $O = [O_1 \dots O_n]$  si y sólo si*

- *$A$  percibe  $O = [O_1 \dots O_n]$*
- *Cada  $i$  es cerrado sobre la notificación.*  $\square$

El siguiente Teorema establece que si un sistema aceptablemente percibe un vector de observación que es incompleto con respecto a un compromiso, entonces el sistema aceptablemente percibirá un vector de observación que es completo en relación al compromiso. El Teorema 3 garantiza completitud solamente en relación a los compromisos más fuertes. En otras palabras, el hecho que un vector sea completo para un compromiso particular, por ejemplo para  $C(r, u \wedge v)$ , no garantiza que el mismo vector sea o no sea incompleto con respecto a un compromiso más débil, en este caso respecto a  $C(r, u)$ .

**Teorema 3** *Si un sistema  $A$  aceptablemente percibe un vector de observación  $O$  que es incompleto en relación a  $C(r, u)$  y  $C(r, u)$  es el compromiso incompleto más fuerte, entonces o bien  $A$  aceptablemente percibe un vector de observación  $O'$ , tal que  $O \sqsubset O'$  y  $O'$  es completo en relación a  $C(r, u)$  o  $(\exists i \in \mathcal{A}, C(s, v) : C(s, v) \succ C(r, u)$  y  $\mathcal{S}(O_i) \vdash C(s, v)$ ).  $\square$*

**Demostración.** La demostración es por inducción sobre el largo de la expresión del compromiso, esta expresión es definida como el número de átomos en la expresión. Sea  $C(p, q)$  un compromiso de largo mínimo (es decir de largo dos). Por la Definición 5,  $O^0$  es percibido aceptablemente y completo con respecto a  $C(p, q)$ . Considere cualquier vector completo  $O^c$ . Sea  $O^i$  la extensión aceptable más chica de  $O^c$  que es incompleto en relación con  $C(p, q)$ . Esto significa que la siguiente condición se cumple por la Definición 5:  $\mathcal{S}(O_x^i) \vdash \neg C(p, q) \wedge \mathcal{S}(O_y^i) \vdash C(p, q) \wedge \neg q \wedge \neg \text{remover}(l)$ . Hay dos casos primarios a considerar. (1)  $\mathcal{S}(O_x^{i-1}) \vdash C(p, q)$ , lo que significa que lo que hizo a  $O^i$  incompleto fue que  $x$  terminara de inferir  $C(p, q)$ . En este caso, o bien un Cancelar o una notificación por descarga está en camino a  $y$ , cuando esa notificación sea recibada por  $y$ , hará completo al vector resultante. (2)  $\mathcal{S}(O_y^{i-1}) \vdash \neg C(p, q)$ , lo que significa que lo que hizo a  $O^i$  incompleto

fue que comenzó a inferir  $C(p, q)$ . Si esto fue por un Crear más fuerte, ya estaría listo por el mismo planteo del Teorema. Si es por una desconexión, entonces aceptablemente significa que una notificación está en camino a  $x$ , la cual cuando arribe producirá un vector completo.

Asumiendo que el Teorema se cumple para todos los compromisos de largo  $k$ . Entonces se debe probar que se cumple para compromisos de largo  $k + 1$ . Por inspección se ve que hay dos casos relevantes. Estos corresponden a los postulados de 4.3.4(disjunción) y 4.3.5(conjunción), respectivamente.

Consideremos en primer lugar el postulado de 4.3.4(disjunción). Sea  $C(r, u)$  y  $C(s, v)$  cada uno de largo  $k$  y  $C(r \vee s, u)$  de largo  $k + 1$ . Sabemos que  $O^0$  es completo en relación a  $C(r \vee s, u)$ . Sea  $O^c$  cualquier vector completo en relación a  $C(r \vee s, u)$ . Como antes, sea  $O^i$  la extensión incompleta con respecto a  $C(r, u \wedge v)$  más pequeña que se percibe aceptablemente. Entonces, deseamos probar que el sistema aceptablemente percibe una extensión completa de  $O^i$ . Por la hipótesis inductiva, sabemos que habrá una extensión  $O^j$  de  $O^i$  que será completa en relación a  $C(r, u)$  y  $C(s, v)$ . Esto significa que,  $S(O_x^j) \vdash \neg C(r, u)$  y  $S(O_y^j) \vdash \neg C(r, u) \wedge \neg u \wedge \neg \text{remover}(\iota)$  y  $S(O_x^j) \vdash \neg C(s, v)$  y  $S(O_y^j) \vdash \neg C(s, v) \wedge \neg u \wedge \neg \text{remover}(\iota)$ . De lo cual se infiere que  $S(O_x^j) \vdash \neg C(r \vee s, u)$  y  $S(O_y^j) \vdash \neg C(r \vee s, u) \wedge \neg u \wedge \neg \text{remover}(\iota)$ . En otras palabras,  $O_j$  es completo en relación a  $C(r \vee s, u)$ .  $\square$

# Capítulo 5

## Perspectiva Argumentativa

Este capítulo se basa en el trabajo [AES16] y considera el uso de la argumentación en la comunicación procurando que los participantes progresen en sus diálogos de una manera que facilite el logro de los objetivos individuales. Entre los distintos tipos de diálogos que existen, los que resultan ser de interés en este capítulo son aquellos en los cuales existe un conflicto entre los participantes y a pesar de eso, se intenta llegar a una resolución que sea aceptada por todos. De acuerdo a la clasificación establecida por Walton y Krabbe, este tipo de diálogo corresponde a los diálogos de persuasión [WK95].

En estos diálogos de persuasión, la situación inicial es un conflicto de opinión lo que impide alcanzar una efectiva comunicación; para superar esta circunstancia se procura hallar una solución basada en la comunicación, una que sea dialógica. En esta búsqueda se encuentran los sistemas de diálogo que se caracterizan por establecer una serie de reglas sobre el uso de los actos del habla, permitiendo que los participantes intercambien afirmaciones o reclamos, planteen desafíos sobre las afirmaciones de otro participante<sup>1</sup>, acepten las afirmaciones del otro, puedan retirar las afirmaciones propias hechas anteriormente o presenten argumentos a favor de ellas. Estos sistemas de diálogo han sido abordados de una manera formal desde mediados del siglo pasado; entre los trabajos realizados se destacan como los dos principales enfoques, el propuesto por Hamblin [Ham70] y el propuesto por MacKenzie [Mac79]. Más adelante se describen algunas características importantes de la primera de estas propuestas.

---

<sup>1</sup> Lo que en términos legales se conoce como reconvencción: cuando el demandado inicia acciones contra el demandante.

La idea que la argumentación está relacionada con el diálogo, por lo tanto con las obligaciones y los compromisos en el diálogo, es tan vieja como la lógica misma. Tradicionalmente, el estudio de los argumentos ha pertenecido a la lógica. Indudablemente, entre los primeros libros sobre teoría lógica se encuentran algunos libros clásicos de Aristóteles; en esos volúmenes se presupone la existencia de ciertas formas de diálogo regulado, las que son de utilidad al brindar instrucciones para lograr un comportamiento eficiente en el contexto de un diálogo. De esa época ha quedado registrada la versión original de un juego de dos personas, por un lado Dios, representando a “La Naturaleza”, quien está en posesión de todas las sentencias verdaderas y es asumido como uno de los oponentes. Frente a Él hay un individuo humano - o quizás un individuo como representante de la Humanidad - dedicado al juego de la paciencia: comenzando con algunas pocas sentencias que fueron obtenidas antes desde Dios, o arrebatadas a Él y mediante las reglas de la lógica, el humano va ganando más y más conocimiento. Con el paso del tiempo fueron apareciendo otros ejemplos sobre la íntima conexión entre lógica, diálogo, argumento y compromisos; se sabe de algunos tratados medievales sobre obligaciones entre el vasallo (o servidor) y el señor feudal, otro ejemplo lo constituyen las disputas descritas en *obligationes*: reglas establecidas durante la edad media para resolver conflictos y que fueron empleadas durante los siglos XIII y XIV. Según los registros que documentan esa época, se deduce que probablemente las *obligationes* tengan su origen en los trabajos de Aristóteles mencionados antes.

Más cerca de los tiempos actuales, se puede mencionar la *lógica dialógica* de Lorenzen [Lor58]. Desde sus comienzos, la lógica dialógica fue concebida con reglas destinadas a ordenar las posibilidades de ataque y defensa en base a las propiedades lógicas que poseen las sentencias que son atacadas o que se defienden. En este ámbito controlado, las confrontaciones que ocurren como parte del juego son llamadas *diálogos*.

Aproximadamente en la misma época que aparecieron los trabajos de Lorenzen, se presentaron los *juegos dialécticos* [Ham70], que resultaron ser especialmente importantes para el estudio de las falacias (aquellos argumentos que parecen válidos, pero no lo son), tales como la clásica *suponiendo el punto inicial* (en inglés “Begging the Question”). Esta falacia fue examinada por Hamblin mediante juegos de diálogo formales. Desde el punto de vista de la argumentación, lo que es más interesante de los sistemas o juegos dialécticos es quizás, no tanto que ellos produzcan nuevas “lógicas”, sino más bien que mantengan el foco del diálogo en el conflicto de opiniones que fueron asumidas en la situación inicial.

Además, ya que es de interés para los participantes, es importante llegar a la resolución de los conflictos mediante palabras.

Estos procedimientos formales, son juegos entre oradores, quienes deben presentar diversas sentencias de acuerdo a reglas estrictas que han ido evolucionando en forma de distintos procedimientos, tal es así que desde los años 1990 han ganado la atención de científicos de computación de diversos sub-campos. La metodología de funcionamiento empleada resultó de particular interés para el estudio de los diálogos entre agentes de software autónomos [MP09]. Dentro de las Ciencias de la Computación e Inteligencia Artificial, ellos también han sido aplicados para modelar razonamientos humanos complejos como los que se presentan en dominios legales y también son de utilidad para la especificación de sistemas de software complejos.

En la primer mitad del siglo XX la denominada “lógica formal” se había volcado mayoritariamente, casi exclusivamente, hacia aspectos específicos de la teoría deductiva y la semántica. Por esta razón, para algunos especialistas, la lógica formal llegó a transformarse en un “aburrido juego solitario”. El grado de rigor mostrado por la formulación del tipo de juegos de diálogo de Lorenzen convirtió a estos juegos en objetos adecuados para el estudio matemático: ellos podían ser analizados como juegos en el sentido de la teoría de juegos matemáticos. Si bien estos sistemas de diálogo hacen un vínculo interesante entre los aspectos lógico (estático) y dialógico (dinámico) de la argumentación, ellos tienen una importante limitación. La lógica subyacente es deductiva, así que los participantes no pueden replicar un argumento con un contraargumento, debido a que tal movida presupone una lógica no monotónica o rebatible. A pesar que en los debates reales es muy común atacar los argumentos del oponente con un contraargumento.

Pero, como se dijo antes, la tendencia encabezada por Lorenzen no fue la única de aquella época, otra corriente fue iniciada por el filósofo e investigador australiano Charles Hamblin quien parece haber sido la primer persona en definir una medida formal de admisible (en inglés *plausibility*) como distinto de probabilidad; su prominente planteo se presentó en una publicación de 1959 [Ham59]. Este formalismo alternativo para la incertidumbre ha jugado un rol muy importante en la toma de decisiones económicas, también en Inteligencia Artificial, (particularmente en el diseño de sistemas basados en conocimiento), teniendo en cuenta que los axiomas de probabilidad estándar de Kolmogorov no siempre logran adecuarse a todas las posibles formas de incertidumbre y a su manipulación [McB11].

En [Lou98, Pra01, Vre95] se plantean algunos juegos que modelan el razonamiento rebatible como una disputa entre proponente y oponente de una proposición. El proponente comienza con un argumento para la proposición inicial, después de lo cual, cada uno debe atacar algún argumento previo del otro participante con un contraargumento de fuerza suficiente. Los ataques pueden estar dirigidos directamente a la conclusión, lo que se conoce como *refutar* (en inglés rebutting), o bien estar dirigidos a las premisas que soportan la conclusión, conocido como *socavar* (en inglés undercutting) [Pol95]. La proposición inicial resulta probada si el proponente tiene una estrategia ganadora, es decir, si él puede hacer que el oponente se quede sin movidas para hacer cualquier ataque. Claramente, este entorno dialéctico se corresponde bien con las aplicaciones de sistemas de diálogo mencionadas anteriormente.

Los juegos de diálogo difieren de los juegos de la teoría económica en que el balance del ganador o del perdedor de un juego no es considerado; tal es así que las ideas de ganador y perdedor no siempre se aplican en los juegos de diálogo. Ellos también difieren de los juegos abstractos<sup>2</sup> usados como semánticas en computación interactiva [A<sup>+</sup>97], debido a que estos últimos juegos no comparten la rica estructura de reglas de los juegos de diálogo, ni pretenden tener una interpretación semántica que involucre la coordinación de acciones entre un grupo de agentes.

Vale la pena señalar aquí que más de una noción de compromiso está presente en la literatura sobre juegos de diálogos. Por ejemplo Hamblin trata a los compromisos con un sentido puramente dialógico: “Un orador quien está obligado a mantener consistencia necesita contar con un almacén de sentencias que representa sus compromisos previos, y requiere que cada nueva sentencia que él hace, pueda ser agregada al almacén sin inconsistencias. El almacén representa una persona de creencias; no necesariamente se corresponde con las creencias reales.” En contraste Walton y Krabbe tratan a los compromisos como obligaciones para ejecutar (o bien iniciar o incurrir en) un curso de acción, lo que ellos llaman compromiso de acción. Estas acciones pueden ser frases en un diálogo, como cuando un orador afirma una proposición y es forzado a defenderla del ataque de los otros. De esa manera Walton y Krabber consideran a los compromisos proposicionales como un caso especial de compromiso de acción. Tanto en el enfoque de Hamblin como en el Walton & Krabbe, los compromisos dialécticos con las proposiciones pueden no necesariamente

---

<sup>2</sup>Bajo este enfoque, un juego puede ser visto como una especificación de posibles interacciones entre un Sistema y su Entorno.



representar una creencia verdadera del participante. Por su parte la semántica social de Singh, requiere que los participantes de una interacción expresen públicamente sus creencias e intenciones, estas expresiones son llamadas compromisos sociales. Los compromisos sociales incluyen dos tipos de expresiones, las que indican la creencia en determinadas proposiciones y las referidas a la intención de realizar futuras acciones.

Nuestra principal motivación es el uso de los juegos de diálogos como la base para protocolos de interacción entre agentes autónomos. Debido a que tales agentes típicamente comenzarán estas interacciones con la finalidad de lograr algún objetivo más amplio, y no solamente por disfrutar de la interrelación en sí misma, creemos que es razonable definir compromisos en términos de acciones futuras o proposiciones externas al diálogo. Por ejemplo, en un diálogo por una transacción comercial, la frase de un oferente puede expresar la complacencia del orador para emprender una transacción subsecuente sobre los términos contenidos en la oferta. Por esta razón podemos ver a los compromisos como una correspondencia semántica entre las locuciones y ciertas frases que expresan acciones o creencias externas al diálogo.

## 5.1. Los sistemas de diálogo

Esta sección describe un marco formal para regular los diálogos de persuasión<sup>3</sup>, los cuales se podrían incorporar a los agentes para permitirles solucionar conflictos de opinión durante la comunicación.

A continuación se plantea una situación que podría darse en el contexto del escenario de Holistan, presentado en la sección 2.3. Esta circunstancia particular resulta conveniente para evaluar un intercambio de actos del habla y los conflictos que pueden surgir durante el mismo. La situación consiste en un diálogo entre dos soldados de distintas fuerzas, quienes deben operar en el campo minado (CM). Ver la Figura 2.4.

El soldado que realiza la primer afirmación es Juan, un miembro de la fuerza militar. Su interlocutor es José, que pertenece a la fuerza humanitaria, quien debe realizar una misión solidaria en una región cercana al campo minado. Juan ha estado trabajando para eliminar todo tipo de riesgo del lugar y desde su perspectiva el objetivo ha sido alcanzado. Por otro lado, para José su vida misma estará en riesgo si el trabajo de Juan

---

<sup>3</sup>Esta sección se basa en [Pra09].

no ha sido bien hecho. El siguiente diálogo presenta algunas de las características típicas de persuasión y será usado para clarificar algunas ideas y para motivar la lectura.

Juan: El CM está desminado. (*hace un reclamo*)

José: ¿Por qué el CM está desminado?.(*consulta sobre fundamentos del reclamo*)

Juan: Porque trabajaron los dragaminas.(*brinda fundamentos del reclamo*)

José: Eso es verdad,(*acepta el reclamo*) pero eso no hace al CM desminado. (*establece una reconvencción*)

Juan: ¿Por qué no lo hace desminado?.(*consulta sobre fundamentos del reclamo*)

José: Porque no buscaron minas marinas en el río. (*establece un contrargumento dando sus fundamentos para la reconvencción*)

Juan: Si, es cierto, no buscamos minas marinas; (*acepta el reclamo*) pero no hace falta porque el Alto Mando dice que aquí no existen minas marinas.(*socava el contrargumento*)

José: De todas maneras creo que el CM no está desminado, porque el informante dijo que hay más minas. (*contrargumento alternativo*)

Juan: De acuerdo, estaba equivocado en que el CM está desminado.

Al analizar el intercambio de frases entre los soldados se pueden rescatar algunas consideraciones importantes. Los participantes en un diálogo de persuasión no solamente intercambian argumentos y contraargumentos, también expresan varias actitudes proposicionales, como reclamos, desafíos, concesiones y retractaciones de las proposiciones presentadas.

Respecto de los argumentos y los contraargumentos, se observan las siguientes características.

-A veces un argumento es atacado mediante un argumento que soporta lo opuesto a una conclusión (como en los dos contraargumentos de José) y otras veces, diciendo que en las actuales circunstancias las premisas del argumento no soportan su conclusión (como en el contraargumento de Juan). Esta es la bien conocida distinción de Pollock entre refutar y socavar contraargumentos [Pol95].

-A veces los contraargumentos son establecidos de manera inmediata o directa (como el socavamiento de Juan y la última movida de José) y a veces son introducidos haciendo una reconvencción (como en la segunda y tercer movida de José).

-Los argumentos en lenguaje natural dejan implícita una generalización de sentido común “El CM está desminado porque trabajaron los dragaminas”.

Como estructura de diálogo, el ejemplo ilustra las siguientes características.

-Los participantes del diálogo pueden retomar decisiones anteriores y hacer réplicas alternativas: en su última movida, José establece un contraargumento alternativo después de ver que Juan tenía un fuerte contraataque en contra de su primer contraargumento. Notar que él también podría haber presentado el contraargumento alternativo inmediatamente después del primero, para dejar a Juan con dos ataques para considerar.

-Los participantes pueden posponer sus réplicas, incluso indefinidamente: con su segundo argumento de porque el CM no está desminado, José pospone su réplica al contraataque de Juan sobre el primer argumento de ese reclamo; si Juan no puede atacar con éxito el segundo argumento de José, tal respuesta resultaría superflua.

Un poco más adelante, el diálogo entre Juan y José será modelizado siguiendo un método formal (en inglés *framework*) especialmente creado para resolver aquellas conversaciones iniciadas entre agentes con un propósito específico [Pra06].

Por esa razón es conveniente presentar ahora un resumen que describa, de manera abreviada, en que consiste el sistema de diálogo escogido: tiene un *objetivo* y al menos dos *participantes*, quienes pueden desempeñar diferentes *roles*. Los sistemas de diálogo tienen dos lenguajes, un *lenguaje de comunicación* usado para armar la frase, que incluye una parte especificada en el *lenguaje del tópico*. Ocasionalmente, el diálogo se desarrolla en un *contexto* de conocimiento fijo e indiscutible, tales como las leyes relevantes en una disputa legal. El corazón del sistema de diálogo está formado por un *protocolo*, el cual especifica las “movidas” (o jugadas) legales en cada instante del diálogo, las *reglas de efecto* que especifican los efectos de las frases sobre los compromisos de los participantes y las *reglas de salidas*, que definen el resultado del diálogo. También dos clases de reglas de protocolo que habitualmente son definidas de manera separada: los *turnos* y las *reglas de terminación*.

A continuación se hace una descripción más detallada de cada elemento que integra un sistema de diálogo:

- Un lenguaje específico para el tema o tópico sobre el cual se desarrolla la comunicación, llamado  $\mathcal{L}_t$ . Este lenguaje debe ser cerrado bajo la negación clásica.
- Un lenguaje de comunicación  $\mathcal{L}_c$ , consistente de un conjunto de actos del habla con un contenido. El conjunto de diálogos posibles, denotado por  $M^{\leq\infty}$ , es el conjunto de todas las secuencias que es posible formar a partir de  $\mathcal{L}_c$ . El conjunto de diálogos finitos, denotado por  $M^{<\infty}$ , es el conjunto de todas las secuencias finitas generadas

a partir de  $\mathcal{L}_c$ . Para cualquier diálogo  $d = m_1, \dots, m_n, \dots$ , la secuencia  $m_1, \dots, m_i$  es denotada por  $d_i$ .

- Un *propósito* del diálogo. Distintos tipos de diálogo persiguen diferentes objetivos.
- Un conjunto  $\mathcal{A}$  de participantes (o “actores”) y un conjunto  $\mathcal{R}$  de *roles*, definidos sobre subconjuntos disjuntos de  $\mathcal{A}$ . Un participante  $a$  puede o no tener una *base de creencias*, posiblemente inconsistente,  $\Sigma_a \in Pot(\mathcal{L}_t)$ , la cual puede o no cambiar durante el diálogo. Además, cada participante acumula un conjunto de *compromisos* durante el diálogo,  $C_a \subseteq \mathcal{L}_t$  usualmente va cambiando y puede ser vacío.
- Un *contexto*  $\mathcal{K} \subseteq \mathcal{L}_t$ , que contiene el conocimiento que se da por sentado y que debe ser respetado durante el diálogo. El contexto se asume que es consistente y permanece sin cambiar durante el diálogo.
- Una *lógica*  $\mathcal{L}$  para  $\mathcal{L}_t$ , la cual puede o no ser nomotónica y la cual puede o no estar basada en argumentos.
- Un conjunto de reglas de efectos  $\mathcal{C}$  para  $\mathcal{L}_c$ , que especifica para cada frase  $\Phi \in \mathcal{L}_c$  sus efectos sobre los compromisos de los participantes. Estas reglas son especificadas como funciones

$$C_a : M^{<\infty} \rightarrow Pot(\mathcal{L}_t)$$

Los cambios en los compromisos están completamente determinados por la última movida en el diálogo y por aquellos compromisos adquiridos previamente a realizar dicha movida:

$$\text{Si } d = d' \text{ entonces } C_a(d, m) = C_a(d', m)$$

- Un protocolo  $Pr$  para  $\mathcal{L}_c$ , especifica las movidas permitidas (o “legales”) en cada etapa del diálogo. Formalmente un protocolo en  $\mathcal{L}_c$  es una función  $Pr$  con dominio en el contexto más un subconjunto  $D$  de  $M^{<\infty}$  tomando subconjuntos de  $\mathcal{L}_c$  como valores. Esto es:

$$Pr : Pot(\mathcal{L}_t) \times D \rightarrow Pot(\mathcal{L}_c)$$

tal que  $D \subseteq M^{<\infty}$ . Los elementos de  $D$  son llamados *diálogos legales finitos*. Los elementos de  $Pr(d)$  son llamadas las movidas permitidas después de  $d$ . Si  $d$  es un diálogo legal y  $Pr(d) = \emptyset$ , entonces se dice que  $d$  es un diálogo terminado.  $Pr$  debe satisfacer la siguiente condición: para todo diálogo finito  $d$  y movidas  $m$ ,  $d \in D$  y  $m \in Pr(d)$  si y sólo si  $d, m \in D$ .

Es útil (aunque no estrictamente necesario) distinguir explícitamente los elementos de un protocolo que regulan los cambios de turnos y la terminación.

- ★ Una función de *cambio de turno*, es una función  

$$T : D \times Pot(\mathcal{L}_t) \longrightarrow Pot(\mathcal{A}).$$

Un turno de un diálogo es definido como la máxima secuencia de movidas del diálogo en la cual el mismo actor puede seguir moviendo. Notar que  $T$  puede designar a más de un actor como siguiente para mover.

- ★ *Terminación* está definida como el caso donde ningún movimiento es legal. De acuerdo a eso, una definición explícita de terminación debiera especificar las condiciones bajo las cuales  $Pr$  retorna el conjunto vacío.
- *Reglas de resultado*  $O^K$ , define las salidas de un diálogo dado un contexto. Por ejemplo, en negociación la salida es una asignación de recursos, en deliberación es una decisión de un curso de acción, y en un diálogo de persuasión es un ganador y un perdedor del diálogo de persuasión. La salida debe ser definida para diálogos terminados y puede ser definida para unos no terminados; en este caso las reglas de salida capturan la idea de salida en “cualquier momento”.

Notar que no se asume ninguna relación entre los compromisos de los participantes y su base de conocimiento. Los compromisos son un punto de vista públicamente declarado del agente sobre una proposición, lo cual no necesita coincidir con las creencias internas del agente.

Lo que sigue se refiere a las particularidades del entorno para persuasión.

### 5.1.1. Sistema de persuasión

Según Walton [Wal84] los diálogos de persuasión son diálogos donde el objetivo es resolver un conflicto entre los puntos de vista de al menos dos participantes. Un *punto de vista* en relación a una proposición puede ser positivo(a favor), negativo (en contra) o dudoso. Los participantes procuran persuadir a el(los) otro(s) participante(s) para que acepte(n) su punto de vista.

Un sistema de diálogo para persuasión requiere las siguientes instanciaciones.

El *propósito del diálogo* es resolver un conflicto de opinión acerca de una o más proposiciones, llamadas *tópicos*,  $\mathcal{T} \subseteq \mathcal{L}_t$ . Este propósito de diálogo da origen a roles de participantes y reglas de resultado según se detalla a continuación.

Los participantes pueden tener los siguientes *roles*. Comenzando con  $prop(t) \subseteq \mathcal{A}$ , los *proponentes* del tópico  $t \in \mathcal{T}$ , es el conjunto (no vacío) de todos los participantes con un punto de vista positivo hacia  $t$ . De la misma manera  $opon(t) \subseteq \mathcal{A}$ , es el conjunto (no vacío) de todos los participantes con un punto de vista indeciso hacia el tópico  $t$ . Si fuese negativo, correspondería a una disputa. Para cualquier  $t$ , los conjuntos  $prop(t)$  y  $opon(t)$  son disjuntos y juntos no siempre agotan  $\mathcal{A}$ .

Las *reglas de resultado* para diálogos de persuasión definidas para un diálogo  $d$ , contexto  $K$  y tópico  $t$  son los *ganadores* y los *perdedores* de  $d$  con respecto al tópico  $t$ . Más precisamente, el resultado  $O$  consiste de dos funciones parciales  $w$  y  $l$ :

$$w : D \times Pot(\mathcal{L}_t) \times \mathcal{L}_t \rightarrow Pot(\mathcal{A})$$

$$l : D \times Pot(\mathcal{L}_t) \times \mathcal{L}_t \rightarrow Pot(\mathcal{A})$$

tal que ellas están definidas al menos para todos los diálogos terminados pero solamente para aquellos  $t$  que son un tópico de  $d$ . Estas funciones se escribirán como  $w_t(d)$  y  $l_t(d)$ . Ellas además satisfacen las siguientes condiciones:

$$w_t(d) \cap l_t(d) = \emptyset$$

$$w_t(d) = \emptyset \text{ si, y sólo si, } l_t(d) = \emptyset$$

Si  $|\mathcal{A}| = 2$ , entonces  $w_t(d)$  y  $l_t(d)$  a lo más tienen un elemento.

Finalmente, en persuasión una de las partes del diálogo se va a dar por vencido.

$$w_t(d) \subseteq prop(t) \text{ o } w_t(d) \subseteq opon(t); \text{ y}$$

Si  $a \in w_t(d)$  entonces

$$\text{Si } a \in prop(t) \text{ entonces } t \in C_a(d)$$

$$\text{Si } a \in opon(t) \text{ entonces } t \notin C_a(d)$$

Estas condiciones aseguran que un ganador no cambiará su punto de vista. Además, de acuerdo a estas condiciones, es posible afirmar que un diálogo de persuasión entablado entre dos personas es un juego de suma cero, ya que mantiene el equilibrio entre la ganancia de un participante y la pérdida del otro. Esta característica establece la diferencia entre los diálogos de persuasión y otros tipos de diálogo como búsqueda de información, deliberación e investigación (inquiry).

El *lenguaje de comunicación* está formado por algunos actos del habla clásicos, los que pueden aparecer como una locución espontánea o como alguna de las réplicas típicas.

La Tabla 5.1 presenta algunos ejemplos de elementos del lenguaje de comunicación. Se puede observar que cada *locución* iniciada por el orador (cualquiera de las que aparecen en la primer columna de la tabla), da lugar a distintas respuestas posibles por parte del oyente. Una *réplica* consiste en alguno de los actos del habla que aparecen en la segunda columna de la fila correspondiente a la locución realizada. Se puede observar que un par de locuciones no requieren respuesta.

Locuciones	Réplicas
<i>reclamo</i> $\varphi$	<i>¿por qué</i> $\varphi?$ , <i>reclamo</i> $\bar{\varphi}$ , <i>concedo</i> $\varphi$
<i>¿por qué</i> $\varphi?$	$\varphi$ <i>dado</i> $S$ (alternativamente: <i>reclamo</i> $S$ ), <i>retiro</i> $\varphi$
<i>concedo</i> $\varphi$	
<i>retiro</i> $\varphi$	
$\varphi$ <i>dado</i> $S$	<i>¿por qué</i> $\psi?$ ( $\psi \in S$ ), <i>concedo</i> $\psi$ ( $\psi \in S$ ), $\varphi'$ <i>dado</i> $S'$
<i>consulta</i> $\varphi$	<i>reclamo</i> $\varphi$ , <i>reclamo</i> $\bar{\varphi}$ , <i>retiro</i> $\varphi$

Tabla 5.1: Locuciones y réplicas típicas

Con este lenguaje de comunicación, el diálogo de Juan y José, sería:

$P_1$ : <i>reclamo</i> desminado	$O_2$ : <i>¿por qué</i> desminado?
$P_3$ : desminado <i>dado</i> barreminas	$O_4$ : <i>concedo</i> barreminas
$P_6$ : <i>¿por qué</i> $\neg$ desminado?	$O_5$ : <i>reclamo</i> $\neg$ desminado
$P_8$ : <i>concedo</i> minas marinas	$O_7$ : $\neg$ desminado <i>dado</i> minas marinas
$P_9$ : <i>¿y qué?</i> <i>dado</i> Alto Mando	$O_{10}$ : $\neg$ desminado <i>dado</i> informante
$P_{11}$ : <i>retiro</i> desminado	

Como se observa en el ejemplo, durante un diálogo de persuasión es posible que un mismo jugador realice sus movidas sin que intervenga el otro jugador, esto califica al sistema como de *múltiple respuesta*.

Como reglas de compromisos, las siguientes son generalmente aceptadas en la literatura. Si *pl* designa al orador, los cambios producidos por los actos del habla sobre los compromisos serían:

- Si  $pl(m) = \text{reclamo } \varphi$  entonces  $C_{pl}(d, m) = C_{pl}(d) \cup \varphi$
- Si  $pl(m) = \text{¿por qué } \varphi?$  entonces  $C_{pl}(d, m) = C_{pl}(d)$
- Si  $pl(m) = \text{concedo } \varphi$  entonces  $C_{pl}(d, m) = C_{pl}(d) \cup \varphi$
- Si  $pl(m) = \text{retiro } \varphi$  entonces  $C_{pl}(d, m) = C_{pl}(d) - \varphi$
- Si  $pl(m) = \varphi \text{ dado } S$  entonces  $C_{pl}(d, m) \supseteq C_{pl}(d) \cup \text{prem}(Arg)$

La regla para *dado* usa  $\supseteq$  porque tal movida puede comprometer a más que solamente las premisas del argumento movido. Por ejemplo en [Pra05] la movida también compromete a  $\varphi$ , debido a que los argumentos también pueden ser movidos como contrargumentos en lugar de réplicas a desafíos de reclamos.

Continuación del ejemplo: De acuerdo a estas reglas, los conjuntos de compromisos de Juan y José al final del diálogo de ejemplo son:

- $C_P(d_{11}) \supseteq \{\text{dragaminas, minas marinas, } \neg\text{minas marinas aquí}\}$
- $C_O(d_{11}) \supseteq \{\neg\text{desminado, minas marinas, informante}\}$

### 5.1.2. Una instancia de sistema de persuasión

En esta sección se instancia el entorno de diálogos de persuasión para lograr resolver la situación planteada como ejemplo. En el marco de Prakken los participantes tienen los roles de proponente y oponente, y sus creencias son irrelevantes para los protocolos, razón por la cual, las creencias tienen una semántica pública. Los diálogos no tienen contexto. El sistema se abstrae del lenguaje de comunicación excepto que requiere la existencia de una estructura de respuesta explícita. También se abstrae del lenguaje lógico y la lógica, excepto que la lógica se asume que es basada en argumentos, que responde a semánticas fijas (en inglés *grounded*) y que los argumentos son árboles de inferencias rebatibles o *árbol de diálogos*.

El principal interés de este marco es enfocarse en los diálogos, mientras se deja suficiente libertad como para realizar réplicas alternativas y para posponer las réplicas. Esta libertad se consigue mediante dos características principales del sistema. Primero,  $\mathcal{L}_c$  tiene una estructura de respuesta explícita, donde cada movida o bien ataca o admite a la



movida objetivo. Un  $\mathcal{L}_c$  de ejemplo con este formato se muestra en la tabla 5.2. Segundo, el ganador se define para cada uno de los diálogos, ya sea que esté terminado o no; el resultado es determinado por el *estado dialógico* de la raíz del árbol. Para resolver el diálogo, se hace uso de las características del árbol de diálogos generado por la estructura de respuesta del  $\mathcal{L}_c$ , ya que dicho árbol mantiene el estado de las distintas movidas. Una movida está *adentro* si es admitida o si todas las réplicas que la atacan están *afuera*. (Esto implica que una movida sin réplicas está *adentro*). Y una movida está *afuera* si tiene una réplica que esté *adentro*. Entonces un diálogo está siendo ganado (actualmente) por el proponente si su movida inicial está *adentro*, mientras que está (actualmente) siendo ganada por el oponente en otro caso.

Acto	Ataques	Admisiones
reclamo $\varphi$	¿por qué $\varphi$ ?	concedo $\varphi$
$\varphi$ dado $S$	¿por qué $\psi$ ? ( $\psi \in S$ ) $\varphi'$ dado $S'$ ( $\varphi'$ dado $S'$ derrota $\varphi$ dado $S$ ) se supone $\varphi'$ ( $\Rightarrow \varphi'$ derrota $\varphi$ dado $S$ )	concedo $\psi$ ( $\psi \in S$ ) concedo $\varphi$
¿por qué $\varphi$ ?	$\varphi$ dado $S$ se supone $\varphi$	retiro $\varphi$
se supone $\varphi$	$\varphi'$ dado $S$ se supone $\varphi'$	concedo $\varphi$
concedo $\varphi$		
retiro $\varphi$		

Tabla 5.2: Un  $L_c$  ejemplo, en el sistema de Prakken

En combinación estas dos características del sistema sustentan la idea de *relevancia* que garantiza que el foco se mantenga en el diálogo, mientras deja cierto grado de libertad: una movida de ataque es relevante cuando deja a su objetivo *afuera* haciendo que el orador sea el actual ganador. Por su parte, una movida de admisión es relevante si es una réplica a una movida relevante. La terminación es definida como la situación en la que un jugador debe mover pero no tiene movidas legales para hacer.

El lenguaje de comunicación propuesto es el mostrado en la Tabla 5.2 y está restringido por el requerimiento de que cada movida debe ser *relevante*. Esto hace que el protocolo sea de *respuesta inmediata*, lo cual implica que cada turno consista de cero o más admisiones

seguidas por un ataque. Dentro de estos límites la posposición de respuesta está permitida, a veces incluso indefinidamente.

Seguidamente se discute un ejemplo, donde el  $\mathcal{L}_t$  se instancia con el lenguaje de programación lógica extendida. (El siguiente ejemplo debiera hablar por sí mismo y se puede consultar el Apéndice A para una explicación sencilla. Las reglas tienen etiquetas, necesarias para expresar prioridades entre las reglas). Se considera el siguiente conjunto de reglas rebatibles:

$$D = \{r_0 := a, r_1 : a \Rightarrow b, r_2 : \sim b \Rightarrow c, r_3 := \neg a, \\ r_4 := r_0 < r_3, r_5 := r_3 < r_0, r_6 := r_5 < r_4\}$$

Entonces las siguientes son movidas legales en el protocolo llamado de movida *relevante* (el objetivo de cada movida es indicado entre corchetes):

$$\begin{array}{l|l} P_1[-]: & \text{reclamo } c \\ P_3[O_2]: & c \text{ dado } \sim b(r_2) \\ P_5[O_4]: & \text{¿por qué } a? \\ P_7[O_6]: & \text{se supone } \neg a(r_3) \\ P_9[O_8]: & \text{se supone } r_5 < r_4(r_6) \\ O_2[P_1]: & \text{¿por qué } c? \\ O_4[P_3]: & b \text{ dado } a(r_1) \\ O_6[P_5]: & \text{se supone } a(r_0) \\ O_8[P_7]: & \text{se supone } r_3 < r_0(r_5) \\ O_{10}[P_7]: & \text{acepto } \neg a \\ O_{11}[P_1]: & \text{acepto } c \end{array}$$

Notar que los argumentos pueden ser construidos de manera escalonada en varias movidas, para hacer eso más visible, en cada movida se incluye entre paréntesis la regla aplicada. Se puede ver que en  $D$  hay varias reglas que no tienen premisas, estos son “casi hechos”, eso quiere decir que en esos casos es imposible atacar el antecedente. Por ejemplo, la única posibilidad de derrotar las reglas  $r_0$  y  $r_3$  es refutar su conclusión. Las reglas de prioridad se usan en las movidas  $O_8$  y  $P_9$ ; en la primera de ellas, el oponente presenta una regla con un argumento prioritario que le permite derrotar a  $P_7$  y a continuación el proponente ataca con la regla que define el conflicto entre las reglas de prioridad. Las admisiones como no son ataques no derrotan a las movidas destino, por eso las últimas movidas del oponente no cambian el estado de las movidas a las que van dirigidas.

Finalmente es posible modelizar mediante este protocolo el ejemplo considerado. La figura 5.1 muestra el árbol de diálogo, donde las movidas con cajas sólidas están *adentro* y las movidas con cajas de puntos están *afuera*.

De acuerdo a lo que se muestra en la Figura 5.1, un acto informar con el contenido “El CM está desminado” no sería realmente exitoso cuando el receptor tiene incorporado

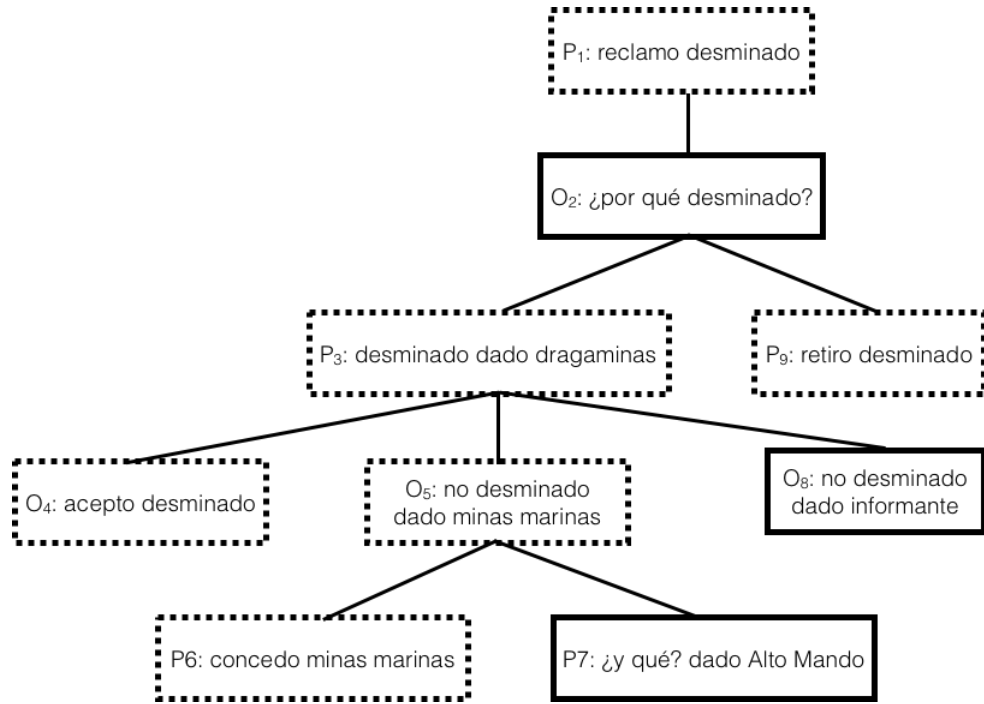


Figura 5.1: El ejemplo en un árbol de Prakken .

un sistema de diálogo de persuasión. Los agentes racionales debieran incorporar sistemas de diálogos como el presentado en este capítulo, si desean colaborar entre si para lograr el objetivo del sistema multi-agente. Para mostrar la flexibilidad del formalismo se puede considerar el uso del lenguaje DeLP [Gar00] como  $L_t$ . Esta idea ha sido presentada en el trabajo [AES17].



# Capítulo 6

## Conclusiones

Abordar la especificación de los actos del habla en los sistemas multi-agente, es recorrer un camino que comienza en el estudio de la comunicación entre humanos, hasta discutir como un agente computacional puede influir sobre otros agentes mediante acciones de comunicación. En el caso de los seres humanos se ha llegado a reconocer la necesidad de establecer un conjunto de condiciones bajo las cuales ciertas frases logran transformar nuestro mundo. Pero sin duda, es la capacidad de adaptación y flexibilidad humana lo que nos permite entendernos. Por esa razón resulta difícil enumerar los requisitos que debe tener una pieza de código de manera que pueda interactuar con otras piezas similares. En las hojas que resumen nuestro recorrido no hay un aporte concreto sobre esos requisitos, sino que se consideran algunos enfoques importantes a modo de visión general del tema.

La intención planteada al inicio del trabajo de estudio e investigación sobre el tema, fue llegar a una perspectiva integradora sobre la comunicación de los agentes. Desde un comienzo se constató la existencia de una gran cantidad de publicaciones que abordan el tema desde el aporte proveniente de la comunicación entre los seres humanos y se consideró importante aprovechar esta perspectiva, especialmente usando la teoría de los actos del habla. Efectivamente se verificó que la consideración adoptada es correcta, ya que la teoría de los actos del habla es tenida en cuenta en muchos aspectos, desde la creación de lenguajes hasta la caracterización del entorno de los sistemas multi-agente.

La comunicación es un proceso social complejo subyacente a todos los sistemas multi-agente ya sean estos computacionales puros, que involucren solamente humanos o bien híbridos. Como todo proceso en un sistema multi-agente realista, se debe tener en cuenta cómo el mismo se visualiza desde el punto de vista de los agentes individuales, del sistema

multi-agente en su conjunto y del contexto en que dicho proceso se lleva a cabo. En este sentido, las tres perspectivas analizadas en los capítulos previos han hecho foco y/o resaltado cada una de estas dimensiones de la comunicación.

En la perspectiva mental, se hizo énfasis en el concepto de actos del habla y en cómo estos actos permiten formalizar los cambios que se producen en los estados mentales de los agentes como las creencias, deseos e intenciones. La formalización se consigue por medio de *operadores* que se combinan dentro de planes; usados siempre que se verifiquen las condiciones requeridas por los operadores y cuando los efectos producidos sirvan para aproximarse al objetivo del plan. Las condiciones se establecen como creencias específicas en el estado mental del orador y los efectos se registran como nuevos estados mentales en el oyente. La consideración de esta perspectiva se resume en una parte de un primer artículo publicado [AES08]. Allí se plantean algunos ejemplos de cómo la argumentación puede asistir en la justificación de la elección de los actos del habla individuales, y de cómo la teoría de los actos del habla hace un aporte a la semántica y actualización de las creencias, deseos e intenciones de los agente argumentativos individuales participantes. Las ideas presentadas en ese trabajo fueron consideradas sobre un entorno específico: un sistema donde los agentes deben colaborar para informar accesos no autorizados dentro de una LAN. Usando DeLP los agentes aplican razonamiento rebatible y reglas de sentido común para determinar si hay presencia de intrusos en la red. El enfoque mental, es un área de investigación importante ya que es empleada en uno de los lenguajes de comunicación de agentes más conocidos, ACL de FIPA. Sin embargo, en este lenguaje la especificación de los operadores usa lógica multimodal, lo que puede ser una limitación al momento de ser implementada computacionalmente. La programación en lógica rebatible es una alternativa que ha probado superar esta última limitación [CC00]. Desde el punto de vista mental, la comunicación no permite demasiada flexibilidad y los mensajes intercambiados son considerados sin emplear racionalidad. Es decir, se debe seguir una receta fija, provista en la forma de un protocolo formado por una secuencia bien definida de pasos, en donde a lo sumo hay que elegir por una de dos alternativas posibles para continuar. El estudio de este tipo de sistema también ha sido motivo de estudio por parte de nuestro grupo de investigación dentro de la gran variedad de alternativas de comunicación de agentes [GLAE12].

La perspectiva social, por su parte, toma un enfoque normativo que pone el énfasis en el sistema multi-agente en su conjunto, y en que condiciones, los agentes deberían

observar sus actos comunicativos, para mantener una comunicación efectiva. El concepto fundamental en este caso es el *compromiso* que permite hacer un seguimiento global y compartido en todo el sistema multiagente de la información social necesaria para una coordinación efectiva de los agentes. Un aspecto muy interesante es la administración de los compromisos generados durante los diálogos, porque es necesario que cada agente mantenga consistencia lógica entre sus afirmaciones. Una forma de lograrlo es obligando a que cada agente respete un conjunto de postulados mediante los cuales se agregan y eliminan las proposiciones del almacén de compromisos. Al tomar en consideración una visión amplia de los sistemas multi-agente, se abordan aspectos como las latencias en la propagación de los mensajes, las diferencias en la percepción de los eventos y las consecuencias que esto trae desde el punto de vista de los sistemas distribuidos. Por eso se estudiaron las condiciones ideales bajo las cuales se requiere que los compromisos estén *alineados*, lo que evita que un agente entienda que es acreedor de un compromiso cuando el deudor no se reconoce como tal. En la propuesta original del uso de los postulados para administrar los compromisos, se usa lógica proposicional para especificarlos. Una alternativa en la que se aplique lógica rebatible para especificar los postulados, permitirá que los agentes razonen sobre los compromisos de una manera más elaborada.

Por último, la perspectiva argumentativa parte del supuesto más realista que la comunicación se realiza en un contexto donde la información intercambiada puede ser parcial, errónea o incompleta y que los agentes deberán contar necesariamente con un mecanismo que les permita justificar, contradecir y en última instancia llegar a un consenso respecto a que piezas de la información intercambiada puede ser racionalmente considerada como válida. Para lograr eso y con la finalidad de ordenar la interacción durante los diálogos, resulta conveniente que los interlocutores se sometan a las reglas establecidas mediante ciertos *protocolos*. Estas reglas establecen, entre otras cosas, que los mensajes tendrán dos capas: una externa de comunicación asociada a un acto del habla y una interna, específica del contexto, la cual emplea un lenguaje con ciertas cualidades. Estas cualidades deben permitir por ejemplo, que un agente plantee un desafío sobre alguna afirmación de su interlocutor, pueda justificar algún reclamo previo, o presente un contraargumento en el momento que sea su turno para “hablar”. Para lograr estas habilidades se requiere emplear una lógica no monótona. También como parte del contexto de la comunicación, es conveniente establecer que tipo de diálogo se está desarrollando. Los distintos tipos de diálogo presentan características diversas y persiguen objetivos distintos. En el capítulo 5 se consideró el tipo de diálogo de persuasión, pero resulta interesante considerar estudios

para ahondar en las particularidades de otros tipos de diálogos. Ganar experiencia en el trabajo individual sobre cada tipo de diálogo es importante, pero concretamente, detectar las transiciones entre un tipo y otro, tiene un interés particular ya que según Walton & Krabbe, en estos cambios es posible detectar *falacias*. Nuestra intención es continuar investigando en esta forma de identificar falacias empleando argumentación rebatible.

Desde nuestro punto de vista, cualquier enfoque realista a la comunicación para ser efectivo, debe tener en cuenta tanto la perspectiva individual, como la social y la de contexto. En este sentido, consideramos que la perspectiva argumentativa puede ser el mejor camino a partir del cual las restantes dimensiones pueden ser integradas y justificadas.

Es importante resaltar que los enfoques considerados en esta tesis son aquellos fundamentados esencialmente en una modelización lógica del problema. Sin embargo, y en concordancia con otras líneas importantes de la inteligencia artificial, los fenómenos asociados a la comunicación entre agentes también han sido analizados desde la perspectiva de la utilidad de los actos del habla, de la incertidumbre asociada a estos actos o a sus estados de creencia; de cómo las decisiones sobre estos actos pueden ser derivadas desde estas creencias y preferencias. En estos casos, los modelos de comunicación entre los agentes están basados usualmente en aquellos que permiten representar la incertidumbre sobre el estado o las acciones de los agentes individuales. Dichos modelos deben hacer una combinación: por un lado considerar la observación hecha sobre el ambiente y por otro, estimar el resultado de la ejecución de las acciones. Este tipo de problemas muchas veces se conocen como *problemas de contingencia* [RN03]. En esta clase de problemas la transición entre estados no puede determinarse mediante una función sino por medio de una distribución de probabilidad [GD01]. Estos modelos matemáticos se conocen como *procesos de decisión Markov* (MDPs) [Bel57, BD62]. De acuerdo a estos enfoques, entonces, los agentes debieran adoptar *políticas* que en cada evolución del tiempo y para un estado o ambiente determinado, elijan aquella acción que maximise el *retorno esperado*. Uno de los métodos estudiados para aproximar una política MDP óptima fue Value Iteration [RN03]. En MDP el modelo del ambiente que usa cada agente, debiera incluir al agente con el cual se comunica. La incertidumbre proviene de considerar una representación aproximada del ambiente y de que no hay garantía de conseguir efectivamente el efecto esperado de las acciones. Si se considera la situación más realista en la cual el agente no puede determinar precisamente el estado actual, el problema se clasifica como un MDP *parcialmente observable* (POMDP) [Ast65, SS73]. En estos casos las observaciones



del agente se relacionan de manera estocástica con el modelo MDP subyacente. Yendo un paso más allá en esa dirección, están los sistemas que buscan modelar un ambiente dinámico dentro del cual es necesario representar las modificaciones introducidas por los otros agentes, son los MDP *multiagente* (MMDP) [Err03]. Con este enfoque, que se ve como muy promisorio para continuar investigando, se debe considerar que incluida dentro de la realidad, aparecen los compromisos que realizan los otros agentes como resultado del accionar del sistema multi-agente. Los resultados, que pudieran ser conseguidos con esta aproximación, resultarían una base con la cual contrastar los resultados logrados usando razonamiento rebatible.

También se estudió un enfoque diferente para representar el estado del sistema multi-agente. Una aproximación basada en técnicas de chequeo del modelo simbólico [CGP99] para explorar el espacio de estado de un sistema de agentes. Sobre ese modelo se procura establecer un conjunto de comportamientos coherentes para los agentes, considerando que cierta incertidumbre pudiera aparecer; el no determinismo en el espacio de estados surge por problemas de comunicación y la acción de otros agentes. El proceso de exploración puede ser realizado en base a la información rebatible que el agente posee sobre la dinámica del espacio de estados, de esa manera, debe continuar progresando en la búsqueda de elementos de su conjunto de metas; a pesar que, dentro de ese conjunto posiblemente existan inconsistencias [Tan12]. Al igual que en el caso de MDP, parece interesante comparar los resultados obtenidos al usar argumentación con model checking versus argumentación con la alineación de compromisos del enfoque social.



# Apéndice A

## El sistema argumentativo PraTor

Este apéndice reseña brevemente las principales características del sistema argumentativo rebatible (que aquí llamaremos PraTor) que fuera propuesto por Prakken & Sartor [PS97]. PraTor fue inspirado a partir del razonamiento usado en los entornos legales y la versión presentada en la Sección A.2 de este anexo, combina el lenguaje de la lógica default de Reiter con las semánticas fijas planteadas por Dung [Dun95]. El principal aporte que hace PraTor al estudio de la argumentación rebatible tiene que ver con el estudio de la relación entre los ataques por *refutar* y *socavar* [Pol95], también sobre la formalización de la argumentación empleada por el criterio de derrota.

### A.1. Introducción

Procurando que sea compacta y lo más genérica posible, esta sección incluye algunos de los principales conceptos y definiciones abstractos, necesarios para entender el sistema argumentativo PraTor.

Los sistemas de argumentación se construyen alrededor de un lenguaje lógico y la idea de consecuencia lógica; ambos elementos se combinan para definir la noción de argumento. En los sistemas que aquí se están considerando, la consecuencia lógica es del tipo monotónica: las nuevas premisas no pueden invalidar argumentos como argumentos, solamente dan origen a contraargumentos.

Un argumento es una entidad abstracta, cuyo papel se determina únicamente por su relación con otros argumentos [Dun95].

Un marco de argumentación es un par, formado por un conjunto de argumentos y una relación de derrota entre los elementos de ese conjunto.

Se consideran dos tipos de conflictos o disputas entre argumentos. El primero, llamado *refutar*, es cuando los argumentos sostienen conclusiones contradictorias; el otro es cuando un argumento hace una suposición no demostrada y un segundo argumento demuestra lo que se suponía improbable por el primero. En PraTor este último conflicto se conoce como ataque a la *suposición* y es una variante de la idea de socavar en términos de Pollock.

La idea de *derrota* es una relación binaria sobre el conjunto de argumentos. Es importante destacar que esta relación no dice que argumentos pueden ganar una disputa; solamente habla de la fuerza relativa entre un par de argumentos en conflicto.

Dentro de los sistemas de argumentación rebatibles, un elemento importante es la definición de los estados de los argumentos. Los argumentos pueden adoptar tres estados posibles: *justificado*, *defendible* y *anulado*. Los argumentos justificados son aquellos con los cuales se puede ganar una disputa. Por su parte los argumentos anulados son aquellos con los que seguramente se perderá una disputa. Finalmente los defendibles son aquellos argumentos que dejan la disputa sin decidir. El marco de argumentación produce como salida el estado que le corresponde a los distintos argumentos.

Para explicar como se determina el estado de los argumentos se puede usar una forma “declarativa”, que habitualmente emplea definiciones de punto fijo para lo cual es conveniente comenzar aclarando lo que es un argumento aceptable.

**Definición 15** (*Argumento aceptable*) *Un argumento  $A$  es aceptable respecto a un conjunto  $S$  de argumentos si y solo si cada argumento que derrota a  $A$  es derrotado por un argumento de  $S$ .*

Se debe considerar que los argumentos del conjunto  $S$  pueden *reinsertar* al argumento  $A$ , en caso que el estado de  $A$  fuese *derrotado*. Por eso, la noción de aceptabilidad no es suficiente para entender lo que es un argumento justificado. Por ejemplo, si  $S = \{A\}$  y  $A$  es aceptable para  $S$ ,  $A$  se estaría reinsertando. Claramente, no es deseable que un argumento pueda reinsertarse a sí mismo; esa es la razón por la que un operador de punto fijo debe ser usado; aplicando el operador propuesto por Dung [Dun95] a cada conjunto de argumentos se consigue el conjunto de argumentos aceptables para él.

**Definición 16** (*Operador*) Dado  $Args$ , un conjunto de argumentos ordenados por una relación binaria de derrota, y dado  $S \subseteq Args$ . Entonces el operador  $F$  se define como:

$$\blacksquare F(S) = \{A \in Args \mid A \text{ es aceptable con respecto a } S\}$$

Dung probó que el operador  $F$  tiene un último punto fijo. (La idea básica es que si un argumento es aceptable con respecto a  $S$ , también será aceptable respecto a cualquier superconjunto de  $S$ , es decir que  $F$  es monotonóxico). De esta manera la auto reinstalación puede ser evitada, definiendo el conjunto de argumentos justificados como el último punto fijo de  $F$ .

En general tenemos que si ningún argumento es inderrotable, entonces  $F(\emptyset) = \emptyset$ . Estas observaciones permiten la siguiente definición de argumento justificado.

**Definición 17** *Un argumento está justificado si es miembro del último punto fijo de  $F$ .*

Es posible considerar algunas alternativas a la definición 16, son aproximaciones al último punto fijo de  $F$ . Comenzando desde abajo y siempre que los argumentos tengan una cantidad finita de derrotadores, es posible obtener dicho punto fijo por aplicación iterada de  $F$  al conjunto vacío. Lo que sigue es una propuesta para la construcción recursiva de esta semántica fija.

**Proposición 1** *Considere la siguiente secuencia de argumentos.*

$$\blacksquare F^0 = \emptyset$$

$$\blacksquare F^{i+1} = \{A \in Args \mid A \text{ es aceptable con respecto a } F^i\}$$

*Las siguientes observaciones se cumplen:*

- *Todos los argumentos en  $\cup_{i=0}^{\infty}(F^i)$  están justificados.*
- *Si cada argumento es derrotado a lo más por un número finito de argumentos, entonces un argumento está justificado si está en  $\cup_{i=0}^{\infty}(F^i)$ .*

Antes de comenzar a considerar las particularidades de PraTor, es necesario mencionar que existe una clase especial de argumento, aquellos que se derrotan a sí mismo. Sea  $L$  un argumento, tal que  $L$  derrota a  $L$ . No se darán aquí más explicaciones solamente comentar que los distintos sistemas argumentativos tratan de manera particular este tipo de argumento.



Para cualquier argumento  $A$ ,  $\phi \in A$  es una *conclusión* de  $A$  syss  $\phi$  es una fórmula de primer orden.  $\phi \in A$  es una *suposición* de  $A$  syss  $\phi$  es una suposición de un default en  $A$ .  $A$  es *estricto* si  $A$  no contiene defaults.  $A$  es *defendible* en otro caso.

Prakken & Sartor distinguen un argumento especial llamado “vacío” ( $[\ ]$ ), el cual no es derrotado por ningún argumento y por definición derrota a cualquier argumento que se derrota a sí mismo.

El conjunto de conclusiones de un argumento  $A$  se denota con  $\text{CONC}(A)$  y el conjunto de suposiciones  $\text{SUP}(A)$ .

Por ejemplo si el argumento  $A$  es  $[a, r1 : a \wedge \sim \neg b \Rightarrow c, c, a \wedge c, r2 : a \wedge c \Rightarrow d, d, d \vee e]$  quiere decir que  $\text{CONC}(A) = \{a, c, a \wedge c, d, d \vee e\}$  y  $\text{SUP}(A) = \{b\}$

La presencia de suposiciones en una regla origina dos clases de conflictos entre argumentos, el ataque *conclusión-a-conclusión*: la conclusión de un argumento ataca la conclusión de otro argumento (proviene de la idea de refutar o *rebutts* (en inglés)) y el ataque *conclusión-a-suposición*: la conclusión de un argumento ataca algún supuesto de otro argumento (proviene de la idea de socavar o *undercuts* (en inglés)).

**Definición 19**  $A$  ataca a  $B$  syss

- $\text{CONC}(A) \cup \text{CONC}(B) \cup F_n \vdash \perp$ , o
- $\text{CONC}(A) \cup F_n \vdash \neg \phi$  para cualquier  $\phi \in \text{SUP}(B)$

La derrota entre argumentos se construye en base a las ideas de “refutar” y “socavar”. Un argumento  $A$  *refuta* un argumento  $B$  syss  $A$  realiza un ataque conclusión a conclusión y, o bien,  $A$  es estricto y  $B$  defendible o las reglas default involucradas por  $A$  en el conflicto, no tienen menor prioridad que las reglas rebatibles involucradas por  $B$ . Hay que aclarar que un conjunto de reglas es menor que otro syss existe alguna regla  $r$  en el primero, tal que para toda regla  $r'$  del segundo, se verifica que  $r < r'$ .

Un argumento  $A$  *socava* un argumento  $B$  precisamente en la segunda clase de conflicto, cuando realiza un ataque a una suposición. Hay que destacar que no es necesario que el default responsable del ataque a la suposición no tenga menor prioridad que el default que contiene la suposición.

**Definición 20** (*derrota*) Un argumento  $A$  derrota a un argumento  $B$  syss  $A \neq []$  y  $B$  se ataca a sí mismo, o sino, si

- $A$  socava  $B$  o
- $A$  refuta  $B$  y  $B$  no socava  $A$ .

Se establece que  $A_1$  derrota *estrictamente* a  $A_2$  syss  $A_1$  derrota a  $A_2$ , pero  $A_2$  no derrota a  $A_1$ .

La razón por la cual el socavar tiene prioridad sobre el rebatir se explica mediante el siguiente ejemplo.

**Ejemplo 6** *Se consideran dos reglas:*

- $r_1: \sim \neg Brutus \text{ es inocente} \Rightarrow Brutus \text{ es inocente}$ .
- $r_2: \phi \Rightarrow \neg Brutus \text{ es inocente}$ .

Se asume que  $r_2$  no tiene prioridad sobre  $r_1$  y se consideran los argumentos  $[r_1]$  y  $[\dots, r_2]$ . Entonces aun cuando  $[r_1]$  refuta a  $[\dots, r_2]$ ,  $[r_1]$  no derrota a  $[\dots, r_2]$  debido a que  $[\dots, r_2]$  socava a  $[r_1]$ . De esa manera  $[\dots, r_2]$  derrota estrictamente a  $[r_1]$ .

Por qué esto debiera ser así? De acuerdo a Prakken y Sartor, la clave está en considerar la suposición de una regla como una de sus condiciones (aunque de una clase especial) para su aplicación. Entonces la única forma de aceptar ambas reglas es creer que no es cierto que Brutus es inocente: en ese caso la condición de  $r_1$  no es satisfecha. Por el contrario, si se cree que Brutus es inocente entonces  $r_2$  tiene que ser rechazada, en el sentido que sus condiciones son creidas pero su consecuente no (aquí 'creer una suposición' significa no creer su negación)

Esta línea de razonamiento no se aplica de manera natural al socavamiento en el estilo de Pollock. Esta es la razón que explica porque en [Pol95] los argumentos que derrotan por refutar y sacavar son colocados en igualdad de condiciones.

El planteo de Prakken y Sartor sobre el estado de los argumentos adhiere a la definición de Dung de semántica fija (grounded) que usa punto fijo. Sin embargo PraTor difiere en el origen de las prioridades default con las cuales se comparan los argumentos en conflicto.



En el ámbito de la investigación en Inteligencia Artificial, la consideración del lugar asignado a estas prioridades no es tratado como un tema de razonamiento de sentido común. En general, o bien se toma un ordenamiento fijo o se usa un ordenamiento de *especificidad*, leído a partir de la semántica o la sintaxis de una teoría de entrada. En PraTor se considera que existen muchos dominios de razonamiento de sentido común, como podrían ser las leyes o algunas burocracias, donde la prioridad es un aspecto que forma parte del dominio. Esto se cumple incluso con la especificidad, porque si bien verificar si un argumento es más específico puede ser una actividad lógica, preferir el argumento más específico, es una decisión extra-lógica. Además de variar de un dominio a otro, las fuentes de prioridad también pueden ser incompletas o inconsistentes, de la misma forma que sucede en los dominios de información “ordinarios”. En otras palabras, el razonamiento sobre las prioridades es razonamiento rebatible.

Por estas razones, Prakken & Sartor sostienen que los estados de los argumentos no solamente dependen de las prioridades, sino que también las determinan. De acuerdo a eso, la prioridad de las conclusiones puede ser derivada rebatiblemente dentro del sistema, de la misma forma que se llega a conclusiones como “Tweety vuela”.

Para formalizar esto, Prakken y Sartor recurren a unos pocos tecnicismos. Primero, la parte de primer orden del lenguaje se extiende con un predicado binario especial “ $\prec$ ”. Que  $x \prec y$  significa que  $y$  tiene prioridad sobre  $x$ . Las variables  $x$  e  $y$  pueden ser instanciadas con nombres default. Este nuevo símbolo predicado debiera denotar un orden parcial estricto sobre el conjunto de defaults que es asumido por la metateoría del sistema. Por esta razón, el conjunto  $\mathcal{F}_n$  debe contener los axiomas de un orden parcial estricto:

- Transitividad:  $\forall x, y, z. x \prec y \wedge y \prec z$  implica que  $x \prec z$
- Asimetría:  $\forall x, y. x \prec y$  implica que  $\neg y \prec x$

Por simplicidad, se asumen ciertas restricciones en la forma sintáctica de las expresiones de prioridad.  $\mathcal{F}_c$  no puede contener expresiones de prioridad, además las expresiones default de prioridad sólo pueden ocurrir en el consecuente, y exclusivamente en la forma de conjunción de literales (un literal es una fórmula atómica o una fórmula atómica negada). Esto excluye por ejemplo expresiones disyuntivas de prioridad.

En consecuencia, las relaciones de rebatimiento y derrota se realizan considerando una relación de orden que puede variar durante el proceso de razonamiento.

**Definición 21** *Para todo conjunto de argumentos  $S$ :*

- $<_S = \{r < r' \mid r \prec r' \text{ es una conclusión de algún } A \in S\}$
- $A$   $S$ -derrota (estrictamente)  $B$  *sys*, asumiendo el ordenamiento  $<_S$  en  $\Delta$ ,  
 $A$  derrota (estrictamente)  $B$ .

La idea es que cuando se debe determinar si un argumento es aceptable en relación a un conjunto de argumentos  $S$ , las relaciones de derrota sean verificadas de acuerdo a las conclusiones de prioridad determinadas por los argumentos de  $S$ .

**Definición 22** *Un argumento  $A$  es aceptable en relación a un conjunto de argumentos  $S$  *sys* todos los argumentos que  $S$ -derrotan  $A$ , son  $S$ -derrotados estrictamente por algún argumento en  $S$ .*

Prakken y Sartor emplean esta última definición de argumento aceptable en la construcción de la semántica fija de la Proposición 1. Prueban que el conjunto de argumentos justificados que se obtiene, es único y no tiene conflictos, y que cuando ese conjunto es  $S$ , el ordenamiento  $<_S$  es un orden parcial estricto.

### A.3. Formas dialécticas de argumentación

En la sección anterior se considera exclusivamente el aspecto semántico de los sistemas de argumentación, haciendo hincapié en las propiedades de conjuntos de argumentos. El tema central de esta sección es la prueba de la teoría o los aspectos procedurales de la argumentación, donde se busca establecer el estado de los argumentos individuales. La idea puede ser explicada en términos de un juego de diálogo entre dos participantes, un proponente y un oponente a un argumento. Un diálogo es una serie alternada de movidas realizadas por los dos participantes. El proponente comienza con un argumento que es puesto a prueba y cada movida subsecuente consiste en un argumento que, con cierta fuerza mínima, ataca alguna movida del contrincante. Probablemente el argumento inicial mantenga un determinado estado si el proponente emplea una estrategia ganadora, es decir, si puede lograr que el oponente agote las movidas posibles. Las reglas precisas del juego dependen de la semántica que se quiere capturar. Una idea natural es la asimetría

dialéctica. Por ejemplo si el juego refleja razonamiento escéptico, es decir, si pretende comprobar si un argumento está justificado, los argumentos del proponente deben derrotar de manera estricta; mientras que las movidas del oponente solamente deben derrotar. Si, por otro lado, el juego refleja razonamiento crédulo, esas reglas se invierten. [Prakken 1999]

A continuación se introduce el concepto de disputa más formalmente, haciendo uso de una versión adaptada de lo que se denomina *diálogo* en [Prakken & Sartor 1996] y *juego de argumentos* en [loui 1998]. El objetivo es capturar el razonamiento escéptico.

**Definición 23** (*Disputa*) Una disputa sobre un argumento  $A$ , es una secuencia no vacía de movidas  $s_i = (Jugador_i, A_i)$  ( $i > 0$ ) con  $A_i = A$ , en el cual un participante llamado PRO usa las movidas impares para tratar de sostener  $A$  y el otro participante, denominado CON, usa las movidas pares para tratar de impedir que  $Jugador_1$  tenga éxito.

1.  $Jugador_i = PRO$  si  $i$  es impar y  $Jugador_i = CON$  si  $i$  es par
2. Si  $Jugador_i = Jugador_j = PRO$  y  $i \neq j$  entonces  $A_i \neq A_j$
3. Si  $Jugador_i = PRO$  ( $i > 1$ ) entonces  $A_i$  derrota estrictamente  $A_{i-1}$
4. Si  $Jugador_i = CON$  entonces  $A_i$  derrota  $A_{i-1}$

La primer condición establece que comienza PRO y luego los jugadores se van turnando, por su parte la segunda condición impide que el proponente repita sus ataques. Las dos condiciones restantes forman el corazón de la definición: establecen los requerimientos de demostración para PRO y CON. De esa manera, se requiere que PRO sustente  $A$  mientras que CON solamente necesita molestar con derrotadores.

La disputa presentada corresponde a una versión simple, donde cada participante puede responder solamente una vez a cada argumento presentado por el oponente y si ese argumento no es efectivo no hay una otra oportunidad para replicar sobre el mismo argumento, o sea, no está permitido hacer regresión. Esta característica hace que el juego de argumentos corresponda con lo que se conoce oficialmente como *juego de suma cero entre dos participantes*, incluyendo los conceptos asociados, el más importante de los cuales es la estrategia.

Una clase de disputa más flexible que el juego de argumentos es la llamada disputa exhaustiva. Una disputa exhaustiva es un diálogo en el que cada participante puede intentar refutar cualquiera de los argumentos de su oponente. Si un participante descubre

que ha presentado un argumento erróneo, puede recuperarse de su error intentando con otro argumento, siempre que exista una alternativa.

Prakken & Sartor [1997b] extendieron su demostración de teoría incluyendo prioridades rebatibles. El problema principal es en base a *que* prioridades se debe determinar la fuerza derrotadora de las movidas. Unas pocas condiciones simples son suficientes para resolver el problema. CON puede ignorar completamente las prioridades: es suficiente que sus movidas derroten la movida previa de PRO. Para PRO se consideran sólo las prioridades establecidas en la movida misma. Es decir que para PRO es válido hacer una movida con un argumento  $A$ , siempre que  $A$   $A$ -derrote estrictamente la movida previa de CON; además PRO tiene otra movida disponible, mover un argumento de prioridad  $A$  de manera que la última movida de CON no  $A$ -derrote la movida anterior de PRO.

Esto resulta en los siguientes cambios en las condiciones 3 y 4 de la Definición 23.

3. Si  $Jugador_i = \text{PRO}$  ( $i > 1$ ) entonces  
 $Args_i$   $Args_i$ -derrota estrictamente a  $Args_{i-1}$  o  
 $Args_{i-1}$  no  $Args_i$ -derrota a  $Args_{i-2}$
4. Si  $Jugador_i = \text{CON}$  entonces  $Args_i$   $\emptyset$ -derrota a  $Args_{i-1}$

Prakken & Sartor [1997b] mostraron que sus resultados también fueron correctos y completos con esta definición (aunque en este caso es necesaria la asimetría dialéctica) La principal característica de su sistema, la cual asegura esos resultados, es la siguiente propiedad de la relación de derrota: si  $A$   $S$ -derrota  $B$  y  $S' \subseteq S$  entonces  $A$   $S'$ -derrota  $B$ .

# Bibliografía

- [A<sup>+</sup>97] ABRAMSKY, S., ET AL. Semantics of interaction: an introduction to game semantics. *Semantics and Logics of Computation* 14, 1 (1997).
- [AES08] AGUIRRE, G., ERRECALDE, M., AND SIMARI, G. Actos del habla en la comunicación. una propuesta usando programación lógica rebatible. In *XIV Congreso Argentino de Ciencias de la Computación (CACIC)* (Octubre 2008).
- [AES16] AGUIRRE, G., ERRECALDE, M., AND SIMARI, G. Una propuesta para mejorar la comunicación mediante persuasión. In *XXII Congreso Argentino de Ciencias de la Computación (CACIC)* (Octubre 2016).
- [AES17] AGUIRRE, G., ERRECALDE, M., AND SIMARI, G. Un formalismo para persuasión con delp. In *XXIII Congreso Argentino de Ciencias de la Computación (CACIC)* (Octubre 2017).
- [Ast65] ASTROM, K. Optimal control of markov processes with incomplete state information. *Journal of Mathematical Analysis and Applications* 10, 1 (1965), 174 – 205.
- [Aus62] AUSTIN, J. L. *How to Do Things with Words*. Harvard University Press, Cambridge, MA, Cambridge, MA, 1962.
- [BD62] BELLMAN, R. E., AND DREYFUS, S. E. *Applied Dynamic Programming*. Princetown University Press, 1962.
- [Bel57] BELLMAN, R. *Dynamic Programming*, 1 ed. Princeton University Press, Princeton, NJ, USA, 1957.

- [BIP88] BRATMAN, M. E., ISRAEL, D. J., AND POLLACK, M. E. Plans and resource-bounded practical reasoning, 1988.
- [BN78] BRUCE, B., AND NEWMAN, D. Interacting plans. *Cognitive Science* (1978), 195–233. (Also published in *Readings in Distributed Artificial Intelligence*, Alan H. Bond and Les Gasser, editors, pages 248–267, Morgan Kaufmann, 1988.).
- [CC00] CAPOBIANCO, M., AND CHESNEVAR, C. Using logic programs to model an agent’s epistemic state. In *II Workshop de Investigadores en Ciencias de la Computación* (Octubre 2000).
- [CGP99] CLARKE, E., GRUMBERG, O., AND PELED, D. *Model Checking*. The MIT Press, 1999.
- [CL85] COHEN, P. R., AND LEVESQUE, H. J. Speech acts and rationality. In *23rd Annual Meeting of the Association for Computational Linguistics, 8-12 July 1985, University of Chicago, Chicago, Illinois, USA, Proceedings*. (1985), pp. 49–60.
- [CS09] CHOPRA, A. K., AND SINGH, M. P. Multiagent commitment alignment. In *Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems-Volume 2* (2009), International Foundation for Autonomous Agents and Multiagent Systems, pp. 937–944.
- [CS15] CHOPRA, A. K., AND SINGH, M. P. Generalized commitment alignment. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems* (2015), International Foundation for Autonomous Agents and Multiagent Systems, pp. 453–461.
- [Dun95] DUNG, P. M. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n–person games. *Artif. Intell.* 77, 2 (1995), 321–358.
- [Err03] ERRECALDE, M. *Aprendizaje Basado en Múltiples Fuentes de Experiencia*. PhD thesis, Universidad Nacional del Sur, Departamento de Ciencias de la Computación, 11 2003.

- [FWW<sup>+</sup>93] FININ, T., WEBER, J., WIEDERHOLD, G., GENESERETH, M., FRITZSON, R., MCKAY, D., MCGUIRE, J., PELAVIN, R., SHAPIRO, S., AND BECK, C. Draft specification of the kqml agent-communication language, 1993.
- [Gar00] GARCÍA, A. J. *Programación en Lógica Rebatible: Lenguaje, Semántica Operacional y Paralelismo*. PhD thesis, Departamento de Ciencias de la Computación, Universidad Nacional del Sur, Bahía Blanca, Argentina, December 2000.
- Esta tesis define el lenguaje de la Programación en Lógica Rebatible (DeLP) y especifica su semántica operacional. El procedimiento de prueba de DeLP consiste de un análisis dialéctico el cual considera argumentos a favor o en contra de la consulta efectuada. Se presenta un modelo de computo secuencial y otro en paralelo. Se desarrollan aplicaciones para el lenguaje definido.
- [GD01] GMYTRASIEWICZ, P. J., AND DURFEE, E. H. Rational communication in multi-agent environments. *Autonomous Agents and Multi-Agent Systems* 4, 3 (2001), 233–272.
- [GFB<sup>+</sup>92] GENESERETH, M., FIKES, R. E., BRACHMAN, R., GRUBER, T., HAYES, P., LETSINGER, R., LIFSCHITZ, V., MACGREGOR, R., MCCARTHY, J., NORVIG, P., AND PATIL, R. Knowledge interchange format version 3.0 reference manual, 1992.
- [GLAE12] GIL, C., LORENZO, A., AGUIRRE, G., AND ERRECALDE, M. Método de mediación entre agentes y servicios web. In *XVIII Congreso Argentino de Ciencias de la Computación (CACIC)* (Octubre 2012).
- [Ham59] HAMBLIN, C. L. The modal "probably". *Mind* 68, 270 (1959), 234–240.
- [Ham70] HAMBLIN, C. L. *Fallacies, by C. L. Hamblin*. Methuen [London], 1970.
- [Kib06] KIBBLE, R. Speech acts, commitment and multi-agent communication. *Computational & Mathematical Organization Theory* 12, 2-3 (2006), 127–145.
- [Lor58] LORENZEN, P. Logik und agon. *Acta del XII Congresso Internazionale de Filosofia* (1958), 187–194.
- [Lou98] LOUI, R. P. Process and policy: Resource-bounded nondemonstrative reasoning. *Computational intelligence* 14, 1 (1998), 1–38.

- [Mac79] MACKENZIE, J. D. Question-begging in non-cumulative systems. *Journal of Philosophical Logic* 8, 1 (1979), 117–133.
- [McB11] MCBURNEY, P. An salute to charles hamblin. url <http://www.vukutu.com/blog/2011/01/a-salute-to-charles-hamblin/>, 2011.
- [MP09] MCBURNEY, P., AND PARSONS, S. Dialogue games for agent argumentation. *Argumentation in Artificial Intelligence* (2009), 261–280.
- [New82] NEWELL, A. The knowledge level.
- [Pol95] POLLOCK, J. *Cognitive Carpentry: A Blueprint for how to Build a Person*. Bradford Bks. MIT Press, 1995.
- [Pra01] PRAKKEN, H. Relating protocols for dynamic dispute with logics for defeasible argumentation. *Synthese* 127, 1-2 (2001), 187–219.
- [Pra05] PRAKKEN, H. Coherence and flexibility in dialogue games for argumentation. *Journal of logic and computation* 15, 6 (2005), 1009–1040.
- [Pra06] PRAKKEN, H. Formal systems for persuasion dialogue. *The knowledge engineering review* 21, 02 (2006), 163–188.
- [Pra09] PRAKKEN, H. Models of persuasion dialogue. In *Argumentation in artificial intelligence*, I. Rahwan and G. R. Simari, Eds. Springer, 2009, pp. 281–300.
- [PS97] PRAKKEN, H., AND SARTOR, G. Argument-based extended logic programming with defeasible priorities. *Journal of applied non-classical logics* 7, 1-2 (1997), 25–75.
- [PV02] PRAKKEN, H., AND VREESWIJK, G. *Logics for Defeasible Argumentation*. Springer Netherlands, Dordrecht, 2002, pp. 219–318.
- [RLV07] ROBERTS, D., LOCK, G., AND VERMA, D. C. Holistan: A Futuristic Scenario for International Coalition Operations. *Integration of Knowledge Intensive Multi-Agent Systems, 2007. KIMAS 2007. International Conference on* (2007), 423–427.
- [RN03] RUSSELL, S. J., AND NORVIG, P. *Artificial Intelligence: A Modern Approach*, 2 ed. Pearson Education, 2003.



- [Sad91] SADEK, M. D. Dialogue acts are rational plans. In *Proceedings of the ESCA/ETRW Workshop on the Structure of Multimodal Dialogue* (Maratea, Italy, 1991), pp. 1–29.
- [Sea69] SEARLE, J. R. *Speech acts: An essay in the philosophy of language*, 1969.
- [Sin98] SINGH, M. P. Agent communication languages: Rethinking the principles. *Computer* 31, 12 (December 1998), 40–47.
- [Sin00] SINGH, M. P. A social semantics for agent communication languages. In *Issues in Agent Communication* (2000), F. Dignum and M. Greaves, Eds., vol. 1916 of *Lecture Notes in Computer Science*, Springer, pp. 31–45.
- [Sin08] SINGH, M. P. Semantical considerations on dialectical and practical commitments. In *Proceedings of the Twenty-Third Conference on Artificial Intelligence, 2008, Chicago, Illinois, USA, July 13-17, 2008* (2008), pp. 176–181.
- [SLB09] SHOHAM, Y., AND LEYTON-BROWN, K. *Multiagent Systems - Algorithmic, Game-Theoretic, and Logical Foundations*. Cambridge University Press, 2009.
- [SS73] SMALLWOOD, R. D., AND SONDIK, E. J. The optimal control of partially observable markov processes over a finite horizon. *Operations Research* 21, 5 (1973), 1071–1088.
- [Tan12] TANG, Y. *A Symbolic Exploration of the Joint State Space and the Underlying Argumentation-based Reasoning Processes for Multiagent Planning*. PhD thesis, City University of New York, USA, 2012.
- [Ver97] VERHAREN, E. *A Language-action Perspective on the Design of Cooperative Information Agents*. E.M. Verharen, 1997.
- [Vre95] VREESWIJK, G. A. The computational value of debate in defeasible reasoning. *Argumentation* 9, 2 (1995), 305–342.
- [Wal84] WALTON, D. N. *Logical Dialogue-Games and Fallacies*. University Press of America, Lanham, Maryland, 1984.
- [WK95] WALTON, D., AND KRABBE, E. *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning*. SUNY series in Logic and Language. State University of New York Press, 1995.

- [Woo09] WOOLDRIDGE, M. *An introduction to multiagent systems*. John Wiley & Sons, 2009.