Universidad Nacional del Sur

Tesis de Magister en

Ciencias de la Computación

Un Modelo Dialéctico para la Deliberación Multiagente

Alejandro G. Stankevicius

Bahía Blanca – Argentina

2004

Universidad Nacional del Sur

Tesis de Magister en

Ciencias de la Computación

Un Modelo Dialéctico para la Deliberación Multiagente

Alejandro G. Stankevicius

Bahía Blanca – Argentina

2004

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 e Ingeniería de la Computación, durante el período comprendido entre junio de 1999 y agosto de 2004, bajo la dirección del Dr. Guillermo R. Simari, profesor Titular del Departamento de Ciencias e Ingeniería de la Computación y la codireccion del Dr. Alejandro J. García, profesor Adjunto de Departamento de Ciencias e Ingeniería de la Computación.

Lic. Alejandro G. Stankevicius

Departamento de Ciencias e Ingeniería de la Computación

Universidad Nacional del Sur

Bahía Blanca, 24 de agosto de 2004

Resumen

Los *sistemas multiagente* deben su gran atractivo a las propiedades que manifiestan los sistemas organizados bajo esta concepción. En lugar de tener que resolver todos los aspectos del problema en consideración de manera centralizada, se apela a un sencillo principio de delegación de responsabilidades, de eficacia comprobada en el mundo empresarial, esta vez en el contexto de un conjunto de entidades independientes, denominadas *agentes*, que cuentan con capacidades individuales. Al indagar las razones por las que los sistemas multiagente manifiestan estas propiedades se arriba casi indefectiblemente a la conclusión que la interacción entre los distintos agentes es la llave que da acceso a ese nuevo horizonte de posibilidades.

Por otra parte, la *argumentación rebatible*, un refinado marco conceptual que permite tanto representar conocimiento como razonar en base al mismo, presenta una clara analogía con el escenario anterior en donde multiples entidades interactúan entre sí. De hecho, el razonamiento argumentativo de base dialéctica, aquél en el cual el intercambio de razones, denominadas *argumentos*, asemeja a un prolijo diálogo entre contendientes, prácticamente parece describir o modelar la interacción entre estos agentes.

De la combinación de estas apreciaciones, la investigación reportada en esta tesis define un *modelo abstracto* para describir la interacción entre agentes elaborado a partir de otro modelo, también de naturaleza abstracta, que captura las principales caractirísticas del razonamiento argumentativo de base dialéctica. El modelo obtenido puede ser instanciado en función de las necesidades del caso, permitiendo abordar problemas de diversa índole, en los más variados contextos.

Abstract

Multiagent Systems owe their great success to properties observed in those systems organized under this conception. Instead of solving all the issues associated to the problem at hand in a centralized manner, a simple principle of task delegation, whose effectivity has already been establish in the industry, is applied this time to the context of a set of independent entities, called *agents*, each possesing individial capabilities. Exploring the reasons that justify why multiagent systems show these properties, one usually conclude that the interaction among different agents is the key aspect granting access to this new horizon of possibilities.

Defeasible argumentation, a refined framework for knowledge representation and reasoning, involves a set of notions clearly analogous to the setting where multiple entities interact with each other. In particular, argumentative reasoning, especialy when conducted in a dialectical fashion, where the exchange of reasons—called in this context *arguments*—resembles a dialog between contenders, almost describing or even modeling the interaction among these agents.

From the combination of these two fields, the research reported in this thesis formally defines an *abstract model* capable of describing this interaction among agents, developed starting from another model, also of abstract nature, capable of capturing the main characteristics of dialectics-oriented argumentative reasoning. The resulting model can be instantiated according our needs, allowing the treatment of diverse problems, from dissimilar contexts.

Agradecimientos

Imaginar poder estar escribiendo estas palabras en el marco de la presente Tesis, fruto de varios años de investigación en Ciencias de la Computación, sin la orientación, el apoyo, el aliento, las críticas (constructivas, desde ya), y la amistad de mi director, Guillermo R. Simari, sería una utopía. Por esta razón, siento la obligación moral de reservar este espacio distinguido para agredecer su rol siempre incondicional de mi lado durante todo este tiempo.

De igual forma, no puedo dejar de reconocer el importante papel de mi codirector, Alejandro J. García, quien volcó su basta experiencia aportando esos consejos invaluables, imposibles obtener de fuentes estáticas de conocimiento, pues sólo se manifiestan como producto de la experiencia en primera persona del desarrollo de la ciencia.

Muchas otras personas estuvieron involucradas en el desarrollo y la maduración de las ideas aquí presentadas, ya sea forma directa o inderecta. Por caso, mis compañeros de oficina, tanto los actuales como los anteriores, contribuyeron generando una atmósfera cálida y hogareña, siempre dispuestos a debatir algún tópico, en ocasiones teórico y relativo a la presente Tesis, o bien a veces enteramente informales.

Por caso, el resto de los integrantes del LIDIA, el Laboratorio de Investigación y Desarrollo en Inteligencia Artificial, con quienes he compartido cursos y materias de posgrado, han aportado críticas constructivas que muchas veces condujeron a reconsiderar la formulación de los conceptos y definiciones vertidos en esta Tesis.

En síntesis, debo extender el agradecimiento a la totalidad del Departamento de Ciencias e Ingeniería de la Computación, tanto a su personal docente como no

docente, con quienes evidentemente he compartido gran cantidad de momentos, y en ocasiones una charla casual era todo lo que hacía falta para que ese concepto difícil de capturar pudiera ser formalizado.

Finalmente, esta Tesis o incluso el poder estar formando parte de este Departamento, lugar en donde he llevado adelante las investigaciones aquí reportadas, no serían una realidad de no existir Marcela, a quien debo todo lo hoy soy.

Índice general

1. Introducción y Motivaciones **1**

1.1. La lógica y el razonamiento . . . 1

1.2. Argumentación rebatible . . . 7

1.3. Sistemas multiagente . . . 8

1.4. Motivaciones . . . 10

1.4.1. Principales aportes . . . 10

1.4.2. Publicaciones previas . . . 12

1.5. Estructura de la tesis . . . 14

2. Sistemas Argumentativos **17**

2.1. Introducción . . . 17

2.1.1. Primeros pasos hacia la argumentación rebatible . . . 18

2.1.2. El rol de la dialéctica . . . 22

2.1.3. Características generales . . . 27

2.2. Principales sistemas argumentativos . . . 31

2.2.1. Marcos Agumentativos . . . 31

2.2.2. El aporte de J. Pollock y su arquitectura Oscar . . . 38

2.2.3. Sistemas Argumentativos Abstractos . . . 47

2.2.4. El sistema de H. Prakken y G. Sartor . . . 62

2.2.5. El sistema Simari-Loui . . . 81

2.2.6. Programación en Lógica Rebatible . . . 100

2.3. Conclusiones . . . 119

3. Sistemas Multiagente **123**

3.1. Motivaciones 124

3.2. Generalidades 127

3.3. Principales aplicaciones 130

3.4. Taxonomía de los Sistema Multiagente 136

3.5. Aspectos a resolver en la etapa de diseño 142

3.6. La interacción entre los agentes 146

3.6.1. Lenguajes de comunicación entre agentes 147

3.6.2. Técnicas de coordinación 158

3.6.3. Mecanismos de cooperación 165

3.6.4. Protocolos de negociación 173

3.7. Conclusiones 182

4. Un Modelo para la Deliberación Multiagente **185**

4.1. Introducción 185

4.2. Primeros acercamientos 187

4.2.1. La propuesta de Ron Loui 187

4.2.2. El modelo de Hendrik Prakken 195

4.3. Un modelo para el razonamiento dialéctico 209

4.3.1. Aspectos primordiales 210

4.3.2. Protocolos Dialécticos 212

4.4. Un modelo abstracto para la deliberación 221

4.5. Ejemplos de instanciaciones concretas 230

4.5.1. Modelando la Programación en Lógica Rebatible 230

4.5.2. Modelando prioridades entre las reglas 238

4.6. Conclusiones 244

5. Conclusiones y Resultados Obtenidos **247**

Glosario **253**

Índice de figuras

2.1. Estructura de un argumento según Toulmin 18

2.2. Una disputación de acuerdo al modelo de Rescher 25

2.3. Ataque directo e indirecto 29

2.4. Representación gráfica de un marco argumentativo 33

2.5. Grafo asociado al marco argumentativo del ejemplo 2.5 36

2.6. Tipos de argumentación circular en un marco argumentativo 36

2.7. Árbol de diálogos asociado al ejemplo 2.18 74

2.8. Análisis dialéctico de los argumentos del ejemplo 2.26 93

2.9. Marcado de los árboles de dialéctica del ejemplo 2.27 94

2.10. Derrotadores recíprocos en el sistema Simari-Loui 96

2.11. Una línea de argumentación en el sistema Simari-Loui 98

2.12. Análisis dialéctico del ejemplo 2.35 115

2.13. Marcado del árbol de dialéctica del ejemplo 2.35 117

3.1. Razones más frecuentes para optar por un sistema multiagente 126

3.2. Una taxonomía de sistemas multiagente 141

3.3. Rol de los ACL en la interacción entre agentes 149

3.4. Un mensaje expresado en KQML 152

3.5. Comunicación entre agentes sincrónica y asincrónica via KQML . . . 153

3.6. Un acto comunicativo expresado en el lenguaje FIPA-ACL 157

3.7. Ejemplo de una convención asociada a la revisión de compromisos . . 164

3.8. Interacción entre los componentes de un sistema de pizarrón 170

4.1. Relevancia de las jugadas en una disputa 208
4.2. Distribución del conocimiento en un contexto de agente único 213
4.3. Distribución del conocimiento en un contexto de agentes múltiples . . 223
4.4. Arbol dialéctico asociado a un estado de disputa 234

Capítulo 1

Introducción y Motivaciones

Este capítulo tiene por objeto situar el tópico abordado en la presente tesis en el marco del estado actual de la investigación científica, describiendo, asímismo, las principales motivaciones que alentaron su emprendimiento. Se introducen someramente los principales conceptos involucrados, los cuales naturalmente serán atendidos con mayor consideración en los restantes capítulos. Por otra parte, se presenta una síntesis de las contribuciones, reseñando las diversas publicaciones confeccionadas a lo largo de los estudios aquí abarcados. Finalmente, concluye con un bosquejo de la estructuración de los restantes capítulos.

1.1. La lógica y el razonamiento

Desde los albores mismos de la civilización tal como hoy la conocemos existe un interrogante que ha venido obsesionando a nuestros más grandes pensadores: *¿qué es la ciencia?*, o también *¿en qué consiste hacer ciencia?* De hecho filósofos de la talla de Sócrates (469–399 AC) o Platón (427–348 AC) dedicaron gran parte de sus vidas a explorar los diversos matices de estas preguntas, arribando eventualmente a la interpretación actual de qué entendemos por ciencia, esto es, *el estudio de las cosas necesarias*, en ocasiones también parafraseado como *el estudio de las cosas verdaderas*. Sin embargo, esta definición a pesar de su aparente simpleza no hace más que introducir un nuevo desafío. Para hacer ciencia debemos restringir nuestras atención sólo a aquellas cosas que son ciertas; pero ¿cómo distinguir a éstas de aquellas que no lo son? Esta nueva pregunta por trivial que resulte no deja de plantear una problemática la cual incluso casi veinticinco siglos después seguimos intentando conquistar.

Puede afirmarse que para poder conducir una investigación científica seria y responsable resulta necesario contar con algún método que permita distinguir las cosas verdaderas de aquellas que no lo son. En otras palabras, estamos obligados a estructurar de forma metódica nuestra incesante cruzada en pos del descubrimiento de nuevas verdades. La primer propuesta en este sentido vino de la mano del mismo Platón, a través de su tan mentada teoría de la división según los géneros. Este acercamiento si bien permite el estudio estructurado de disciplinas tales como la biología (sin duda adecuado para el estadío de ésta en aquellos tiempos), resultó posteriormente criticada por su discípulo predilecto, Aristóteles (384–322 AC), quien señaló que no resultaba propicia para abordar a la totalidad de la ciencia en su conjunto. La crítica de Aristóteles vino acompañada de una nueva propuesta: los *silogismos*, una particular forma de estructurar y ordenar el razonamiento de forma tal que nos podamos despreocupar de la validez de los argumentos formulados, ya que la razonabilidad de éstos resultará a todas luces evidente. Por caso, el silogismo quizás más conocido sea aquel que señala que en virtud de que Sócrates es un hombre y que todo hombre es mortal, Sócrates necesariamente debe ser mortal. De hecho, la estructura de todo silogismo guarda esa forma: cuenta con una conclusión obtenida en base a dos premisas, una de ellas inicial y la otra conectiva. La aceptabilidad de la conclusión descansa tanto en su vínculo con la premisa intermedia como en la de ésta con la inicial. No obstante, no todo silogismo conduce a descubrir nuevas verdades anteriormente desconocidas, sólo aquellos que partan a su vez de premisas ya establecidas como válidas conducirán a nuevas premisas igualmente válidas. Aristóteles distingue a este particular tipo de silogismo como *silogismo científico*.

Este filósofo no descansó ni aún habiendo enumerado y analizado a fondo una gran cantidad de formas silogísticas, sino que avanzó un paso más allá, señalando que incluso el silogismo científico no simplifica la tarea de aquellos en pos de nuevas verdades, ya que las infinitas posibilidades a la hora de recombinar una y otra vez estas diversas formas suelen agobiarnos antes de podamos alcanzar algún resultado nóvel o de utilidad práctica. En este sentido, introdujo lo que posiblemente constituya una de sus contribuciones más preciadas: *la dialéctica*. Esta técnica, a la cual apelaremos una y otra vez a lo largo de la presente tesis, permite comprobar sistemáticamente la fortaleza de una dada propuesta, estudiando si presenta alguna flaqueza en base a la cual cuestionar su aceptabilidad. Según Aristóteles, la dialéctica combinada con el silogismo científico constituye el camino más adecuado para explorar nuevas ideas, para luego una vez que estas nuevas ideas son unánimemente aceptadas como válidas, expandir nuestro entendimiento de la misma ciencia.

Lamentablemente, pocos fueron los que pudieron comprender el sentido real de la propuesta de Aristóteles, principalmente debido a que los primeros ensayos concretos con este tipo de dialéctica inquisitiva adolecieron de un dejo de informalidad que invalidó rápidamente toda conclusión formada. Mejores resultados se obtuvieron a partir de la idea precursora, el silogismo científico, pero recién cuando la lógica simbólica alcanzó la maduración necesaria para formalizar este concepto de una forma fiel a la visión de su artífice. En este sentido poco sucedió desde la época dorada de la Grecia clásica hasta la revolución en el razonamiento simbólico, iniciada posiblemente por René Descartes (1596–1650) con su propuesta para reinterpretar problemas geométricos como meras ecuaciones algebraicas.

El primero en venturar el crucial rol que la lógica puede desempeñar en la ciencia fue Gottfried Wilhelm Leibnitz (1640-1710), quien aseguró que era enteramente factible concebir un formalismo lógico capaz de modelar incluso nuestro propio razonamiento. No obstante, sus consideraciones en este sentido no resultaron tan influyentes como se hubiera esperado, posiblemente debido una vez más a la falta de madurez de la lógica simbólica como disciplina científica. Para apreciar un avance significativo en esta dirección debemos apelar a los aportes de eminencias tales como Augustus de Morgan (1806–1871) o George Boole (1815–1864), cientos de años después de que Leibnitz enunciara lo que hoy se conoce como *su sueño*. De Morgan extendió el concepto de silogismo introduciendo rudimentarios conectivos lógicos proposicionales junto con sus respectivas reglas de comportamiento (hoy conocidas como tablas de verdad). Boole, debido a su vasta experiencia con los métodos algebráicos simbólicos, reinterpretó la incipiente lógica proposicional en sus propios términos, siendo el primero que logró llevar adelante un razonamiento de base lógica desde una perspectiva puramente algebráica. De más está enfatizar que el aporte de estos y otros renombrados científicos a la lógica simbólica fue mucho más allá de lo que esta somera introducción permite reseñar.

Con el estudio de las propiedades formales de la lógica proposicional, se concluyó rápidamente que no contaba con el poder expresivo necesario para modelar situaciones lo suficientemente interesantes como para que las eventuales conclusiones allí formadas tuvieran alguna relevancia en el mundo real (es decir, para que sirviera como herramienta a través de la cual expandir la frontera de la ciencia). El principal inconveniente radica en la contundencia de las proposiciones; éstas aseveran propiedades genéricas, que no permiten hacer una referencia explícita a los individuos involucrados ni tampoco a las modalidades o cuantificaciones que puedan afectarlas. No eran representables ni siquiera aquellas ya contempladas en los tiempos de Aristóteles.[1] El próximo avance significativo en la lógica simbólica—tanto en poder expresivo como en complejidad del formalismo—vino de la mano de Gottlob Frege (1848–1925), quien proveyó la primer formalización completa de lo que hoy conocemos como cálculo de predicados. En ese entonces había un creciente interés en lograr capturar toda la matemática dentro de un sistema formal, con el objeto último de estructurar un poco mejor al área y evitar así las indeseables paradojas y falacias, producto de teorías asentadas en bases no del todo sólidas.

La frustración que causó el por muchos inesperado descubrimiento que condena a la incompletitud o a la inutilidad a todo sistema formal lo suficientemente expresivo,

[1] por caso el silogismo acerca de Sócrates hace un uso implícito de la cuantificación universal.

sintetizado de forma brillante por Kurt Gödel (1906–1978), en su famoso teorema de la incompletitud, asestó un duro golpe al avance del área. El resultado demostrado por Gödel, si bien un tanto desalentador, no hace más que enfatizar un aspecto con el cual debemos convivir: *ningún formalismo de interés podrá ser completo*, en el sentido de que algunas verdades eludirán eternamente a su maquinaria de inferencia. La lógica simbólica entró en un nuevo impasse, esta vez hasta el advenimiento de las computadoras y su correspondiente ciencia de la computación. En particular, la disciplina conocida como Inteligencia Artificial, aquella que persigue que las computadores logren manifestar un comportamiento inteligente, rescató muchos de estos avances ya que el tipo de razonamiento que induce una teoría lógica formal parece constituir el camino más evidente hacia el objetivo trazado. Cabe aclarar, en este contexto por razonar se entiende el proceso que permite obtener nuevas conclusiones en base a premisas ya conocidas y a sus particulares interrelaciones.

La combinación de estos dos factores, lógicas lo suficientemente expresivas con computadoras lo suficientemente poderosas, dio a lugar al boom que la Inteligencia Artificial vivenció en sus orígenes, donde incluso genios de la talla de Alan Mathison Turing (1912–1954)—prácticamente el padre de la Inteligencia Artificial, cuando no de la misma ciencia de la computación—llegaron a realizar predicciones excesivamente optimistas en relación a la magnitud de los problemas que iban a poder ser resueltos a futuro. Poco de esto fue verdad, principalmente debido a que la complejidad de los problemas por resolver es mucho mayor de lo que un análisis superficial permite apreciar. Escenarios tan ingenuos como el mundo de bloques terminaban representando un desafío infranqueable desde un punto de vista computacional, situación que terminó por decepcionar a la mayoría de los investigadores abocados al área.

Por suerte no todos se vieron frustrados por la riqueza conceptual de la realidad o lo profundo del más simple de los razonamientos. Casualmente John McCarthy (1927–), otro de los fundadores de la Inteligencia Artificial como disciplina científica, fue uno de los pioneros en señalar los infinitos matices que parece presentar el razonamiento de sentido común al que todos apelamos al llevar adelante incluso tareas harto triviales. Uno de los aspectos más problemáticos de las lógicas disponibles en ese entonces no era ya su capacidad representacional, sino la excesiva persistencia de las consecuencias sancionadas por éstas. Concretamente, el conjunto de las conclusiones sustentadas por la maquinaria de inferencia de las lógicas clásicas crece monótonamente, debido a que una vez establecida una dada consecuencia, no importa que nuevas premisas sean tenidas a partir de allí en cuenta, ésta seguirá siendo sancionada por la teoría. Esta otrora deseable propiedad (denominada *monotonía*), se constituyó en una pesada carga difícil de evadir al intentar utilizar este tipo de lógica como sustrato en base al cual razonar, puesto que nuestro razonamiento intrínseco pareciera comportarse en ocasiones de forma no monótona (sobre todo, al confrontar información incompleta o potencialmente contradictoria).

El comienzo de la década de los ‘80 marcó una explosión en el desarrollo de alternativas que subsanen esta contingencia, bajo la forma de un abanico de formalismos cuya principal característica en común era precisamente comportarse de forma *no monótona*. Estas propuestas fueron recopiladas en el memorable volumen número 13 de *Artificial Intelligence*, presentación en sociedad de importantes teorías como la *lógica default* de R. Reiter [Reiter, 1980], la *circunscripción* del mismo J. McCarthy [McCarthy, 1980] o la *lógica no monótona* de D. McDermott y J. Doyle [McDermott y Doyle, 1980] entre otras. Como estos formalismos permitían retractar conclusiones aún luego de haber sido éstas establecidas, un nuevo tipo de razonamiento pudo ser llevado a cabo, el cual fue acertadamente denominado *razonamiento rebatible.* [2] El razonamiento rebatible, si bien más adecuado para capturar el tipo de razonamiento requerido para manifestar un comportamiento inteligente, introdujo a su vez nuevos desafíos. Uno de los inconvenientes más observados fue que los primeros acercamientos resultaron excesivamente complejos y poco intuitivos, lo que a su vez poco colaboraba a la hora de trocar papel y lápiz por un dominio de aplicación en concreto.

[2] término atribuído al filósofo John Pollock, posiblemente en virtud del artículo [Pollock, 1987].

1.2. Argumentación rebatible

En el marco del creciente interés por los modelos no monótonos de razonamiento es que toma fuerza la idea de revivir el espíritu original de la propuesta de Aristóteles, en el sentido de apelar a la crítica dialéctica de las ideas como base, un tanto más intuitiva, sobre la cual erigir una teoría para el razonamiento rebatible: la *argumentación rebatible* estaba en sus albores. Cabe señalar que en esta sucinta introducción tan sólo exploraremos los conceptos básicos, puesto que más adelante todo un capítulo estará encargado de repasar los aspectos más relevantes de esta disciplina.

Los sistemas argumentativos suelen representar el conocimiento que tienen acerca del mundo a través de un conjunto de *reglas*, inicialmente inspiradas en la implicación material de la lógica clásica, es decir, en el concepto de inclusión de categorías (base por otra parte de los propios silogismos). Son construcciones de la forma $\alpha \rightarrow \beta$, donde el contar con un razón para aceptar a α nos compromete en cierto sentido a tener que aceptar a β, si bien por lo general, y a diferencia de la implicación material, las razones para aceptar $\neg\beta$ suelen no tener relación alguna con la aceptabilidad de $\neg\alpha$. Cuando la conexión entre antecedente y consecuente de una dada regla sea un tanto menos directa, se suele apelar a un segundo conectivo para denotar esta nueva situación más sutil, por caso $\alpha \Rightarrow \beta$. En este escenario se podría decir que aquellas razones para creer en α son a su vez buenas razones para creer en β. El cambio en el énfasis de la conección entre antecedente y consecuente intenta reflejar la diferencia de corte pragmático en estos dos constructores, si bien la diferencia exacta entre '$\rightarrow$' y '$\Rightarrow$' dependerá de la semántica del sistema considerado. En general la tendencia actual es interpretar estas reglas como si fueran meras *reglas de inferencia*, las cuales permiten obtener nuevas conclusiones a partir de la evidencia ya conocida.

El sistema argumentativo utiliza estos conjuntos de reglas para construir *argumentos*, básicamente simples encadenamientos sucesivos de la aplicación de estas reglas. Un argumento denota una pieza tentativa de razonamiento a favor de una cierta conclusión. Considerando que es factible que existan argumentos tanto a favor como en contra de una tesis dada, la noción de argumento no resulta propicia para caracterizar el conjunto de conclusiones sancionadas por una teoría argumentativa: es necesario considerar un nuevo nivel de abstracción. A tal efecto, todos los argumentos que pueden ser construídos a partir de una base de conocimiento deben ser cotejados entre sí, a fin de determinar cuáles resultan aceptables y cuáles deben ser descartados. Se suelen introducir un conjunto de relaciones entre argumentos las cuáles capturan las diferentes situaciones de interés, por caso, subargumento, desacuerdo, contraargumentación o derrota. Esta última relación, caracterizada usualmente en base a las anteriores, resulta crucial ya que la misma semántica del sistema argumentativo puede en general ser reformulada en términos del conjunto de conclusiones sustentadas por aquellos argumentos *no derrotados*.

1.3. Sistemas multiagente

Si bien los sistemas argumentativos en virtud de su poder expresivo permitieron atacar problemas más complejos, la mayoría de los escenarios que hoy en día se consideran atractivos no sólo comparten su complejidad como característica en común, sino que además suelen presentar requerimientos de control físicamente distribuido, poder acceder a un conocimiento no centralizable o bien precisar que varios cómputos se lleven adelante de forma asincrónica. Esta nueva demanda motivó que en las últimas décadas el desarrollo de los *sistemas multiagente* haya monopolizado la atención de gran parte de la comunidad académica. Ahora el énfasis parece haberse dirigido hacia aquellos sistemas en los cuales existan no uno sino varios componentes con capacidades individuales, que interactúen entre sí o con su entorno. Cada uno de estos actores con capacidades individuales se denominan *agentes*. Nuevamente, el tratamiento brindado a esta temática en la presente introducción lejos de ser exhaustivo es tan sólo una mera presentación en sociedad del tópico, el cuál será posteriormente abordado en mayor profundidad en un capítulo aparte.

Aún cuando no existe un consenso generalizado sobre qué constituye una caracterización precisa del concepto de sistema multiagente, a nuestros efectos consideraremos que se trata de una colección de agentes racionales y autónomos, capaces de coordinar su conocimiento, objetivos, habilidades y estrategias a fin de resolver un problema en común. En otras palabras un sistema multiagente es un conjunto de entidades independientes que cuentan con un conjunto de creencias y de capacidades propias, que interactúan entre sí y con su entorno (a veces simplemente modelado como un agente más) a fin de alcanzar algún objetivo en común previamente acordado. El aspecto más controversial de la definición anterior suele ser el requerimiento de racionalidad, si bien en el contexto que nos atañe, denota sencillamente el optar por las acciones más apropiadas que tiendan a satisfacer los objetivos actuales (lo que en términos de la teoría de juegos se define como la maximización de la función beneficio).

Una de las particularidades más llamativas de este tipo de sistema es que su éxito se basa en que la versatilidad del sistema multiagente en su conjunto parece superar a la suma de las capacidades individuales de sus integrantes. Sin duda este incremento en la capacidad del sistema sólo puede ser atribuída a aquellos aspectos inherentes a la presencia de múltiples agentes. En este sentido, es la misma interacción entre los agentes la que permite manifestar aquellos comportamientos más elaborados tales como el intercambio de información, la coordinación de tareas y comportamientos, la persuación de otros agentes para que reconozcan las virtudes de la posición propia, o incluso la negociación de las contraprestaciones deseadas a cambio de llevar a cabo una dada tarea. Esta situación justifica la multitud de modelos recientemente propuestos para describir esta interacción, con el objeto tanto de ahondar nuestro propio entendimiento de su dinámica así como para explorar nuevas avenidas en lo que a las aplicaciones concretas respecta. Cabe acotar que es factible que uno de los modelos de interacción más estudiados sea aquel inspirado en la teoría de juegos, disciplina a la cual eminencias de la talla de John von Neumann o John Nash entre otros dedicaron gran parte de su atención, si bien recientemente se han ensayado otros enfoques alternativos.

1.4. Motivaciones

Es en este contexto histórico, someramente reseñado en las secciones anteriores, donde se encuadra la investigación llevada adelante en la presente tesis. La propuesta aquí desarrollada consiste, en pocas palabras, en extrapolar un concepto desarrollado y probado en el ámbito del razonamiento rebatible al marco de los sistemas multiagente. Es decir, la idea básica consiste en capitalizar la experiencia adquirida al trabajar con los diversas teorías para la argumentación rebatible, específicamente en sus concepciones dialécticas, con el objeto de reinterpretar estas nociones en el ámbito de los sistemas multiagente, donde el razonamiento de carácter introspectivo típico de los sistemas argumentativos se vea transformado en un juego dialógico entre las diversas entidades que conforman un sistemas multiagente. Como señaláramos en la sección anterior, el modelar la interacción entre los agentes constituye una problemática en pleno estudio. Por otra parte, la estrategia sugerida brinda en principio una guía conceptual que permite abordar este tópico en una forma metódica y ordenada.

A continuación sintetizaremos los diversos aportes generados a los largo de la consecución de este objetivo, así como aquellas publicaciones previas en las que se exploraron las ideas fundacionales que sustentan el presente estudio.

1.4.1. Principales aportes

El aporte central de este tesis consiste en proponer un modelo formal para caracterizar la interacción entre los agentes de una sistema multiagente, tomando como punto de partida la similitud estructural entre el razonamiento de tipo dialéctico y el juego dialógico que llevan adelante los agentes al interactuar entre sí. No obstante, otras contribuciones de diversa índole fueron a su vez aportadas al perseguir esta meta, a saber:

- Se propuso una estrategia en particular que permite obtener, de una manera relativamente directa, al modelo para describir en términos formales a la interacción entre agentes. La problemática abordada constituye, como se verá más adelante, una de las facetas en las cuales actualmente se requiere un mayor desarrollo desde el terreno teórico, con el objeto de cimentar sólidamente aquel conjunto de características deseables, pero lamentablemente hoy ausentes en la mayoría de las implementaciones concretas.

- Se condujo una extensiva y profunda revisión de los principales exponentes en la literatura de las teorías que modelan la argumentación de tipo rebatible, con el objeto de identificar cuáles son las caractirísticas recurrentes en este tipo de formalismo, considerando a su vez tanto sus aspectos atractivos como sus falencias. En particular, en relación a aquellos aspectos controversiales de cada uno de los formalismos considerados, se analizó en detalle cuáles fueron las razones que condujeron a cuestionar tales características.

- Se elaboró un modelo abstracto para caracterizar el proceso de razonamiento dialéctico unilateral, básicamente asociado a lo que conocemos como argumentación rebatible de carácter dialéctico. El modelo propuesto tiene la capacidad de capturar el comportamiento de la mayoría de los sistemas argumentativos estudiados, aún a pesar de las aparentes diferencias entre estos. La experiencia sugiere que detrás de esas divergencias meramente superficiales existe un estilo de organización del razonamiento afín a todos estos sistemas.

- Se reinterpretó el modelo elaborado anteriormente incorporando nuevos actores al juego diálogico unilateral, para así obtener un modelo capaz de caracterizar a la dialéctica multilateral, es decir, para obtener un marco conceptual que permita estudiar la interacción entre los diversos agentes que componen un sistema multiagente. El modelo abstracto por este medio obtenido puede ser instanciado en una multitud de protocolos, los cuales a su vez permiten gobernar los diversos tipos de interacción entre agentes, tales como la deliberación, la persuación, etc.

1.4.2. Publicaciones previas

Las nociones y conceptos estudiados en la presente tesis fueron desarrollados y refinados a lo largo de un conjunto de publicaciones previas, en las cuales es posible atestiguar la evolución a lo largo del tiempo de las principales intuiciones. Por caso, el primer bosquejo fue esbozado en el artículo *Modelling Negotiation Protocols in a Dialectical Framework* [Stankevicius y García, 1999], donde se abordó la posibilidad de apelar a un modelo dialéctico de argumentación rebatible como marco conceptual para guiar el desarrollo de una negociación entre dos partes. El principal aporte de este trabajo consistió en la definición de un protocolo para gobernar el intercambio de información entre las partes involucradas. El punto de partida en este caso era la caracterización dialéctica para el sistema Simari-Loui presentada en [Simari et al., 1994]. Otro interesante aspecto aquí considerado es la identificación de los principales desafíos que deben ser atendidos al tener en cuenta un dominio de aplicación en concreto, tales como la selección de un adecuado criterio de comparación entre las razones expuestas, el grado de coherencia entre las creencias propias y las del contrincante o incluso los diversos grados de confianza entre las partes involucradas. Por último, también se exploró el abanico de alternativas existentes para afrontar cada uno de estos desafíos, dando a lugar a una familia de protocolos en función de cómo se resuelvan cada uno de ellos.

La línea de investigación por aquel entonces en desarrollo fue sintetizada en el resumen extendido titulado *Could negotiations among agents be regarded as an argumentative process?* [Stankevicius y García, 2000]; aquí se fundamentan las razones que sugieren que la negociación entre agentes inteligentes puede en general ser reinterpretada como si en realidad fuera un mero proceso argumentativo. La intuición reseñada consiste en reconocer el carácter dialógico de toda negociación, puesto que en esencia consiste del intercambio de razones, algunas a favor de la posición propia y otras en desmedro de la del interlocutor de turno. Sin duda bajo esta concepción el paralelo entre la negociación y un proceso argumentativo de base dialéctica resulta un tanto más evidente.

Elaborando sobre la base delineada en los artículos anteriores, el trabajo titulado *A Framework for Multiagent Deliberation Based on Dialectical Argumentation* [Stankevicius y Simari, 2000] extiende el interés inicial por modelar el proceso de negociación a una forma más sutil de interacción entre agentes: *la deliberación*. A tal efecto, se caracterizó formalmente el marco de conceptual en el cual llevar adelante la deliberación, así como el respectivo protocolo de intercambio de información. En este contexto, una deliberación abarca cuatro etapas: en primer lugar la toma de la decisión de deliberar con otro agente, luego el contacto inicial con el contendiente de turno (puesto que la deliberación no debe ser de carácter compulsivo), posteriormente el desarrollo de la deliberación en sí misma y finalmente la incorporación a la base de conocimiento del resultado consensuado. Como conclusión del análisis de estos estadíos intermedios se desprende que contrariamente a lo esperado es el último de éstos el que supone un desafío de consideración, en particular para el contendiente que no logre imponerse en la deliberación, ya que es factible que para honrar el resultado de la misma deba apelar a alguna forma de revisión de creencias.

Al experimentar con el marco conceptual introducido, se concluyó que resultaría más flexible contar con la posibilidad de adecuar su dinámica en función del dominio aplicación en consideración. En consecuencia, esta nueva vertiente de investigación fue reseñada en el resumen extendido titulado *Modeling Multiagent Deliberation from an Abstract Standpoint* [Stankevicius, 2001]. Este trabajo en particular delinea una estrategia tendiente a obtener el modelo abstracto buscado, esencialmente la misma estrategia luego desarrollada a lo largo de esta tesis. Por otra parte, también se analizaron otras propuestas tomadas de la literatura tales como la de R. P. Loui [Loui, 1998] o la de H. Prakken [Prakken, 2001], en las cuales se había comenzado a explorar esta misma idea.

Por último, una aproximación bastante fiel a los contenidos recopilados en la presente tesis fueron resumidos en el artículo *An Abstract Model for the Process of Deliberation within Multiagent Systems* [Stankevicius y Simari, 2001]. Este trabajo pone en práctica la anterior estrategia: a partir de un análsis en profundidad de los principales exponentes de la argumentación rebatible se postula un modelo que factoriza sus principales características, donde el modelo resultante es posteriormente reinterpretado como modelo del proceso de deliberación entre agentes. El modelo por este medio obtenido resulta de carácter abstracto, lo que permite que sea luego instanciado de manera que permita describir fielmente el comportamiento de diversos protocolos de deliberación. Por otra parte, las principales definiciones que constituyen el andamiaje teórico sobre el cual descansa la totalidad del modelo introducido fueron también motivadas y estudiadas a lo largo de este artículo.

1.5. Estructura de la tesis

Para concluir el presente capítulo, a continuación brindaremos un bosquejo de la organización de los contenidos por capítulo, a saber:

Capítulo 1: Introducción y Motivaciones

Este capítulo situa los estudios realizados en la presente tesis en el marco del estado actual de la investigación científica, enfatizando las motivaciones que impulsaron su desarrollo.

Capítulo 2: Sistemas argumentativos

Aquí se presenta un detallado análisis de los principales exponentes de la argumentación rebatible, identificando tanto virtudes como falencias de los formalismos considerados. También se enfatiza cuáles son los aspectos recurrentes que suelen estar presentes en la mayoría de las propuestas consideradas.

Capítulo 3: Sistemas multiagente

Este capítulo introduce las principales nociones de las distintas líneas de investigación existentes en el ámbito de los sistemas multiagente, cotejando y contrastando las diversas propuestas introducidas en la literatura. En particular, se resalta el preponderante rol desempeñado por la interacción entre agentes en este tipo de sistemas.

Capítulo 4: Un Modelo Dialéctico para la Interacción Multiagente

Aquí se plasma paso a paso el principal aporte: el desarrollo de un modelo formal que provea el marco de referencia para la interacción entre los diversos agentes que integran los sistemas multiagente. El modelo propuesto fue obtenido siguiendo la estrategia diseñada a tal efecto, que permite descomponer la meta trazada en objetivos mucho más asequibles.

Capítulo 5: Conclusiones y Resultados Obtenidos

Este capítulo sintetiza las conclusiones y los diversos resultados obtenidos a lo largo del desarrollo de la temática elegida, resumiendo esencialmente las conclusiones y resultados parciales reseñados al final de cada uno de los capítulos anteriores.

Glosario:

Considerando que no existe un consenso generalizado en relación a cómo traducir muchos de los términos más frecuentes en ciencias de la computación, se decidió incorporar como referencia una sencilla tabla conteniendo los términos originales en inglés y los correspondientes vocablos en castellano empleados a lo largo de la presente tesis.

Bibliografía:

En este apartado se recopilan, ordenadas alfabéticamente por autor, las diversas referencias bibliográficas citadas en los distintos capítulos. Cada una se encuentra acompañada de un breve resumen que sintetiza los principales tópicos abordados en el artículo o trabajo en cuestión, reseñando sus aportes más trascendentes.

Capítulo 2

Sistemas Argumentativos

Los *sistemas argumentativos* constituyen una alternativa atractiva para formalizar el tipo de razonamiento denominado *rebatible*. En estos sistemas una conclusión resulta aceptable si es capaz de sobrevivir a un proceso de debate en donde razones a su favor y en su contra son contrastadas. El presente capítulo aborda los principales aspectos de estos sistemas, desde sus comienzos hasta sus más recientes tendencias, reseñando a su vez las propuestas más significativas presentes en la literatura.

2.1. Introducción

Esta sección elabora una introducción a la argumentación rebatible, reseñando sus orígenes y describiendo sus principales componentes. La sección 2.1.1 analiza los primeros acercamientos a la argumentación, mencionando sus aportes más relevantes. Luego, la sección 2.1.2 explora el rol de la dialéctica en el contexto de los sistemas argumentativos, justificando su reciente preponderancia. Por último, la sección 2.1.3 presenta una visión general de los elementos constitutivos de un sistema argumentativo desde una perspectiva abstracta e independiente de las formalizaciones preexistentes.

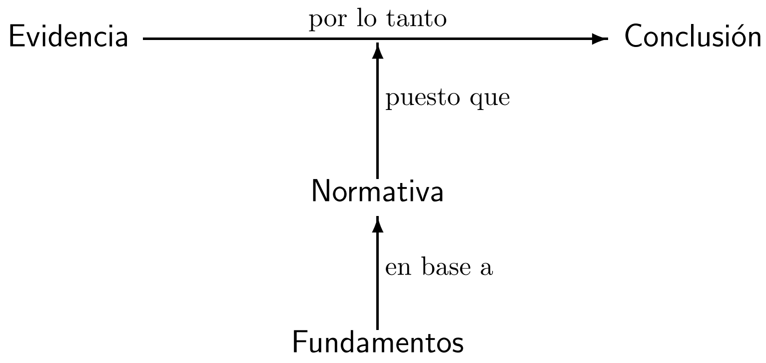


Figura 2.1: Estructura de un argumento según Toulmin

2.1.1. Primeros pasos hacia la argumentación rebatible

Los orígenes de la argumentación rebatible se remontan a la Grecia clásica, donde Platón describe el particular estilo de persuación que Sócrates adoptaba para con sus interlocutores de turno. Por caso, cuando él deseaba acaparar la atención de algún plebeyo solía realizar una serie de preguntas tendientes a explicitar la ignorancia de su par, manifestada como una contradicción evidente en los dichos de este último. Otro estilo de persuación reseñado en los escritos de Platón consistía en un *duelo de discursos*, también conocido como *certamen de palabras*, donde discursos y contradiscursos representando opiniones contrarias o complementarias matizaban el tema en discusión. Aristóteles, fuertemente influenciado por esta tradición en la estructuración del razonamiento, sentó las bases de una *teoría epistémica* inspirada en la *dialéctica conversacional*. En ésta dos contendientes tomaban parte en una argumentación ordenada con el objeto de debatir acerca del estado de una tesis. La semilla para desarrollar un modelo procedural del razonamiento inteligente había sido sembrada.

A pesar del importante avance que representó la propuesta de Aristóteles, el auge del *deduccionismo matemático* (incentivado por otra de las contribuciones de este filósofo, los silogismos), relegó la germinación de esta idea por casi veinticinco siglos. No es hasta mediados de la década del '50 que resurge el interés en el estudio de la argumentación, particularmente como sustrato formal sobre el cual erigir una teoría epistémica. En este sentido, el trabajo llevado a cabo por filósofo Stephen Toulmin en torno al rol de los argumentos [Toulmin, 1958] resultó especialmente influyente. Para Toulmin, un argumento—noción esencial a todo sistema de argumentación rebatible—involucra los siguientes componentes:

Conclusión: La conclusión representa el propósito del argumento. La razón de ser de todo argumento es sustentar su conclusión.

Evidencia: La evidencia es el conjunto de información que en base a una normativa permite inferir la conclusión.

Normativa: La normativa es la razón no demostrativa que permite alcanzar la conclusión como resultado de la observación de la evidencia.

Fundamentos: Los fundamentos representan el respaldo de la normativa y suelen depender del dominio de discurso actual.

La intuición de Toulmin refleja que las conclusiones son alcanzadas en base a la existencia de la evidencia, donde la relación entre conclusión y evidencia se encuentra justificada a partir de la normativa y a su vez la validez de la normativa queda establecida por los fundamentos. Por citar un ejemplo, la evidencia de que cierta ave es un cuervo permite concluir que su color es negro. La normativa empleada en este caso es que todo cuervo es de color negro. Los fundamentos que sustentan esta normativa son las observaciones anteriores de cuervos negros, en número suficiente como para asegurar la validez de la generalización capturada en la normativa. La interacción entre estos componentes se encuentra esquematizada en la figura 2.1.

En este acabado estudio se establece una clara distinción entre dos tipos de argumentos: los *argumentos analíticos* y los *argumentos demostrablemente válidos*. Los argumentos analíticos son aquellos cuya conclusión es consecuencia directa de la unión de los fundamentos con la evidencia. En contraste, en los argumentos demostrablemente válidos la conclusión depende de la aplicación de una normativa, puesto que guardan la forma "evidencia, normativa, *ergo* conclusión". En palabras del propio Toulmin, la diferencia esencial es que cualquier argumento puede ser ajustado al esquema de un argumento demostrablemente válido, simplemente explicitando toda normativa invocada, pero pocos argumentos guardan la forma de los argumentos analíticos. Cabe señalar, no se considera factible el explicitar los fundamentos de un argumento con el objeto único de amoldarse al esquema de los argumentos analíticos; en cierto sentido, los fundamentos se asumen como previamente acordados y no pueden ser cuestionados, razón por la cual carece de sentido explicitarlos.

Sin desmedro del acabado análisis acerca de la estructura y el rol de los argumentos, Toulmin no aborda cómo determinar el *estado epistémico* de éstos; simplemente se limita a identificar los posibles puntos de ataque que presenta cada tipo de argumento, sin formular una metodología que permita establecer su viabilidad. Para encontrar un adecuado tratamiento del siguiente componente de los sistemas de argumentación rebatible (*i.e.*, una teoría que permita determinar qué argumentos deben aceptarse y cuáles no), debemos apelar al trabajo *"Defeat among Arguments"* de Ronald P. Loui [Loui, 1987]. Este autor es en cierto sentido el responsable de introducir la argumentación rebatible al ámbito de la Inteligencia Artificial, hasta ese entonces tópico abordado sólo en los círculos filosóficos.

Loui erige su propuesta en torno a una lógica $\mathcal{L}$ de primer orden, estructurando sus fórmulas mediante una relación metalingüística denominada *regla de inferencia rebatible*, notado en este formalismo como $\Phi \succ\!\!- \Psi$. La semántica asociada a estas reglas señala que ante la falta de información que manifieste lo contrario, Φ constituye una razón valedera para sustentar a Ψ. En este sistema la información se representa a través de un conjunto EK de hechos (los cuales deben ser fórmulas de $\mathcal{L}$) y un conjunto R de reglas de inferencia rebatible. En base a estos conceptos, Loui parafrasea la noción de argumento en términos de un grafo de prueba en el cual la información *fluye* desde sus *premisas* (fuentes) hasta su *conclusión* (desagote). Considerando que el argumento tiene por objeto establecer una conclusión puntual, se impone como restricción adicional que el grafo sea acícliclo, dirigido y que posea un único desagote. En este contexto, cada nodo del grafo denotará una de las fórmulas de $\mathcal{L}$, por lo que dos nodos distintos no podrán denotar la misma fórmula. En términos formales, para que un cierto grafo sea considerado argumento con respecto a una base de conocimiento (EK, R), deben cumplirse que las fuentes del grafo sean elementos de EK y, además, que para cada arco desde un nodo α hacia un nodo β sea tal que, o bien β se pueda obtener clásicamente a partir de $EK \cup \alpha$, o bien exista una regla rebatible en R de la forma $\alpha \succ \beta$.

El autor identifica dos importantes relaciones entre argumentos, las nociones de *contraargumentación* y de *derrota*. Un argumento A se dice que *derrota* a otro argumento B cuando las conclusiones de A y de B son contradictorias entre sí y A resulta preferido a B de base a un cierto criterio de comparación entre argumentos previamente acordado. En contraste, un argumento A se dice que *contraargumenta* a otro argumento B si existe alguna fórmula δ asociada a algún nodo del grafo de inferencia de B, tal que la conclusión de A y δ resultan contradictorios entre sí. Finalmente, en base a estas relaciones surge de manera bastante natural otro de los conceptos preponderantes en todo sistema argumentativo, el de argumento *justificado* o *garantizado*, el cual permite describir la clase de aquellos argumentos que son capaces de hacer que sus conclusiones prevalezcan al ser confrontados con los restantes argumentos existentes. En este sentido, se dice que un argumento A *justifica* a su conclusión si no existe un argumento B, tal que B derrote a A y para cada contraargumento de A, existe a su vez otro argumento que derrota a este último. La semática del sistema resultante se encuentra caracterizada precisamente en base al conjunto de las conclusiones de los argumentos justificados.

Si bien este formalismo manifiesta un conjunto de aspectos controversiales, oportunamente señalados en la literatura, merece ser reconocido como una de las primeras teorías argumentativas que incluye casi la totalidad de los componentes de un sistema argumentativo contemporáneo.

2.1.2. El rol de la dialéctica

Los sistemas argumentativos se han inspirado en sus comienzos en los principios que gobiernan la argumentación entre la gente común. No obstante, los primeros acercamientos formales parecen renegar su origen, centrándose en torno a intuiciones que no capitalizan esta analogía. Por ejemplo, nociones definidas mediante puntos fijos [Dung, 1995, Prakken, 1993, Prakken y Sartor, 1996] o capturadas en oscuras caracterizaciones recursivas [Pollock, 1987, Pollock, 1995, Simari y Loui, 1992] eran preferidas a aquellas que presentaban un cariz disputacional o dialéctico, mucho más afín a sus raíces.

Posteriormente, a medida que la argumentación rebatible alcanzó el reconocimiento dentro de la Inteligencia Artificial como disciplina establecida, los diversos sistemas argumentativos existentes comenzaron a ser reformulados en términos *dialécticos*. Por caso, el sistema Simari-Loui [Simari y Loui, 1992] fue luego dotado de una *semántica de corte dialéctico* [Simari et al., 1994], o el sistema Prator [Prakken y Sartor, 1996], al que se le anexó posteriormente un *procedimiento de prueba dialéctico* [Prakken y Sartor, 1997].

Aún años antes de consensuar esta tendencia, el filósofo Nicholas Rescher ya había reconocido el importante rol que la dialéctica desempeña en la *investigación científica* [Rescher, 1977]. Postuló que los avances en la ciencia se llevaban a cabo a través de un proceso por el cual el proponente de una nueva tesis debía defender su posición ante el cuestionamiento de uno o varios oponentes. Para dimensionar cuan inherente resulta la dialéctica al estudio de la argumentación rebatible es necesario destacar uno de los principales resultados obtenidos por Rescher: la existencia de una equivalencia en forma y estructura, prácticamente un "isomorfismo", entre la *dialéctica multilateral* (*i.e.*, el debate entre proponentes y oponentes) y la *dialéctica unilateral* (*i.e.*, el razonamiento introspectivo). En otras palabras, no será posible defender ante nuestros pares tesis alguna que no supere nuestra propia autocrítica.

Rescher respalda sus resultados definiendo un marco semi-formal de disputación, donde un *proponente* y un *oponente*, supeditados a un *determiner*,[1] siguen un conjunto de reglas estipuladas con el propósito de estructurar la disputación. Estas reglas definen un juego dialógico en el cual se cruzan razones a favor y contra de la tesis sobre la cual se debate. Presentaremos a continuación un resumen de las principales características de este marco de disputación.

Los citados contendientes disponen del siguientes conjunto de movidas básicas:

- *Afirmación categórica*, denotada $!P$, que se debe interpretar como "es el caso que P". Esta movida sólo puede ser realizada por el proponente.

- *Afirmación cauta*, denotada $\dagger P$, que se debe interpretar como "P resulta compatible con lo hasta aquí expuesto". Esta movida sólo puede ser realizada por el oponente.

- *Afirmación condicionada*, denotada P/Q, que se debe interpretar como "cuando es el caso que Q, usualmente también es el caso que P". Esta movida debe estar acompañada por una afirmación cauta o categórica para Q. Puede ser realizada tanto por el proponente como por el oponente.

Las afirmaciones condicionadas no deben ser intepretadas como reglas estrictas (modeladas usualmente mediante la implicación matemática), pues para ser exactos están inspiradas en la noción de regla rebatible. Esta distición se encuentra fundamentada con especial ahínco por Rescher, ya que por aquel entonces la comunidad científica todavía no estaba del todo convencida de la importancia de esta noción.

Una disputación comienza con el proponente proveyendo una afirmación categórica que soporte la tesis a ser debatida. Una afirmación categórica de la forma $!P$ puede a su vez ser cuestionada mediante las siguientes movidas:

- *Refutación cauta*, denotada $\dagger{\sim}P$, que se debe interpretar como "aparentemente no es el caso que P". Esta movida sólo puede ser realizada por el oponente.

[1] este vocablo carece de una traducción exacta al castellano, si bien puede ser interpretado como "juez" o "árbitro".

- *Refutación condicionada*, denotada $(\sim P/Q)\ \&\ \dagger Q$ para algún Q, que se debe interpretar como "cuando es el caso que Q, usualmente también es el caso que $\sim P$, y además Q resulta compatible con lo hasta aquí expuesto". Esta movida sólo puede ser realizada por el oponente.

Las refutación cautas tienen por objeto poner en duda la veracidad de una afirmación categórica e invitan al destinatario de la refutación a proveer una justificación explícita para la tesis cuestionada. En particular, las refutaciones condicionadas proveen de una razón para controvertir una afirmación categórica anterior.

Una afirmación cauta de la forma $\dagger P$ puede ser posteriormente cuestionada mediante las siguientes movidas:

- *Contraafirmación categórica*, denotada $!\sim P$, que se debe interpretar como "no es el caso que P". Esta movida sólo puede ser realizada por el proponente.

- *Contraafirmación condicionada*, denotada $(\sim P/Q)\ \&\ !Q$ para algún Q, que se debe interpretar como "cuando es el caso que Q, usualmente también es el caso que $\sim P$, y en estos momentos es el caso que Q". Esta movida sólo puede ser realizada por el proponente.

En el caso de que la afirmación cauta fuera en referencia a la negación de una tesis (*i.e.*, $\dagger\sim P$), es posible responder con una contra-afirmación categórica de la forma $!\sim\sim P$ (equivalentemente $!P$) o con una contra-afirmación condicionada $(P/Q)\ \&\ !Q$, para algún Q. Como puede observarse, las contra-afirmaciones categórica y condicionadas para el caso de afirmaciones cautas sobre tesis negadas coinciden con las afirmaciones categóricas y condicionadas respectivamente.

Por último, resta considerar las alternativas disponibles para contrarrestar las afirmaciones o refutaciones cautas. Sea $(P/Q)\ \&\ Q$ una movida donde Q es una afirmación no condicionada de la forma $!Q$ (si fue provista por el proponente) o de la forma $\dagger Q$ (si fue provista por el oponente). Esta movida puede ser cuestionada atacando la parte no condicionada Q a través de las jugadas consideradas anteriormente, o bien atacando la parte condicionada P/Q mediante las siguientes movidas:

	proponente	oponente	
(1)	!vuelan	†~vuelan	(2)
(3)	(vuelan/aves) & !aves	~vuelan/(aves & pingüinos) & †(aves & pingüinos)	(4)

Figura 2.2: Una disputación de acuerdo al modelo de Rescher

- *Distinción débil*, denotada $(\sim P/(Q \,\&\, R)) \,\&\, \dagger(Q \,\&\, R)$, que se debe interpretar como "cuando es el caso que Q en conjunción con R, usualmente también es el caso que $\sim P$, y por otra parte Q en conjunción con R resulta compatible con lo hasta aquí expuesto". Esta movida sólo puede ser realizada por el oponente.

- *Distinción fuerte*, denotada $(\sim P/(Q \,\&\, R)) \,\&\, !(Q \,\&\, R)$, que se debe interpretar como "cuando es el caso que Q en conjunción con R, usualmente también es el caso que $\sim P$ y en estos momentos es el caso que Q en conjunción con R". Esta movida sólo puede ser realizada por el proponente.

Considerando que Rescher asume como simplificación la correctitud de las afirmaciones o refutaciones condicionadas, no es factible cuestionar a una movida conteniendo el condicionamiento P/Q argumentando directamente que $\sim(P/Q)$.

Finalmente, es posible derivar los ataques a las movidas compuestas (*e.g.*, afirmaciones condicionadas, refutaciones condicionadas, etc.), mediante los cuestionamientos a las movidas elementales (*e.g.*, afirmaciones categóricas, refutaciones cautas, etc.), que ya han sido consideradas. Básicamente, una movida compuesta es atacada detractando a alguno de sus componentes elementales. Por ejemplo, para atacar una refutación condicionada de la forma $(\sim P/Q) \,\&\, \dagger Q$ se debe cuestionar la afirmación cauta $\dagger Q$ mediante una contra-afirmación categórica $!\sim Q$ o una contra-afirmación condicionada $(\sim Q/R) \,\&\, !R$, o bien cuestionar la afirmación condicionada $\sim P/Q$ mediante la distinción fuerte $(P/(Q \,\&\, S)) \,\&\, !(Q \,\&\, S)$.

Ejemplo 2.1. El siguiente ejemplo versa sobre la habilidad de volar con la que cuentan las aves. Asumiendo como válidas las siguientes afirmaciones,

1. las aves usualmente vuelan,

2. los pingüinos usualmente no vuelan, y

3. los pingüinos son un tipo de ave.

una disputa sobre la capacidad de volar de los pingüinos puede adoptar la siguiente forma: el proponente sostiene que los pingüinos vuelan; acto seguido el oponente cuestiona el por qué; el proponente responde que las aves usualmente vuelan y es el caso que es posible considerar a los pingüinos como aves; el oponente señala que los pingüinos son un tipo de ave que no vuela, observación que el fuerza al proponente a conceder la disputa. La figura 2.2 presenta una reconstrucción de esta disputa en términos del modelo de Rescher.

En síntesis, una disputa consiste de una sucesión de movidas que respetan las reglas estipuladas. Como ha sido señalado en la literatura [Loui, 1998], es posible derivar versiones simplificadas de estas reglas ignorando la distinción entre movimientos categóricos y cautos, puesto que las movidas categóricas están disponibles sólo para el proponente y las movidas cautas sólo para el oponente. Las concesiones se llevan a cabo de manera tácita: una afirmación que no es cuestionada se asumen como concedida. Si al término de la disputa no se consensúa el estado de la tesis debatida, el *determiner* debe analizar la evidencia expuesta y dirimir la situación.

Rescher aborda incluso el controversial tópico de asegurar la *terminación* de las disputas. En este sentido, restringe su atención sólo a disputas de carácter progresivo. Una disputa se dice de carácter progresivo si cada movida realizada aporta nueva evidencia. Por caso, no se acepta que un participante reitere una movida anterior, pues esta situación no aportará nueva evidencia y de hecho puede conducir a una eventual disputación infinita. En el contexto de los sistemas argumentativos, se han arribado a conclusiones similares al considerar esta misma problemática.

Otro aspecto interesante que se explora mediante este esquema de disputación es la noción de *"peso de la prueba"*. Por "peso de la prueba" se entiende el esfuerzo extra a ser realizado por alguno de los participantes en función de cierto aspecto.

Por ejemplo, en los sistemas legales de la mayoría de las democracias el "peso de la prueba" recae en la parte acusadora, ya que se debe respetar el principio de que una persona es inocente hasta que se demuestre lo contrario. En este modelo de disputación se pueden identificar dos tipos de "peso de la prueba". El de mayor exigencia recae en el proponente, ya que es él quien desea justificar la nueva pieza de conocimiento. Este "peso de la prueba" se manifiesta en la obligación del proponente de realizar movidas categóricas (*i.e.*, con fundamento), mientras que el oponente puede limitarse a objetar movidas anteriores de manera cauta. Otro "peso de la prueba", un tanto más evidente, es aquél que recae alternativamente en el proponente y el oponente, estableciendo quien está obligado a realizar la siguiente movida.

Las nociones abordadas por Rescher constituyen hoy en día aspectos primordiales no sólo en el ámbito de la argumentación rebatible, sino de la Inteligencia Artificial en general. Tal como se manifiesta en [Chesñevar et al., 2000]:

> *"...Rescher predijo la noción de regla rebatible (caracterizada como una afirmación condicionada), que fue reinventada por la comunidad de razonamiento no monótono. Incluso reconoció el carácter ampliativo del razonamiento no demostrativo y acotado en recursos, en comparación a los paradigmas no ampliativos. Su trabajo abordó áreas en las cuales el razonamiento rebatible iba a estar involucrado por una década, tales como la búsqueda, la no monotonía y los criterios de derrota."*

2.1.3. Características generales

Al estudiar los diversos sistemas argumentativos propuestos en la literatura se pueden apreciar un conjunto de características que se reiteran. Estos aspectos en común constituyen el núcleo de todo sistema argumentativo. Por caso, en el contexto de su artículo de discusión del estado del arte en la argumentación rebatible, H. Prakken y G. Vreeswijk [Prakken y Vreeswijk, 2002] analizan estas características generales antes de abordar en concreto los diversos sistemas de argumentación. Los autores distinguen a los siguientes componentes como esenciales:

1. La lógica subyacente.

2. La definición de argumento.

3. La noción de conflicto entre argumentos.

4. La noción de derrota entre argumentos.

5. La especificación del estado final de los argumentos.

Seguidamente analizaremos el rol de cada uno de estos componentes en el contexto de los sistemas de argumentación rebatible.

En la actualidad todo sistema argumentativo incluye la caracterización de una lógica cuya respectiva relación de consecuencia inducirá la noción de argumento deseada. La descripción de esta lógica puede no ser exhaustiva, quedando en algunas oportunidades incluso completamente sin especificar. Paradójicamente, la mayoría de estas lógicas suelen ser monótonas, aún al estar formando parte de sistemas de razonamiento no monótono. La intuición del caso es que la incorporación de información permite formular nuevos argumentos, en vez de invalidar antiguas deducciones. Los argumentos usualmente sustentan conclusiones apelando al concepto de demostración o prueba en esta lógica subyacente. En consecuencia, un argumento puede adoptar diversas manifestaciones tales como secuencias de deducción, árboles de prueba, conjuntos de suposiciones, etc. No obstante, no toda deducción en la lógica subyacente resulta directamente asociada a un argumento, en realidad tienden a existir condicionamientos adicionales (*e.g.*, consistencia de las premisas empleadas, minimalidad de la derivación, etc.).

Dentro de un sistema de argumentación rebatible es factible formular argumentos en desacuerdo, pues resulta razonable suponer que la lógica subyacente admite la derivación de metas en conflicto. Esto constituye el tercer elemento común a todo sistema argumentativo, la caracterización formal de una relación entre argumentos que denote la existencia de un *conflicto* entre éstos. La relación de conflicto también se conoce en la literatura bajo otras denominaciones tales como *ataque* o *contra-*

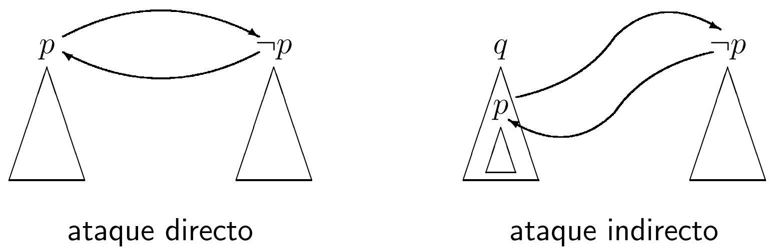


Figura 2.3: Ataque directo e indirecto

argumentación. El conflicto entre argumentos más evidente es aquel producto de sustentar conclusiones complementarias, por caso, un argumento para $vuela(tweety)$, basado en que *tweety* es un ave, y otro argumento para $\neg vuela(tweety)$, basado en que *tweety* es un pingüino. Este tipo de conflicto se conoce como *ataque por rebatimiento*. Otro conflicto posible se presenta entre un argumento basado en un conjunto de suposiciones y otro argumento que evidencia a alguna de estas suposiciones como incorrecta, por caso, un argumento para $\neg llevar(paragua)$, basado en la suposición $\neg llueve(parque)$, y otro argumento para $llueve(parque)$. Este tipo de conflicto se denomina *ataque a las suposiciones*. Existe un tercer tipo de conflicto, introducido por el filósofo J. Pollock, que se basa en cuestionar la utilización de una cierta regla de inferencia en la formación de un argumento. El ejemplo más conocido involucra un argumento a favor de que el color de un cubo es rojo, basado en que parece ser de color rojo, y otro argumento señalando que el cubo se encuentra iluminado por un haz de luz roja, por lo tanto el parecer de color rojo no asegura que efectivamente sea de ese color. Este tipo de conflicto entre argumentos se conoce como *ataque por socavamiento*.

Para cada tipo de conflicto entre argumentos es posible identificar dos variantes, una *directa* y otra *indirecta*, en función del aspecto que uno de los argumentos cuestione del otro. El ataque se denomina directo cuando se cuestiona la cabeza del argumento (*i.e.*, la conclusión sustentada o la última regla empleada) e indirecto cuando se cuestiona el cuerpo del argumento (*i.e.*, una conclusión o regla intermedia).

En este sentido, la figura 2.3 esquematiza las variantes mencionadas a través de un ejemplo de ataque por rebatimiento.

Teniendo en cuenta que la relación de conflicto sólo captura el desacuerdo entre argumentos, es evidente que esta relación no será capaz de distinguir los ataques satisfactorios de aquellos que no lo sean. La relación de *derrota*, cuarto elemento común a todo sistema argumentativo, es un refinamiento de la relación de conflicto, que en particular denota sólo a aquellos cuestionamientos satisfactorios. En otras palabras, la derrota entre argumentos describe bajo qué circunstancias un argumento está en condiciones de negar la fuerza conclusiva del argumento objetado, entendiendo por fuerza conclusiva a la capacidad de un argumento para sustentar a su conclusión. La relación de derrota también se la conoce en la literatura bajo otras denominaciones tales como *ataque* o *interferencia*. Por lo general, se suele adoptar algún criterio de comparación entre argumentos para determinar la direccionalidad de la derrota en aquellos casos de conflicto recíproco. Esto se debe a que la relación de conflicto entre argumentos tiende a ser simétrica, pero es preferible que la relación de derrota no lo sea. Naturalmente, también es posible extrapolar las nociones de conflicto directo e indirecto al marco de la relación de derrota.

Por otra parte, debemos tener en cuenta que la derrota no sanciona cuál será el *estado final* de un argumento, simplemente expresa la fuerza conclusiva relativa de los argumentos involucrados. El estado final de un argumento no depende de una derrota en particular, sino de la interacción concurrente de todas ellas. En particular, la especificación del estado final de los argumentos constituye el último elemento común a todo sistema argumentativo. Debido al importante rol que la especificación formal del estado final de los argumentos desempeña en todo sistema argumentativo se han ensayado una gran cantidad de alternativas (definiciones de punto fijo, caracterizaciones recursivas, descripciones procedurales, etc.). Si bien es difícil consensuar cuáles pueden ser los estados posibles de un argumento, se suelen identificar por lo menos dos categorías: los *argumentos garantizados* y aquellos que no lo están. En la literatura se refiere a los argumentos garantizados bajo otras denominaciones tales como *justificados*, *activos* o incluso *no derrotados*. El conjunto de los argumentos garantizados suele estar en íntima relación con el conjunto de conclusiones sancionadas por el sistema.

Para concluir, también debemos señalar la existencia de escenarios conflictivos que complican o hacen imposible la determinación del estado último de los argumentos. Un caso típico es la existencia de más de una asignación de estados a los argumentos razonables en un contexto dado. Como la semántica del mismo sistema depende de poder determinar el estado último de los argumentos, se han ensayado diversas alternativas para proveer un adecuado tratamiento de estas situaciones problemáticas. Por caso, se ha propuesto considerar todas las asignaciones de estado posibles, determinando el estado final de un argumento en función de los estados asociados en cada una de las posibles asignaciones. Esta propuesta, si bien soluciona el inconveniente anterior, incrementa notablemente el costo computacional asociado a la determinación del estado final, ya que todas las asignaciones posibles deben ser tenidas en cuenta antes de poder descubrir el estado final de un argumento.

2.2. Principales sistemas argumentativos

Esta sección desarrolla un análisis de los sistemas argumentativos más representativos, resumiendo sus conceptos primordiales y destacando las virtudes o falencias identificadas en cada caso.

2.2.1. Marcos Agumentativos

Phan M. Dung desarrolla un distintivo acercamiento a la argumentación rebatible ignorando la estructura interna de los argumentos, hasta entonces incuestionable protagonista de todo sistema argumentativo. El énfasis en este formalismo se traslada a la determinación del estado final de los argumentos. Esta particular concepción presenta un alto grado de abstracción, incrementando en consecuencia su capacidad para modelar situaciones.

El autor desarrolla sobre una serie de artículos [Dung, 1993a, Dung, 1993b, Dung, 1995] a la noción de *marco argumentativo*, su modelo para capturar los mecanismos que gobiernan la argumentación cotidiana, con el objeto último de obtener una teoría lo suficientemente expresiva como para atacar la problemática del razonamiento de sentido común. Es en virtud de este original punto de vista que los conceptos introducidos por el autor fueron bien recibidos por la comunidad academica, resultando altamente influyentes en los desarrollos posteriores.

Todo marco argumentativo pivota en torno a un *conjunto de argumentos* de estructura y origen no relevante, y la *relación de conflicto* entre ellos. Diversas nociones de consecuencia son posteriormente introducidas para dotar de significado a los marcos argumentativos. Estas nociones son caracterizadas mediante conjuntos de argumentos que satisfacen ciertas propiedades, denominados *extensiones*. En lo que resta de esta sección, comentaremos las principales definiciones que integran el formalismo ideado por Dung.

Definición 2.1. **[Dung, 1995]**
(marco argumentativo)
Un *marco argumentativo* es un par $AF = \langle Args, attacks \rangle$, donde $Args$ es un conjunto de argumentos y $attacks$ es una relación binaria sobre $Args$. ■

Aquí interpretaremos $(A, B) \in attacks$ como que el argumento A ataca al argumento B. Por otra parte, diremos que un conjunto de argumentos S ataca a un argumento A, si A es atacado por algún argumento contenido en S. Finalmente, con el objeto de no sobrecargar innecesariamente la notación asumiremos la existencia en todo momento, salvo mención explícita de lo contrario, de un marco argumentativo arbitrario pero fijo $\langle Args, attacks \rangle$.

Definición 2.2. **[Dung, 1995]**
(conjunto libre de conflictos)
Un conjunto de argumentos S se dice *libre de conflictos* si todo par de argumentos $A, B \in S$, verifica que A no ataca a B. ■

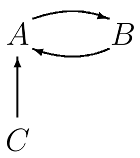


Figura 2.4: Representación gráfica de un marco argumentativo

Un marco argumentativo puede ser representado pictóricamente mediante un grafo, donde los nodos representan a los argumentos y los arcos denotan la relación de derrota (asumiendo que un arco de A a B denota que B es atacado por A). Por ejemplo, la figura 2.4 denota al siguiente marco argumentativo:

$$AF = \langle \{A, B, C\}, \{(A, B), (B, A), (C, A)\} \rangle$$

Dung postula que un agente racional considerará aceptable a un cierto argumento si empleando su conocimiento actual, representado a través de un conjunto de argumentos, resulta capaz de defender al argumento en cuestión contra todo ataque posible. De esta tesitura se desprende una noción central a los marcos argumentativos:

Definición 2.3. **[Dung, 1995]**
(argumento aceptable)
Un argumento $A \in Args$ se dice *aceptable* con respecto a un conjunto de argumentos S si, y sólo si, para cada argumento $B \in Args$ se cumple que cuando B ataca a A, entonces B es atacado por S. ■

Ejemplo 2.2. En el marco del marco argumentativo denotado por la figura 2.4, $\{A\}$ es un conjunto libre de conflictos pero $\{A, B\}$ no lo es; B es aceptable con respecto a $\{A\}$ pero A no es aceptable con respecto a $\{B\}$.

Esta definición provee un sencillo criterio para determinar si un conjunto de argumentos resulta "razonable" o "admisible".

Definición 2.4. **[Dung, 1995]**
(conjunto admisible)

Un conjunto de argumentos libre de conflictos S se dice *admisible* si, y sólo si, cada argumento $A \in S$ es aceptable con respecto a S. ■

A partir de esta idea puede definirse una semántica crédula, la cual por construcción siempre existirá.

Definición 2.5. **[Dung, 1995]**
(extensión preferida)
Denominaremos *extensión preferida* de un marco argumentativo AF a todo conjunto admisible maximal (con respecto a la inclusión de conjuntos) de AF. ■

Otra semántica que resulta útil para modelar sistemas que admiten múltiples extensiones (*e.g.*, lógica default [Reiter, 1980], la lógica autoepistémica [Moore, 1985], etc.) es la basada en la noción de modelos estables, adaptada al contexto de los marcos argumentativos en la siguiente definición:

Definición 2.6. **[Dung, 1995]**
(extensión estable)
Denominaremos *extensión estable* de un marco argumentativo AF a todo conjunto libre de conflictos S tal que ataca a todo argumento no contenido en S. ■

Esta definición puede ser reinterpretada como una ecuación de punto fijo de apariencia más familiar, entendiendo que un conjunto de argumentos S será una extensión estable si, y sólo si, $S = \{A \mid A \text{ no es atacado por } S\}$.

Ejemplo 2.3. Continuando con el ejemplo 2.2, $\{B\}$ es un conjunto admisible pero $\{A\}$ no lo es; el conjunto $\{A, C\}$ constituye tanto una extensión preferida como una extensión estable.

La contrapartida de la semántica preferida, de carácter crédulo, es la semántica escéptica conocida como semántica fija. Esta semántica está definida en términos de una función característica que involucra a la noción de aceptabilidad.

Definición 2.7. **[Dung, 1995]**
(función característica)
La *función característica* de un marco argumentativo AF, denotada F_{AF}, es una función $F_{AF} : \mathcal{P}(Args) \longrightarrow \mathcal{P}(Args)$ tal que para todo conjunto S de argumentos se verifica que $F_{AF}(S) = \{A \mid A \text{ es aceptable con respecto a } S\}$. ■

Entre las diversas propiedades que pueden ser demostradas acerca de esta función característica, cabe destacarse en particular el hecho de que se trata de una función monótona con respecto a la inclusión de conjuntos, garantizando en consecuencia la existencia de su menor punto fijo.

Definición 2.8. **[Dung, 1995]**
(extensión fija)
Denominaremos *extensión fija* de un marco argumentativo AF al menor punto fijo de la función característica F_{AF}. ■

Por último, una cuarta semántica derivada a partir de la semántica fija provee el puente entre las semánticas crédula y escéptica:

Definición 2.9. **[Dung, 1995]**
(extensión completa)
Un conjunto admisible de argumentos S se denomina *extensión completa* si, y sólo si, cuando un cierto argumento es aceptable con respecto a S, entonces ese argumento ya pertenece a S. ■

Contando con la definición de la función característica F_{AF}, es posible ensayar una formulación alternativa para el principio de aceptabilidad, ya que un conjunto de argumentos libre de conflictos S será admisible si se verifica que $S \subseteq F_{AF}(S)$. Como puede apreciarse, la semántica completa constituye la situación límite de la definición anterior, pues un conjunto admisible de argumentos será una extensión completa si $S = F_{AF}(S)$. Esta interrelación entre las diversas semánticas consideradas se debe al hecho de que se encuentras definidas a partir de un mismo punto de partida (*i.e.*, la noción de aceptabilidad).

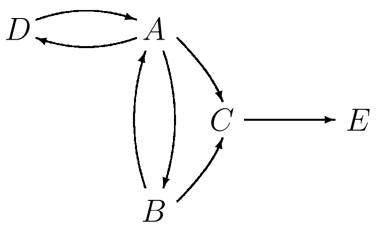


Figura 2.5: Grafo asociado al marco argumentativo del ejemplo 2.5

ciclo de longitud par ciclo de longitud impar

Figura 2.6: Tipos de argumentación circular en un marco argumentativo

Ejemplo 2.4. Continuando con el ejemplo 2.2, el conjunto $\{A, C\}$ constituye tanto una extensión fija como una extensión completa.

En los ejemplos anteriores las diversas semánticas consideradas contaban con una única extensión. También resulta factible que un marco argumentativo induzca múltiples extensiones, tal como se ilustra en el siguiente ejemplo:

Ejemplo 2.5. En el contexto del siguiente marco argumentativo,

$$\langle\{A, B, C, D, E\}, \{(A, B), (A, C), (A, D), (B, A), (B, C), (C, E), (D, A)\}\rangle$$

representado gráficamente en la figura 2.5, las extensiones preferidas son $\{B, D, E\}$ y $\{A, E\}$, las extensiones estables coinciden con las extensiones preferidas, la única extensión fija es $\emptyset$ y las extensiones completas son $\emptyset$, $\{B, D, E\}$ y $\{A, E\}$.

Tal como lo sugiere los ejemplos anteriores, es posible formular marcos argumentativos que presenten una o más extensiones bajo las diversas semánticas introducidas. Resta considerar un tercer escenario más controversial, donde la semántica resulta indefinida precisamente debido a la inexistencia de extensiones. En este sentido, resulta instructivo considerar los casos de *argumentación circular* ejemplificados en la figura 2.6.

Para el caso de los ciclos que contienen una cantidad par de argumentos, el resultado es el esperado: las semánticas preferida, estable y completa razonan crédulamente sancionando dos extensiones, $\{A, C\}$ y $\{B, D\}$, y la semántica fija no toma partido en el ciclo argumentativo sancionando como extensión a $\emptyset$. Desafortunadamente, el resultado para el caso de los ciclos que contienen una cantidad impar de argumentos es problemático: la semántica estable—una de las más expresivas—se indefine y $\emptyset$ es la única extensión preferida, fija y completa. Sin embargo, cabe acotar que la argumentación circular constituye un tipo de argumentación inválida conocido como *argumentación falaz*, donde la dificultad encontrada a la hora de formular una interpretación apropiada para estos casos es precisamente una característica común a las diversas falacias argumentativas.[2]

Esta indefinición de la semántica, manifestada como la ausencia de extensiones, constituye un ubicuo inconveniente enfrentado por las teorías que modelan el razonamiento no monótono. Consciente de esta situación, Dung a su vez exploró aquellos marcos argumentativos que al satisfacer ciertas propiedades aseguran la existencia de cada una de las diversas semánticas. De este análisis se desprenden las siguientes conclusiones:[3]

- Para cada conjunto admisible de argumentos existe una extensión preferida que lo contiene.
- Todo marco argumentativo posee al menos una extensión preferida.
- Toda extensión preferida es una extensión completa, pero no viceversa.
- Toda extensión estable es una extensión preferida, pero no viceversa.

[2][Martínez y García, 1999] presenta un detallado análisis del impacto de las diversas falacias en los marcos argumentativos.

[3]estos resultados se encuentran formalmente fundamentados en [Dung, 1995].

- La extensión fija es la menor (con respecto a la inclusión de conjuntos) extensión completa.

A pesar de los cuestionamientos semánticos observados, Dung parafrasea sin inconvenientes a la mayoría de los formalismos para modelar el razonamiento inteligente (*e.g.*, la lógica default [Reiter, 1980], la programación en lógica [Lloyd, 1987], el sistema argumentativo de J. Pollock [Pollock, 1987]), en términos de marcos argumentativos. De hecho, reconstruye a la lógica default de manera tan fiel que incluso la indefinición de su semántica (*i.e.*, la no existencia de extensiones), es modelada mediante la indefinición en la semántica del marco argumentativo asociado.

Las ideas desarrolladas por Dung evolucionaron posteriormente bajo la forma de un sistema argumentativo abstracto [Bondarenko et al., 1997]. Este formalismo constituye una generalización del sistema Theorist desarrollado por D. Poole [Poole, 1988]. Se asemeja a un marco argumentativo en donde los argumentos representan los conjuntos de hipótesis que permiten derivar una cierta conclusión y la noción de ataque se define a partir de la derivación del complemento de alguna hipótesis. Qué se entiende por una hipótesis y cómo se obtiene su complemento son los aspectos sin especificar que deben ser instanciados de acuerdo al comportamiento que se desee obtener.

2.2.2. El aporte de J. Pollock y su arquitectura Oscar

La interacción entre las disciplinas científicas que adoptan metodologías dispares suele dar a lugar a novedosas contribuciones. La teoría epistémica desarrollada en los círculos filosóficos por John Pollock constituye un claro ejemplo de esta situación, pues este formalismo representó un sustancial aporte al campo de la Inteligencia Artificial. La propuesta de Pollock fue refinada a lo largo de quince años antes de ser descubierta por la comunidad de razonamiento no monótono, principalmente debido a la trascendencia que obtuvo su artículo titulado *"Defeasible Reasoning"* [Pollock, 1987]. Pollock en cierto sentido es el mentor del empleo de sistemas argumentativos para razonar rebatiblemente, ya que este trabajo postula modelar este tipo de razonamiento a través de un sistema basado en argumentos. Cabe indicar que Pollock también se benefició de la interacción con la computación, pues en la implementación de su teoría argumentativa, conocida como OSCAR,[4] encontró un mecanismo de prueba y refinamiento para sus ideas.

Este formalismo evolucionó a lo largo del tiempo [Pollock, 1987, Pollock, 1992, Pollock, 1994, Pollock, 1995], atendiendo a las objeciones que el mismo Pollock descubría como producto de la retroalimentación provista por su implementación. Este filósofo centró su atención en modelar fielmente diversas paradojas del razonamiento tales como la paradoja de la lotería o la paradoja del prefacio, las cuales no habían recibido un adecuado tratamiento en los restantes sistemas de razonamiento [Pollock, 1994]. El refinamiento de los resultados obtenidos para estos casos de prueba guiaron las modificaciones introducidas en la sucesivas revisiones del sistema. En particular, a continuación abordaremos los principales conceptos que integran su última manifestación [Pollock, 1995].

Pollock sostiene que el razonamiento inteligente se lleva a cabo construyendo *argumentos*, donde por argumento se entiende un conjunto estructurado de *razones*. Las razones a su vez pueden ser *conclusivas* o *prima facie*. Una razón es conclusiva cuando implica lógicamente a su conclusión, siendo prima facie en caso contrario; esto es, las razones prima facie denotan razones rebatibles. En este sistema las razones se representan como un par ordenado $\langle \Gamma, p \rangle$, donde Γ es un conjunto de premisas y p es una conclusión.

Una razón prima facie puede ser cuestionada a través de un *derrotador*. Pollock manifiesta que el tipo de derrota entre razones más conocido es el *rebatimiento*, donde una razón ataca a otra al sustentar el complemento de su conclusión. En este contexto, se denota con el símbolo '$\sim$' a la negación clásica, y mediante el símbolo '$\neg$' al complemento en términos de ésta, donde $\neg\varphi = \ulcorner\vartheta\urcorner$, en el caso que φ guarde la forma $\ulcorner{\sim}\vartheta\urcorner$ para algún ϑ, y $\neg\varphi = \ulcorner{\sim}\varphi\urcorner$, en el caso contrario. El autor

[4]en realidad OSCAR constituye una arquitectura para modelar agentes racionales que adopta como maquinaria de inferencia la teoría argumentativa creada por Pollock.

relega los detalles que considera irrelevantes apelando al operador $\lceil \cdot \rceil$ para denotar la codificación de una dada sentencia en el lenguaje objeto del sistema. De esta forma el rebatimiento entre razones se formaliza de la siguiente manera:

Definición 2.10. **[Pollock, 1995]**
(rebatimiento)
Sea $\langle \Gamma, p \rangle$ una razón prima facie. Diremos que una razón $\langle \Lambda, q \rangle$ es un *derrotador por rebatimiento* de $\langle \Gamma, p \rangle$ si, y sólo si, $q = \lceil \neg p \rceil$. ■

Este sistema incluye como aspecto novel un segundo tipo de ataque entre razones denominado *socavamiento*. El ataque por socavamiento no ataca a la conclusión de una razón, sino que cuestiona la conección de las premisas con la conclusión. Pollock sostiene que este tipo de ataque es el más importante en el contexto de un sistema argumentativo. Este característica distingue a esta propuesta de las restantes, ya que si bien existen otros sistemas que incluyen alguna forma de ataque por socavamiento (*e.g.*, el sistema Prator), éstos sólo comparten la denominación pero no el espíritu de este tipo de ataque. Denotando como $p \gg q$ al hecho que es razonable concluir q en base a p, y como $\Pi\Gamma$ a la conjunción de los elementos de Γ, el ataque por socavamiento puede caracterizarse como:

Definición 2.11. **[Pollock, 1995]**
(socavamiento)
Sea $\langle \Gamma, p \rangle$ una razón prima facie. Diremos que una razón $\langle \Lambda, q \rangle$ es un *derrotador por socavamiento* de $\langle \Gamma, p \rangle$ si, y sólo si, $q = \lceil \sim(\Pi\Gamma \gg p) \rceil$. ■

A manera de simplificación, denotaremos a la expresión $\lceil \sim(p \gg q) \rceil$ como $\lceil p \otimes q \rceil$. El ejemplo paradigmático de la derrota por socavamiento señala que percibir un objeto como de color rojo es razón prima facie para concluir que efectivamente es de color rojo, pero apreciar que este objeto está iluminado por luz de color roja socava a la razón prima facie anterior, pues no es razonable seguir afirmando sobre esa base que el objeto es efectivamente de color rojo. Pollock postula que el socavamiento modela fielmente la falibilidad de la percepción humana, situación no representable a través del rebatimiento de razones.

Esta teoría epistémica postula que todo razonamiento se retrotrae a un conjunto inicial de premisas, denominado $\mathsf{inicial}$, que se asume como previamente especificado. Las eventuales conclusiones se elaboran a partir de este conjunto mediante la aplicación de diversos patrones de inferencia. En este contexto, los argumentos son estructuras que mantienen piezas de razonamiento. Si bien la mayoría de los sistemas argumentativos sólo focalizan su atención en los argumentos elaborados a partir de una cierta base de conocimiento, el sistema argumentativo ideado por Pollock identifica una segunda clase de argumento la cual permite expresar un razonamiento de tipo suposicional. Pollock manifiesta que prescindiendo del razonamiento suposicional será imposible arribar a aquellas conclusiones *a priori* independendientes del conjunto $\mathsf{inicial}$, tales como $p \vee {\sim}p$ ó $(p \wedge q) \rightarrow q$. Los argumentos tradicionales se denominan *argumentos lineales* (ya que codifican un razonamiento de tipo lineal), y los nuevos argumentos se denominan *argumentos suposicionales*.

En este sistema, la noción de argumento es asociada con secuencias de razones relacionadas entre sí mediante la aplicación de un determinado conjunto de patrones de inferencia, donde el argumento sanciona al conjunto de conclusiones sustentadas por esta secuencia de razones. Pollock estructura los diversos argumentos construíbles a partir de $\mathsf{inicial}$ en un grafo de derivación que abarca a las diversas secuencias de inferencias. Cada nodo del grafo de inferencia representa una razón, y los enlaces entre estos nodos representan una inferencia realizada siguiendo el conjunto de patrones disponibles. Bajo esta concepción, las reglas de inferencia pueden interpretarse como una especificación de las alternativas posibles para extender un dado grafo de inferencia. Pollock considera a las siguientes reglas de inferencia, en esta ocasión expresadas como criterios de expansión del grafo de inferencia:

Premisa: Si $p \in \mathsf{inicial}$ y $\mathcal{G}$ es un grafo de inferencia, entonces para cualquier conjunto finito de suposiciones X es posible crear un nuevo grafo de inferencia agregando en $\mathcal{G}$ un nodo sin ancestros conteniendo la razón $\langle X, p \rangle$.

Suposición: Si $\mathcal{G}$ es un grafo de inferencia y X es un conjunto finito de suposiciones, entonces es posible crear un nuevo grafo de inferencia agregando en $\mathcal{G}$ un nodo sin ancestros conteniendo la razón $\langle X, p\rangle$, para algún $p \in X$.

Razonamiento: Si $\mathcal{G}$ es un grafo de inferencia conteniendo los nodos $\alpha_1, \ldots, \alpha_n$, asociados a las razones $\langle X, p_1\rangle, \ldots, \langle X, p_n\rangle$, y $\langle\{p_1, \ldots, p_n\}, q\rangle$ constituye una razón conclusiva o prima facie, entonces es posible crear un nuevo grafo de inferencia agregando en $\mathcal{G}$ un nodo conteniendo la razón $\langle\{p_1, \ldots, p_n\}, q\rangle$, con los nodos $\alpha_1, \ldots, \alpha_n$ como sus únicos ancestros.

Condicionalización: Si $\mathcal{G}$ es un grafo de inferencia conteniendo un nodo α, asociado a la razón $\langle X \cup \{p\}, q\rangle$, entonces es posible crear un nuevo grafo de inferencia agregando en $\mathcal{G}$ un nodo conteniendo la razón $\langle X, p \rightarrow q\rangle$, con el nodo α como su único ancestro.

Dilema: Si $\mathcal{G}$ es un grafo de inferencia conteniendo los nodos α, β y γ, asociados a las razones $\langle X, (p \vee q)\rangle$, $\langle X \cup \{p\}, r\rangle$ y $\langle X \cup \{q\}, r\rangle$ respectivamente, entonces es posible crear un nuevo grafo de inferencia agregando en $\mathcal{G}$ un nodo conteniendo la razón $\langle X, r\rangle$, con los nodos α, β y γ como sus únicos ancestros.

Pollock señala que toda teoría epistémica lo suficientemente seria debe permitir la construcción argumentos con diversos grados de fuerza conclusiva. No obstante, el autor critica la utilización de probabilidades para inducir un orden de preferencia sobre las conclusiones, ilustrando diversas situaciones en las cuales el resultado obtenido mediante la aplicación de este método no es el deseado. En este sentido, Pollock postula que a su entender el criterio más adecuado es el *principio del enlace más débil*, el cual puede ser sintetizado en los siguientes términos:

> *"El grado de soporte de una conclusión de un argumento deductivo equivale al mínimo entre los grados de soporte de sus premisas"*

Las definiciones anteriores se adecúan al principio del enlace más débil de la siguiente manera:

Definición 2.12. **[Pollock, 1995]**
(rebatimiento revisado)
Sea $\langle \Gamma, p \rangle$ una razón prima facie de fuerza conclusiva ξ. Diremos que una razón $\langle \Lambda, q \rangle$ de fuerza conclusiva η es un *derrotador por rebatimiento* de $\langle \Gamma, p \rangle$ si, y sólo si, se verifica que $q = \lceil \neg p \rceil$, donde $\Lambda \subseteq \Gamma$ y $\eta \geq \xi$. ■

Si bien las caracterizaciones anteriores de este sistema argumentativo ignoraban la fuerza conclusiva a la hora de determinar el socavamiento entre razones, la última manifestación del sistema refleja que el autor ha cambiado de opinión al respecto, según se aprecia a continuación:

Definición 2.13. **[Pollock, 1995]**
(socavamiento revisado)
Sea $\langle \Gamma, p \rangle$ una razón prima facie de fuerza conclusiva ξ. Diremos que una razón $\langle \Lambda, q \rangle$ de fuerza conclusiva η es un *derrotador por socavamiento* de $\langle \Gamma, p \rangle$ si, y sólo si, se verifica que $q = \lceil {\sim}(\Pi\Gamma \gg p) \rceil$, donde $\Lambda \subseteq \Gamma$ y $\eta \geq \xi$. ■

Uno de los aspectos que más ha cambiado a lo largo de la evolución de este sistema es la caracterización del estado final de los argumentos. En este sentido, resulta instructivo considerar cada una de las propuestas ensayadas, analizando las razones que motivaron las diversas modificaciones. La primer especificación introdujo una definición inductiva por niveles [Pollock, 1987], la cual influenció varios desarrollos posteriores tales como el sistema Simari-Loui [Simari y Loui, 1992] o los sistemas argumentativos abstractos [Vreeswijk, 1993b]. Esta definición puede ser reformulada prescindiendo de los tan mentados niveles de la siguiente manera:

Definición 2.14. **[Pollock, 1995]**
(asignación de estados)
Sea S un conjunto de argumentos basados en inicial. Los estados *no derrotado*, *provisionalmente derrotado* o *totalmente derrotado* pueden ser asignados a los argumentos contenidos en S de acuerdo al siguiente detalle:

1. Todo argumento contenido en S construído mediante la aplicación reiterada de la regla "premisa" está *no derrotado.*

2. Un argumento A contenido en S está *no derrotado* si, y sólo si, se verifica que todo subargumento propio de A está no derrotado, y por otra parte todo argumento que derrota a A está a su vez totalmente derrotado.

3. Un argumento A contenido en S está *totalmente derrotado* si, y sólo si, se verifica que algún subargumento propio de A está derrotado, y por otra parte algún argumento que derrota a A está no derrotado.

Todo argumento contenidos en S que no esté totalmente derrotado ni no derrotado está *provisionalmente derrotado.* ■

Esta caracterización parece arrojar interesantes resultados en los diversos casos de estudio discutidos en la literatura. Sin embargo, Pollock se percató de que su caracterización de la noción de argumento permitía situaciones paradójicas: para cualquier argumento sustentando una conclusión p, resulta factible construir un argumento inconsistente para $\sim p$, simplemente apelando a cualquiera de las conocidas técnicas de derivación de una conclusión arbitraria a partir de una contradicción. El impacto de este inconveniente en la semántica del sistema es colosal, ya que toda conclusión se torna provisionalmente derrotada. La solución a este problema—producto de no requerir consistencia en los argumentos—fue fácilmente alcanzada estipulando *ad-hoc* que todo argumento inconsistente está totalmente derrotado, donde un argumento resulta inconsistente si alguno de sus subargumentos ataca a otro de sus subargumentos.

Pollock modificó una vez más la asignación de estado en su última formulación del sistema, principalmente con el objeto de obtener un mejor resultado en las situaciones donde existen conclusiones flotantes (*e.g.*, la argumentación circular). Actualmente ha abandonado la asignación única de estados en favor de una asignación múltiple, favorecientdo sólo aquellas conclusiones sancionadas por todas las asignaciones posibles. El autor captura cada una de estas asignaciones posibles a través de la noción de asignación de estados parcial.

Definición 2.15. **[Pollock, 1995]**
(asignación de estados parcial)
Sea S un conjunto de argumentos basados en inicial. Los estados *derrotado* o *no derrotado* pueden ser asignados a los argumentos contenidos en S', donde $S' \subseteq S$, de acuerdo al siguiente detalle:

1. Todo argumento contenido en S' construído mediante la aplicación reiterada de la regla "premisa" está *no derrotado.*

2. Un argumento A contenido en S' está *no derrotado* si, y sólo si, se verifica que todo subargumento propio de A está no derrotado y por otra parte todo argumento que derrota a A está a su vez derrotado.

3. Un argumento A contenido en S' está *derrotado* si, y sólo si, se verifica que algún subargumento propio de A está derrotado y por otra parte algún argumento que derrota a A está no derrotado.

■

La asignación de estados parcial determina el estado de un subconjunto de los argumentos existentes. Pollock acota que sólo aquellas asignaciones basadas en subconjuntos maximales resultan relevantes para determinar el estado último de los argumentos. Estas asignaciones de estado parciales maximales se denominan simplemente asignaciones de estado. Por ende, el estado final de los argumentos queda establecido en función del estatus obtenido en cada asignación de estado posible.

Definición 2.16. **[Pollock, 1995]**
(estado final de un argumento)
Sea S un conjunto de argumentos basados en inicial. Diremos que un argumento está *no derrotado en base a S* si, y sólo si, toda asignación de estado para los

argumentos en S lo sanciona como no derrotado. De los argumentos que no están no derrotados en base a S, diremos que un argumento está *totalmente derrotado en base a S* si, y sólo si, ninguna asignación de estado para los argumentos en S lo sanciona como *no derrotado*. Los argumentos que no están totalmente derrotados en base a S ni no derrotados en base a S están *provisionalmente derrotados en base a S*. ■

El autor, no satisfecho con el análisis anterior, incluso identifica dos tipos de semánticas: una idealizada, elaborada en base a un conjunto potencialmente infinito de argumentos, y la otra práctica, computable en tiempo finito. La semántica idealizada, de formulación natural pero no implementable computacionalmente, se define en términos del conjunto de todos los argumentos basados en inicial.

Definición 2.17. **[Pollock, 1995]**
(conclusiones idealmente garantizadas)
Sea S el conjunto de todos los argumentos basados en inicial. Diremos que una conclusión está *idealmente garantizada en base a inicial* si, y sólo si, está sustentada por algún argumento no derrotado en base a S. ■

La segunda semántica representa una aproximación a la semántica idealizada, con la ventaja de ser computacionalmente viable. Esta semántica se define en términos de una secuencia de conjuntos $\{S_k\}_{k=0}^{\infty}$, tal que para todo $i < j$ se verifica que $S_i \subset S_j$, donde la unión de los S_k coincide con el conjunto de todos los argumentos basados en inicial. La intuición detrás de esta secuencia es modelar la consideración sucesiva por parte de un agente inteligente de nuevos argumentos a medida que progresa en su razonamiento. En este sentido, diremos que una conclusión está garantizada en un nivel k si, y sólo si, existe algún argumento que sustenta a esa conclusión que está no derrotado en base al conjunto S_k. Por último, la nueva semántica intenta capturar el estado de los argumentos luego de las eventuales fluctuaciones producto del carácter incremental de los conjuntos contenidos en la secuencia.

Definición 2.18. **[Pollock, 1995]**
(conclusiones garantizadas)

Diremos que una conclusión está *garantizada en base a* ***inicial*** si, y sólo si, existe un n tal que para todo $k \geq 0$ se verifica que algún argumento sustentando esa conclusión está no derrotado en base al conjunto S_{n+k}. ■

Pollock destaca que a pesar de las evidentes similitudes entre las nociones de conclusión garantizada y de conclusión idealmente garantizada, las mismas no se encuentran relacionadas entre sí, si bien sostiene que restringiendo las posibles configuraciones del grafo de derrota sería posible derivar un resultado de equivalencia entre estas nociones.

2.2.3. Sistemas Argumentativos Abstractos

El sistema argumentativo ideado por Lin y Shoham [Lin y Shoham, 1989] tuvo por objeto proveer un marco unificador en el cual estudiar el comportamiento bajo diversas situaciones de los principales formalismos de razonamiento. Sin embargo, este sistema no contaba con todos los elementos propios de una teoría argumentativa, ya que carecía de criterios de comparación entre argumentos. Gerard A. W. Vreeswijk elabora en su tesis doctoral [Vreeswijk, 1993b] sobre la base sentada por Lin y Shoham, intentando corregir los aspectos más controversiales del sistema argumentativo introducido por estos autores. A tal efecto, desarrolla la noción de *sistema argumentativo abstracto*, en esencia un formalismo de argumentación rebatible que incorpora explícitamente el elemento faltante en el sistema de Lin y Shoham (*i.e.*, la comparación entre argumentos).

Vreeswijk señala que su objetivo consiste en estudiar los principios que gobiernan la argumentación rebatible empleando como modelo a los sistemas argumentativos abstractos. Se puede afirmar que comparte en algún sentido las aspiraciones que Lin y Shoham tuvieron para con su sistema, intentando definir una teoría tan genérica como sea posible. Cabe aceptar que el artículo [Vreeswijk, 1997] resume los principales resultados recopilados por Vreeswijk en su tesis doctoral. En lo que resta de la presente sección repasaremos las nociones y conceptos que componen todo sistema argumentativo abstracto.

Definición 2.19. **[Vreeswijk, 1997]**
(sistema argumentativo abstracto)
Un *sistema argumentativo abstracto* es una terna $(\mathcal{L}, \mathcal{R}, \leq)$, donde $\mathcal{L}$ es un lenguaje conteniendo al elemento distinguido '$\perp$', $\mathcal{R}$ es un conjunto de reglas de inferencia y '$\leq$' es un orden sobre argumentos reflexivo y transitivo. ■

El lenguaje objeto de un sistema argumentativo abstracto queda completamente sin especificar, sólo asumiendo la existencia del elemento distinguido '$\perp$' que representa a una proposición contradictoria. El segundo componente del sistema modela la maquinaria de inferencia asociada a $\mathcal{L}$, caracterizada a través de un conjunto de reglas de inferencia.

Definición 2.20. **[Vreeswijk, 1997]**
(regla de inferencia)
Sea $\mathcal{L}$ un lenguaje. Diremos que:

1. Una *regla de inferencia estricta* es una expresión de la forma $\phi_1, \ldots, \phi_n \rightarrow \phi$, $n \geq 0$, donde ϕ y todos los ϕ_i, $1 \leq i \leq n$, son fórmulas de $\mathcal{L}$.

2. Una *regla de inferencia rebatible* es una expresión de la forma $\phi_1, \ldots, \phi_n \Rightarrow \phi$, $n \geq 0$, donde ϕ y todos los ϕ_i, $1 \leq i \leq n$, son fórmulas de $\mathcal{L}$.

Una *regla de inferencia* puede ser tanto una regla de inferencia estricta o una regla de inferencia rebatible. ■

Las reglas de inferencia constituyen una relación metalingüística entre fórmulas de $\mathcal{L}$, tratamiento similar al adoptado en otros sistemas argumentativos tales como [Simari y Loui, 1992, Simari et al., 1994, Prakken y Sartor, 1997]. Vreeswijk señala que las reglas de inferencia de un sistema argumentativo abstracto se componen de principios lógicos generales independientes del dominio. Esta restricción es consecuencia de no poder indagar en la estructura interna de $\mathcal{L}$, producto a su vez del nivel de abstracción con que se representa el conocimiento en este sistema. Sin embargo, si bien es posible encontrar diversos principios lógicos generales de carácter estricto (*e.g.*, *modus ponendo ponens* o *modus tollendo ponens*), no queda claro que existan principios lógicos generales que sean de carácter rebatible.

Al igual que en la mayoría de los sistemas argumentativos, las reglas de inferencia se estructuran en demostraciones, la cuales dan a lugar a la noción de argumento.

Definición 2.21. **[Vreeswijk, 1997]**
(argumento)
Sea $\mathcal{R}$ un conjunto de reglas. Un *argumento* tiene asociado premisas, una conclusión, sentencias, suposiciones, subargumentos, un argumento último, una longitud y un tamaño. Formalmente, un argumento σ es:

1. un elemento de $\mathcal{L}$, en cuyo caso $prem(\sigma) = \{\sigma\}$, $conc(\sigma) = \sigma$, $sent(\sigma) = \{\sigma\}$, $sup(\sigma) = \emptyset$, $suba(\sigma) = \{\sigma\}$, $ult(\sigma) = \{\sigma\}$, $long(\sigma) = 1$ y $tam(\sigma) = 1$; o bien

2. una expresión de la forma $\sigma_1, \ldots, \sigma_n \rightarrow \phi$, $n \geq 0$, donde $\sigma_1, \ldots, \sigma_n$ es una secuencia finita de argumentos, tal que existe una regla de inferencia en $\mathcal{R}$, de la forma $\phi_1, \ldots, \phi_n \rightarrow \phi$, y se verifica que $conc(\sigma_1) = \phi_1, \ldots, conc(\sigma_n) = \phi_n$, con $\phi \notin sent(\sigma_1) \cup \ldots \cup sent(\sigma_n)$. De ser así,

$$
\begin{aligned}
&prem(\sigma) = prem(\sigma_1) \cup \ldots \cup prem(\sigma_n),\\
&conc(\sigma) = \phi,\\
&sent(\sigma) = sent(\sigma_1) \cup \ldots \cup sent(\sigma_n) \cup \{\phi\},\\
&sup(\sigma) = sup(\sigma_1) \cup \ldots \cup sup(\sigma_n),\\
&suba(\sigma) = suba(\sigma_1) \cup \ldots \cup suba(\sigma_n) \cup \{\sigma\},\\
&ult(\sigma) = \{\tau_1, \ldots, \tau_n \rightarrow \phi \mid \tau_1 \in ult(\sigma_1), \ldots, \tau_n \in ult(\sigma_n)\} \cup \{\phi\},\\
&long(\sigma) = \text{máx}(long(\sigma_1), \ldots, long(\sigma_n)) + 1, \text{ y}\\
&tam(\sigma) = tam(\sigma_1) + \ldots + tam(\sigma_n) + 1;
\end{aligned}
$$

   o bien

3. una expresión de la forma $\sigma_1, \ldots, \sigma_n \Rightarrow \phi$, $n \geq 0$, donde $\sigma_1, \ldots, \sigma_n$ es una secuencia finita de argumentos, tal que existe una regla de inferencia en $\mathcal{R}$, de la forma $\phi_1, \ldots, \phi_n \Rightarrow \phi$, y se verifica que $conc(\sigma_1) = \phi_1, \ldots, conc(\sigma_n) = \phi_n$,

con $\phi \notin sent(\sigma_1) \cup \ldots \cup sent(\sigma_n)$. De ser así, los diversos componentes se definen de manera análoga al caso anterior, salvo $sup(\sigma)$ que incorpora a ϕ (*i.e.*, $sup(\sigma) = sup(\sigma_1) \cup \ldots \cup sup(\sigma_n) \cup \{\phi\}$).

■

Un aspecto en el cual la noción de argumento provista por Vreeswijk se diferencia de otros formalismos es que no aborda la problemática asociada a la existencia de argumentos redundantes o no minimales.

Ejemplo 2.6. Considerando el sistema argumentativo abstracto $(\mathcal{L}, \mathcal{R}, \emptyset)$, donde

$$\mathcal{L} = \{p;\ q;\ r\} \cup \{\perp\} \qquad \mathcal{R} = \{p, q \Rightarrow r\}$$

En este contexto, para el argumento $\sigma = p, q \Rightarrow r$ se verifica que $prem(\sigma) = \{p;\ q\}$, $conc(\sigma) = \{r\}$, $sent(\sigma) = \{p;\ q;\ r\}$, $sup(\sigma) = \{s\}$, $suba(\sigma) = \{q;\ r;\ p, q \Rightarrow r\}$, $ult(\sigma) = \{p, q \Rightarrow r;\ r\}$, $long(\sigma) = 2$ y $tam(\sigma) = 3$.

Seguidamente asumiremos la existencia de un sistema argumentativo abstracto $(\mathcal{L}, \mathcal{R}, \leq)$ arbitrario pero fijo, salvo mención explícita de lo contrario. A manera de simplificación notaremos como $\sigma_1, \ldots, \sigma_n \rightarrow \sigma$ al argumento construído a partir de la regla $\phi_1, \ldots, \phi_n \rightarrow \phi$ y los argumentos $\sigma_1, \ldots, \sigma_n$, donde $conc(\sigma_1) = \phi_1, \ldots, conc(\sigma_n) = \phi_n$ y $conc(\sigma) = \phi$. Por otra parte, denominaremos *última regla* de σ a la regla $\sigma_1, \ldots, \sigma_n \Rightarrow \sigma$. Adicionalmente, en el caso de los argumentos que finalizan aplicando una regla de inferencia rebatible se adoptaran también las mismas convenciones. Finalmente, para todo argumento σ y τ notaremos $\tau \sqsubseteq \sigma$ toda vez que $\tau \in sub(\sigma)$.

Definición 2.22. **[Vreeswijk, 1997]**
(argumento estricto, argumento rebatible)
Un argumento σ se dice *estricto* si, y sólo si, $\sigma \in \mathcal{L}$, o bien σ es de la forma $\sigma_1, \ldots, \sigma_n \rightarrow \sigma$, con $\sigma_1, \ldots, \sigma_n$ argumentos estrictos. Un argumento se dice *rebatible* cuando no es estricto. ■

Considerando que un argumento sustenta a su conclusión proveyendo una demostración para ésta, resulta necesario especificar sobre qué fundamentos descansa un argumento dado:

Definición 2.23. **[Vreeswijk, 1997]**
(base de un argumento)
Sea $\mathcal{L}$ un lenguaje y P un subconjunto de $\mathcal{L}$. Diremos que un argumento A está *basado* en P si, y sólo si, toda premisa de A está contenida en P. ■

La relación de comparación entre argumentos '$\leq$', tercer componente de todo sistema argumentativo abstracto, representa la fuerza relativa entre los diversos argumentos. En este sentido, $\sigma \leq \tau$ denota que τ es tan bueno como σ, y $\sigma < \tau$ que τ es mejor que σ, para todo par de argumentos σ y τ. Vreeswijk acota que no toda relación de comparación resulta razonable en el contexto de un sistema argumentativo abstracto, por lo que impone las siguientes restricciones adicionales:

Definición 2.24. **[Vreeswijk, 1997]**
(orden de fuerza conclusiva)
Un *orden de fuerza conclusiva* es una relación reflexiva y transitiva entre argumentos que satisface las siguientes condiciones:

1. No existen cadenas infinitas ascendentes de la forma $\sigma_1 < \ldots < \sigma_n < \ldots$.

2. Para todo par de argumento σ y τ, si $\sigma \sqsubseteq \tau$, entonces $\tau \leq \sigma$.

3. Si $\sigma_1, \ldots, \sigma_n \to \sigma$, entonces $\sigma_i \leq \sigma$, para algún $1 \leq i \leq n$.

■

La primer condición tiene por objeto que la resolución de un conflicto apelando a la relación de comparación tenga fin. Las últimas dos condiciones aseguran la uniforme distribución del orden de fuerza conclusiva sobre los argumentos. En particular, la segunda condición estipula que no es posible reforzar un argumento incorporando más información. Cabe señalar que esta tesitura se contrapone a la noción de "acopio de razones" introducida por Bart Verheij en [Verheij, 1996], donde la presencia de múltiples razones a favor de una conclusión permite construir un argumento de mayor fuerza conclusiva. Por último, la tercer condición impide la fuga de fuerza conclusiva como producto del encadenamiento de reglas de inferencia estrictas.

En los sistemas argumentativos abstractos la relación de derrota se define en términos del conflicto entre argumentos y del orden de fuerza conclusiva. La siguiente definición caracteriza la porción que involucra al orden de fuerza conclusiva.

Definición 2.25. **[Vreeswijk, 1997]**
(debilitamiento)
Un argumento τ es un *debilitador* de un conjunto de argumentos Σ si, y sólo si, se verifica que $\sigma \leq \tau$, para algún $\sigma \in \Sigma$. ■

El autor menciona una paradoja que se suscita a partir de esta noción, ya que al incrementar la cantidad de argumentos contenidos en un conjunto en vez de capitalizar fuerza conclusiva se obtiene un conjunto de argumentos al que resulta más sencillo debilitar. Otros autores señalan que este resultado lejos de ser paradójico es en realidad el esperado, tal como lo reflejan al imponer restricciones de minimalidad al concepto de argumento con el objeto de evitar la formulación de aquellos que expongan una cantidad innecesaria de puntos de ataque.

La segunda mitad de la caracterización de la derrota es una generalización del concepto de inconsistencia presente en la lógica clásica. Para enfatizar esta distinción, se adopta el término *'compatibilidad'* en vez de *'consistencia'*.

Definición 2.26. **[Vreeswijk, 1997]**
(compatibilidad)
Sea P un subconjunto de $\mathcal{L}$. Diremos que P es *incompatible* si, y sólo si, existe algún argumento estricto σ basado en P tal que $conc(\sigma) = \perp$. En caso contrario diremos que P es *compatible*. ■

Ejemplo 2.7. Considerando el sistema argumentativo abstracto $(\mathcal{L}, \mathcal{R}, \emptyset)$, donde

$$\mathcal{L} = \{p;\ q;\ r;\ s\} \cup \{\perp\} \qquad \mathcal{R} = \{p \Rightarrow q;\ p, q \rightarrow s;\ r, s \rightarrow \perp\}$$

En este contexto, todos los subconjuntos de $\{p;\ q;\ s\}$, $\{p;\ r\}$ y $\{q;\ r\}$ son compatibles. En cambio, todos los superconjuntos de $\{p;\ q;\ r\}$ y $\{r;\ s\}$ son incompatibles.

Vreeswijk extiende la noción de compatibilidad postulando que un conjunto de argumentos Σ es compatible si, y sólo si, el conjunto $conc(\Sigma)$ es compatible, donde por $conc(\Sigma)$ entendemos al conjunto formado por las conclusiones de los argumentos en Σ (*i.e.*, $conc(\Sigma) = \{\phi \mid \sigma \in \Sigma \wedge conc(\sigma) = \phi\}$). Por último, para caracterizar a la relación de derrota entre argumentos es necesario indicar el contexto en el cual tiene a lugar la derrota. Este contexto, denominado *conjunto base*, se representa a través de un conjunto compatible de fórmulas de $\mathcal{L}$ (este requerimiento de compatibilidad tiene por objeto evitar el conflicto entre argumento estrictos).

Definición 2.27. **[Vreeswijk, 1997]**
(conjunto base)
Sea P un subconjunto finito de $\mathcal{L}$. Diremos que P es un *conjunto base* si, y sólo si, P es compatible. ■

La relación de derrota en los sistemas argumentativos abstractos recombina las nociones de incompatibilidad y debilitamiento de la siguiente manera:

Definición 2.28. **[Vreeswijk, 1997]**
(derrota)
Sea P un conjunto base y σ un argumento. Diremos que el conjunto de argumentos Σ es un *derrotador* de σ si, y sólo si, Σ resulta incompatible con σ, pero σ no constituye un debilitador de Σ. ■

Si bien la relación de derrota usualmente se define entre argumentos, este formalismo presenta como particularidad que la derrota se modela como una relación entre un argumento y un conjunto de argumentos. En este sentido, H. Prakken y G. Vreeswijk indican en [Prakken y Vreeswijk, 2002] que esta atípica caracterización es consecuencia del nivel de abstracción manejado en el sistema, el cual impide reconocer la inconsistencia entre argumentos de forma directa.

Ejemplo 2.8. Considerando el sistema argumentativo abstracto $(\mathcal{L}, \mathcal{R}, \leq)$, donde

$$\mathcal{L} = \{p;\ q;\ r;\ s\} \cup \{\bot\} \qquad \mathcal{R} = \{p \Rightarrow q;\ p \Rightarrow s;\ q \Rightarrow r;\ r \Rightarrow s;\ q, r, s \rightarrow \bot\}$$

y '$\leq$' representa el orden de fuerza conclusiva que prefiere los argumentos que emplean menor cantidad de reglas rebatibles. En este contexto, el argumento $\sigma = (p \Rightarrow q) \Rightarrow r$ no es derrotado por $\Sigma_1 = \{p;\ p \Rightarrow q;\ p \Rightarrow s;\ ((p \Rightarrow q) \Rightarrow r) \Rightarrow s\}$, pues σ debilita a Σ_1. En cambio, tanto $\Sigma_2 = \{p \Rightarrow q;\ p \Rightarrow s\}$ como $\Sigma_3 = \{p;\ p \Rightarrow q;\ p \Rightarrow s\}$ derrotan a σ. Por último $\Sigma_4 = \{p \Rightarrow q\}$ y $\Sigma_5 = \{p \Rightarrow s\}$ son compatibles con σ, y en consecuencia no lo derrotan.

El ejemplo anterior ilustra la importancia de la relación de inclusión entre conjuntos al determinar el alcance de la derrota entre argumentos. En particular, Σ_1 no derrota a σ, pero $\Sigma_2 \subset \Sigma_1$ logra derrotar a σ. Sin embargo, $\Sigma_4 \subset \Sigma_2 \subset \Sigma_1$ no es capaz de derrotar a σ. La siguiente definición captura estas facetas, sintetizando el papel de los subargumentos a nivel de derrota entre argumentos.

Definición 2.29. **[Vreeswijk, 1997]**
(habilitamiento)
Sea P un conjunto base y sea Σ un conjunto de argumentos. Diremos que un argumento σ es *habilitado* por Σ en base a P, denotado $\Sigma \mathrel{|\!\!\sim}_P \sigma$, si, y sólo si:

1. σ está contenido en P, o bien
2. existen argumentos $\sigma_1, \ldots, \sigma_n$, tales que $\Sigma \mathrel{|\!\!\sim}_P \sigma_1, \ldots, \Sigma \mathrel{|\!\!\sim}_P \sigma_n$ y por otra parte σ guarda la forma $\sigma_1, \ldots, \sigma_n \rightarrow \sigma$, o bien
3. existen argumentos $\sigma_1, \ldots, \sigma_n$, tales que $\Sigma \mathrel{|\!\!\sim}_P \sigma_1, \ldots, \Sigma \mathrel{|\!\!\sim}_P \sigma_n$, σ guarda la forma $\sigma_1, \ldots, \sigma_n \Rightarrow \sigma$ y por otra parte Σ no contiene derrotadores de σ.

■

La intuición detrás de la noción de habilitamiento es que un argumento σ es habilitado por un conjunto de argumentos Σ si todo subconjunto de Σ es compatible con σ, o bien es debilitado por éste.

Ejemplo 2.9. Considerando el sistema argumentativo abstracto $(\mathcal{L}, \mathcal{R}, \emptyset)$, donde

$$\mathcal{L} = \{p;\ q;\ r;\ s\} \cup \{\perp\} \qquad \mathcal{R} = \{p \Rightarrow q;\ q \Rightarrow r;\ p \Rightarrow s;\ q, s \rightarrow \perp\}$$

En este contexto, deseamos determinar si el argumento $\sigma = p \Rightarrow q$ está habilitado por el conjunto de argumentos $\Sigma = \{p;\ p \Rightarrow q;\ p \Rightarrow s\}$, a partir del conjunto base $P = \{p\}$. Ya que σ termina con la aplicación de una regla rebatible, la definición indica que p debe estar habilitado por Σ en base a P, y que Σ no debe contener derrotadores de σ. Sin embargo, $\{p \Rightarrow s\} \subseteq \Sigma$ es un derrotador para σ, lo que implica que $\Sigma \not\approx_P \sigma$.

La noción de habilitamiento captura el resultado de considerar el efecto inmediato de la relación de derrota sobre un conjunto de argumentos. La siguiente definición explicita este concepto.

Definición 2.30. **[Vreeswijk, 1997]**
(operador de habilitamiento)
Sea P un conjunto base y Σ un conjunto de argumentos. El *operador de habilitamiento*, notado $habilitados_P(\Sigma)$, denota al conjunto $\{\sigma \mid P \approx_\Sigma \sigma\}$. ■

Del análisis de las propiedades abstractas satisfechas por el operador de habilitamiento se desprenden las siguientes conclusiones:

- Sea P un conjunto base y Σ un conjunto compatible de argumentos. De ser así, necesariamente se verifica que $\Sigma \subseteq habilitados_P(\Sigma)$.

- Sea P un conjunto base y sean Σ_1 y Σ_2 dos conjuntos de argumentos tales que $\Sigma_1 \subseteq \Sigma_2$. De ser así, necesariamente se verifica que $habilitados_P(\Sigma_2) \subseteq habilitados_P(\Sigma_1)$.

Vreeswijk ensaya dos alternativas para determinar el estado final de los argumentos, ámbas elaboradas a partir de la noción de habilitamiento. La primer caracterización, adaptada de [Pollock, 1987], se basa en el conjunto de argumentos activos luego de aplicar el operador de habilitamiento una cierta cantidad de veces.

Definición 2.31. **[Vreeswijk, 1997]**
(argumentos activos por niveles)

Sea P un conjunto base. Un argumento σ está *activo en el nivel 1 en base a P* si, y sólo si, σ está basado en P. Por otra parte, un argumento σ está *activo en el nivel k en base a P*, denotado $P \mathrel{\mid\!\sim_k} \sigma$, si se verifica alguna de las siguientes condiciones:

1. El conjunto P contiene a σ, o bien
2. existen argumentos $\sigma_1, \ldots, \sigma_n$, tales que $P \mathrel{\mid\!\sim_k} \sigma_1, \ldots, P \mathrel{\mid\!\sim_k} \sigma_n$ y a su vez σ guarda la forma $\sigma_1, \ldots, \sigma_n \rightarrow \sigma$, o bien
3. existen argumentos $\sigma_1, \ldots, \sigma_n$, tales que $P \mathrel{\mid\!\sim_k} \sigma_1, \ldots, P \mathrel{\mid\!\sim_k} \sigma_n$, σ guarda la forma $\sigma_1, \ldots, \sigma_n \Rightarrow \sigma$ y a su vez se verifica que ningún conjunto de argumentos activos en el nivel $k-1$ en base a P derrota a σ.

Notaremos mediante $info_k(P)$ al conjunto $\{\sigma \mid P \mathrel{\mid\!\sim_k} \sigma\}$. ■

A partir de la noción de argumentos activos por niveles es posible determinar el estado final de un argumento, de acuerdo a la siguiente definición:

Definición 2.32. **[Vreeswijk, 1997]**
(estados de un argumento)
Sea P un conjunto base y σ un argumento basado en P. Diremos que σ está:

- *definitivamente no derrotado* si, y sólo si, existe un $n \geq 1$, tal que para todo $k \geq 0$ se verifica que $P \mathrel{\mid\!\sim_{n+k}} \sigma$.
- *provisionalmente derrotado* si, y sólo si, para todo $n \geq 1$, se verifica que $P \mathrel{\mid\!\sim_{n+k}} \sigma$ para algún $k \geq 0$, y $P \not\mathrel{\mid\!\sim}_{n+k'} \sigma$ para otro $k' \geq 0$.
- *definitivamente derrotado* si, y sólo si, existe un $n \geq 1$, tal que para todo $k \geq 0$ se verifica que $P \not\mathrel{\mid\!\sim}_{n+k} \sigma$.

■

La taxonomía anterior se extiende de forma natural a conclusiones: diremos que una conclusión esta definitivamente no derrotada (resp. provisionalmente derrotada, definitivamente derrotada) si, y sólo si, se encuentra sustentada por un argumento definitivamente no derrotado (resp. un argumento provisionalmente derrotado, un argumento definitivamente derrotado).

Ejemplo 2.10. Considerando el sistema argumentativo abstracto $(\mathcal{L}, \mathcal{R}, \leq)$, donde

$$\mathcal{L} = \{p;\ q;\ r;\ s;\ t\} \cup \{\bot\}$$

$$\mathcal{R} = \{p \Rightarrow q;\ q \Rightarrow r;\ p \Rightarrow s;\ p \Rightarrow t;\ q, s \rightarrow \bot;\ r, t \rightarrow \bot\}$$

y '$\leq$' representa el orden que prefiere los argumentos que emplean menor cantidad de reglas rebatibles. En este contexto, los conjuntos $\{(p \Rightarrow q) \Rightarrow r;\ p \Rightarrow t\}$ y $\{p \Rightarrow q;\ p \Rightarrow s\}$ capturan los conflictos que surgen entre los argumentos basados en P. En el primer conflicto, el argumento $p \Rightarrow t$ derrota al argumento $(p \Rightarrow q) \Rightarrow r$, pues emplea menos reglas rebatibles. En el segundo conflicto, ninguno de los argumentos resulta mejor que el otro. Por consiguiente, los diversos argumentos activos en cada nivel son

$$info_1(P) = \{p;\ p \Rightarrow q;\ (p \Rightarrow q) \Rightarrow r;\ p \Rightarrow s;\ p \Rightarrow t\}$$
$$info_2(P) = \{p;\ p \Rightarrow t\}$$
$$info_3(P) = \{p;\ p \Rightarrow q;\ p \Rightarrow s;\ p \Rightarrow t\}$$
$$info_4(P) = \{p;\ p \Rightarrow t\}$$
$$info_5(P) = \{p;\ p \Rightarrow q;\ p \Rightarrow s;\ p \Rightarrow t\}$$

y así sucesivamente.

Finalmente, p y $p \Rightarrow t$ son argumentos definitivamente no derrotados, $p \Rightarrow q$ y $p \Rightarrow s$ son argumentos provisoriamente derrotados y $(p \Rightarrow q) \Rightarrow r$ es un argumento definitivamente derrotado.

La segunda caracterización para determinar el estado final de los argumentos se asemeja a una ecuación de punto fijo, si observamos que la definición hace referencia a sí misma.

Definición 2.33. **[Vreeswijk, 1997]**
(relación de inferencia rebatible)
Sea P un conjunto base. Una relación entre P y un argumento σ basados en P,

notado $P \mid\!\sim \sigma$, es una *relación de inferencia rebatible* si, y sólo si, se verifica alguna de las siguientes condiciones:

1. El conjunto P contiene a σ, o bien
2. existen argumentos $\sigma_1, \ldots, \sigma_n$, tales que $P \mid\!\sim \sigma_1, \ldots, P \mid\!\sim \sigma_n$ y por otra parte σ guarda la forma $\sigma_1, \ldots, \sigma_n \rightarrow \sigma$, o bien
3. existen argumentos $\sigma_1, \ldots, \sigma_n$, tales que $P \mid\!\sim \sigma_1, \ldots, P \mid\!\sim \sigma_n$, σ guarda la forma $\sigma_1, \ldots, \sigma_n \Rightarrow \sigma$ y por otra parte ningún conjunto de argumentos sancionados por la relación de inferencia rebatible derrota a σ.

■

A diferencia de la semántica anterior, la cual se encuentra unívocamente definida en toda situación, la nueva caracterización puede no estar definida, o bien sancionar múltiples interpretaciones para la misma situación. En general, se suelen denominar extensión a cada una de las diferentes interpretaciones que surgen toda vez que la semántica de un sistema no queda unívocamente determinada.

Definición 2.34. **[Vreeswijk, 1997]**
(extensión)
Sea P un conjunto base. Diremos que un conjunto de argumentos Σ es una *extensión* de P si, y sólo si, para alguna relación de inferencia rebatible se verifica que $\Sigma = \{\sigma \mid P \mid\!\sim \sigma\}$. ■

Ejemplo 2.11. Considerando el sistema argumentativo abstracto $(\mathcal{L}, \mathcal{R}, \emptyset)$, donde:

$$\mathcal{L} = \{p;\ q;\ r\} \cup \{\bot\} \qquad \mathcal{R} = \{p \Rightarrow q;\ p \Rightarrow r;\ q, r \rightarrow \bot\}$$

En este contexto, los únicos argumentos en conflicto son $\{p \Rightarrow q\}$ y $\{p \Rightarrow r\}$. Como este conflicto no es dirimido por el orden de fuerza conclusiva provisto, existen dos relaciones de inferencia rebatible, una sancionando $\{p;\ p \Rightarrow q\}$ y la otra $\{p;\ p \Rightarrow r\}$. Por ende, este sistema argumentativo abstracto admite dos extensiones.

En la mayoría de los formalismos que admiten múltiples extensiones surgen naturalmente dos tipos de semánticas, una de carácter escéptico (obtenida al intersectar todas las extensiones) y otra de carácter crédulo (obtenida al unir todas las extensiones). Ya que el autor conjetura la coincidencia de la semántica escéptica con el conjunto de los argumentos definitivamente no derrotados (que pueden ser obtenidos en base a la semántica anterior por niveles), sólo incorpora en este sistema una variante de la semántica crédula.

Definición 2.35. **[Vreeswijk, 1997]**
(argumento garantizado)
Sea P un conjunto base. Un argumento σ basado en P está *garantizado* si, y sólo si, en base a alguna relación de inferencia rebatible se verifica que $P \mid\!\sim \sigma$. ■

Las semánticas propuestas para dotar de significado a los sistemas argumentativos abstractos son de corte netamente no constructivo, lo cual dificulta su implementación concreta. Si bien el autor atiende este cuestionamiento estudiando por caso aquellos procedimientos que permiten obtener cada una de las diversas extensiones, paralelamente explora una interesante teoría de prueba dialéctica semiformal, la cual permite debatir acerca del estado de una tesis de manera constructiva [Vreeswijk, 1993a]. Seguidamente desglosaremos los principales conceptos asociados a esta teoría de prueba.

Un *debate rebatible* en el marco de un sistema argumentativo abstracto involucra a un *proponente* notado **P** y un *oponente* notado **O**. Este debate tiene por objeto analizar una cierta *tesis* a partir de una *base compartida* P, donde P es un conjunto base. Un debate rebatible es en principio una secuencia alterna de movidas de **P** y **O**, que siguen un determinado patrón que dictamina las movida y contramovidas aceptables para cada configuración. El proponente y el oponente cuentan con tres tipos de movidas: pueden *postular una tesis*, o bien *formular un pregunta*, o bien *proveer una respuesta*. El contendiente que postula una tesis es su *defensor*, notado **D**. Su contrincante es el *atacante*, notado **A**. Tanto **P** como **O** puede adoptar cualquiera de estas actitudes.

Todo debate rebatible comienza con el proponente postulando la tesis principal sobre la cual se debatirá, acompañada de un conjunto de argumentos de sustento:

$| \, 1. \, \mathbf{P} : \sigma_1, \ldots, \sigma_n \, !$

En esta oportunidad **P** asume el rol **D**, por lo que se compromete con los argumentos $\sigma_1, \ldots, \sigma_n$, y debe responder por ellos ante cualquier objeción de **A**. Vreeswijk señala que el signo de exclamación simplemente denota una movida libre (pues no hay restricciones en relación a qué argumentos proveer), puesto que no fue provista en respuesta a una pregunta (en cuyo caso si se está circunscripto a responder precisamente esa pregunta). Por otra parte, el número de barras a la izquierda de la movida denota el nivel de anidamiento del debate.

Toda tesis debe ser seguida de un cuestionamiento en la forma de una pregunta. Si **D** postula $\sigma_1, \ldots, \sigma_n$, **A** debe seleccionar un subargumento de algún σ_i, $1 \leq i \leq n$, a ser atacado.

$| \ldots | \, m. \, \mathbf{D} : \sigma_1, \ldots, \sigma_n \quad !$

$| \ldots | \, m+1. \, \mathbf{A} : \sigma'_i \sqsubseteq \sigma_i \quad ?$

En esta situación, **A** se ve obligado a explicitar a este subargumento a través del siguiente movimiento:

$| \ldots | \, m. \, \mathbf{A} : \sigma \sqsubseteq \tau \quad ?$

$| \ldots | \, m+1. \, \mathbf{D} : \sigma$

Cabe aclarar que la respuesta a la pregunta no va acompañada del signo de exclamación, pues se trata de una respuesta forzada por la pregunta anterior.

Luego de que **D** suministre su respuesta, **A** cuenta con dos alternativas. En primer lugar, **A** puede refutar la respuesta de **D** a través de un conjunto de tesis locales que constituyan un derrotador:

$| \ldots | \, m. \, \mathbf{D} : \sigma$

$| \ldots || \, m+1. \, \mathbf{A} : \tau_1 \ldots \tau_n \quad !$

En este contexto, $\{\tau_1 \ldots \tau_n\}$ es un derrotador de σ. Este movimiento tiene como efecto colateral que **P** y **O** intercambien sus roles, ya que **A** asumirá una postura defensiva para con $\tau_1 \dots \tau_n$ y es **D** quien deberá atacar a alguno de sus subargumentos. Nótese que esta movida agrega un nuevo nivel al debate, representado en el diálogo rebatible como una barra adicional.

En el caso que **A** no fuera capaz de encontrar un derrotador para la última tesis de **D**, su única opción es conceder. Al conceder, el debate retrocede dos niveles hasta el nivel próximo anterior donde **A** desempeñaba el mismo rol, ya que su último ataque queda anulado por el ataque de su contrincante en el anteúltimo nivel. Dos niveles atrás, **A** cuenta con las mismas alternativas que tuvo antes de conceder, si bien es otra la tesis que debe ser atacada. Si no es capaz de refutar esta nueva tesis, debe conceder tantas veces como sea necesario o posible. Caso contrario, habiendo retrocedido una cantidad par de niveles, el debate sigue su curso natural cuestionando otro subargumento (de ser esto posible).

$$| \dots | \dots | \, m. \, \mathbf{D} : \sigma$$
$$| \dots | \, m+1. \, \mathbf{A} : \tau \sqsubseteq \upsilon \quad ?$$

Esta movida cuestiona al argumento υ postulado en un movimiento anterior y denota la concesión de σ. En caso de que no queden tesis por ser cuestionadas (*i.e.*, si **A** retrocede hasta el nivel 0), **A** pierde el debate. Por otra parte, para evitar reiteraciones sin sentido en el seno del debate se impide refutar un argumento múltiples veces mediante el mismo derrotador.

Vreeswijk acota en relación al resultado de un debate que quien realice el último movimiento ganará, si bien identifica dos formas de ganar:

- Se gana sobre una *base sustancial* al proveer una respuesta tal que fuerza a su contrincante a conceder en todos los subdebates abiertos.
- Se gana sobre una *base procedural* si el contrincante estaba en condiciones de proseguir con el debate pero no lo hizo, o bien porque le impidieron hacerlo (*e.g.*, se agotó el tiempo, intervino un árbitro, etc.), o bien porque por alguna razón decidió retirarse.

Aquellas tesis acerca de las cuales es posible ganar un debate rebatible se relacionan con las conclusiones sustentadas por argumentos garantizados de la forma esperada, como puede observarse en el siguiente ejemplo.

Ejemplo 2.12. Considerando el sistema argumentativo abstracto $(\mathcal{L}, \mathcal{R}, \leq)$, donde

$$\mathcal{L} = \{p;\ q;\ r\} \cup \{\bot\} \qquad \mathcal{R} = \{p \Rightarrow q;\ p \rightarrow r;\ q, r \rightarrow \bot\}$$

y '$\leq$' representa el orden de fuerza conclusiva que prefiere a los argumentos estrictos por sobre los argumentos rebatibles. En base al conjunto base $P = \{p\}$, un debate rebatible acerca de q puede adoptar la siguiente forma:

$|\ 1.\ \mathbf{P} : p \Rightarrow q \quad !$

$|\ 2.\ \mathbf{O} : (p \Rightarrow q) \sqsubseteq (p \Rightarrow q) \quad ?$

$|\ 3.\ \mathbf{P} : p \Rightarrow q$

$||\ 4.\ \mathbf{O} : p \rightarrow r \quad !$

Luego de que **P** sostiene a q mediante $p \Rightarrow q$, **O** ataca al argumento de sustento mediante el argumento estricto $p \rightarrow r$. Como resultado, **P** se ve obligado a conceder, retrocede dos niveles, y finalmente pierde el debate.

El proceso de disputación recién introducido admite una subsecuente simplificación, la cual consiste en descartar las movidas irrelevantes, pasando de la postulación de una tesis directamente a su refutación, de ser esta posible.

2.2.4. El sistema de H. Prakken y G. Sartor

Hendrik Prakken y Giovanni Sartor desarrollan en [Prakken y Sartor, 1997] un sistema argumentativo cuya particular característica consiste en permitir la modificación dinámica del criterio de comparación entre argumentos adoptado. El sistema Prator[5] constituye el primer formalismo que admite un comportamiento no monótono tanto en las creencias sancionadas por el sistema como en las prioridades empleadas en su derivación.

[5]denominación interna empleada por H. Prakken y G. Sartor para este sistema.

La estructura actual del sistema es el producto de sucesivos refinamientos. Sus orígenes se remontan a la tesis doctoral de Hendrik Prakken [Prakken, 1993], un acabado estudio cuyo propósito principal era extrapolar las ideas fundacionales desarrolladas en el contexto de la Inteligencia Artificial al dominio del razonamiento legal. A tal efecto, formula un sistema argumentativo que al combinar las principales características de las teorías de razonamiento no monótono de la época estaría en condiciones de sentar las bases para una formalización del razonamiento legal. El sistema enunciado por Prakken se asemeja en gran medida al sistema Simari-Loui [Simari y Loui, 1992], si bien adopta como base en contraste al lenguaje de la lógica default [Reiter, 1980], y caracteriza su semántica en términos de un operador de punto fijo. No obstante, esta primer versión no contaba con la posibilidad de modificar dinámicamente al critério de comparación entre argumentos. La elección de este lenguaje de base acarreó principalmente dos inconvenientes: por un lado no permite representar suposiciones, y por otro requiere caracterizar las conclusiones de cada argumento mediante un operador de punto fijo.

Una posterior reformulación del sistema [Prakken y Sartor, 1996] solucionó los problemas derivados de esta elección, empleando en su lugar al lenguaje de la programación en lógica extendida [Gelfond y Lifschitz, 1990]. El cambio a nivel de lenguaje de representación de conocimiento fue acompañado por un cambio análogo en la semántica del sistema, adoptando en esta ocasión una semántica basada en argumentos formulada específicamente para la programación en lógica extendida [Dung, 1993a]. En esta oportunidad el sistema se presenta en dos versiones: una versión simplificada que recibe como parámetro al criterio de comparación entre argumentos a ser utilizado, y una versión mejorada en la cual el criterio de comparación es a su vez derivado rebatiblemente dentro del sistema. Cabe acotar que la nueva semántica aún se caracterizaba en términos de un operador de punto fijo.

Finalmente, la manifestación más reciente del sistema [Prakken y Sartor, 1997] atiende el cuestionamiento anterior incorporando un mecanismo de prueba dialéctico, el cual resulta un tanto más intuitivo que la ecuación de punto fijo previa.

A continuación repasaremos las principales definiciones que comprenden el sistema Prator, presentando en primer lugar la variante del sistema con prioridades estáticas, repasando posteriormente su extensión a prioridades rebatibles.

El sistema Prator adopta como lenguaje objeto al de la *programación lógica extendida*. Las reglas se construyen a partir de fórmulas atómicas en un lenguaje de primer orden, por virtud del conectivo binario unidireccional '$\Rightarrow$'. Los *literales fuertes* representan fórmulas atómicas precedidas o no por el símbolo '$\neg$' de la negación fuerte, y los *literales defaults* representan literales fuertes precedidos del símbolo de la negación default '$\sim$'. El complemento de una fórmula atómica se define en términos de la negación fuerte.

Definición 2.36. **[Prakken y Sartor, 1997]**

(regla)

Una *regla* es una expresión de la forma

$$r : L_0 \wedge \ldots \wedge L_j \wedge {\sim}L_k \wedge \ldots \wedge {\sim}L_m \Rightarrow L_n$$

donde r, un término de primer orden, es el nombre de la regla y cada L_i, $0 \leq i \leq m$, es un literal fuerte. Como suele ser usual en el ámbito de la programación lógica, una regla con variables libres denota al conjunto de todas sus instancias fijas. ■

En este formalismo, la *base de conocimiento* consiste de un conjunto de reglas en el cual se identifican dos subconjuntos disjuntos: el conjunto S de reglas estrictas y el conjunto D de reglas rebatibles. Esta distinción—ya presente en el sistema Simari-Loui—permite separar aquel conocimiento indisputable S del conocimiento de carácter tentativo D. Se asume que las reglas presentes en S no contienen literales default. En lo que resta del análisis, simplificaremos la notación empleando el conectivo '$\rightarrow$' para las reglas estrictas y el conectivo '$\Rightarrow$' para las reglas rebatibles, si bien desde un punto de vista formal tal distinción no existe. En esta versión del sistema la base de conocimiento se complementa con una relación predefinida de preferencia entre reglas, representada a través de un orden estricto parcial sobre los elementos de D.

Definición 2.37. **[Prakken y Sartor, 1997]**
(base de conocimiento)
Denominaremos *base de conocimiento*[6] KB a toda terna $(S, D, <)$, donde S y D son conjuntos de reglas, ninguna regla de S contiene literales default, y '$<$' es un orden estricto parcial sobre los elementos de D. ■

La base de conocimiento es empleada para construir *argumentos*. Los argumentos se caracterizan a través del encadenamiento sucesivo de reglas.

Definición 2.38. **[Prakken y Sartor, 1997]**
(argumento)
Un *argumento* A es una secuencia finita $[r_0, \ldots, r_n]$ de instancias fijas de reglas tal que:

1. Para cada i, $0 \leq i \leq n$, se verifica que para todo literal fuerte L_j que figura en el antecedente de r_i, existe alguna regla r_k, $k < i$, tal que L_j es el consecuente de r_k.

2. Para todo par de reglas r_i y r_j, si $i \neq j$, entonces la conclusión de r_i es distinta de la conclusión de r_j.

Un argumento A se dice *basado* en KB, donde $KB = (S, D, <)$, si y sólo si, todas las reglas de A están contenidas en $S \cup D$. ■

La primer condición establece que los argumento se forman encadenando reglas, e ignorando a los literales default. En cierto sentido, se está asumiendo la existencia de dos reglas de inferencia, una que permite derivar la conjunción de literales fuertes, y la otra una especie de *modus ponens* rebatible, de la forma:

$$\frac{\begin{array}{l} r : L_0 \wedge \ldots \wedge L_j \wedge {\sim}L_k \wedge \ldots \wedge {\sim}L_m \Rightarrow L_n \\ \quad L_0 \wedge \ldots \wedge L_j \end{array}}{L_n}$$

[6] *teoría ordenada* según la terminología empleada en [Prakken y Sartor, 1997].

Por otra parte, la segunda condición tiene por objeto evitar la formación de ciclos en el encadenamiento de reglas.

A diferencia de las nociones de argumento introducidas en otros formalismos argumentativos, la definición de argumento utilizada en el sistema Prator representa de forma explícita tanto la información tentativa como la estricta, si bien los argumentos jamás podrán ser atacados en base a esta última. Otra particularidad de la presente definición es admitir como válidos incluso a aquellos argumentos no minimales, tal como lo ilustra el siguiente ejemplo:

Ejemplo 2.13. Considerando la base de conocimiento $KB = (S, D, \emptyset)$, donde

$$S = \{r_1 :\rightarrow a\} \qquad D = \{r_2 : a \Rightarrow b;\ r_3 : b \Rightarrow c;\ r_4 : a \Rightarrow c\}$$

es posible construir los siguientes argumentos basados en KB:

$A_1 = [r_1 :\rightarrow a;\ r_4 : a \Rightarrow c]$.

$A_2 = [r_1 :\rightarrow a;\ r_2 : a \Rightarrow b;\ r_3 : b \Rightarrow c]$.

$A_3 = [r_1 :\rightarrow a;\ r_2 : a \Rightarrow b;\ r_4 : a \Rightarrow c]$.

En el contexto de este sistema, se denota al conjunto de todos los argumentos construíbles a partir de una base de conocimiento KB como $Args_{KB}$. A su vez, se dice que un argumento A es estricto si no contiene reglas rebatibles, que es rebatible en caso contrario, y que L es una suposición del argumento A si alguna regla de A contiene al literal default $\sim\overline{L}$, donde $\overline{L}$ es el complemento de L.

El propósito de todo argumento es sustentar a su conclusión. No obstante, en este sistema es posible que un argumento soporte múltiples conclusiones, tal como se refleja a continuación:

Definición 2.39. **[Prakken y Sartor, 1997]**
(conclusión de un argumento)
Sea A un argumento. Diremos que un literal L es una *conclusión* de A si, y sólo si, L figura como consecuencia de alguna regla en A. ■

El siguiente componente del sistema argumentativo, la noción de conflicto entre argumentos, captura el hecho de que un argumento A puede ser atacado en una conclusión o suposición L por cualquier argumento capaz de soportar a $\overline{L}$, pero incluso con la eventual ayuda del conocimiento estricto de la base de conocimiento. En primer lugar, es necesario formalizar como se puede *"ayudar"* a un argumento.

Definición 2.40. **[Prakken y Sartor, 1997]**
(extensión de un argumento)
Sea A un argumento y T una secuencia de reglas. La *extensión* de A por T, denotada $A + T$, es el resultado de concatenar T al final de A. ■

En base a este concepto es posible formalizar la noción de ataque, la cual captura el conflicto entre un par de argumentos. Seguidamente se asumirá la existencia en todo momento de una base de conocimiento arbitraria pero fija $KB = (S, D, <)$.

Definición 2.41. **[Prakken y Sartor, 1997]**
(ataque)
Sean A_1 y A_2 dos argumentos y sea S_1 una secuencia de reglas estrictas tal que $A_1 + S_1$ constituye un argumento sustentando una conclusión L. El argumento A_1 *ataca* al argumento A_2 si, y sólo si, se verifica que:

- existe una secuencia de reglas estrictas S_2 tal que $A_2 + S_2$ constituye un argumento sustentando una conclusión $\overline{L}$, o bien
- $\overline{L}$ es una de las suposiciones del argumento A_2.

■

Ejemplo 2.14. Considerando la base de conocimiento $KB = (S, D, \emptyset)$, donde

$$S = \{r_1 : casado \rightarrow \neg soltero;\ r_2 : soltero \rightarrow \neg casado\}$$

$$D = \{r_3 :\Rightarrow casado;\ r_4 :\Rightarrow soltero\}$$

es posible construir los argumentos $A_1 = [r_3 :\Rightarrow casado]$ y $A_2 = [r_3 :\Rightarrow soltero]$. En este contexto, los argumentos A_1 y A_2 se atacan recíprocamente:

- A_1 ataca a A_2, pues $A_1 + []$ sustenta a *casado*, y $A_2 + [r_2 : soltero \rightarrow \neg casado]$ sustenta a $\neg casado$.

- A_2 ataca a A_1, pues $A_2 + []$ sustenta a *soltero*, y $A_1 + [r_1 : casado \rightarrow \neg casado]$ sustenta a $\neg soltero$.

Como la definición de argumento adoptada en este formalismo (def. 2.38, pág. 65), no requiere consistencia interna, resulta viable formular argumentos que se ataquen a sí mismo. Este tipo anómalo de argumento se denomina *incoherente*. Por caso, tanto el argumento $[r_i :\rightarrow a;\ r_j : a \Rightarrow \neg a]$ como el argumento $[r_s : \sim a \Rightarrow b;\ r_t : b \rightarrow a]$ resultan incoherentes.

Resta aún refinar la relación de ataque entre argumentos a través del conjunto de prioridades entre reglas, para capturar bajo qué circunstancias un ataque resulta satisfactorio. En aquellas situaciones en las cuales el conflicto entre argumentos se circunscriba a un par de reglas rebatibles será posible aplicar el orden estricto parcial disponible. No obstante, en situaciones más complejas donde el conflicto surja de la extensión de un argumento a través de reglas estrictas, quizás haya que contrastar conjuntos de reglas. En primer lugar, se formaliza la noción de conjunto de reglas relevante a una conclusión:

Definición 2.42. **[Prakken y Sartor, 1997]**
(reglas relevantes a una conclusión)
Sea A un argumento y S un conjunto de reglas estrictas tales que $A + S$ sustentan una conclusión L. El conjunto de reglas rebatibles relevante a L en el argumento $A + S$, notado $R_L(A + S)$, se define como:

- $\{r_k\}$, en el caso que A contenga una regla rebatible r_k cuya consecuencia es L,

- $R_{L_1}(A + S) \cup \ldots \cup R_{L_n}(A + S)$, en el caso que A sea rebatible, y S contenga una regla estricta $r : L_1 \wedge \ldots \wedge L_n \rightarrow L$.

■

Finalmente, para poder comparar conjuntos de reglas rebatibles hay que extender el criterio de comparación entre pares de reglas a pares de conjuntos de reglas.

Definición 2.43. **[Prakken y Sartor, 1997]**
(comparación entre conjuntos)
Sea R y R' dos conjuntos de reglas rebatibles. Diremos que $R < R'$ si, y sólo si, existe alguna regla $r \in R$, tal que para toda regla $r' \in R'$, se verifica que $r < r'$. ■

La definición anterior captura la intuición de los autores en relación a cómo comparar conjuntos de reglas rebatibles. Señalan que un conjunto de reglas rebatibles R' resulta preferible a otro conjunto R si es el caso que R puede ser mejorado reemplazando alguna de sus reglas rebatibles por cualquiera de las reglas de R'.

Antes de caracterizar la relación de derrota entre argumentos se debe incorporar en la noción de ataque aquella información suplida por el orden estricto parcial, tercer componente de toda base de conocimiento.

Definición 2.44. **[Prakken y Sartor, 1997]**
(rebatimiento, socavamiento)
Sean A_1 y A_2 dos argumentos y sea S_1 una secuencia de reglas estrictas, tal que $A_1 + S_1$ constituye un argumento sustentando una conclusión L. Diremos que:

- A_1 *rebate* a A_2 si, y sólo si, existe una secuencia de reglas estrictas S_2, tal que $A_2 + S_2$ constituye un argumento sustentando una conclusión $\overline{L}$, y se verifica que $R_L(A_1 + S_1) \not< R_{\overline{L}}(A_2 + S_2)$.
- A_1 *socava* a A_2 si, y sólo si, $\overline{L}$ es una de las suposiciones del argumento A_2.

■

Nótese que los ataques a las suposiciones son resueltos sin apelar a la información sobre prioridades entre reglas, puesto que el estatus epistémico de toda suposición es inferior al estatus de cualquier otro argumento. En este sentido, Prakken y Sartor indican que su noción de socavamiento constituye un caso particular de la noción homónima definida por Pollock.

En consecuencia, la relación de derrota queda caracterizada en términos de los ataques por rebatimiento y por socavamiento de la siguiente manera:

Definición 2.45. **[Prakken y Sartor, 1997]**
(derrota)
Sean A_1 y A_2 dos argumentos. Diremos que A_1 *derrota* a A_2 si, y sólo si, A_1 es el argumento vacío y A_2 es un argumento incoherente, o bien:

- A_1 socava a A_2, o
- A_1 rebate a A_2, y A_2 no socava a A_1.

Diremos que A_1 *derrota estrictamente* a A_2 si, y sólo si, A_1 derrota a A_2, pero A_2 no derrota a A_1. ■

Una propiedad interesante de la definición anterior es que si A_1 ataca a A_2 entonces A_1 derrota a A_2, o A_2 derrota a A_1. En otras palabras, todo conflicto entre argumentos es dirimible.

Ejemplo 2.15. Considerando la base de conocimiento $KB = (S, D, \emptyset)$, donde

$$S = \{\} \qquad D = \{r_1 : {\sim}\neg inocente \Rightarrow inocente;\ r_2 :\Rightarrow \neg inocente\}$$

es posible formular un argumentos $A_1 = [r_1 : {\sim}\neg inocente \Rightarrow inocente]$ sustentando a *inocente*, y un argumento $A_2 = [r_2 :\Rightarrow \neg inocente]$ sustentado a $\neg inocente$. En este escenario, si bien A_1 rebate a A_2, A_1 no es capaz de derrotar a A_2, ya que A_2 socava a A_1. Por consiguiente, A_2 derrota estrictamente a A_1.

Finalmente, el estado último de los argumentos se determina apelando a la semántica basada en argumentos para la programación en lógica extendida desarrollada por Dung [Dung, 1993a]. La noción de aceptabilidad, central en el trabajo de Dung, se adapta al presente sistema de la siguiente manera:

Definición 2.46. **[Prakken y Sartor, 1997]**
(argumento aceptable)

Un argumento A se dice *aceptable* con respecto a un conjunto $args$ de argumentos si, y sólo si, cada argumento que derrota a A, es estrictamente derrotado por algún argumento contenido en $args$. ■

En particular, la semántica escéptica definida en [Dung, 1993a] (análoga a la introducida en la definición 2.8, pág. 35), es empleada para determinar el conjunto de argumentos justificados. Esta semántica se obtiene a partir de la función característica asociada a la base de conocimiento.

Definición 2.47. **[Prakken y Sartor, 1997]**
(función característica)
La *función característica* de una base de conocimiento KB, denotada F_{KB}, es una función $F_{KB} : \mathcal{P}(Args_{KB}) \longrightarrow \mathcal{P}(Args_{KB})$, donde $\mathcal{P}(Args_{KB})$ representa el conjunto de todos los subconjuntos de $Args_{KB}$, tal que para todo conjunto de argumentos S se verifica que:

$$F_{KB}(S) = \{A \in Args_{KB} \mid A \text{ es aceptable con respecto a } S\}$$

■

Esta función característica entre otras propiedades satisface el comportarse de manera monótona en relación a la inclusión de conjuntos. En consecuencia, posee un menor punto fijo unívocamente determinado, el cual denota la semántica buscada.

Definición 2.48. **[Prakken y Sartor, 1997]**
(justificación)
Sea KB una base de conocimiento y A un argumento basado en KB. En esta situación, diremos que:

- A está *justificado* si, y sólo si, A está contenido en el menor punto fijo de F_{KB}.
- A está *denegado* si, y sólo si, A es atacado por un argumento justificado.
- A es *defendible* si, y sólo si, A no está justificado ni denegado.

Por otra parte, extenderemos esta terminología a literales diciendo que un literal está *justificado* si se encuentra sustentado como conclusión de algún argumento justificado, que es *defendible* si no está justificado pero se encuentra sustentado como conclusión de algún argumento defendible, y que está *denegado* si no está justificado ni es defendible, pero se encuentra sustentado como conclusión de algún argumento denegado. ■

Ejemplo 2.16. Considerando la base de conocimiento $KB = (S, D, \{(r_1 < r_4)\})$, donde

$$S = \{\} \qquad D = \{r_1 :\Rightarrow a;\ r_2 : a \Rightarrow b;\ r_3 : \sim b \Rightarrow c;\ r_4 :\Rightarrow \neg a\}$$

se aprecia que el argumento $A_1 = [r_1 :\Rightarrow a;\ r_2 : a \Rightarrow b]$ derrota al argumento $A_2 = [r_3 : \sim b \Rightarrow c]$, ya que A_1 socava a A_2. A su vez, el argumento $A_3 = [r_4 :\Rightarrow \neg a]$ derrota por rebatimiento al argumento A_1 en virtud de las prioridades entre reglas estipuladas. En consecuencia, el menor punto fijo del operador F_{KB} contiene, entre otros, a los argumentos A_3 y A_2, por lo que A_3 y A_2 resultan justificados, y A_1 resulta denegado.

Prakken y Sartor interpretan esta definición basada en la existencia de un punto fijo como la semántica declarativa del sistema. Sin embargo, como para determinar el estado de un argumento se debe calcular el estado de la totalidad de los argumentos, los autores ensayaron una semántica alternativa basada en un dialógo entre dos contendientes.

Definición 2.49. **[Prakken y Sartor, 1997]**
(diálogo)
Un *diálogo* es una secuencia finita de movidas $(jugador_i, arg_i)$, $i > 0$, donde:

1. Si i es par, entonces $jugador_i = P$. Caso contrario, $jugador_i = O$.

2. Si $jugador_i = jugador_j = P$, $i \neq j$, entonces $arg_i \neq arj_j$.

3. Si $jugador_i = P$, $i > 1$, entonces arg_i es el menor argumento (con respecto a la inclusión de conjuntos), que derrota estrictamente a arg_{i-1}.

4. Si $jugador_i = O$, entonces arg_i es un argumento que derrota a arg_{i-1}.

■

La primer condición estipula que un diálogo se compone de una secuencia alterna de movidas introducidas por el proponente P, o el oponente O. La segunda condición evita la reiteración de movidas, con el objeto de descartar eventuales diálogos infinitos. Las últimas dos condiciones hacen recaer el "peso de la prueba" sobre P, indicando que toda movida de P debe derrotar estrictamente los argumentos introducidos por O, pero las movidas de O pueden simplemente derrotar a los argumentos introducidos por P.

Ejemplo 2.17. Considerando la base de conocimiento del ejercicio 2.16, un diálogo comenzando a partir del argumento $[r_3 : {\sim}b \Rightarrow c]$, puede discurrir como se indica a continuación:

$$\begin{array}{ll} P: & [r_3 : {\sim}b \Rightarrow c] \\ O: & [r_1 :\Rightarrow a;\ r_2 : a \Rightarrow b] \\ P: & [r_4 :\Rightarrow \neg a] \end{array}$$

Los diversos diálogos que pueden originarse a partir de un mismo argumento se estructuran a manera de árbol de diálogos de la siguiente manera:

Definición 2.50. **[Prakken y Sartor, 1997]**
(árbol de diálogos)
Un *árbol de diálogos* es un árbol finito de movidas, tal que todo camino desde la raíz a una hoja es un diálogo, y todo nodo conteniendo la movida $(jugador_i, arg_i)$, donde $jugador_i = P$, tiene por sucesores a todos los derrotadores del argumento arg_i. ■

Un árbol de diálogos explora todas las posibles interacciones entre los argumentos que aparecen en los diversos diálogos. El resultado de un árbol de diálogos se define en términos del resultado de cada uno de los diálogos que lo componen.

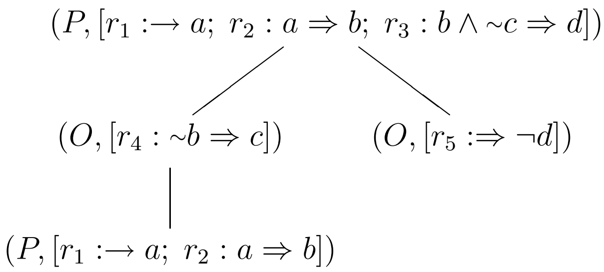


Figura 2.7: Árbol de diálogos asociado al ejemplo 2.18

Definición 2.51. **[Prakken y Sartor, 1997]**
(resultado de un diálogo)
Diremos que un jugador *gana* un diálogo cuando su contrincante agota el conjunto de sus posibles movidas. ■

La noción de resultado de un diálogo es generalizada a árboles de diálogo de la manera esperada:

Definición 2.52. **[Prakken y Sartor, 1997]**
(resultado de un árbol de diálogos)
Diremos que un jugador *gana* un árbol de diálogos si gana todos los diálogos que lo componen. ■

A partir de esta última definición, es posible parafrasear la caracterización del estado final de los argumentos (def. 2.48, pág. 72).

Definición 2.53. **[Prakken y Sartor, 1997]**
(justificación demostrable)
Un argumento A está *demostrablemente justificado* si, y sólo si, existe un árbol de diálogos, ganado por el proponente, con A en su raíz. Al igual que antes, el literal L está *demostrablemente justificado* si, y sólo si, existe algún argumento demostrablemente justificado que sustente a L. ■

Ejemplo 2.18. Considerando la base de conocimiento $KB = (S, D, \{(r_1 < r_5)\})$, donde

$$S = \{r_1 :\rightarrow a\} \qquad D = \{r_2 : a \Rightarrow b;\ r_3 : b \wedge {\sim}c \Rightarrow d;\ r_4 : {\sim}b \Rightarrow c;\ r_5 :\Rightarrow \neg d\}$$

es posible formular los siguientes argumentos:

$$A_1 = [r_1 :\rightarrow a;\ r_2 : a \Rightarrow b;\ r_3 : b \wedge {\sim}c \Rightarrow d]$$

$$A_2 = [r_4 : {\sim}b \Rightarrow c]$$

$$A_3 = [r_1 :\rightarrow a;\ r_2 : a \Rightarrow b]$$

$$A_4 = [r_5 :\Rightarrow \neg d]$$

Considerando que A_1 derrota a A_2 y A_2 derrota a A_1, que A_3 derrota estrictamente a A_2, y que A_4 derrota estrictamente a A_1, es posible formular partiendo del argumento A_1 dos diálogos, $[(P, A_1), (O, A_2), (P, A_3)]$ y $[(P, A_1), (O, A_4)]$, que pueden estructurarse en el árbol de diálogos presente en la figura 2.7. Ya que uno de los diálogos no pudo ser ganado por el proponente, se puede afirmar que A_1 no está demostrablemente justificado.

Las nociones de argumento justificado y demostrablemente justificado se relacionan de la siguiente manera:

- Para toda base de conocimiento, se verifica que los argumentos demostrablemente justificados están justificados.
- Para toda base de conocimiento finitaria,[7] se verifica que los argumentos justificados están demostrablemente justificados.

La extensión a prioridades dinámicas comparte naturalmente gran parte de los desarrollos previos, si bien las definiciones centrales del sistema deben ser adecuadas

[7]una base de conocimiento se dice finitaria si todo argumento es derrotado a lo sumo por una cantidad finita de argumento.

al nuevo escenario. Por caso, el orden estricto parcial presente en toda base de conocimiento se torna redundante, ya que el mismo será derivado rebatiblemente en el seno del sistema.

Los conceptos de regla y de argumento no requieren ser revisados. Con respecto a la representación de conocimiento, se asume la existencia de un predicado binario infijo '$\prec$', a través del cual el lenguaje objeto permite expresar las prioridades rebatibles deseadas. De acuerdo a los autores, los siguientes esquemas de axioma describen el comportamiento esperado para el predicado '$\prec$':

$$\begin{aligned}
t_1 :&\quad X \prec Y \wedge Y \prec Z \rightarrow X \prec Z \\
t_2 :&\quad X \prec Y \wedge \neg(X \prec Z) \rightarrow \neg(Y \prec Z) \\
t_3 :&\quad Y \prec Z \wedge \neg(X \prec Z) \rightarrow \neg(X \prec Y) \\
a :&\quad X \prec Y \rightarrow \neg(Y \prec X)
\end{aligned}$$

La noción de base de conocimiento es revisada prescindiendo de su tercer componente e incorporando el conjunto de axiomas anterior.

Definición 2.54. **[Prakken y Sartor, 1997]**
(base de conocimiento priorizada)
Denominaremos *base de conocimiento priorizada* a toda dupla $KB = (S, D)$, donde S y D son conjunto de reglas, S incluye a los axiomas t_1, t_2, t_3 y a, y ninguna regla de $S - \{t_1, t_2, t_3, a\}$ contiene literales default o al predicado $\prec$. ■

Las principales definiciones del sistema con prioridades estáticas dependen de la existencia de un orden estricto parcial. La siguiente definición indica como reconstruir esta información a partir de un dado conjunto de argumentos.

Definición 2.55. **[Prakken y Sartor, 1997]**
(prioridades sancionadas)
Sea $args$ un conjunto de argumentos. Diremos que el conjunto de *prioridades sancionadas* por $args$, notado $<_{args}$, es el siguiente conjunto:

$$\{(r < r') \mid r \prec r' \text{ es la conclusión de algún } A \in args\}$$

■

Esta caracterización permite adaptar la mayoría de las definiciones anteriores apelando al sistema con prioridades estáticas ya formalizado. Por ejemplo, en el contexto de una base de conocimiento priorizada (S, D) diremos que en base a un conjunto de argumentos $args$ un argumento A $args$-derrota (resp. $args$-derrota estrictamente) a un argumento B en base a (S, D) si aplicando la definición 2.45 A derrota (resp. derrota estrictamente) a B en base a $(S, D, <_{args})$. Por otra parte, como caso particular notaremos A-derrota en lugar de $\{A\}$-derrota.

Con el objeto de completar la definición de la semántica para el sistema extendido, la definición de aceptabilidad es adecuada de una manera análoga, tal como se indica a continuación:

Definición 2.56. **[Prakken y Sartor, 1997]**
(argumento priorizado aceptable)
Un argumento A se dice *priorizado aceptable* con respecto a un conjunto $args$ de argumentos si, y sólo si, cada argumento que $args$-derrota a A, es estrictamente $args$-derrotado por algún argumento contenido en $args$. ■

Prakken y Sartor encontraron que al extrapolar la noción de función característica no era posible asegurar la crucial propiedad de monotonía. Sin embargo, comprobaron que para no perder esta propiedad era necesario restringir el dominio de la función a conjuntos libre de conflictos.[8]

Definición 2.57. **[Prakken y Sartor, 1997]**
(función característica priorizada)
La *función característica priorizada* de una base de conocimiento priorizada KB, denotada G_{KB}, es una función $G_{KB} : \mathcal{C}(Args_{KB}) \longrightarrow \mathcal{P}(Args_{KB})$, donde $\mathcal{C}(Args_{KB})$ representa al conjunto de todos los subconjuntos libres de conflicto de $Args_{KB}$, y $\mathcal{P}(Args_{KB})$ representa al conjunto de todos los subconjuntos de $Args_{KB}$, tal que para todo conjunto de argumentos libre de conflictos S, se verifica que:

$$G_{KB}(S) = \{A \in Args_{KB} \mid A \text{ es priorizado aceptable con respecto a } S\}$$

■

[8]un conjunto se dice libre de conflictos si ningún par de miembros es tal que el primero ataca al segundo.

Considerando que el operador G_{KB} restringido a conjuntos libres de conflicto se comporta de manera monótona, la existencia y unicidad de su menor punto fijo está garantizada.

Definición 2.58. **[Prakken y Sartor, 1997]**
(justificación revisada)
Sea KB una base de conocimiento priorizada. Diremos que el argumento A está *justificado* en base a KB si, y sólo si, A está contenido en el menor punto fijo del operador G_{KB}. ■

Las nociones de argumento defendible y argumento denegado se definen en base a la noción de argumento justificado de manera análoga que en el sistema con prioridades estáticas (def. 2.48, pág. 72).

Ejemplo 2.19. Considerando la base de conocimiento priorizada $KB' = (S', D')$ que extiende a la base de conocimiento $KB = (S, D, \{(r_1 < r_5)\})$ del ejemplo 2.18, donde

$$S' = S \cup \{t_1, t_2, t_3, a\}$$

$$D' = D \cup \{r_6 :\Rightarrow r_1 \prec r_5;\ r_7 :\Rightarrow r_5 \prec r_1;\ r_8 :\Rightarrow r_7 \prec r_6\}$$

se aprecia que el argumento $A = [r_8 :\Rightarrow r_7 \prec r_6]$ no es $\emptyset$-derrotado por argumento alguno, estando en consecuencia presente en el menor punto fijo del operador $G_{KB'}$. Ya que A está justificado, $r_7 \prec r_6$ será una conclusión justificada, causando que $r_7 < r_6$ forme parte del orden estricto parcial final. Por lo tanto, el argumento $[r_6 :\Rightarrow r_1 \prec r_5]$ derrotará al argumento $[r_7 :\Rightarrow r_5 \prec r_1]$, tornando justificada a la conclusión $r_1 \prec r_5$. Finalmente, al incorporar $r_1 < r_5$, el ejemplo continúa de forma análoga al caso estático.

Uno de los objetivo de diseño al concebir el sistema de prueba dialéctico era no tener que calcular previamente el orden estricto parcial inducido por la totalidad de los argumentos justificados. Los autores señalan que mediante un par de sencillas restricciones es posible evitar este inconveniente, evitando explorar el estado de aquellos argumentos irrelevantes al dialogo actual.

Definición 2.59. **[Prakken y Sartor, 1997]**
(diálogo priorizado)
Un *diálogo priorizado* es una secuencia finita de movidas $(jugador_i, arg_i)$, $i > 0$, donde:

1. Si i es par, entonces $jugador_i = P$. Caso contrario, $jugador_i = O$.
2. Si $jugador_i = jugador_j = P$, $i \neq j$, entonces $arg_i \neq arj_j$.
3. Si $jugador_i = P$, $i > 1$, entonces arg_i es el menor argumento (con respecto a la inclusión de conjuntos), tal que
   - arg_{i-1} es estrictamente arg_i-derrotado por arg_i, o
   - arg_{i-1} no arg_i-derrota a arg_{i-2}.
4. Si $jugador_i = O$, entonces arg_i es un argumento que $\emptyset$-derrota a arg_{i-1}.

■

Un aspecto interesante de la definición anterior es que la derrota se determina en función del orden estricto parcial inducido sólo por el argumento actual (en el caso de P) o por $\emptyset$ (en el caso de O), sustancial mejora en relación al requerimiento inicial de determinar el orden estricto parcial inducido por la totalidad de los argumentos justificados. La noción de árbol de diálogos se extiende de la misma manera.

Definición 2.60. **[Prakken y Sartor, 1997]**
(árbol de diálogos priorizado)
Un *árbol de diálogos priorizado* es un árbol finito de movidas, tal que todo camino desde la raíz a una hoja es un diálogo priorizado, y todo nodo conteniendo la movida $(jugador_i, arg_i)$, donde $jugador_i = P$, tiene por sucesores a todos los $\emptyset$-derrotadores del argumento arg_i. ■

Las nociones de resultado de un diálogo priorizado y resultado de un árbol de diálogos priorizado se definen igual que en el sistema con prioridades estáticas. Finalmente, el conjunto de argumentos sancionados por el sistema de prueba dialéctico extendido queda cristalizado en la siguiente definición:

Definición 2.61. **[Prakken y Sartor, 1997]**
(justificación demostrable revisada)
Un argumento A está *demostrablemente justificado* si, y sólo si, existe un árbol de diálogos priorizado ganado por el proponente, con A en su raíz. De forma análoga, el literal L está *demostrablemente justificado* si, y sólo si, existe algún argumento demostrablemente justificado que sustente a L. ■

Ejemplo 2.20. Considerando la base de conocimiento priorizada $KB = (S, D)$, donde

$$S = \{t_1, t_2, t_3, a\} \qquad D = \{r_1 :\Rightarrow a;\ r_2 :\Rightarrow \neg a;\ r_3 :\Rightarrow r_2 \prec r_1\}$$

resulta posible formular los siguientes argumentos:

$$A_1 = [r_1 :\Rightarrow a]$$

$$A_2 = [r_2 :\Rightarrow \neg a]$$

$$A_3 = [r_3 :\Rightarrow r_2 \prec r_1]$$

En base a que A_3 está justificado, en virtud de no tener $\emptyset$-derrotadores, el argumento A_1 también se encuentra justificado, ya que A_1 derrota a A_2, pero A_2 no puede derrotar a A_1 en base al orden estricto parcial inducido por el argumento A_3. Por otra parte, el argumento A_1 también se encuentra demostrablemente justificado, pues el siguiente diálogo ganado por el proponente

$$\begin{array}{ll} P: & [r_1 :\Rightarrow a] \\ O: & [r_2 :\Rightarrow \neg a] \\ P: & [r_3 :\Rightarrow r_2 \prec r_1] \end{array}$$

constituye la única rama de su árbol de diálogos priorizado asociado.

Las relaciones identificadas en el sistema con prioridades estáticas entre los argumento justificados y los argumentos demostrablemente justificados se preservan en el sistema extendido, esto es:

- Para toda base de conocimiento priorizada, se verifica que los argumento demostrablemente justificados están justificados.

- Para toda base de conocimiento priorizada que sea finitaria, se verifica que los argumentos justificados están demostrablemente justificados.

2.2.5. El sistema Simari-Loui

La tesis doctoral de Guillermo R. Simari introdujo un distintivo sistema de razonamiento rebatible basado en la argumentación, que recombinaba las ideas de J. Pollock [Pollock, 1987] y D. Poole [Poole, 1985]. Si bien este sistema apuntaba a ser más práctico que innovador, generó mucha repercusión en la comunidad de aquellos tiempos, influenciando incluso diversos desarrollos posteriores. Los principales aportes de esta tesis fueron sintetizados en el artículo [Simari y Loui, 1992], razón por la cual actualmente se suele referenciar a este formalismo de argumentación rebatible como sistema Simari-Loui [Prakken, 1993, Vreeswijk, 1993b].

Las inferencias en el sistema Simari-Loui están sustentadas por argumentos, donde un argumento es una pieza tentativa de razonamiento que provee de evidencia a favor de su conclusión. No obstante, es factible que algunos argumentos sustenten conclusiones en desacuerdo: resulta necesario instaurar un mecanismo que permita resolver estos conflictos. En este sentido, se incluye un procedimiento definido a tal efecto que determina cuáles son los argumentos que están justificados. Sucintamente, un argumento está justificado si los argumentos con los cuales está en conflicto no lo están. En lo que resta de la presente sección analizaremos las principales nociones que integran este sistema argumentativo.

Como punto de partida se asume la existencia de un lenguaje de primer orden $\mathcal{L}$, provisto de sus correspondientes axiomas y reglas de inferencia (por ejemplo, los axiomas y las reglas de inferencia discutidos en [Davis, 1989] o en [Fitting, 1996]). Este lenguaje se emplea para codificar dos tipos de conocimiento: el *conocimiento estricto*, de carácter incuestionable, y el *conocimiento rebatible*, de carácter tentativo. El conocimiento estricto se representa mediante fórmulas de $\mathcal{L}$. En contraste, el conocimiento rebatible se representa mediante una relación binaria metalingüistica, con la particularidad de que sólo involucra fórmulas de $\mathcal{L}$ con variables libres en común,[9] denotada a través del símbolo infijo '$\succ\!-$'. Los componentes de esta relación se conocen como *reglas rebatibles*. La interpretación pretendida para una regla rebatible '$\alpha \succ\!- \beta$' es *"razones para creer en el antecedente α proveen razones para creer en el consecuente β"*.

Definición 2.62. **[Simari y Loui, 1992]**
(base de conocimiento)
Denominaremos *base de conocimiento*[10] KB a todo par $(\mathcal{K}, \Delta)$, donde $\mathcal{K}$ es un conjunto consistente de fórmulas, y Δ es un conjunto finito de reglas rebatibles. ■

El conjunto $\mathcal{K}$ puede a su vez ser particionado en dos subconjunto disjuntos: el conjunto de conocimiento necesario $\mathcal{K}_N$, conteniendo las fórmulas de $\mathcal{K}$ con variables libres, y el conjunto de conocimiento contingente $\mathcal{K}_C$, conteniendo las restantes fórmulas de $\mathcal{K}$. A lo largo de esta sección, simplificaremos la notación asumiendo en todo momento la existencia de una base de conocimiento $KB = (\mathcal{K}, \Delta)$, arbitraria pero fija. La base de conocimiento tiene por objeto sustentar *argumentos*. Para definir formalmente esta noción, es necesario describir qué conjunto de literales se derivan a partir de una dada base de conocimiento.

Definición 2.63. **[Simari y Loui, 1992]**
(derivación rebatible)
Sea Γ un subconjunto de $\mathcal{K} \cup \Delta^{\downarrow}$, donde $\Delta^{\downarrow}$ denota el conjunto de instancias fijas de Δ. Diremos que h se *deriva rebatiblemente* a partir de Γ si, y sólo si, existe una secuencia finita $\alpha_1, \ldots, \alpha_n$, tal que $\alpha_n = h$, y para todo α_i, $1 \leq i < n$, se verifica que $\alpha_i \in \Gamma$, o bien α_i es un axioma de $\mathcal{L}$, o bien α_i es una consecuencia directa de la aplicación de alguna de las reglas de inferencia de $\mathcal{L}$ a elementos anteriores de esta secuencia. En este sentido, las instancias fijas de las reglas rebatibles serán

[9]si relacionara fórmulas fijas (*i.e.*, sin variables), no representaría conocimiento tentativo.
[10]*estructura lógica rebatible* según la terminología empleada en [Simari y Loui, 1992].

consideradas como implicaciones materiales en lo que a la aplicación de reglas de inferencia concierne. ■

Ejemplo 2.21. Sea $KB = (\mathcal{K}, \Delta)$ una base de conocimiento, donde

$$\mathcal{K} = \{ave(tweety);\ pinguino(pengo);\ pinguino(X) \rightarrow ave(X)\}$$

$$\Delta = \{ave(X) \succ vuela(X);\ pinguino(X) \succ \neg vuela(X)\}$$

En este contexto, es posible derivar rebatiblemente $vuela(tweety)$, pero no es posible derivar rebatiblemente $\neg vuela(tweety)$. En contraste, es posible derivar tanto $vuela(pengo)$ como $\neg vuela(pengo)$.

Por lo general, los sistema argumentativos asocian a los argumentos con algún tipo de prueba cuya fuerza conclusiva es inferior a la de una deducción lógica tradicional. El sistema Simari-Loui equipara argumentos con derivaciones rebatibles, las cuales admiten el uso de información tentativa.

Definición 2.64. **[Simari y Loui, 1992]**
(argumento, estructura de argumento)
Diremos que $\mathcal{A} \subseteq \Delta^{\downarrow}$, donde $\Delta^{\downarrow}$ denota el conjunto de instancias fijas de Δ, constituye un *argumento* para h si, y sólo si, se satisfacen las siguientes condiciones:

1. $\mathcal{K} \cup \mathcal{A} \mathrel{|\!\sim} h$.
2. $\mathcal{K} \cup \mathcal{A} \not\mathrel{|\!\sim} \bot$.
3. $\nexists\ \mathcal{A}' \subseteq \mathcal{A}$ tal que $\mathcal{K} \cup \mathcal{A}' \mathrel{|\!\sim} h$.

Si $\mathcal{A}$ es un argumento para h, diremos que $\langle \mathcal{A}, h \rangle$ constituye una *estructura de argumento.* ■

Cabe rescatar que esta definición toca dos aspectos controversiales. En primer lugar, los argumento son caracterizados mediante el conjunto de reglas rebatibles requerido para derivar (rebatiblemente) una conclusión. Otros autores han sostenido que esta decisión es cuestionable [Prakken y Vreeswijk, 2002]:

> "...esta noción de argumento parece fuera de lo común, pues no hace referencia a un arból o a una secuencia de reglas de inferencia. Por el contrario, la definición [de argumento en el sistema Simari-Loui] apenas requiere de un argumento que sea una colección de reglas sin un orden preestablecido que en conjunto implican a una cierta conclusión."

Por el contrario, podría objetarse que es el árbol de deducción o la secuencia de prueba los que resultan irrelevantes en un sistema argumentativo, ya que sólo la porción de razonamiento de caracter no deductivo podrá ser cuestionada en el posterior proceso argumentativo llevado a cabo para determinar el estado de los argumentos.

En segundo lugar, el requerir minimalidad en la definición anterior también suele ser objetado, pues verificar esta condición resulta costoso desde un punto de vista computacional. Las deducciones en una lógica convencional no impone esta restricción, puesto que toda demostración, independientemente de su longitud, presenta la misma fuerza conclusiva. En contraste, la fuerza conclusiva de una derivacíon rebatible depende directamente de la cantidad de información tentativa que se haya utilizado. En consecuencia, resulta razonable considerar como válidos sólo aquellos argumentos que hagan el menor uso posible de las reglas rebatibles.

Ejemplo 2.22. Sea $KB = (\mathcal{K}, \Delta)$ una base de conocimiento, donde

$$\mathcal{K} = \{p(a);\ p(X) \rightarrow q(X);\ r(X) \vee s(X) \rightarrow t(X)\}$$

$$\Delta = \{q(X) \succ\!\!- r(X);\ q(X) \succ\!\!- s(X);\ q(X) \succ\!\!- \neg p(X)\}$$

En este contexto, se pueden construir dos estructuras de argumentos para $t(a)$, $\langle \mathcal{A}_1, t(a)\rangle$ y $\langle \mathcal{A}_2, t(a)\rangle$, donde $\mathcal{A}_1 = \{q(a) \succ\!\!- r(a)\}$ y $\mathcal{A}_2 = \{q(a) \succ\!\!- s(a)\}$. Sin embargo, $\{q(a) \succ\!\!- \neg p(a)\}$ no constituye un argumento para $t(a)$, pues no satisface el requerimiento de consistencia. Por otra parte, $\{q(a) \succ\!\!- r(a);\ q(a) \succ\!\!- s(a)\}$ tampoco constituye un argumento para $t(a)$, pues no satisface el requerimiento de minimalidad.

La derrota entre argumentos se caracteriza en este sistema combinando la noción de conflicto con la aplicación de un criterio de preferencia. En particular, se adopta una variante del *criterio de especificidad* [Poole, 1985] como mecanismo de contrastación entre argumentos.

Definición 2.65. **[Simari y Loui, 1992]**
(especificidad)
Dadas dos estructuras de argumento $\langle \mathcal{A}_1, h_1 \rangle$ y $\langle \mathcal{A}_2, h_2 \rangle$, el argumento $\mathcal{A}_1$ para h_1 resulta *estrictamente más específico* que el argumento $\mathcal{A}_2$ para h_2 si, y sólo sí:

1. Para toda fórmula fija e de $\mathcal{L}$, tal que $\mathcal{K}_N \cup \{e\} \cup \mathcal{A}_1 \mid\!\sim h_1$ y $\mathcal{K}_N \cup \{e\} \not\mid\!\sim h_1$, entonces $\mathcal{K}_N \cup \{e\} \cup \mathcal{A}_2 \mid\!\sim h_2$.

2. Existe una fórmula fija e de $\mathcal{L}$, tal que $\mathcal{K}_N \cup \{e\} \cup \mathcal{A}_2 \mid\!\sim h_2$ y $\mathcal{K}_N \cup \{e\} \not\mid\!\sim h_2$, entonces $\mathcal{K}_N \cup \{e\} \cup \mathcal{A}_1 \not\mid\!\sim h_1$.

Cuando la estructura de argumento $\langle \mathcal{A}_1, h_1 \rangle$ sea estrictamente más específica que la estructura de argumento $\langle \mathcal{A}_2, h_2 \rangle$, notaremos $\langle \mathcal{A}_1, h_1 \rangle \succ \langle \mathcal{A}_2, h_2 \rangle$. ■

Esta definición puede ser reformulada con un cariz semántico mediante la introducción de terminología auxiliar: diremos que una fórmula e activa a un argumento $\langle \mathcal{A}, h \rangle$ si $\mathcal{K}_N \cup \{e\} \cup \mathcal{A} \mid\!\sim h$, o, en otras palabras, cuando e contenga todo el conocimiento contingente necesario para reconstruir la derivación rebatible de h. Además, diremos que un activador e para un argumento $\langle \mathcal{A}, h \rangle$ es no trivial si $\mathcal{K}_N \cup \{e\} \not\mid\!\sim h_2$, o, en otras palabras, cuando e no vuelva innecesario al aporte del argumento $\mathcal{A}$ en relación a la derivación rebatible de h. En resumen, la definición 2.65 determina que un argumento $\mathcal{A}_1$ para h_1 es estrictamente más específico que el argumento $\mathcal{A}_2$ para h_2 si para toda fórmula fija de $\mathcal{L}$ que activa de forma no trivial a $\mathcal{A}_1$, también activa a $\mathcal{A}_2$, y por otra parte existe una fórmula fija de $\mathcal{L}$ que activa de forma no trivial a $\mathcal{A}_2$, pero no activa a $\mathcal{A}_1$.

El criterio de especificidad captura dos principios relativamente intuitivos, que han sido ampliamente estudiados en el contexto de las redes semánticas.[11] El primer principio premia a los argumentos que involucran una mayor cantidad de información en la sanción de sus conclusiones.

[11]también conocidas como jerarquías de herencia [Etherington y Reiter, 1983].

Ejemplo 2.23. Sea $KB = (\mathcal{K}, \Delta)$ una base de conocimiento, donde

$$\mathcal{K} = \{u(a);\ v(a)\}$$

$$\Delta = \{u(a) \wedge v(a) \succ w(a);\ u(a) \succ \neg w(a)\}$$

En este contexto, la estructura de argumento $\langle \{u(a) \wedge v(a) \succ w(a)\}, w(a) \rangle$ resulta estrictamente más específica que $\langle \{u(a) \succ \neg w(a)\}, \neg w(a) \rangle$, pues todo activador del primer argumento también activa al segundo argumento, pero $u(a)$ activa al segundo sin activar al primero. Por lo tanto:

$$\langle \{u(a) \wedge v(a) \succ w(a)\}, w(a) \rangle \prec \langle \{u(a) \succ \neg w(a)\}, \neg w(a) \rangle$$

El segundo principio reconciliado en el seno del criterio de especificidad favorece a aquellos argumentos que arriban a sus conclusiones de manera más directa, encadenando una menor cantidad de reglas rebatibles.

Ejemplo 2.24. Sea $KB = (\mathcal{K}, \Delta)$ una base de conocimiento, donde

$$\mathcal{K} = \{u(a)\}$$

$$\Delta = \{u(a) \succ w(a);\ u(a) \succ v(a);\ v(a) \succ \neg w(a)\}$$

En este contexto, la estructura de argumento $\langle \{u(a) \succ w(a)\}, w(a) \rangle$ resulta estrictamente más específica que $\langle \{u(a) \succ v(a);\ v(a) \succ \neg w(a)\}, \neg w(a) \rangle$, pues todo activador del primer argumento también activa al segundo argumento, pero $v(a)$ activa al segundo sin activar al primero. Por lo tanto:

$$\langle \{u(a) \succ w(a)\}, w(a) \rangle \prec \langle \{u(a) \succ v(a);\ v(a) \succ \neg w(a)\}, \neg w(a) \rangle$$

Para establecer la relación de derrota se debe formalizar qué se entiende por conflicto entre argumentos. En primer lugar, se debe definir una noción previa que captura el *desacuerdo* entre argumentos.

Definición 2.66. **[Simari y Loui, 1992]**
(desacuerdo)
Diremos que dos argumento $\mathcal{A}_1$ para h_1 y $\mathcal{A}_2$ para h_2 están en *desacuerdo* si, y sólo si, $\mathcal{K} \cup \{h_1, h_2\} \mathrel{|\!\sim} \perp$. ■

El conflicto entre argumentos, denominado *contra-argumentación* en este sistema, se define como una generalización de la noción de desacuerdo entre argumentos. La idea básica consiste en capturar no sólo los conflictos a nivel de la última conclusión de un argumento, sino abarcar a todas sus conclusiones. A tal efecto, debemos apelar al concepto auxiliar *subargumento*: diremos que $\langle \mathcal{A}, h \rangle$ es un subargumento de $\langle \mathcal{B}, h' \rangle$ si, y sólo si, $\mathcal{A} \subseteq \mathcal{B}$.

Definición 2.67. **[Simari y Loui, 1992]**
(contra-argumentación)
Diremos que el argumento $\mathcal{A}_1$ para h_1 *contra-argumenta* al argumento $\mathcal{A}_2$ para h_2 en el punto h si, y sólo si, existe un subargumento $\langle \mathcal{A}, h \rangle$ de $\langle \mathcal{A}_2, h_2 \rangle$ tal que $\langle \mathcal{A}_1, h_1 \rangle$ y $\langle \mathcal{A}, h \rangle$ están en desacuerdo. ■

La relación de desacuerdo resulta simétrica por definición. En contraste, la relación de contra-argumentación no es simétrica, si bien es factible que un par de argumentos se contra-argumenten recíprocamente. El criterio de especificidad permite refinar la noción de contra-argumentación para caracterizar a la relación de *derrota* entre argumentos.

Definición 2.68. **[Simari y Loui, 1992]**
(derrota)
Un argumento $\mathcal{A}_1$ para h_1 *derrota* al argumento $\mathcal{A}_2$ para h_2 si, y sólo si, existe un subargumento $\langle \mathcal{A}, h \rangle$ de $\langle \mathcal{A}_2, h_2 \rangle$ tal que $\langle \mathcal{A}_1, h_1 \rangle$ contra-argumenta a $\langle \mathcal{A}_2, h_2 \rangle$ en el punto h y $\langle \mathcal{A}_1, h_1 \rangle$ resulta estrictamente más específico que $\langle \mathcal{A}, h \rangle$. ■

Ejemplo 2.25. Sea $KB = (\mathcal{K}, \Delta)$ una base de conocimiento, donde

$$\mathcal{K} = \{ave(tweety);\ débil(tweety);\ pinguino(pengo);\ pinguino(X) \rightarrow ave(X)\}$$

$$\Delta = \{ave(X) \succ\!\!- vuela(X);\ débil(X) \succ\!\!- \neg vuela(X);\ pinguino(X) \succ\!\!- \neg vuela(X)\}$$

En base a KB pueden formularse los siguientes argumentos:

$\mathcal{A}_1 = \{ave(fred) \succ\!\!- vuela(fred)\}$, para $vuela(fred)$.

$\mathcal{A}_2 = \{débil(fred) \succ\!\!- \neg vuela(fred)\}$, para $\neg vuela(fred)$.

$\mathcal{A}_3 = \{ave(pengo) \succ\!\!- vuela(pengo)\}$, para $vuela(pengo)$.

$\mathcal{A}_4 = \{pinguino(pengo) \succ\!\!- \neg vuela(pengo)\}$, para $\neg vuela(pengo)$.

En este contexto, se verifican entre otras las siguiente relaciones: el argumento $\langle \mathcal{A}_1, vuela(fred)\rangle$ está en desacuerdo con el argumento $\langle \mathcal{A}_2, \neg vuela(fred)\rangle$, y el argumento $\langle \mathcal{A}_2, \neg vuela(fred)\rangle$ contra-argumenta al argumento $\langle \mathcal{A}_1, vuela(fred)\rangle$. No obstante, los argumentos $\langle \mathcal{A}_1, vuela(fred)\rangle$ y $\langle \mathcal{A}_2, \neg vuela(fred)\rangle$ no están relacionados por el criterio de especificidad. Por otra parte, el argumento $\langle \mathcal{A}_4, \neg vuela(pengo)\rangle$ es estrictamente más específico que el argumento $\langle \mathcal{A}_3, vuela(pengo)\rangle$. En consecuencia, el argumento $\langle \mathcal{A}_4, \neg vuela(pengo)\rangle$ derrota al argumento $\langle \mathcal{A}_3, vuela(pengo)\rangle$.

El estado final de un argumento depende del estado final de sus derrotadores. Como los derrotadores son a su vez argumentos, también debe determinarse el estado de los derrotadores que a su vez derrotan a estos derrotadores. Esta situación puede reiterarse un cantidad arbitraria de veces. El sistema Simari-Loui estructura este análisis recursivo a través de una definición basada en niveles análoga a la introducida por J. Pollock (def. 2.14, pág. 44).

Definición 2.69. **[Simari y Loui, 1992]**
(argumentos activos por niveles)

Diremos que los argumentos construíbles a partir de una base de conocimiento se encuentran *activos* en los diversos niveles como argumentos de *soporte* o de *interferencia*, de acuerdo a las siguientes restricciones:

1. Todos los argumentos construíbles están activos en el (nivel 0) como argumentos de *soporte* y como argumentos de *interferencia*.

2. Un argumento $\langle \mathcal{A}_1, h_1 \rangle$ está activo en el (nivel $i{+}1$) como argumento de *soporte* si, y sólo si, no existe argumento alguno $\langle \mathcal{A}_2, h_2 \rangle$, activo en el (nivel i) como argumento de interferencia, tal que el argumento $\langle \mathcal{A}_2, h_2 \rangle$ contra-argumente al argumento $\langle \mathcal{A}_1, h_1 \rangle$ en el punto h, para algún literal h.

3. Un argumento $\langle \mathcal{A}_1, h_1 \rangle$ está activo en el (nivel $i + 1$) como argumento de *interferencia* si, y sólo si, no existe argumento alguno $\langle \mathcal{A}_2, h_2 \rangle$, activo en el (nivel i) como argumento de interferencia, tal que el argumento $\langle \mathcal{A}_2, h_2 \rangle$ derrote al argumento $\langle \mathcal{A}_1, h_1 \rangle$.

■

Si un argumento logra eventualmente permanecer activo a partir de un nivel k, ninguno de sus derrotadores estará activo en los niveles superiores a k. Esta situación captura la intuición inicial con respecto a qué argumentos deben ser aceptados como sólidas justificaciones de sus conclusiones.

Definición 2.70. **[Simari y Loui, 1992]**
(justificación)
Diremos que una estructura de argumento $\langle \mathcal{A}, h \rangle$ constituye una *justificación* para h si, y sólo si, existe un k a partir del cual para todo $k' \geq k$, $\langle \mathcal{A}, h \rangle$ está activo en el (nivel k') como argumento de soporte. ■

Ejemplo 2.26. Sea $KB = (\mathcal{K}, \Delta)$ una base de conocimiento, donde:

$$\mathcal{K} = \{gallina(clueca);\ asustada(clueca);\ gallina(X) \rightarrow ave(X)\}$$

$$\Delta = \{ave(X) \succ vuela(X);\ gallina(X) \succ \neg vuela(X),$$
$$gallina(X) \wedge asustada(X) \succ vuela(X)\}$$

En base a KB, pueden formularse los siguientes argumentos:

$\mathcal{A}_1 = \{ave(clueca) \succ vuela(clueca)\}$, para $vuela(clueca)$.

$\mathcal{A}_2 = \{gallina(clueca) \succ \neg vuela(clueca)\}$, para $\neg vuela(clueca)$.

$\mathcal{A}_3 = \{gallina(clueca) \wedge asustado(clueca) \succ vuela(clueca)\}$, para $vuela(clueca)$.

Considerando que $\mathcal{A}_2$ resulta estrictamente más específico que $\mathcal{A}_1$, y a su vez $\mathcal{A}_3$ resulta estrictamente más específico que $\mathcal{A}_2$, los argumentos activos en cada nivel como argumentos de soporte y de interferencia son los siguientes:

Nivel	Soporte	Interferencia
0	$\mathcal{A}_1$, $\mathcal{A}_2$, $\mathcal{A}_3$	$\mathcal{A}_1$, $\mathcal{A}_2$, $\mathcal{A}_3$
1	–	$\mathcal{A}_3$
2	$\mathcal{A}_1$, $\mathcal{A}_3$	$\mathcal{A}_1$, $\mathcal{A}_3$
3	$\mathcal{A}_1$, $\mathcal{A}_3$	$\mathcal{A}_1$, $\mathcal{A}_3$
$\vdots$	$\vdots$	$\vdots$

Resulta evidente que los argumentos $\mathcal{A}_1$ y $\mathcal{A}_3$ permanecen activos como argumentos de soporte en todo nivel $k \geq 2$. En consecuencia, se puede afirmar que los argumentos $\mathcal{A}_1$ y $\mathcal{A}_3$ están justificados.

La caracterización para los argumentos justificados depende directamente de la definición de argumentos activos por niveles. No obstante, esta definición resulta a los estándares de hoy en día un tanto oscura y difícil de aplicar. Por caso, H. Prakken en [Prakken y Vreeswijk, 2002] confunde esta definición por niveles—que necesariamente mantiene dos conjuntos de argumentos en cada nivel—con aquella que asocia tan sólo un conjunto de argumentos por nivel, la cual establece que:

- Todos los argumentos están activos en el (nivel 0).
- Un argumento esta activo en el (nivel $i + 1$) si, y sólo si, no es derrotado por argumento activo en el (nivel i).
- Un argumento está justificado si existe un k, tal que para todo $k' \geq k$ se sanciona a este argumento como activo en el (nivel k').

Si bien ambas definiciones presentan un espíritu análogo, la caracterización sugerida por Prakken sólo modela el conjunto de argumentos activos como argumentos de interferencia, los cuales efectivamente se caracterizan por la relación de derrota. En contraste, la justificación en el sistema Simari-Loui está basada en el conjunto de argumentos activos como argumentos de soporte, no capturados en esta definición, pues dependen de la relación de contra-argumentación y no de la relación de derrota.

Esta falta de naturalidad en lo que respecta a la caracterización del estado de los argumentos es atendida en la reformulación dialéctica del sistema Simari-Loui [Simari et al., 1994]. La enrevezada definición por niveles es reemplazada por la novel noción de árbol de dialéctica, la cual resulta más natural y afín al propósito del sistema (en concordancia con lo comentado en la sección 2.1.2). Por otra parte, esta reformulación es una de las primeras formalizaciones de la argumentación rebatible en incorporar un análisis acerca de la argumentación falaz, un tipo de argumentación capaz de justificar conclusiones carentes de fundamento. Considerando que esta reformulación constituye un refinamiento del sistema Simari-Loui, resulta instructivo contrastar sus principales definiciones con aquellas presentes en el sistema original.

La reformulación del sistema mantiene la estructura de la base de conocimiento (def. 2.62, pág. 82), presente en el sistema original, así como las nociones de derivación rebatible (def. 2.63, pág. 83), y argumento (def. 2.64, pág. 83). En particular, se extiende la noción de derrota para incorporar dentro de ésta a parte del protagonismo que la relación de contra-argumentación tenía en la definición de argumentos activos por niveles (def. 2.69, pág. 89).

Definición 2.71. **[Simari et al., 1994]**
(derrotador propio, derrotador de bloqueo)

Un argumento $\mathcal{A}_1$ para h_1 derrota al argumento $\mathcal{A}_2$ para h_2 si, y sólo si, existe un subargumento $\langle \mathcal{A}, h \rangle$ de $\langle \mathcal{A}_2, h_2 \rangle$ tal que $\langle \mathcal{A}_1, h_1 \rangle$ contra-argumenta a $\langle \mathcal{A}_2, h_2 \rangle$ en el punto h y se verifica que:

(1) el $\langle \mathcal{A}_1, h_1 \rangle$ resulta estrictamente más específico que $\langle \mathcal{A}, h \rangle$, o bien

(2) el $\langle \mathcal{A}_1, h_1 \rangle$ no está relacionado por especificidad con $\langle \mathcal{A}, h \rangle$.

Si $\langle \mathcal{A}_1, h_1 \rangle$ derrota a $\langle \mathcal{A}_2, h_2 \rangle$ por (1), el argumento $\mathcal{A}_1$ constituye un *derrotador propio*. Caso contrario, si $\langle \mathcal{A}_1, h_1 \rangle$ derrota a $\langle \mathcal{A}_2, h_2 \rangle$ por (2), el argumento $\mathcal{A}_1$ constituye un *derrotador por bloqueo.* ■

La principal modificación en términos estructurales se aprecia en la formalización de la noción de justificación mediante un análisis dialéctico, eliminando las confusiones que causaba la definición original por niveles. El análisis dialéctico sólo involucra los argumentos que intervienen en la determinación del estado de un dado argumento, en franco contraste a la definición 2.69, la cual necesariamente determinaba el estado de todos los argumentos construíbles antes de poder establecer el estado de un argumento en particular. Este análisis dialéctico es estructurado a través de la noción de árbol de dialéctica.

Definición 2.72. **[Simari et al., 1994]**
(árbol de dialéctica)
Sea $\langle \mathcal{A}, h \rangle$ una estructura de argumento. Un *árbol de dialéctica* para $\langle \mathcal{A}, h \rangle$, denotado como $\mathcal{T}_{\langle \mathcal{A}, h \rangle}$, se construye la siguiente manera:

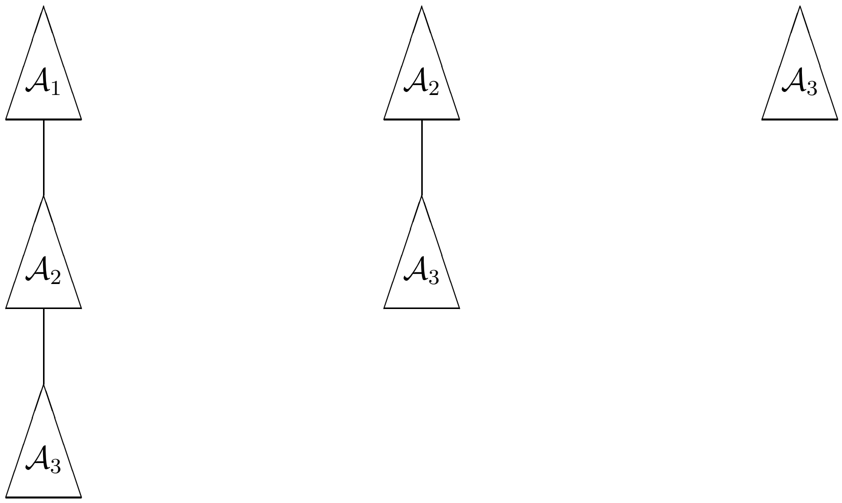
- Si la estructura de argumento $\langle \mathcal{A}, h \rangle$ no posee derrotadores propios ni por bloqueo, entonces su árbol de dialéctica asociado se compone de un único nodo conteniendo a $\langle \mathcal{A}, h \rangle$.

- Si la estructura de argumento $\langle \mathcal{A}, h \rangle$ posee como derrotadores propios o por bloqueo a los argumentos $\langle \mathcal{A}_1, h_1 \rangle, \ldots, \langle \mathcal{A}_n, h_n \rangle$, entonces su árbol de dialéctica asociado se obtiene colocando a los nodos raíces de los árboles de dialéctica $\mathcal{T}_{\langle \mathcal{A}_1, h_1 \rangle}, \ldots, \mathcal{T}_{\langle \mathcal{A}_n, h_n \rangle}$ como sucesores de un nodo conteniendo a $\langle \mathcal{A}, h \rangle$.

■

Ejemplo 2.27. Continuando con el ejemplo 2.26, el análisis dialéctico de los argumentos $\mathcal{A}_1$, $\mathcal{A}_2$ y $\mathcal{A}_3$ puede apreciar en la figura 2.8.

En un árbol de dialéctica se relaciona explícitamente cada argumento con la totalidad de sus derrotadores. Considerando que el estado de un argumento depende



$\mathcal{T}_{\langle \mathcal{A}_1, vuela(clueca) \rangle}$ $\mathcal{T}_{\langle \mathcal{A}_2, \neg vuela(clueca) \rangle}$ $\mathcal{T}_{\langle \mathcal{A}_3, vuela(clueca) \rangle}$

Figura 2.8: Análisis dialéctico de los argumentos del ejemplo 2.26

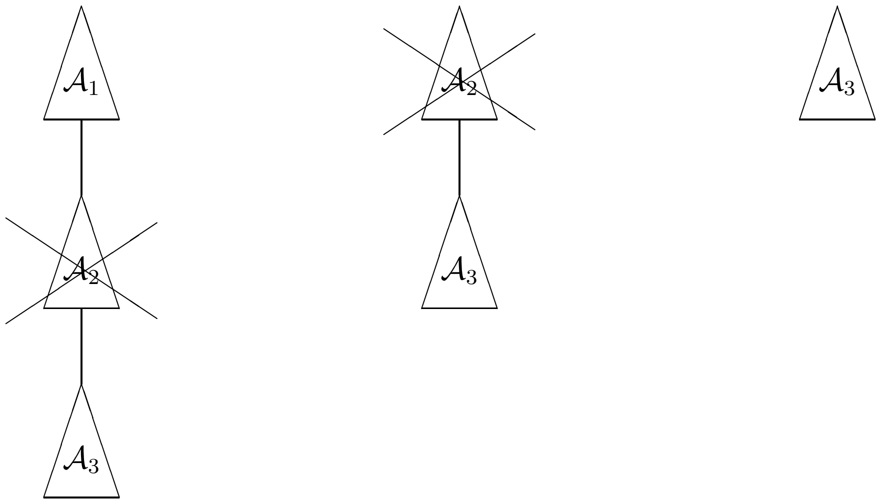
únicamente del estado de sus derrotadores, es posible establecer el estado del argumento contenido en la raíz de árbol de dialéctica empleando sólo la información ya presente en el árbol. La siguiente definición describe el procedimiento de marcado recursivo que lleva a cabo esta tarea:

Definición 2.73. **[Simari et al., 1994]**
(marcado de un árbol de dialéctica)
Sea $\mathcal{T}_{\langle \mathcal{A},h \rangle}$ el árbol de dialéctica asociado a la estructura de argumento $\langle \mathcal{A}, h \rangle$. Los nodos presentes en $\mathcal{T}_{\langle \mathcal{A},h \rangle}$ pueden ser marcados como nodos *derrotados* o *no derrotados* de acuerdo a la siguiente caracterización:

- Un nodo será marcado como *nodo derrotado* si, y sólo si, al menos uno de sus sucesores está marcado como nodo no derrotado.

- Un nodo será marcado como *nodo no derrotado* si, y sólo si, todos sus sucesores están marcados como nodos derrotados.

■



$\mathcal{T}_{\langle \mathcal{A}_1, vuela(clueca)\rangle}$ $\mathcal{T}_{\langle \mathcal{A}_2, \neg vuela(clueca)\rangle}$ $\mathcal{T}_{\langle \mathcal{A}_3, vuela(clueca)\rangle}$

Figura 2.9: Marcado de los árboles de dialéctica del ejemplo 2.27

Ejemplo 2.28. Los árboles de dialéctica desarrollados en el ejemplo 2.27 pueden ser marcados tal como se aprecian en la figura 2.9. En esta ocasión, los argumentos derrotados son distinguidos de los argumentos no derrotados a través de un par líneas cruzadas.

Considerando que los nodos hoja del árbol de dialéctica pueden ser marcados inmediatamente como nodos no derrotados (pues "todos" sus sucesores están marcados como nodos derrotados), esta definición en esencia sugiere un procedimiento de marcado desde las hojas hacia la raíz.

Por último, es posible parafrasear la noción de argumento justificado (def. 2.70, pág. 89), en términos del análisis dialéctico estructurado mediante árboles de dialéctica y sus correspondientes marcados.

Definición 2.74. **[Simari et al., 1994]**
(justificación)
Sea $\langle \mathcal{A}, h\rangle$ una estructura de argumento. Diremos que $\langle \mathcal{A}, h\rangle$ constituye una *justificación* si, y sólo si, es posible marcar al nodo raíz de su árbol de dialéctica asociado $\mathcal{T}_{\langle \mathcal{A}, h\rangle}$ como nodo no derrotado. ■

Otra innovación introducida en la reformulación del sistema Simari-Loui es el refinamiento su la relación de consecuencia, producto del análisis dialéctico llevado a cabo y de la distinción entre derrotadores propios y de bloqueo. En este contexto, una consulta h puede obtener las siguientes respuestas:

- YES: si existe una estructura de argumento $\langle \mathcal{A}, h \rangle$ que constituya una justificación.
- NO: si para cada estructura de argumento $\langle \mathcal{A}, h \rangle$, en su correspondiente árbol de dialéctica $\mathcal{T}_{\langle \mathcal{A}, h \rangle}$ algún derrotador propio del nodo raíz está marcado como no derrotado.
- UNDECIDED: si para cada estructura de argumento $\langle \mathcal{A}, h \rangle$, en su correspondiente árbol de dialéctica $\mathcal{T}_{\langle \mathcal{A}, h \rangle}$ todos los derrotadores propios del nodo raíz están marcados como derrotados, pero existe al menos un derrotador de bloqueo del nodo raíz marcado como no derrotado.
- UNKNOWN: si no es posible construir una estructura de argumento para h.

Ejemplo 2.29. De acuerdo al marcado de los árboles de dialéctica desarrollados en el ejemplo 2.28, la consulta $vuela(clueca)$ recibe como respuesta YES, ya que se encuentra sustentada por las estructuras de argumento justificadas $\langle \mathcal{A}_1, vuela(clueca) \rangle$ y $\langle \mathcal{A}_3, vuela(clueca) \rangle$. En contraste, la consulta $\neg vuela(clueca)$ recibe como respuesta NO, ya que el único argumento que la sustenta está derrotado por un derrotador propio no derrotado. Por último, la consulta $vuela(teté)$ recibe como respuesta UNKNOWN, ya que no es posible construir un argumento que sustente a esta consulta.

Esta reformulación constituye uno de los primeros formalismos en analizar el controversial tópico de la argumentación falaz. Se distinguen tres tipos de argumentación no deseada:

Argumentación circular: este tipo de falacia ha sido largamente estudiada en el campo de la filosofía. Constituye una instancia del razonamiento inválido

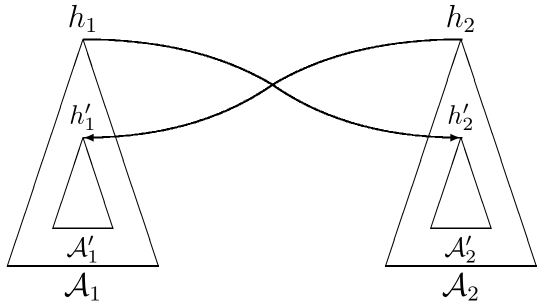


Figura 2.10: Derrotadores recíprocos en el sistema Simari-Loui

conocido como *petitio principii* (que refuta sus bases), donde la tesis que se está demostrando es reintroducida como evidencia en su favor.

Argumentación contradictoria: razonamiento claramente inválido, conocido bajo la denominación *argumentum non sequitur* (la conclusión no se deriva de la evidencia), donde se sostienen por verdaderas y en simultáneo a un conjunto inconsistente de razones.

Argumentación recíproca: este tipo de falacia sólo se manifiesta en el sistema Simari-Loui, que dada su particular definición de derrota permite la formulación de argumentos que se derroten recíprocamente.

En primer lugar consideraremos el tipo de falacia más simple de evitar (la argumentación recíproca) y posteriormente analizaremos como evitar los otros casos de razonamiento falaz (la argumentación circular y la argumentación contradictoria). La argumentación recíproca atañe a la relación de derrota; si bien la relación de especificidad constituye un orden parcial, aún es posible formular argumentos que se derroten recíprocamente. Por ejemplo, la figura 2.10 esquematiza los argumentos $\mathcal{A}_1$ y $\mathcal{A}_2$ que se derrotan mutuamente,[12] pues el argumento $\mathcal{A}_1$ es estrictamente más específico que el subargumento $\mathcal{A}'_2$ del argumento $\mathcal{A}_2$ y a su vez, el argumento $\mathcal{A}_2$ es estrictamente más específico que el subargumento $\mathcal{A}'_1$ del argumento $\mathcal{A}_1$. Ya que la relación de derrota es un refinamiento de la relación de contra-argumentación, la

[12]una instanciación concreta de estos argumentos es discutida en el artículo [Simari et al., 1994].

solución propuesta modifica la noción de contra-argumentación incorporando una restricción adicional que evita a este tipo de falacia.

Definición 2.75. **[Simari et al., 1994]**
(contra-argumentación revisada)
Sean $\langle \mathcal{A}_1, h_1 \rangle$ y $\langle \mathcal{A}_2, h_2 \rangle$ dos estructuras de argumento. Diremos que el argumento $\mathcal{A}_1$ para h_1 *contra-argumenta* al argumento $\mathcal{A}_2$ para h_2 en el punto h' si, y sólo sí:

1. Para cada subargumento propio $\langle \mathcal{A}, h \rangle$ de $\langle \mathcal{A}_1, h_1 \rangle$, se verifica que $\langle \mathcal{A}_2, h_2 \rangle$ no contra-argumenta en punto alguno a $\langle \mathcal{A}, h \rangle$.

2. Existe un subargumento $\langle \mathcal{B}, h' \rangle$ de $\langle \mathcal{A}_2, h_2 \rangle$ tal que $\langle \mathcal{A}_1, h_1 \rangle$ y $\langle \mathcal{B}, h' \rangle$ están en desacuerdo.

■

Los restantes tipos de razonamiento falaz serán evitados condicionando las posibles líneas de argumentación a ser exploradas dentro de un árbol de dialéctica. Una líneas de argumentación constituye cada una de las secuencias de ataques y contra-ataques considerados en un análisis de dialéctica.

Definición 2.76. **[Simari et al., 1994]**
(línea de argumentación)
Sea $\mathcal{T}_{\langle \mathcal{A}_0, h_0 \rangle}$ el árbol de dialéctica asociado a la estructura de argumento $\langle \mathcal{A}_0, h_0 \rangle$. Cada camino λ en $\mathcal{T}_{\langle \mathcal{A}_0, h_0 \rangle}$ que une la raíz $\langle \mathcal{A}, h \rangle$ con una hoja $\langle \mathcal{A}_n, h_n \rangle$, denotado $\lambda = [\langle \mathcal{A}_0, h_0 \rangle, \langle \mathcal{A}_1, h_1 \rangle, \ldots, \langle \mathcal{A}_n, h_n \rangle]$, constituye una *línea de argumentación* para $\langle \mathcal{A}_0, h_0 \rangle$. ■

Dada una línea de argumentación arbitraria como la esquematizada en la figura 2.11, se aprecia que el primer argumento $\mathcal{A}_0$ es esgrimido a favor de la tesis discutida h_0; por otra parte, el argumento $\mathcal{A}_1$ ataca al argumento anterior, cuestionando la aceptabilidad de la tesis h_0. A su vez, considerando que el tercer argumento $\mathcal{A}_2$ ataca al atacante del argumento que sustenta a h_0, es lógico concluir que $\mathcal{A}_2$ sustenta

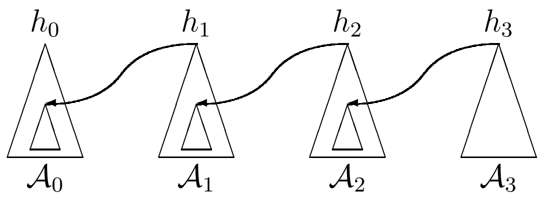


Figura 2.11: Una línea de argumentación en el sistema Simari-Loui

(indirectamente) a h_0. Un anális similar arroja como resultado que el argumento $\mathcal{A}_3$ interfiere (nuevamente, de forma indirecta) con la tesis h_0. En síntesis, los argumentos que ocupan posiciones pares soportan la tesis h_0, y, en contraste, los argumentos que ocupan posiciones impares interfieren con ésta.

Definición 2.77. **[Simari et al., 1994]**
(argumentos de soporte, argumentos de interferencia)
Sea $\mathcal{T}_{\langle \mathcal{A}_0, h_0 \rangle}$ un árbol de dialéctica, y sea $\lambda = [\langle \mathcal{A}_0, h_0 \rangle, \langle \mathcal{A}_1, h_1 \rangle, \ldots, \langle \mathcal{A}_n, h_n \rangle]$ una línea de argumentación en $\mathcal{T}_{\langle \mathcal{A}_0, h_0 \rangle}$. Los argumentos que integran λ puede separarse en argumentos de *soporte* o de *interferencia* de acuerdo a la siguiente caracterización recursiva:

- El argumento $\langle \mathcal{A}_0, h_0 \rangle$ constituye un argumento de *soporte* de λ.
- El argumento $\langle \mathcal{A}_i + 1, h_i + 1 \rangle$, $0 \leq i \leq n$, constituye un argumento de *soporte* (resp. de *interferencia*) de λ si el argumento $\langle \mathcal{A}_i, h_i \rangle$ constituye un argumento de interferencia (resp. de soporte) de λ.

Denotaremos λ_S (resp. λ_I) al conjunto de argumentos de soporte (resp. de interferencia) de *lambda*. ■

Contando con estas definiciones, es posible erradicar aquellos razonamientos inválidos productos de la argumentación falaz. La argumentación circular se elimina evitando la reiteración de argumentos o subargumentos a lo largo de una línea de argumentación. Por otra parte, para no caer en la argumentación contradictoria basta con preservar la consistencia entre los argumentos de soporte y de interferencia presentes a lo largo de cada línea de argumentación. Estas intuiciones son capturadas en la noción de línea de argumentación aceptable.

Definición 2.78. **[Simari et al., 1994]**
(línea de argumentación aceptable)
Una línea de argumentación λ de la forma $[\langle \mathcal{A}_0, h_0\rangle, \langle \mathcal{A}_1, h_1\rangle, \ldots, \langle \mathcal{A}_n, h_n\rangle]$ constituye una *línea de argumentación aceptable* si, y sólo sí:

1. Los argumentos de soporte (resp. de interferencia) presentes en λ son concordantes de a pares, esto es, para todo par de argumentos $\langle \mathcal{A}_i, h_i\rangle, \langle \mathcal{A}_j, h_j\rangle \in \lambda_S$ (resp. $\langle \mathcal{A}_i, h_i\rangle, \langle \mathcal{A}_j, h_j\rangle \in \lambda_I$) se verifica que $\mathcal{K} \cup \mathcal{A}_i \cup \mathcal{A}_j \not\vdash \perp$.

2. Los argumentos de soporte (resp. de interferencia) presentes en λ no derrotan a argumento alguno de interferencia (resp. de soporte) que ocurra posteriormente en la secuencia, esto es, para todo argumento $\langle \mathcal{A}_i, h_i\rangle \in \lambda_S$ (resp. $\langle \mathcal{A}_i, h_i\rangle \in \lambda_I$) se verifica que ningún argumento $\langle \mathcal{A}_j, h_j\rangle \in \lambda_I$ (resp. $\langle \mathcal{A}_j, h_j\rangle \in \lambda_S$) tal que $i < j$ es derrotado por $\langle \mathcal{A}_i, h_i\rangle$.

■

Restringiendo el análisis dialéctico sólo a líneas de argumentación aceptables se evita caer en los casos considerados de argumentación falaz. Por consiguiente, la noción de árbol de dialéctica se revisa de la siguiente manera:

Definición 2.79. **[Simari et al., 1994]**
(árbol de dialéctica aceptable)
Una árbol de dialéctica $\mathcal{T}_{\langle \mathcal{A}, h\rangle}$ constituye un *árbol de dialéctica aceptable* si, y sólo si, todas sus líneas de argumentación son aceptables. ■

Los árboles de dialéctica aceptables en virtud de ser árboles de dialéctica en general pueden utilizar el procedimiento de marcado anterior (def. 2.73, pág. 93), de forma inmediata. Por último, para determinar si un argumento está justificado en este contexto se puede reutilizar la definición 2.74, requiriendo simplemente que el árbol de dialéctica asociado al argumento en cuestión sea aceptable.

2.2.6. Programación en Lógica Rebatible

Alejandro J. García introduce en su tesis de magister [García, 1997] al paradigma de la *programación en lógica rebatible*, un formalismo que, elaborando sobre de la base delineada en [Simari y Loui, 1992, Simari et al., 1994], combina a la programación en lógica convencional con la argumentación rebatible, capitalizando las principales virtudes de ambas disciplinas. Si bien la expresividad de la programación en lógica convencional ha sido cuestionada en reiteradas oportunidades [Gelfond y Lifschitz, 1990, Pereira et al., 1991], la programación en lógica rebatible incorpora los medios necesarios para subsanar estos cuestionamientos, permitiendo modelar sin mayores inconvenientes *información incompleta* y *potencialmente inconsistente*.

Al igual que en la reformulación dialéctica del sistema Simari-Loui (analizado en la sección 2.2.5), las conclusiones en este sistema se sustentan por argumentos, los argumentos compiten entre sí en un análisis dialéctico, el análisis dialéctico se estructura como un árbol de dialéctica, y finalmente se sancionan sólo aquellas conclusiones sustentadas por argumentos que prevalezcan al proceso de argumentación. En lo que resta de la presente sección se describirá a la programación en lógica rebatible, también conocida como DeLP,[13] en base a lo desarrollado en la tesis doctoral de A. J. García [García, 2000].

El lenguaje de DeLP está cuidadosamente definido, partiendo de una *signatura* de base $\langle \mathcal{V}, \mathcal{F}, \mathcal{P} \rangle$. Esta signatura denota al conjunto de *variables* ($\mathcal{V}$), *símbolos funcionales* ($\mathcal{F}$) y *símbolos predicativos* ($\mathcal{P}$) considerados. Cada símbolo funcional tiene asociado una determinada aridad, distinguiendo a los símbolos funcionales de aridad cero bajo la denominación de *constante*. Estos elementos se combinan entre sí para conformar términos de la manera usual.

Definición 2.80. **[García, 2000]**
(término)

[13]por su sigla en inglés, *Defeasible Logic Programming*.

Sea $\sigma = \langle \mathcal{V}, \mathcal{F}, \mathcal{P} \rangle$ una signatura. Un *término* basado en σ, se caracteriza inductivamente de la siguiente forma:

1. Toda variable $v \in \mathcal{V}$ constituye un término.

2. Si $f/n \in \mathcal{F}$ es símbolo funcional de aridad n y $t_1, \ldots, t_n$ son términos basados en σ, entonces $f(t_1, \ldots, t_n)$ también será un término basado en σ.

■

Como es usual, denominaremos *fijo* a aquellos términos que no contengan variables. Los términos se combinan con los símbolos predicativos para conformar átomos de la manera esperada.

Definición 2.81. **[García, 2000]**
(átomo)
Sea $\sigma = \langle \mathcal{V}, \mathcal{F}, \mathcal{P} \rangle$ una signature y $t_1, \ldots, t_n$ términos basados en σ. Para todo símbolo predicativo $p/n \in \mathcal{P}$ de aridad n, diremos que $p(t_1, \ldots, t_n)$ es un *átomo* basado en σ. ■

Con el objeto de simplificar la notación prescindiremos de la explicitación de la signatura en curso, asumiendo en todo momento la existencia de una signatura $\langle \mathcal{V}, \mathcal{F}, \mathcal{P} \rangle$ arbitraria pero fija. El adjetivo fijo se aplica de manera análoga a los átomos, denominando átomos fijos a aquellos conformados únicamente por términos fijos. En el lenguaje de la programación en lógica rebatible se adopta el símbolo '$\sim$' para denotar a la *negación fuerte.*[14] Esta noción permite introducir el concepto de *literal*, el cual abarca a todo átomo A o su negación fuerte $\sim A$. Por último, entenderemos por *complemento* de un literal a su complemento en relación a la negación fuerte, esto es, si $L = A$, para algún átomo A, entonces el complemento de L, notado $\overline{L}$, será $\sim L$, mientras que en caso contrario, si $L = \sim A$, entonces $\overline{L} = A$.

Los literales constituyen la piedra basal de la representación de conocimiento en este sistema. Se estructuran de acuerdo a tres patrones como *hechos*, *reglas estrictas*

[14]en el sentido otorgado a esta denominación en [Alferes et al., 1998].

o *reglas rebatibles*, en función del tipo de conocimiento que se desee modelar. En primer lugar, los hechos representan el conjunto de observaciones a partir del cual se elabora todo razonamiento. Se representan mediante literales fijos (*i.e.*, átomos fijos o su negación fuerte), y tienen por objeto permitir la obtención de nuevas conclusiones a partir del encadenamiento de reglas tanto estrictas como rebatibles.

Definición 2.82. **[García, 2000]**
(regla estricta)
Una *regla estricta* es un par ordenado, denotado '$Cabeza \leftarrow Cuerpo$', donde $Cabeza$ es un literal fijo, y $Cuerpo$ es un conjunto finito de literales fijos distinto de vacío. Una regla estricta con cabeza L_0 y cuerpo $\{L_1, \ldots, L_n\}$, $n > 0$, se notará también $L_0 \leftarrow L_1, \ldots, L_n$ ■

Las reglas estrictas sólo codifican conocimiento seguro, en el sentido de que la certeza de los literales presentes en su cuerpo debe asegurar la certeza del literal presente en su cabeza. Considerando que este nivel de compromiso es poco frecuente en los dominios realistas, el formalismo cuenta con un segundo tipo de regla más flexible.

Definición 2.83. **[García, 2000]**
(regla rebatible)
Una *regla rebatible* es un par ordenado, denotado '$Cabeza \prec Cuerpo$', donde $Cabeza$ es un literal fijo, y $Cuerpo$ es un conjunto finito de literales fijos distinto de vacío. Una regla estricta con cabeza L_0 y cuerpo $\{L_1, \ldots, L_n\}$, $n > 0$, se notará también $L_0 \prec L_1, \ldots, L_n$ ■

Las reglas rebatibles, en contraste a las reglas estrictas, sólo establecen una presunción a favor del literal presente en su cabeza una vez aceptada la certeza de los literales presentes en su cuerpo. La interpretación intuitiva para una regla rebatible de la forma $\alpha \prec \beta$ es que *"razones para creer en el antecedente α proveen razones para creer en el consecuente β"*. Esta sutil diferencia permite representar situaciones prototípicas o reglas por defecto, las cuales no son expresables mediante reglas estrictas.

Un aspecto controversial en las definiciones anteriores acerca del cual resulta interesante elaborar es que sólo involucran a literales fijos. El sistema Simari-Loui señalaba que toda regla rebatible debía poseer al menos una variable libre en común entre la cabeza y el cuerpo. La intuición reflejada en ese sistema es que una regla resulta rebatible cuando la relación que representa se verifica en la mayoría de sus instanciaciones, pero no en todas. En este sentido, una regla de la forma $a \prec b$ carece de sentido como regla rebatible, ya que será estricta si a está relacionado con b (pues la relación modelada se verifica en todas sus instanciaciones), o, en caso contrario, será incorrecta (pues ninguna de sus instanciaciones verifica la relación modelada). La tesitura adoptada en DeLP es modelar el conocimiento en términos de reglas rebatibles conteniendo variables libres, con la salvedad de interpretar formalmente tales reglas como esquemas que denotan el conjunto de sus instanciaciones fijas. De manera análoga, los hechos o reglas estrictas conteniendo variables serán interpretados como una representación abreviada del conjunto de sus instanciaciones fijas.

La noción de *programa lógico rebatible*, concepto equivalente al de base de conocimiento en otros sistemas argumentativos, se define de la siguiente manera:

Definición 2.84. **[García, 2000]**
(programa lógico rebatible)
Un *programa lógico rebatible* es un conjunto $\mathcal{P}$, posiblemente infinito, de hechos, reglas estrictas y reglas rebatibles. En un programa $\mathcal{P}$, identificaremos con Θ al conjunto de hechos, con Ω al conjunto de reglas estrictas, y con Δ al conjunto de reglas rebatibles. En ocasiones, se denotará también como Π al conjunto $\Theta \cup \Omega$. Esto es, un programa lógico rebatible $\mathcal{P}$ puede ser denotado a través de la tupla (Θ, Ω, Δ), o bien simplemente como (Π, Δ), según convenga. ■

Ejemplo 2.30. Considerando el programa lógico rebatible (Π, Δ), donde:

Π	Δ
`ave(X) <- pingüino(X)`	`vuela(X) -< ave(X)`
`~vuela(X) <- pingüino(X)`	
`pingüino(pengo)`	
`ave(tweety)`	

la regla rebatible $vuela(X) \prec ave(X)$, por caso, denota que un objeto al satisfacer la propiedad de ser un ave, satisface tentativamente la propiedad de volar. De acuerdo a la convención adoptada, debe interpretarse como el conjunto de sus instancias fijas $\{vuela(pengo) \prec ave(pengo);\ vuela(tweety) \prec ave(tweety)\}$.

La noción de derivación rebatible del sistema Simari-Loui posee en DeLP una caracterización mucho más natural, puesto que los programas lógicos rebatibles no contienen variables.

Definición 2.85. **[García, 2000]**
(derivación rebatible)
Sea $\mathcal{P} = (\Pi, \Delta)$ un programa lógico rebatible, y sea L un literal. Una *derivación rebatible* para L a partir de $\mathcal{P}$ consiste de una secuencia finita de literales fijos $L_1, \ldots, L_n = L$, provisto de que para cada L_i, $1 \leq i \leq n$, se verifica alguna de las siguientes condiciones:

1. L_i es un hecho contenido en $\mathcal{S}$, o bien
2. existe en $\mathcal{P}$ una regla estricta o rebatible con cabeza L_i y cuerpo $A_1, \ldots, A_k$, tal que todo A_j, $1 \leq j \leq k$, es un elemento de la secuencia que precede a L_i.

■

La noción anterior de derivación se dice rebatible puesto que pueden existir otras razones en contra de la aceptación de un cierto literal derivado rebatiblemente. En contraste, las derivaciones rebatibles obtenidas en base a sólo reglas estrictas establecen conclusiones que necesariamente deben ser aceptadas, pues están basadas en conocimiento incuestionable.

Definición 2.86. **[García, 2000]**
(derivación estricta)
Sea $\mathcal{P} = (\Pi, \Delta)$ un programa lógico rebatible, y sea L un literal. Diremos que existe una *derivación estricta* para L a partir de $\mathcal{P}$ si, y sólo si, existe es una derivación rebatible para L a partir de $(\Pi, \emptyset)$. ■

La derivación rebatible de literales complementarios a partir de un cierto conjunto de hechos y reglas denota la existencia de un conflicto. La siguiente definición formaliza esta intuición.

Definición 2.87. **[García, 2000]**
(conjunto contradictorio)
Sea $\mathcal{S}$ un conjunto de hechos, reglas estrictas y reglas rebatibles. Diremos que $\mathcal{S}$ es un *conjunto contradictorio* si, y sólo si, es posible derivar rebatiblemente a partir de $\mathcal{S}$ tanto a un literal dado como a su complemento. ■

La programación en lógica rebatible se circunscribe sólo a aquellos programas lógicos rebatibles (Π, Δ) cuyos conjuntos Π sean no contradictorios, si bien el conjunto $\Pi \cup \Delta$ puede resultar contradictorio, tal como lo refleja el siguiente ejemplo:

Ejemplo 2.31. En el contexto del programa lógico rebatible del ejemplo 2.30, la secuencia $[pinguino(pengo), ave(pengo), vuela(pengo)]$ constituye una derivación rebatible para $vuela(pengo)$, y la secuencia $[pinguino(pengo), {\sim}vuela(pengo)]$ constituye una derivación rebatible para ${\sim}vuela(pengo)$. Por consiguiente, el conjunto $\Pi \cup \Delta$ resulta contradictorio, aún cuando el conjunto Π de todas formas satisface el requisito de ser no contradictorio.

Los programas lógicos rebatibles se utilizan para construir *argumentos*. Esta noción, eje de todo sistema argumentativo, se formaliza de la siguiente manera:

Definición 2.88. **[García, 2000]**
(estructura de argumento)
Sea $\mathcal{P} = (\Pi, \Delta)$ un programa lógico rebatible, y sea h un literal. Diremos que $\langle \mathcal{A}, h \rangle$

constituye una *estructura de argumento* para h si, y sólo si, existe un $\mathcal{A} \subseteq \Delta$, tal que satisface las siguientes restricciones:

1. existe una derivación rebatible para h a partir de $(\Pi, \mathcal{A})$,

2. $\Pi \cup \mathcal{A}$ es un conjunto no contradictorio, y además

3. $\mathcal{A}$ es minimal, es decir, no existe un subconjunto propio $\mathcal{A}'$ de $\mathcal{A}$, tal que $\mathcal{A}'$ satisface las dos condiciones anteriores.

■

La discusión que sigue a la definición de argumento en el sistema Simari-Loui (def. 2.64, pág. 83), en relación a por qué representar un argumento a través de un conjunto de reglas rebatibles y las razones que fundamentan cada una de las condiciones impuestas, se aplica al contexto de la programación en lógica rebatible dadas las similitudes entre los formalismos y no será recreada aquí.

Ejemplo 2.32. En el contexto del programa lógico rebatible del ejemplo 2.30, es posible formular una estructura de argumento $\langle \mathcal{A}, vuela(tweety) \rangle$ para $vuela(tweety)$, donde $\mathcal{A} = \{vuela(tweety) \prec ave(tweety)\}$, pero no es factible formular estructura de argumento alguna para $vuela(pengo)$, ya que el conjunto de reglas rebatibles $\{vuela(pengo) \prec ave(pengo)\}$, si bien habilita la derivación rebatible de $vuela(pengo)$, resulta contradictorio al ser considerado en conjunción a Π (pues $\Pi \mathrel{|\!\!\sim} {\sim}vuela(pengo)$). Como consecuencia de esta observación, $\langle \emptyset, {\sim}vuela(pengo) \rangle$ constituye una estructura de argumento válida para ${\sim}vuela(pengo)$.

El solo hecho de contar con una estructura de argumento para una determinada conclusión no asegura su aceptación, ya que puede existir otra estructura de argumento, en conflicto con la primera, sancionando su complemento. La noción de conflicto entre estructuras de argumento se elabora en base al concepto de conjunto contradictorio, tal como se indica a continuación.

Definición 2.89. **[García, 2000]**
(desacuerdo)
Sea (Π, Δ) un programa lógico rebatible. Diremos que una estructura de argumento $\langle \mathcal{A}, h \rangle$ está en *desacuerdo* con otra estructura de argumento $\langle \mathcal{A}', h' \rangle$ si, y sólo si, el conjunto $\Pi \cup \{h, h'\}$ resulta contradictorio. ■

La noción de estructuras de argumento en desacuerdo se generaliza en la relación de *contra-argumentación* con el objeto de incorporar los desacuerdos con las conclusiones internas a una estructura de argumento. Por esta razón, se introduce el concepto auxiliar de *subestructura de argumento*: diremos que $\langle \mathcal{A}', h' \rangle$ es una subestructura de argumento de $\langle \mathcal{A}, h \rangle$ si, y sólo si, $\mathcal{A}' \subseteq \mathcal{A}$.

Definición 2.90. **[García, 2000]**
(contra-argumentación)
Diremos que la estructura de argumento $\langle \mathcal{A}_1, h_1 \rangle$ *contra-argumenta* a la estructura de argumento $\langle \mathcal{A}_2, h_2 \rangle$ en el literal h si, y sólo si, existe una subestructura de argumento $\langle \mathcal{A}, h \rangle$ de $\langle \mathcal{A}_2, h_2 \rangle$ tal que $\langle \mathcal{A}_1, h_1 \rangle$ y $\langle \mathcal{A}, h \rangle$ están en desacuerdo. ■

Al igual que en el sistema Simari-Loui, la contra-argumentación no dictamina si el ataque resulta exitoso. En DeLP se adopta un criterio de comparación entre estructuras de argumentos para determinar la direccionalidad de la derrota implícita en toda contra-argumentación. Se definen dos criterios de comparación, uno inspirado en el criterio de especificidad, y el otro involucrando a un ordenamiento previamente suministrado entre las reglas rebatibles. En primer lugar se aborda la adaptación del criterio de especificidad al contexto de la programación en lógica rebatible, denominado *especificidad generalizada*

Definición 2.91. **[García, 2000]**
(especificidad generalizada)
Sea $\mathcal{P} = (\Theta, \Omega, \Delta)$ un programa lógico rebatible, y sea $\mathcal{F}$ el conjunto de literales que pueden derivarse rebatiblemente a partir de $\mathcal{P}$. Diremos que una estructura de argumento $\langle \mathcal{A}_1, h_1 \rangle$ es *estrictamente más específica* que otra estructura de argumento $\langle \mathcal{A}_2, h_2 \rangle$ si, y sólo si, se satisfacen las siguiente condiciones:

1. para todo conjunto $H \subseteq \mathcal{F}$, tal que
   si existe una derivación rebatible para h_1 a partir de $\Omega \cup H \cup \mathcal{A}_1$,
   y no existe una derivación estricta para h_1 a partir de $\Omega \cup H$,
   entonces existe una derivación rebatible para h_2 a partir de $\Omega \cup H \cup \mathcal{A}_2$.

2. existe al menos un conjunto $H' \subseteq \mathcal{F}$, tal que
   si existe una derivación rebatible para h_2 a partir de $\Omega \cup H' \cup \mathcal{A}_2$,
   y no existe una derivación estricta para h_2 a partir de $\Omega \cup H'$,
   entonces existe una derivación rebatible para h_1 a partir de $\Omega \cup H' \cup \mathcal{A}_1$.

■

Este criterio esencialmente reconcilia dos principios: favorecer aquellas estructuras de argumento que emplean mayor cantidad de información, y preferir las estructuras de argumento que involucran derivaciones más cortas. En este sentido, la discusión que acompaña a la introducción del criterio de comparación entre argumentos basado en el principio de especificidad, en el marco del sistema Simari-Loui (def. 2.65, pág. 85), resulta aplicable de la misma manera al contexto de la programación en lógica rebatible y no será recreada aquí.

El segundo criterio de comparación se elabora a partir de un orden estricto parcial entre las reglas rebatibles del programa lógico en consideración. La viabilidad de erigir el criterio de comparación en torno a un orden previamente suministrado ha sido ensayado en otras propuestas tales como el sistema Prator [Prakken y Sartor, 1997] o los sistemas argumentativos abstractos [Vreeswijk, 1997]. Si bien algunos formalismos se limitan a comparar reglas rebatibles con cabezas complementarias, en DeLP se adopta un criterio más amplio que contrasta la totalidad de las reglas rebatibles de las estructuras de argumento en consideración.

Definición 2.92. **[García, 2000]**
(preferencia)
Sea (Π, Δ) un programa lógico rebatible, y sea '$>$' un orden estricto parcial definido sobre los elementos de Δ. Diremos que una estructura de argumento $\langle \mathcal{A}_1, h_1 \rangle$ resulta

preferida en base a '>' sobre otra estructura de argumento $\langle \mathcal{A}_2, h_2 \rangle$, si, y sólo si, se satisfacen las siguientes condiciones:

1. existe una regla $r_1 \in \mathcal{A}_1$ tal que para alguna regla $r_2 \in \mathcal{A}_2$ se verifica que $r_1 > r_2$, y además
2. para toda regla $r'_1 \in \mathcal{A}_1$ no existe regla alguna $r'_2 \in \mathcal{A}_2$ tal que $r'_2 > r'_1$.

■

Ejemplo 2.33. Considerando el programa lógico rebatible (Π, Δ), donde

Π	Δ
`ave(fred)`	`vuela(X) -< ave(X)`
`débil(fred)`	`~vuela(X) -< débil(X)`

es posible elaborar las estructuras de argumento $\langle \mathcal{A}_1, vuela(fred) \rangle$ para $vuela(fred)$, y $\langle \mathcal{A}_2, {\sim}vuela(fred) \rangle$ para ${\sim}vuela(fred)$, donde $\mathcal{A}_1 = \{vuela(fred) \prec ave(fred)\}$, y $\mathcal{A}_2 = \{{\sim}vuela(fred) \prec débil(fred)\}$. Adoptando un criterio de comparación entre las reglas en Δ tal que $({\sim}vuela(fred) \prec débil(fred)) > (vuela(fred) \prec ave(fred))$, $\langle \mathcal{A}_2, {\sim}vuela(fred) \rangle$ resulta preferida sobre $\langle \mathcal{A}_1, vuela(fred) \rangle$.

Cabe destacar que la especificidad generalizada resulta más adecuada para modelar conocimiento que el criterio basado en el orden sobre las reglas rebatibles, esencialmente por el carácter abarcador de la primera a diferencia del carácter individualista del segundo. Por otra parte, la especificidad generalizada induce un ordenamiento estricto parcial sobre las estructuras de argumento, cuando en contraste el criterio de preferencia basado en el orden entre reglas ni siquiera satisface transitividad (*i.e.*, es posible que $\langle \mathcal{A}_1, h_1 \rangle$ no sea preferido sobre $\langle \mathcal{A}_3, h_3 \rangle$ aún cuando $\langle \mathcal{A}_1, h_1 \rangle$ resulte preferido sobre $\langle \mathcal{A}_2, h_2 \rangle$ y $\langle \mathcal{A}_2, h_2 \rangle$ resulte preferido sobre $\langle \mathcal{A}_3, h_3 \rangle$). Otra razón a favor de la especificidad generalizada es que este criterio no depende del programa lógico rebatible considerado, a diferencia del criterio basado en el ordenamiento entre reglas, el cual se verá afectado por la incorporación de nuevas reglas rebatibles. Es posible que el bajo requerimiento computacional[15] sea el único aspecto a favor del segundo criterio, ya que la especificidad generalizada resulta evidentemente más demandante en términos computacionales.

A partir del criterio de comparación entre estructuras de argumento elegido es posible denotar el resultado de la comparación de manera sucinta como $\langle \mathcal{A}_1, h_1 \rangle \succ \langle \mathcal{A}_2, h_2 \rangle$, indicando que $\langle \mathcal{A}_1, h_1 \rangle$ es mejor que $\langle \mathcal{A}_2, h_2 \rangle$ de acuerdo al criterio seleccionado. La relación de derrota entre estructuras de argumento surge de combinar la noción de contra-argumentación con alguno de los criterios de comparación introducidos. Ya que estos criterios de comparación son parciales, la relación de derrota distingue dos situaciones complementarias. La primer situación identificada, denominada *derrota propia*, describe aquellas contra-argumentaciones en las cuales el criterio de comparación señala a alguna de las estructuras de argumento como preferida.

Definición 2.93. **[García, 2000]**
(derrotador propio)
Sean $\langle \mathcal{A}_1, h_1 \rangle$ y $\langle \mathcal{A}_2, h_2 \rangle$ estructuras de argumento para h_1 y h_2 respectivamente. Diremos que $\langle \mathcal{A}_1, h_1 \rangle$ es un *derrotador propio* para $\langle \mathcal{A}_2, h_2 \rangle$ en el literal h si, y sólo si, existe una subestructura de argumento $\langle \mathcal{A}, h \rangle$ de $\langle \mathcal{A}_2, h_2 \rangle$, tal que $\langle \mathcal{A}_1, h_1 \rangle$ contra-argumenta a $\langle \mathcal{A}_2, h_2 \rangle$ en el literal h, y se verifica por otra parte que, en base al criterio de comparación seleccionado, $\langle \mathcal{A}_1, h_1 \rangle \succ \langle \mathcal{A}, h \rangle$. ■

La segunda situación identificada, denominada *derrota por bloqueo*, caracteriza aquellas contra-argumentaciones en las cuales el criterio de comparación no es capaz de determinar a cuál de las estructuras de argumento debe preferir.

Definición 2.94. **[García, 2000]**
(derrotador de bloqueo)
Sean $\langle \mathcal{A}_1, h_1 \rangle$ y $\langle \mathcal{A}_2, h_2 \rangle$ estructuras de argumento para h_1 y h_2 respectivamente.

[15] apenas polinomial en la cantidad de reglas rebatibles de las estructuras de argumento que se comparan

Diremos que $\langle \mathcal{A}_1, h_1 \rangle$ es un *derrotador de bloqueo* para $\langle \mathcal{A}_2, h_2 \rangle$ en el literal h si, y sólo si, existe una subestructura de argumento $\langle \mathcal{A}, h \rangle$ de $\langle \mathcal{A}_2, h_2 \rangle$, tal que $\langle \mathcal{A}_1, h_1 \rangle$ contra-argumenta a $\langle \mathcal{A}_2, h_2 \rangle$ en el literal h, y se debe verificar por otra parte que, en base al criterio de comparación seleccionado, no es el caso que $\langle \mathcal{A}_1, h_1 \rangle \succ \langle \mathcal{A}, h \rangle$, ni que $\langle \mathcal{A}, h \rangle \succ \langle \mathcal{A}_1, h_1 \rangle$. ■

Las definiciones 2.93 y 2.94 describen las situaciones reconciliadas bajo la noción de derrota, si bien la derrota propia denota un tipo de derrota un tanto más categórica que la derrota por bloqueo. Diremos que una estructura de argumento simplemente *derrota* a otra estructura de argumento si la primera constituye un derrotador propio o un derrotador de bloqueo de la segunda.

Ejemplo 2.34. Considerando el programa lógico rebatible (Π, Δ), donde

Π	Δ
`ave(X) <- pingüino(X)`	`vuela(X) -< ave(X)`
`ave(tweety)`	`~vuela(X) -< débil(X)`
`débil(tweety)`	`~vuela(X) -< pingüino(X)`
`pingüino(pengo)`	

es posible formular las siguientes estructuras de argumento:

$\langle \mathcal{A}_1, vuela(pengo) \rangle$, donde $\mathcal{A}_1 = \{vuela(pengo) \prec ave(pengo)\}$.

$\langle \mathcal{A}_2, {\sim}vuela(pengo) \rangle$, donde $\mathcal{A}_2 = \{{\sim}vuela(pengo) \prec pinguino(pengo)\}$.

$\langle \mathcal{A}_3, vuela(tweety) \rangle$, donde $\mathcal{A}_3 = \{vuela(tweety) \prec ave(tweety)\}$.

$\langle \mathcal{A}_4, {\sim}vuela(tweety) \rangle$, donde $\mathcal{A}_4 = \{{\sim}vuela(tweety) \prec débil(tweety)\}$.

Adoptando como criterio de comparación entre estructuras de argumento a la especificidad generalizada, se verifican las siguientes relaciones: $\langle \mathcal{A}_2, {\sim}vuela(pengo) \rangle$ constituye un derrotador propio para $\langle \mathcal{A}_1, vuela(pengo) \rangle$, ya que la primera es estrictamente más específica que la segunda. Por otra parte, tanto $\langle \mathcal{A}_3, vuela(fred) \rangle$ constituye un derrotador de bloqueo para $\langle \mathcal{A}_4, {\sim}vuela(fred) \rangle$ como $\langle \mathcal{A}_4, {\sim}vuela(fred) \rangle$

un derrotador de bloqueo para $\langle \mathcal{A}_3, vuela(fred) \rangle$, ya que estas estructuras de argumento no se encuentran relacionadas por el criterio de comparación elegido.

El estado final de una estructura de argumento dependerá del estado final de las estructuras de argumento que la derroten. En este sentido, la programación en lógica rebatible caracteriza su semántica de manera análoga que en la reformulación dialéctica del sistema Simari-Loui, esto es, apelando a un análisis estructurado como un árbol de dialéctica. La noción de línea de argumentación modela el hecho de que una estructura de argumento que derrota a otra estructura de argumento puede a su vez ser derrotada por una tercera estructura de argumento, situación que puede reiterarse una cantidad arbitraria de veces.

Definición 2.95. **[García, 2000]**
(línea de argumentación)
Sea $\mathcal{P}$ un programa lógico rebatible, y sea $\langle \mathcal{A}_0, h_0 \rangle$ una estructura de argumento para h_0. Diremos que la secuencia de estructuras de argumento obtenidas a partir de $\mathcal{P}$ $[\langle \mathcal{A}_0, h_0 \rangle, \langle \mathcal{A}_1, h_1 \rangle, \ldots, \langle \mathcal{A}_n, h_n \rangle, \ldots]$ constituye una *línea de argumentación a partir de* $\langle \mathcal{A}_0, h_0 \rangle$, denotada $\lambda^{\langle \mathcal{A}_0, h_0 \rangle}$, si, y sólo si, toda estructura de argumento $\langle \mathcal{A}_i, h_i \rangle$, $i \geq 1$, derrota a su antecesor inmediato $\langle \mathcal{A}_{i-1}, h_{i-1} \rangle$. ■

La reformulación dialéctica del sistema Simari-Loui impone diversas limitaciones sobre la estructura de las líneas de argumentación (def. 2.78, pág. 99), con el propósito de evitar diversos casos de argumentación falaz. La argumentación falaz es una forma de razonamiento no sensato donde las conclusiones obtenidas pueden carecer de sustento. En DeLP se evita el razonamiento falaz de manera análoga, imponiendo un conjunto de restricciones sobre la estructura de las líneas de argumentación. En primer lugar, se distinguen las estructuras de argumento que soportan la tesis inicial de una línea de argumentación de aquellas que interfieren con ésta.

Definición 2.96. **[García, 2000]**
(estructuras de argumento de soporte)
Sea $\lambda^{\langle \mathcal{A}_0, h_0 \rangle} = [\langle \mathcal{A}_0, h_0 \rangle, \langle \mathcal{A}_1, h_1 \rangle, \ldots, \langle \mathcal{A}_n, h_n \rangle, \ldots]$ una línea de argumentación para

$\langle \mathcal{A}_0, h_0 \rangle$. Diremos que el conjunto de las estructuras de argumento que ocupan las posiciones impares de la secuencia $\lambda^{\langle \mathcal{A}_0, h_0 \rangle}$ constituye el conjunto de las *estructuras de argumento de soporte* para $\langle \mathcal{A}_0, h_0 \rangle$, denotado $\lambda_S^{\langle \mathcal{A}_0, h_0 \rangle}$. ■

Definición 2.97. **[García, 2000]**
(estructuras de argumento de interferencia)
Sea $\lambda^{\langle \mathcal{A}_0, h_0 \rangle} = [\langle \mathcal{A}_0, h_0 \rangle, \langle \mathcal{A}_1, h_1 \rangle, \ldots, \langle \mathcal{A}_n, h_n \rangle, \ldots]$ una línea de argumentación para $\langle \mathcal{A}_0, h_0 \rangle$. Diremos que el conjunto de las estructuras de argumento que ocupan las posiciones pares de la secuencia $\lambda^{\langle \mathcal{A}_0, h_0 \rangle}$ constituye el conjunto de las *estructuras de argumento de interferencia* para $\langle \mathcal{A}_0, h_0 \rangle$, denotado $\lambda_I^{\langle \mathcal{A}_0, h_0 \rangle}$. ■

La argumentación circular se evita adoptando una tesitura similar a la tomada por el sistema Simari-Loui, esto es, imponiendo como restricción que la apararición una estructura o subestructura de argumento no pueda reiterarse. En relación a la derrota recíproca—situación en la cual dos estructuras de argumento al derrotarse mutuamente de manera propia generan una línea de argumentación sin fin—a diferencia del sistema anterior que refinaba a la noción de contra-argumentación, en DeLP se considera a este tipo de derrota como una instancia más de la argumentación circular. Cabe resaltar que existe todavía una tercer alternativa posible, considerar a la derrota propia recíproca como derrota por bloqueo, pero esta opción requiere adaptar las definiciones de derrota propia (def. 2.93) y de derrota por bloqueo (def. 2.94) recién introducidas.

La reformulación dialéctica del sistema Simari-Loui evita razonar basado en argumentación contradictoria imponiendo como requisito que los conjunto de argumentos de soporte e interferencia sean consistentes de a pares en conjunción con el conocimiento estricto. En DeLP se refina este requerimiento estipulando que la totalidad del conjunto de estructuras de argumento de soporte e interferencia sean concordantes, donde la noción de concordancia generaliza al concepto de consistencia de la siguiente manera:

Definición 2.98. **[García, 2000]**
(concordancia)

Sea (Π, Δ) un programa lógico rebatible. Diremos que el conjunto de estructuras de argumento $\{\langle \mathcal{A}_1, h_1 \rangle, \ldots, \langle \mathcal{A}_n, h_n \rangle\}$ es *concordante* si, y sólo si, el conjunto de estructuras de argumento $\bigcup_{i=1}^{n} \mathcal{A}_i \cup \Pi$ es no contradictorio. ■

Las restricciones impuestas con el objeto de evitar la argumentación falaz en relación a las líneas de argumetación a ser consideradas como válidas se resumen en la siguiente definición:

Definición 2.99. **[García, 2000]**
(línea de argumentación válida)
Sea $\lambda^{\langle \mathcal{A}_0, h_0 \rangle} = [\langle \mathcal{A}_0, h_0 \rangle, \langle \mathcal{A}_1, h_1 \rangle, \ldots, \langle \mathcal{A}_n, h_n \rangle, \ldots]$ una línea de argumentación para $\langle \mathcal{A}_0, h_0 \rangle$. Diremos que $\lambda^{\langle \mathcal{A}_0, h_0 \rangle}$ constituye una *línea de argumentación válida* si, y sólo si, se satisfacen las siguientes condiciones:

1. $\lambda^{\langle \mathcal{A}_0, h_0 \rangle}$ es finita.

2. Los conjuntos $\lambda_S^{\langle \mathcal{A}_0, h_0 \rangle}$ y $\lambda_I^{\langle \mathcal{A}_0, h_0 \rangle}$ son concordantes.

3. Ninguna $\langle \mathcal{A}_i, h_i \rangle$ es una subestructura de argumento de otra $\langle \mathcal{A}_j, h_j \rangle$, para todo $j < i$.

4. Si $\langle \mathcal{A}_i, h_i \rangle$, $i \geq 2$, es un derrotador de bloqueo de $\langle \mathcal{A}_{i-1}, h_{i-1} \rangle$, entonces $\langle \mathcal{A}_{i-1}, h_{i-1} \rangle$ es un derrotador propio de $\langle \mathcal{A}_{i-2}, h_{i-2} \rangle$.

■

La primer condición en la definición anterior explicita que toda línea de argumentación debe ser finita, la segunda tiene por objeto evitar la argumentación contradictoria, la tercera intenta evitar la argumentación circular y la cuarta enfatiza la diferencia entre la derrota propia y la derrota por bloqueo. Cabe acotar que un aspecto llamativo de la tercer condición es que constituye una variante más lábil de la restricción adoptada con el mismo propósito en el contexto del sistema Simari-Loui, y no logra evitar la totalidad de los casos de argumetación falaz basados en la reintroducción de argumentos.

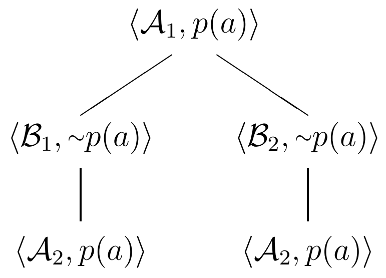


Figura 2.12: Análisis dialéctico del ejemplo 2.35

Finalmente, las diversas líneas de argumentación aceptable se estructuran como un árbol de dialéctica. Este árbol de dialéctica sintetiza el análisis dialéctico llevado a cabo en relación al estado de la estructura de argumento que origina las líneas de argumentación consideradas.

Definición 2.100. **[García, 2000]**
(árbol de dialéctica)
Sea $\langle \mathcal{A}_0, h_0 \rangle$ una estructura de argumento. Un *árbol de dialéctica* para $\langle \mathcal{A}_0, h_0 \rangle$, denotado como $\mathcal{T}_{\langle \mathcal{A}_0, h_0 \rangle}$, se construye de acuerdo al siguiente procedimiento:

1. El nodo raíz se etiqueta $\langle \mathcal{A}_0, h_0 \rangle$.

2. Sea $[\langle \mathcal{A}_0, h_0 \rangle, \langle \mathcal{A}_1, h_1 \rangle, \ldots, \langle \mathcal{A}_n, h_n \rangle]$ la secuencia de etiquetas presentes en una rama desde la raíz hasta un nodo etiquetado $\langle \mathcal{A}_n, h_n \rangle$, y sean las estructuras de argumento $\langle \mathcal{B}_1, q_1 \rangle, \langle \mathcal{B}_2, q_2 \rangle \ldots, \langle \mathcal{B}_k, q_k \rangle$ todos los derrotadores de $\langle \mathcal{A}_n, h_n \rangle$. En este contexto, para cada derrotador $\langle \mathcal{B}_i, q_i \rangle$ de $\langle \mathcal{A}_n, h_n \rangle$, $1 \leq i \leq k$, tal que $[\langle \mathcal{A}_0, h_0 \rangle, \langle \mathcal{A}_1, h_1 \rangle, \ldots, \langle \mathcal{A}_n, h_n \rangle, \langle \mathcal{B}_i, q_i \rangle]$ resulta ser una línea de argumentación válida para $\langle \mathcal{A}_0, h_0 \rangle$, se debe incorporar un nodo etiquetado $\langle \mathcal{B}_i, q_i \rangle$ como hijo del nodo etiquetado $\langle \mathcal{A}_n, h_n \rangle$.

■

Ejemplo 2.35. Considerando el programa lógico rebatible (Π, Δ), donde:

Π	Δ
`q(X) <- r(X)`	`p(X) -< q(X)`
`r(a)`	`~p(X) -< r(X)`
`s(a)`	`~p(X) -< s(X)`
	`p(X) -< r(X), s(X)`

es posible formular las siguientes estructuras de argumento:

$\langle \mathcal{A}_1, p(a) \rangle$ para $p(a)$, donde $\mathcal{A}_1 = \{p(a) \prec q(a)\}$.

$\langle \mathcal{A}_2, p(a) \rangle$ para $p(a)$, donde $\mathcal{A}_1 = \{p(a) \prec r(a), s(a)\}$.

$\langle \mathcal{B}_1, \sim p(a) \rangle$ para $\sim p(a)$, donde $\mathcal{A}_1 = \{\sim p(a) \prec r(a)\}$.

$\langle \mathcal{B}_2, \sim p(a) \rangle$ para $\sim p(a)$, donde $\mathcal{A}_1 = \{\sim p(a) \prec s(a)\}$.

En base al criterio de especificidad generalizada se puede observar que $\langle \mathcal{B}_1, \sim p(a) \rangle$ es un derrotador propio para $\langle \mathcal{A}_1, p(a) \rangle$, que $\langle \mathcal{B}_2, \sim p(a) \rangle$ y $\langle \mathcal{A}_1, p(a) \rangle$ se derrotan por bloqueo mutuamente, y que tanto $\langle \mathcal{B}_1, \sim p(a) \rangle$ como $\langle \mathcal{B}_2, \sim p(a) \rangle$ son derrotados de manera propia por $\langle \mathcal{A}_2, p(a) \rangle$. En consecuencia, el árbol de dialéctica asociado a $\langle \mathcal{A}_1, p(a) \rangle$ presenta la estructura esquematizada en la figura 2.12.

En el contexto de la reformulación dialéctica del sistema Simari-Loui se destacó el hecho de que el árbol de dialéctica contenía todas las estructuras de argumento capaces de influenciar el estado del argumento alojado en su raíz, observación aplicable también al contexto de la programación en lógica rebatible. Por consiguiente, un procedimiento de marcado análogo al descripto en el marco de aquel sistema (def. 2.73, pág. 93), permite establecer en DeLP el estado final de una estructura de argumento empleando sólo la información presente en su árbol de dialéctica asociado.

Definición 2.101. **[García, 2000]**
(árbol de dialéctica marcado)

Sea $\mathcal{T}_{\langle \mathcal{A}, h \rangle}$ el árbol de dialéctica asociado a la estructura de argumento $\langle \mathcal{A}, h \rangle$. El *árbol de dialéctica marcado* asociado a $\mathcal{T}_{\langle \mathcal{A}, h \rangle}$, notado como $\mathcal{T}^{\star}_{\langle \mathcal{A}, h \rangle}$, se obtiene por medio del siguiente procedimiento recursivo:

$$\begin{tikzcd}[column sep=small]
& \langle \mathcal{A}_1, p(a)\rangle^U \arrow[dl, dash] \arrow[dr, dash] & \\
\langle \mathcal{B}_1, \mathord{\sim} p(a)\rangle^D \arrow[d, dash] & & \langle \mathcal{B}_2, \mathord{\sim} p(a)\rangle^D \arrow[d, dash] \\
\langle \mathcal{A}_2, p(a)\rangle^U & & \langle \mathcal{A}_2, p(a)\rangle^U
\end{tikzcd}$$

Figura 2.13: Marcado del árbol de dialéctica del ejemplo 2.35

- Todas las hojas de $\mathcal{T}_{\langle \mathcal{A},h\rangle}$ se marcan con "U" en $\mathcal{T}^{\star}_{\langle \mathcal{A},h\rangle}$.
- Sea N un nodo interno de $\mathcal{T}_{\langle \mathcal{A},h\rangle}$. El nodo N se marca con "U" si todo nodo hijo de N está marcado con "D", y N se marca con "D" si existe al menos un nodo hijo de N marcado con "U".

■

Ejemplo 2.36. En la figura 2.13 se aprecia el árbol de dialéctica marcado asociado al árbol de dialéctica desarrollado en el ejemplo 2.35.

Finalmente, contando con este concepto es posible caracterizar a aquellos literales sancionados por un programa lógico rebatible como producto del análisis dialéctico. La siguiente definición introduce la noción de literal garantizado, término que denota aquellos literales aceptados por un razonador que adopte a la programación en lógica rebatible como su maquinaria de inferencia.

Definición 2.102. **[García, 2000]**
(literales garantizados)
Sea $\mathcal{P} = (\Pi, \Delta)$ un programa lógico rebatible, y sea h un literal. Diremos que h está *garantizado* si, y sólo si, existe alguna estructura de argumento $\langle \mathcal{A}, h\rangle$ para h, tal que la raíz de su correspondiente árbol de dialéctica marcado $\mathcal{T}^{\star}_{\langle \mathcal{A},h\rangle}$ está marcado como "U". ■

Con el objeto de prescindir de los tecnicismos que caracterizan la noción de literal garantizado se introduce un operador modal $\mathcal{B}$, el cual al preceder a un literal denotará que éste está garantizado. La expresión $\mathcal{B}h$ se debe interpretar como "en base al conocimiento actual, es posible creer en h". Se puede conjeturar sobre una base firme que esta lógica modal satisface la propiedad $\mathcal{B}h \Rightarrow \neg\mathcal{B}{\sim}h$, si bien aún no ha podido ser demostrado formalmente. Esta atractiva propiedad representa una forma de sensatez a nivel de estado epistémico, no permitiendo sostener simultáneamente como creencia a un determinado literal y a su complemento. El operador modal $\mathcal{B}$ posibilita a su vez caracterizar las posibles respuestas brindadas por un intérprete para la programación en lógica rebatible ante una determinada consulta.

Definición 2.103. **[García, 2000]**
(respuestas posibles)
Sea $\mathcal{P} = (\Pi, \Delta)$ un programa lógico rebatible, y sea h un literal. Si h pertenece a la signatura de $\mathcal{P}$, entonces una consulta acerca del estado de h recibirá como respuesta SI en el caso que $\mathcal{B}h$, NO en el caso que $\mathcal{B}\overline{h}$ e INDECISO si no es el caso que $\mathcal{B}h$, ni es el caso que $\mathcal{B}\overline{h}$. Caso contrario, si h no pertenece a la signatura de $\mathcal{P}$, entonces toda consulta acerca del estado de h recibirá como respuesta DESCONOCIDO ■

Finalmente, en [García, 2000] también extiende a la programación en lógica rebatible para incorporar un segundo tipo de negación, inspirado en la negación por falla de la programación en lógica convencional, con la particularidad que '`not` h' se interpreta como "no es posible garantizar h", en contraste a su interpretación clásica de "no es posible derivar h". Otro aspecto considerado como extensión es la incorporación de reglas rebatibles sin literales en el cuerpo, denominadas *presuposiciones*, noción explorada inicialmente por Nute en el contexto de su lógica rebatible [Nute, 1987]. Las presuposiciones permiten expresar qué hipótesis resultan razonables adoptar en el contexto de un dado programa lógico rebatible. Si bien resulta interesante contar con el poder expresivo aportado por las presuposiciones para extender el dominio de aplicación de la programación en lógica rebatible, cabe objetar que las presupociones no respetan el espíritu de una regla rebatible, puesto que no están modelando una relación parcial entre un literal y un conjunto de literales. Una posibilidad interesante para explorar a futuro podría ser considerar a las presuposiciones como un nuevo componente de los programas lógicos rebatibles, independiente los conjuntos Π y Δ.

En síntesis, la programación en lógica rebatible provee un lenguaje de representación de conocimiento de gran expresividad, capaz de modelar información incompleta y potencialmente contradictoria, con la particularidad de que su maquinaria de inferencia se basa en una argumentación rebatible de carácter dialéctico que combina nociones relativamente sencillas obteniendo un formalismo altamente robusto.

2.3. Conclusiones

A continuación repasaremos aquellos aspectos comunes a la mayoría de los sistemas argumentativos, analizando posteriormente las características particulares que diferencian cada uno de estos acercamientos del resto de las propuestas.

En principio, todos los sistemas argumentativos considerados compartes las generalidades descriptas en la sección 2.1.3: cuentan con una *lógica subyacente* que permite la construcción de *argumentos*; algunos de estos argumentos están en *conflicto*, relación usualmente simétrica, que apela a la noción de *derrota* con el objeto de determinar la direccionalidad del ataque implícito en el conflicto, estableciendo en último término el *estado final* de los argumentos a partir de esta noción de derrota. Otro aspecto sobre el cual se aprecia un evidente consenso entre las diversas propuestas es en la metodología de representación de conocimiento por defecto, donde generalmente se adopta alguna forma de regla rebatible.

Por otra parte, esta tesis explota la presencia recurrente en los distintos acercamientos de otro patrón cuya existencia fue someramente esbozada en la sección 2.1.2: todas las propuestas[16] admiten reformular sus semánticas mediante alguna forma de caracterización dialéctica. En este sentido, el capítulo 4 elabora un modelo dialéctico

[16]con la notable excepción del sistema argumentativo formulado por J. Pollock.

abstracto en el cual es posible capturar cada una de estas diversas reformulaciones semánticas. El interés en desarrollar una modelo abstracto surge de la gran aceptación que recibió la propuesta de Dung, los marcos agumentativos (sección 2.2.1). De hecho, puede afirmarse que la generalidad obtenida como producto del nivel de abstracción constituye la característica más atractiva de los marcos argumentativos.

En relación a los aspectos distintivos de cada acercamiento, debe señalarse que el sistema argumentativo formulado por J. Pollock (sección 2.2.2) ha influenciado profundamente diversos desarrollos en el área de la argumentación rebatible. Pollock a través de su arquitectura OSCAR establece la importancia de acortar la distancia entre la teoría y su implementación computacional. A tal efecto, propone un ciclo de evolución donde la experimentación con la implementación de una dada teoría permite depurar sus falencias, caracterizando una nueva teoría refinada, que a su vez da a lugar a una nueva implementación, la cual permite continuar este ciclo sucesivamente. Según Pollock, si la separación entre una teoría y su implementación resulta excesiva, la teoría en cuestión carecerá de sentido, ya que constituiría simplemente un ejercicio intelectual sin aplicación a la vida real. Esta tesitura puede resumirse en el principio—no siempre atendido—de que al postular nuevas teorías formales resulta esencial mantener el contacto con la realidad.

El sistema argumentativo propuesto por H. Prakken y G. Sartor (sección 2.2.4) se caracteriza por ser actualmente la única propuesta que incorpora la posibilidad de razonar rebatiblemente acerca de la relación de comparación entre argumentos a ser aplicada en la obtención de las propias conclusiones sancionadas por una base de conocimiento. Si bien las restantes propuestas optan por dejar de lado este aspecto debido a la complejidad inherente a la argumentación rebatible *per se*, Prakken y Sartor consideraron imprescindible poder modelar prioridades rebatible, puesto que uno de los objetivos de diseño fue poder modelar distintos escenarios judiciales, ámbito en el cual resulta natural modificar dinámicamente los criterios de comparación entre argumentos.

El sistema Simari-Loui a través de su reformulación dialéctica (sección 2.2.5) introduce el tratamiento de la argumentación falaz al contexto de los sistema argumentativos. Este tipo no deseado de argumentación, el cual permite obtener conclusiones incorrectas no sustentadas por su base de conocimiento, no había sido tenido en cuenta en las propuestas previas. Otro aspecto que distingue en particular a este sistema es haber reconocido tempranamente las virtudes que capitalizan al adoptar una caracterización dialéctica para describir la semántica de un sistema argumentativo. Si bien hoy en día la mayoría de los restantes acercamientos cuentan con alguna variante dialéctica de sus semánticas, este sistema puede ser sindicado como el pionero en adoptar esta postura en lo que a sistemas argumentativos respecta.

En relación al paradigma de la Programación en Lógica Rebatible (sección 2.2.6), cabe destacar como su principal característica el reconciliar las ideas desarrolladas en el marco del sistema Simari-Loui con el principio postulado por J. Pollock, obteniéndose un sistema argumentativo que presenta un adecuado balance entre poder expresivo e implementabilidad. Por un lado incorpora el poder expresivo del sistema Simari-Loui, tratamiento de la argumentación falaz incluído, y por otro lado hereda el razonamiento guiado por la metas de la programación en lógica,[17] el cual admite implementaciones más eficientes que aquellas que no lo son. Para lograr esta fina mixtura las definiciones del sistema han sido cuidadosamente adaptadas, de forma tal de preservar el citado carácter guiado por las metas. Esta particularidad no se encuentra presente en los restantes acercamientos, lo que necesariamente implicará costes computacionales adicionales para con sus respectivas implementaciones.

[17] en realidad tipo de razonamiento propio del principal exponente de la programación en lógica (*i.e.*, el lenguaje PROLOG).

Capítulo 3

Sistemas Multiagente

A lo largo de la última década el desarrollo de agentes de software *inteligentes* o *racionales* ha cautivado la atención de la comunidad de Inteligencia Artificial. Esta tendencia denota una maduración en el área, ya que se han abandonado aquellos objetivos titánicos de los primeros días en pos de metas mucho más concretas, que posean aplicaciones prácticas directas. Sin embargo, como un sistema inteligente solipsístico (*i.e.*, aislado de su entorno y sólo interesado en sí mismo), carece de interés práctico, la mayor parte de la atención se ha dirigido hacia aquellos sistemas en los cuales existan una multitud de agentes, interactuando entre sí y con su entorno, dando a lugar a una sociedad de agentes. A estas sociedades de agentes se las conocen actualmente como *sistemas multiagente*, o simplemente SMA.

Este capítulo presenta una visión general de las distintas líneas de investigación en el terreno de los sistemas multiagente, describiendo y cotejando los principales acercamientos propuestos en la literatura. Entre otras cosas se reseñan las taxonomías de agente existentes, así como los diversos lenguajes de comunicación y protocolos que dan a lugar a toda forma de interacción entre agentes, desde la coordinación o la cooperación, pasando por la competencia o la persuación, hasta llegar a la deliberación o la negociación. Como la preponderancia de esta interacción es ineludible, puesto que un sistema multiagente con poca o nula interacción se asemeja más al poco atractivo agente solipsístico que a una sociedad de agentes, el presente capítulo concluye repasando las propuestas presentes en la literatura que han intentado caracterizar a esta interacción siguiendo distintos enfoques.

3.1. Motivaciones

La impresión inicial al tomar contacto con un sistema multiagente por primera vez suele ser desconcertante: se mantiene la esperanza de que pueda existir un sistema centralizado que desempeñe el mismo rol, pero presentando una construcción más simple y un mantenimiento menos costoso. Sin embargo, al considerar dominios de aplicación en concreto esta esperanza se diluye, puesto que muchos de los escenarios más atractivos del mundo real presentan un conjunto de entidades independientes que interactúan entre sí, donde cada una posee objetivos propios, no necesariamente libres de conflicto con el resto, que cuentan con la capacidad de almacenar información privada, la cual puede influenciar el proceso de toma de decisiones de cada entidad. En este tipo de escenarios, más allá de que tal vez exista algún modelo centralizado que resulte viable, pueden ser modelados de forma más natural al aplicar el enfoque de agentes múltiples.

Elaborando sobre este cuestionamiento para con los sistemas centralizados, existen actualmente múltiples dominios de aplicación en los que los actores se encuentran distribuidos geográficamente, o que generan o procesan un importante volumen de información, donde la topología de este tipo de sistemas resulta usualmente dinámica, con el contenido de sus componentes cambiando tan rápidamente que resulta difícil o imposible reunir la información suficiente como para mantener actualizado a un modelo centralizado del sistema. Más aún, no sólo la ubicación física de los componentes de un sistema puede impedir la aplicación del acercamiento tradicional, pues incluso en aquellas situaciones en las que el modelo centralizado resulte apropiado es posible que el modelo multiagente sea de todas formas el elegido. Un ejemplo concreto de esta situación son aquellos sistemas centralizados en los que se requiera o bien un alto rendimiento, o bien una confiabilidad superior.

Por lo general, toda empresa de cierta complejidad—como aquelas que resultan atractivas en el mundo real—admiten ser descompuestas en subtareas independientes. En este contexto, cada subtarea puede ser desarrollada por un agente en particular, permitiendo que las subtareas se lleven a cabo en forma simultánea, lo que a su vez puede traducirse en un incremento de la performance. Pero por otra parte, esta distribución de tareas incrementa la confiabilidad del sistema, ya que ante el fallo de uno de los componentes sólo las subtareas asignada a los agentes afectados serán interrumpidas, mientras que el resto continúa con su normal desarrollo. En contraste, en un sistema centralizado toda falla, por aislada que sea, provoca una detención total de las actividades. En síntesis, el paralelismo se obtiene al asignar roles independientes a diferentes agentes, y la confiabilidad al contar con más de un agente asignado al mismo rol, puesto que si el control y las responsabilidades son compartidos en forma adecuada entre ese grupo de agentes, el sistema podrá tolerar la falla de una o más de estas entidades.

Otro aspecto a tener en cuenta es que la modularidad provista por los SMA puede simplificar notablemente la tarea del programador del sistema. En lugar de asignar todo el trabajo a un agente centralizado monolítico, el programador puede identificar grupos de tareas auxiliares, asignando el control de cada una de ellos a un agente diferente. Esta política resuelve, entre otras cosas, el problema de distribuir la atención del único agente al conjunto de tareas a llevar a cabo, cuidando que todas las tareas reciban su cuota de atención para así evitar la inhanición. Por otra parte, la propia subdivisión de tareas entre distintos agentes puede dar a lugar a nuevas soluciones más elegantes (*i.e.*, nuevos algoritmos), que difícilmente hubieran sido descubiertos siguiendo el acercamiento centralizado.

Un cuestionamiento de peso para con el acercamiento centralizado es su pobre capacidad de adaptarse a la expansión natural del sistema modelado. Por caso, aquella ingeniosa solución que daba excelentes resultado con poco volumen quizás pueda resultar desastrosa si se duplica el caudal de información manejado. En contraste, los SMA admiten una escalabilidad sin parangón, puesto que la modularidad mencio-

- el sistema modelado lo requiere	- simplifica la programación
- paralelismo	- permite estudiar la inteligencia
- robustez	- escalabilidad

Figura 3.1: Razones más frecuentes para optar por un sistema multiagente

nada anteriormente permite incorporar nuevas funcionalidades al sistema afectando someramente la estructura preexistente. En contraste, una falta de escalabilidad en un sistema implica necesariamente incurrir en altos costos a la hora de tener que adecuarse a un cambio en su entorno.

Recientemente se ha comenzado a explorar un nuevo dominio de aplicación para los sistema multiagente muy propicio: analizar a los fenómenos sociales producto de la interacción entre entidades autónomas. Uno de los fenómenos sociales más interesantes para considerar es la *inteligencia*, esa escurridiza característica presente en primates superiores, seres humanos y otros mamíferos acuáticos, la cual de ser capturada formalmente tendría innumerables aplicaciones prácticas. De hecho, todos los sistemas inteligentes que se dan de forma espontánea en la naturaleza se caracterizan precisamente por estar integrados a un ambiente en el cual operan múltiples entidades inteligentes simultáneamente. Algunos autores [Dautenhahn, 1995] han propuesto acertadamente que una de las claves para desarrollar una entidad artificial dotada de inteligencia consiste en permitir que estas entidades manifiesten capacidades sociales avanzadas. En este sentido, un sistema multiagente provee el escenario práctico perfecto para verificar nuestras intuiciones en relación a aquellos procesos que requieran un cierto grado de inteligencia para ser llevados a cabo, tales como la representación de conocimiento, la toma de decisiones en base a éste, la construcción de planes para alcanzar objetivos, etc.

Finalmente, el rol de la interacción en el marco de un sistema multiagente no debe ser menoscabado, puesto que es ésta precisamente la que permite que una colección de agentes puede alcanzar objetivos más allá de sus propias capacidades, excediendo incluso las limitaciones individiuales. Como ejemplo paradigmático consideremos la siguiente situación (propuesta inicialmente en [Stankevicius y Simari, 2001]), donde dos empleados de una compañía de mudanzas desean cargar un cierto sofá a un camión, pero este sofá resultó ser lo suficientemente pesado como para que ninguno de ellos lo pueda levantar por sí mismo. Sin duda, la única posibilidad es que coordinen su empuje de forma tal que ambos levanten el sofá al mismo tiempo, combinando sus fuerzas. Quizás no sea evidente el importante rol que la comunicación desempeña en este simple escenario, pero de no mediar contacto entre los empleados—de forma tal de coordinar sus acciones—no habría forma de cumplir con la tarea encomendada. En este simple ejemplo se puede apreciar como en conjunto lograron alcanzar un objetivo que estaba más allá de sus posibilidades individuales. La gran atención que está recibiendo el desarrollo de protocolos de interacción entre agentes responde precisamente a esta realidad, siendo la presente tesis un aporte en este sentido.

3.2. Generalidades

El origen de los sistema multiagente se remonta a la *resolución distribuida de problemas* [Bond y Gasser, 1988], una técnica desarrollada en el seno de la *Inteligencia Artificial Distribuida*. Esta disciplina, una rama dentro de la propia Inteligencia Artificial, buscaba resolver problemas cada vez más complejos apelando a contar con un mayor número unidades de cómputo. A medida de que se dotaba de mayores capacidades a cada una de estas unidades de cómputo, tales como la representación de conocimiento o la toma de decisiones de forma local, un nuevo concepto surgió espontáneamente: el sistema multiagente.

Si bien este concepto ha sido utilizado extensivamente a lo largo del presente capítulo, aún cuando todavía no hemos definido de forma precisa en qué consiste un sistema de este tipo. Antes de introducir nuestra definición, cabe señalar que en la comunidad aún no existe consenso suficiente como para formular una definición concreta. Esta falta de consenso puede ser atribuída principalmente a la naturaleza multidisciplinaria del área, la cual actualmente engloba a una gran cantidad de desarrollos, que han dado a lugar a casi tantas definiciones como sistemas multiagentes fueron introducidos en la literatura. En este sentido, teniendo en cuenta el propósito del presente capítulo, así como las motivaciones vertidas en la sección anterior, formularemos la siguiente definición con el objeto de que sirva como punto de partida para explorar las características de los SMA en general.

> *Un sistema multiagente se define como una colección de agentes racionales y autónomos, insertos en un ambiente en común, capaces de coordinar su conocimiento, objetivos, habilidades y estrategias a fin de resolver un cierto problema.*

En esta definición un agente consiste en esencia de una entidad computacional, tal como un programa de software o un robot que posea la capacidad de percibir su ambiente y actuar sobre el mismo. Esta entidad debe ser autónoma, esto es, poseer control sobre su comportamiento y poder actuar sin intervención humana. También debe ser racional,[1] esto es, debe ser capaz de interactuar con su entorno de forma flexible y en una variedad de circunstancias, siempre tendiendo a satisfacer sus propios objetivos. En la mayoría de los casos, cada agente se verá afectado por las otras entidades o agentes que coexistan en su ambiente. Para aprovechar las potenciales ventajas de esta interacción entre entidades es necesario proveer a los agentes con mecanismos de comunicación adecuados, que permitan coordinar sus comportamientos. Esta interacción se puede realizar en forma indirecta a través del ambiente en el que se encuentran embebidos los agentes (por caso, realizando alguna acción que modifique un parámetro del ambiente), o en forma directa por medio de un lenguaje de comunicación en común.

Como se mencionó anteriormente, el caracter multidisciplinario de los sistemas multiagente hace un tanto difícil la identificación de aspectos en común a todos

[1] en el sentido que Stuart Russell le da al término en [Russell, 1997].

ellos. Sin embargo, existen ciertos patrones que se pueden detectar incluso en las propuestas más dispares. A continuación sintetizaremos estos aspectos recurrentes, señalando las diversas posibilidades ensayadas para cada uno de éstos.

En primer lugar, el componente que no puede faltar en todo SMA es el *conjunto de agentes* empleados para modelar el problema en consideración. Existen principalmente dos variantes: una posibilidad es que todos los agentes cuenten con el mismo conjunto de capacidades, y la otra que el sistema cuente con agentes especializados en determinadas tareas, por lo que sus capacidades se encontrarán adaptadas a la tarea que deban desarrollar. En este sentido, los sistemas del primer tipo se demoninan *homogéneos*, mientras que los del segundo tipo se denominan *heterogéneos*.

Otro componente que debe ser tenido en cuenta en todo sistema multiagente es el *entorno*, en donde todo agente del sistema se encuentra embebido. En ocasiones, el entorno es representado sencillamente como otro agente más, si bien pragmáticamente la distinción entre agentes y entorno se preserva (por caso, no tiene sentido que existan dos agentes que estén modelando al entorno en simultáneo). Existen esencialmente dos concepciones para el entorno: se lo puede considerar estático, esto es, que permanece inmutable a lo largo del tiempo, o bien dinámico, esto es, que se ve influenciado por la misma evolución del sistema.

Los sensores presentes en los agentes de un SMA suelen percibir su entorno de una manera no necesariamente uniforme. Puede que las percepciones no coincidan por corresponder a diferentes locaciones, o bien por haber sido observadas en diferentes ocasiones. Mas aún, es posible que las mismas observaciones hechas en un determinado momento sean interpretadas de diferente forma dependiendo de las creencias de los agentes que realicen el sensado. Por consiguiente, se puede afirmar que el entorno se vuelve *observable sólo de forma parcial* (desde la perpectiva de cada agente), situación que debe ser tenida en cuenta cuando el agente deba tomar decisiones por su cuenta. En este sentido, como esta toma de decisiones también debe ser llevada a cabo de forma distribuida, el control del sistema no podrá ser concentrado en un único lugar. Este es otro de los aspectos comunes a todo sistema multiagente, el control del sistema debe ser altamente *descentralizado*. Por otra parte, esta dispersión de las percepciones presenta otro desafío a la hora de reconciliar las diversas observaciónes parciales con el objeto de formar una imagen acabada del estado actual del entorno.

En un sistema centralizado es razonable suponer que el conocimiento acerca del estado actual del entorno así como del efecto de las acciones llevadas a cabo también se encuentra centralizado. En contraste, en un sistema multiagente no es posible centralizar el conocimiento, ya que como mencionáramos anteriormente cada uno de los agentes cuenta con una perpectiva única del entorno. Sin embargo, dependiendo de las características del problema modelado es posible que exista algún cúmulo de información disponible a todos los agentes, denominado *conocimiento común*. Este conocimiento común puede estar disponible inicialmente, o bien puede ser adquirido por los agentes a través de su percepción o interactuando con los restantes agentes.

Por último, en todo sistema multiagente se asume la existencia de alguna forma de *comunicación*, la cual puede asumir desde la forma de un elemental mecanismo de intercambio de datos hasta un elaborado protocolo de negociación. Naturalmente, para que los agentes se puedan comunicar entre sí de una forma intelegible se debe contar con algún *lenguaje* que sea conocido y entendido por todas las partes involucradas. En general, para que la interpretación del lenguaje de comunicación esté libre de ambigüedades también se tener acceso a su *ontología* asociada.

3.3. Principales aplicaciones

La gran versatilidad que brinda un sistema multiagente se ha traducido a lo largo del tiempo en una vasta librería de problemas resueltos siguiendo este enfoque. Esta sección presenta una reseña de algunos de estos dominios en concreto, comentando las aplicaciones más significativas introducidas en la literatura.

Diversos autores [Chaib-draa, 1995, Wooldridge y Ciancarini, 1999] coinciden en señalar que los primeros sistema multiagente se orientaron mayoritariamente hacia las redes de sensado, esto es, sistemas que modelaban una red de sensores distribuidos en el espacio. Estos sensores debían permitir una interpretación distribuida de la información, así como una planificación de las tareas a ser desarrolladas, indicando los requerimientos de recursos o de materias primas. Uno de los pioneros en implementar este modelo fue Victor Lesser, con su conocido *Distributed Vehicle Monitoring Technique*, que se convirtió en el banco de prueba para muchas de las técnicas modernas de desarrollo de SMA. Entre las aplicaciones iniciales que siguen el modelo de sensado distribuido se pueden destacar a manera de ejemplos ilustrativos los siguientes:

Sistemas para optimizar el flujo del tráfico aéreo: Las compañías de transporte aéreo están enfrentando un constante incremento en el tráfico aéreo. Para satisfacer esta demanda se deben solucionar un conjunto de problemas, que incluyen a la utilización ineficiente del espacio aéreo, el aumento constante del volumen de trabajo de los controladores aéreos, así como el uso de tecnología obsoleta para realizar el control de tráfico (por caso, la comunicación por voz entre las aeronaves y las torres de control). Surgen entonces los sistemas de soporte para los controladores de tráfico, como es el caso de OASIS [Hayton et al., 1998]. OASIS maximiza la utilización de la pista de aterrizaje ordenando las aeronaves en una secuencia óptima de aterrizaje. A tal efecto, asigna una estampa de tiempo a cada avión que le indica cuando debe aterrizar y luego monitorea en tiempo real su desempeño. La arquitectura de este sistema es orientada a agentes, sus componentes son agentes independientes tal que cada uno resuelve una porción del problema, actuando de forma cooperativa. Cabe acotar que actualmente se estan realizando pruebas de este sistema en aeropuertos de Australia.

Otro de los usos de los SMA en el control de tráfico aéreo es el desarrollo de sistemas de control híbrido para facilitar la tarea de los controladores aéreos. Muchas de las normas utilizadas por los controladores están dictadas por la necesidad de simplificar la tarea de los controladores humanos. En general, éstos se encargan esencialmente de la seguridad, esto es, evitar catástrofes, pero no disponen del tiempo suficiente para ocuparse de otras cuestiones tales como optimizar el consumo de combustible. En este sentido se están desarrollando sistemas multiagente para auxiliar a los controladores aéreos en estas tareas, con el objeto de poder aumentar la cantidad de aeronaves que maneja cada uno. Un proyecto un tanto más ambicioso, que se encuentra actualmente en la etapa de desarrollo, persigue el concepto de *free flight*, es decir, permitir que cada piloto elija su propia ruta, altitud y velocidad, que a diferencia del método utilizado en la actualidad, que restringe el espacio aéreo al conjunto de rutas de navegación preestablecidas.

Monitoreo de redes de telecomunicaciones en tiempo real: La complejidad creciente de las redes de telecomunicaciones ha impulsado el desarrollo de herramientas automatizadas para auxiliar a los operadores humanos en sus tareas. Si bien los primeros de estos sistemas eran monolíticos y centralizados, posteriormente evolucionó hacia el enfoque de agentes múltiples, el cual en este contexto resulta mucho más idóneo. Un ejemplo concreto lo constituye el sistema denominado CNFM, por su sigla en inglés *Cooperative Network Fault Managenment*, introducido en [Garijo et al., 1996]. Estos autores proponen un framework genérico orientado a agentes para diagnosticar fallos en una red, donde un conjunto de agentes descentralizados y autónomos cooperan para llevar adelante el control de las fallas en esta red, apelando al conocimiento específico que cada agente tiene de la sección de red que monitorea. Los agentes se comunican entre sí mediante un lenguaje de alto nivel, si bien cuentan también con un conjunto de primitivas de un protocolo de interacción estandard para poder comunicarse con el resto de los componentes del sistema del cual la red monitoreada forma parte.

Planificación de procesos industriales o de producción: La mayor parte de las plantas de manufactura o de procesos industriales pueden abstraerse como una serie de estaciones de trabajo sucesivas que cuentan con funciones específicas (por caso, una línea de ensamblado de automóviles). Un sistema multiagente provee el marco ideal para modelar este tipo de problemas, donde cada agente representa a las distintas estaciones de trabajo. Este área de aplicación posiblemente sea una de las más exploradas, como lo atestigua la multitud de propuestas existentes en la literatura. Por caso, el sistema denominado YAMS, por su sigla en inglés *Yet Another Manufacturing System*, propuesto por Parunak modelaba tanto las diferentes estaciones de ensamblado como las diversas fuentes de materias primas como agentes, utilizando el conocido protocolo de redes de contratos[2] para controlar el flujo de materiales.

Otra propuesta que sigue esta línea fue introducida en [Sycara et al., 1991] por Katia Sycara. Este trabajo describe como llevar adelante la asignación de recursos en una planta de producción, donde la misma planificación se realiza en forma distribuida. En este contexto, la planificación se concibe como un proceso en el cual participan un grupo de agentes, los cuales cuentan con un conjunto acotado de recursos para desarrollar sus tareas y tienen un conocimiento limitado tanto del ambiente como de las intenciones de los otros agentes. El propósito de este sistema es obtener una planificación completa a través de la cooperación de los distintos agentes, los cuales al ir combinando incrementalmente sus planes parciales deben asegurar que la solución encontrada respete los objetivos globales del sistema. En contraste, Michael Wooldridge [Wooldridge et al., 1996] al abordar esta misma problematica propuso dotar a los agentes de un mecanismo de negociación, el cual se utiliza para resolver localmente los conflictos de asignación de recursos que puedan surgir entre las distintas células de manufactura.

El advenimiento del internet comercial en la pasada década se constituyó en un nuevo dominio de aplicación donde el enfoque de agentes múltiples resulta claramente superior, puesto que la red de redes constituye un sistema abierto, altamente distribuido, sin protocolos o lenguajes preestablecidos a nivel de sus componentes,

[2]concepto abordado posteriormente en la sección 3.6, pág. 170.

con incluso una organización interna que se adapta constantemente a los cambios topológicos. Los SMA han sido marco de una gran cantidad de aplicaciones altamente difundidas, de las cuales las que a continuación se exponen son simplemente una muestra:

Búsqueda de información: La *world wide web* sustentada por internet agrupa a una gran cantidad de recursos de infomación físicamente distribuidos en el espacio. Si bien el potencial que brinda éste repositorio está fuera de discusión, han surgido diversos incovenientes a la hora de sacar provecho del mismo. El problema principal es precisamente el exceso de contenido: en ocasiones encontrar una pieza de información en internet termina insumiendo más recursos que una búsqueda convencional. Este problema fue resuelto en parte apelando a grandes índices que facilitan las búsquedas tales como `www.google.com` o `www.yahoo.com`, aunque el estudio de la información indexada puede de todas formas insumir excesivos recursos. Actualmente se han introducido diversas propuestas que siguiendo el enfoque de agentes múltiples tienen por objeto simplificar el acceso a la información deseada. Por caso, el concepto de *metabuscador*, en donde una búsqueda ingresada por el usuario es automáticamente reenviada a múltiples buscadores, componiendo los resultados obtenidos en una única lista de la cual se eliminan las reiteraciones. En este tipo de aplicación cada buscador es modelado como un agente poseedor de cierta información (en este caso, el buscador), el cual responde a un lenguaje de comunicación dado (en este caso, el lenguaje HTML), existiendo un agente adicional a cargo de gestionar la interacción con el usuario y con los restantes agentes. Otro ejemplo concreto de este tipo de aplicación son los generadores de *resúmenes de noticias personalizados*. Aquí la información de las diversas fuentes electrónicas de noticias es filtrada y recompaginada en base a las preferencias particulares del usuario (indicadas de forma explítica, o bien aprendidas observando el comportamiento del usuario).

Comercio electrónico: Año tras año el volumen de las transacciones comerciales que se realizan mediante medios electrónicos en general e internet en particular presenta un crecimiento inusitado. Los analistas económicos coinciden en señalar que el potencial aquí latente está apenas empezando a ser capitalizado. Por caso, el éxito del portal `www.ebay.com`, el cual permite realizar subastas virtuales de prácticamente cualquier mercancía, también se inspira en el modelo de agentes múltiples. El sistema no es del todo autónomo, pues las pujas deben ser introducidas inicialmente por las personas físicas. No obstante, un agente parametrizado con las decisiones del usuario permanece presente el salón virtual, a la espera de realizar las contra-ofertas requeridas para asegurar que su representado gane la subasta, siempre que no se supere el máximo que el usuario estaba dispuesto a pagar.

La literatura provee innumerables ejemplos de este tipo de aplicaciones. Por ejemplo, Wooldridge y Ciancarini consideran el caso de una agencia de viajes [Wooldridge y Ciancarini, 1999]. Ya que los preparativos de todo viaje que cruce el Atlántico suelen ser engorrosos, puesto que existen distintas rutas posibles, varias aerolíneas entre las cuales optar, etc., la mayoría de las personas confían estas tareas a una agencia de viajes de confianza. En primer lugar, el cliente expresa sus requerimientos, por caso si posee alguna preferencia entre las aerolíneas, cuantas horas está dispuesto a esperar en las escalas intermedias, o si desea optar por algún menú especial quizás no disponible en todas las aerolíneas. Luego, el agente de viajes presenta al cliente aquellas posibilidades que respeten sus preferencias. Los autores indican que este modelo, el cual posiblemente refleje la dinámica de gran parte de las agencias de viajes, puede ser parafraseado en términos de un conjunto de agentes autónomos con relativa sencillez.

Otra área comercial que constantemente atrae nuevos desarrollos son las herramientas para asistir a la toma de decisiones. En este sentido, la compra y venta de acciones constituye uno de los tópicos más estudiados, posiblemente debido al volumen de capital involucrado. Por caso, en [García et al., 2000] se propone un modelo multiagente para asistir a este tipo de toma de desición, donde un conjunto de agentes especializados se encargan de monitorear el estado de la bolsa apelando a internet para recopilar las cotizaciones actualizadas, novedades de interés, etc. Estos agentes, denominados de extracción de información, construirán una base de datos con la información obtenida siguiendo el popular modelo de pizarra.[3] Más tarde, esta información es analizada por un segundo tipo de agentes, denominados agentes deliberativos, que son los encargados de tomar las decisiones de compra y venta. Estos agentes cuentan con un módulo de razonamiento que adopta a la programación en lógica rebatible como lenguaje de representación de conocimiento y razonamiento.

Tal como se ha reseñado a lo largo de la presente sección, el enfoque de agentes múltiples permite modelar los más diversos escenarios, desde aplicaciones que abordan problemas localizados hasta sistemas cuya área de efecto puede extenderse a más de un continente. En este sentido, diversos autores [Wooldridge y Ciancarini, 1999, Jennings, 2000] han llegado a proponer esta forma de concebir el mundo como un nuevo paradigma de programación: la *programación orientada a agentes*. En este paradigma, propuesto como sucesor de la programación orientada a objetos, caracteriza a los agentes como estructuras de software capaces de tomar decisiones racionales. Los autores sostienen que estos componentes así caracterizados resultan adecuados para la construcción de ciertos tipos de software en los cuales la ingeniería tradicional no había obtenido el éxito esperado.

3.4. Taxonomía de los Sistema Multiagente

Dada la gran diversidad de sistemas multiagente que han sido propuestos a través de los años resulta natural postular una clasificación general basada en el conjunto de sus características más evidentes. No obstante, puesto que precisamente determinar cuáles son las caracteríscas más evidentes de un sistema es a su vez un aspecto

[3]concepto abordado posteriormente en la sección 3.6, pág. 166.

controversial, en la literatura se han introducido varias clasificaciones alternativas.

La primer taxonomía presentada en el campo de la inteligencia artificial distribuida fue propuesta por Keith Decker. En este trabajo, se distinguen cuatro dimensiones en base a las cuales clasificar a los diversos sistemas:

1. La granularidad de los agentes: puede ser fina o gruesa.
2. La heterogeneidad del conocimiento de los agentes: éste puede ser redundante, o cada agente se puede especializar en un determinado tema.
3. El método para distribuir el control: puede ser cooperativo o competitivo, por grupos o jerárquico, estático o con roles que se modifican dinámicamente.
4. La formas de llevar a cabo la comunicación: ésta puede ser de alto o de bajo nivel. Asimismo se puede utilizar un mecanismo de pasaje de mensajes punto a punto, o utilizar un repositorio común.

Posteriormente, Parunak [Parunak, 1998] presentó una taxonomía para los SMA desde otra perspectiva, en esta ocasión centrándose en las potenciales aplicaciones. Un aspecto interesante de esta alternativa es que las dimensiones propuestas distinguen las características de los agentes de las características del sistema.

En cuanto a los agentes del sistema, Parunak distingue las siguientes cualidades:

- Cómo se realiza el mapeo de los componentes del sistema en concreto a los agente, esto es, que elementos serán modelados por cada agente.
- Cómo modelará cada agente al mundo, ya que en general los agentes difieren en cuanto al grado de sofistifación de la representación conocimiento y razonamiento utilizados.
- La arquitectura de los agentes, donde varía el grado de heterogeneidad y el modelo escogido, que puede ser ser deliberativo, reactivo o mixtos.

La sociedad de agentes posee a su vez sendas características que la distinguen. En particular, Parunak destaca las siguientes:

- Cuántos agentes existen en el sistema: los sistemas difieren en la cantidad de agentes que involucran, y en la cantidad de tipos de agentes que abarque. A su vez, la propia población de agentes puede variar durante la ejecución: el sistema puede crear, eliminar o fusionar agentes como parte de su dinámica.

- Qué canales de comunicación usan los agentes: los canales a través de los cuales la información se comunica de un agente a otro se clasifican en cuanto al medio, el mecanismo de direccionado, la persistencia y la localidad. En este sentido, el medio puede ser un componente de acceso compartido, o una red de comunicaciones. En cuanto al mecanismo de direccionado, se puede usar un mecanismo de difusión en masa, comunicando el mensaje a todos los agentes en simultáneo, o poseer alguna forma de direccionamiento restringido. La persistencia se refiere al tiempo de vida de los mensajes, éstos pueden permanecer en un pizarrón o ser enviados a través de una red y consumirse ni bien sean recibidos. Finalmente, la localidad se refiere a la posibilidad de que los agentes migren dentro del espacio de comunicación, con el objeto de aproximarse a algún componente con el que deseen interactuar.

- Qué protocolos de comunicación usan los agentes: Parunak refiere con este término a lo que se suele denominar protocolos de interacción entre agentes. En esta categoría distingue mecanismos *directivos*, en los que un agente da órdenes a los otros, mecanismos de *votación*, de *negociación* o finalmente *actos de habla*.[4]

- Cómo es la topología del sistema: la configuración relativa de los agentes determina los vecinos de cada uno de ellos, formando una topología a través de la cual fluye la información. Esta estructura puede permanecer invariable (de acuerdo a como fue establecida por el diseñador del sistema), o bien puede cambiar durante a lo largo de la ejecución.

[4]el autor en este caso cita concretamente a la teoría introducida por John Searle [Searle, 1970].

- Qué mecanismos usan los agentes para coordinar sus acciones: ya que los agentes son autónomos, pero no anárquicos, deben coordinar sus acciones apelando a uno o más mecanismos a tal efecto. Parunak distigue entre una coordinación jerárquica, en la cual existen niveles preestablecidos de control, y una coordinación igualitaria, que emerge dinámicamente en base a la interacción entre los agentes.

Un reciente enfoque propuesto por Stone y Veloso en [Stone y Veloso, 2000] se destaca por su singularidad, donde se presenta una clasificación basada esencialmente en dos aspectos: el grado de heterogeneidad y el grado de comunicación. La taxonomía resultante de combinar estos dos atributos resulta simple e intuitiva:

1. Agentes homogéneos que no se comunican
2. Agentes heterogéneos que no se comunican
3. Agentes homogéneos que sí se comunicación
4. Agentes heterogéneos sí se comunicación

Cabe indicar que los autores presentan esta clasificación como marco de referencia en el cual analizar los problemas y desafíos existentes al encarar el desarrollo de un sistema en cada una de estos escenarios. A continuación repasaremos brevemente las conclusiones obtenidas para cada uno de estos contextos (los aspectos considerados se encuentran resumidos en la figura 3.2, reproducida de [Stone y Veloso, 2000]).

En la primer categoría, agentes homogéneos sin comunicación, todos los agentes poseen la misma estructura interna, incluyendo objetivos, conocimiento del dominio y acciones que pueden realizar. La única diferencia entre los agentes es que están situados en distintas locaciones del ambiente, razón por la cual sus sensores pueden percibir distintos estímulos y en consecuencia toman diferentes decisiones. Al diseñar esta clase de sistemas se debe decidir en primer lugar qué arquitectura de agentes utilizar, dependiendo de qué tan sofisticado deba ser el proceso de razonamiento en los mismos. La alternativa más sencilla la constituyen los agentes reactivos, los cuales no almacenan estado interno alguno y como su nombre lo indica simplemente reaccionan a los estímulos del ambiente, respondiendo a un conjunto de reglas establecidas previamente durante la propia etapa de diseño. En contraste, los agentes deliberativos son capaces de mantener un estado interno, lo que les permite predecir las consecuencias de sus acciones. Otra de las decisiones a tomar consiste en qué tipo de visión del mundo se le brindará a cada agente. Si se adopta una perspectiva global, cada agente tiene acceso a toda la infomación disponible, pero si se prefiere una visión local, cada agente se limitará a los parámetros que le competan. Por último, también se debe decidir si el estado de los otros agentes será modelado. Esto resulta de interés puesto que si una agente puede predecir las acciones de los restantes agentes, podrá elegir el curso de acción qué más beneficie a la totalidad del sistema.

La segunda categoría, agentes heterogéneos sin comunicación, brinda al diseñador más posibilidades, dado que los agentes pueden poseer distintos objetivos, modelos del mundo no necesariamente coincidentes, e incluso diferentes capacidades. Una opción importante a considerar es el modelo que seguirá la sociedad de agentes, que puede ser cooperativo o competitivo. Al tener diferentes objetivos pueden surgir conflictos, esto es, los agentes pueden llegar a interferir el uno con el otro. Por otra parte, también es factible que se formen alianzas entre agentes con objetivos comunes, que se organicen en grupos donde cada uno toma un rol diferente, con el propósito de alcanzar el objetivo acordado con mayor celeridad. No obstante, estas alianzas necesariamente serán estáticas, pues en esta categoría se asume que los agentes no se comunican. En este sentido, resulta un tanto controversial que se considere como posible un escenario donde el SMA es tal que sus componentes no se comunican entre sí, ya que de ser este el caso, la necesidad de apelar a un sistema multiagente resulta más que objetable. En diversas fuentes [Wooldridge y Ciancarini, 1999, Jennings, 2000] se destaca a la interacción como un elemento esencial a todo SMA, y no puede existir interacción si no se dispone de alguna forma de comunicación entre los agentes. De hecho, al incorporar la comunicación a los escenarios anteriormente descriptos,

Homogéneos sin comunicación	**Heterogéneos sin comunicación**
- Reactivo o deliberativo	- Cooperativo o competitivo
- Visión local o global	- Estático o dinámico
- Modelar el estado de otros agentes	- Modelar el estado de otros agentes
	- Distribución de los roles
Homogéneos con comunicación	**Heterogéneos con comunicación**
- Sensado distribuido	- Elección del protocolo y lenguaje
- Tipo de información comunicada	- Planificación de los actos de habla
	- Negociación

Figura 3.2: Una taxonomía de sistemas multiagente

dando a lugar a las dos últimas categorías de esta taxonomía, se enriquece la capacidad del modelo, ya que por medio de la interacción es ahora posible coordinar el comportamiento de los agentes de forma más eficiente.

En el caso de agentes homogéneos, hay dos aspectos a ser considerados: el contenido de la información comunicada y el sensado distribuido del ambiente. Con respecto al contenido de la información comunicada existen dos posibilidades, compartir la totalidad de la información disponible, o bien reservar parte de ésta para darle un uso privado. En relación al sensado del ambiente, en este escenario resulta factible que distintos agentes situados en diferentes ubicaciones colaboren para formar un modelo común del ambiente. Por caso, un sistema para informar sobre el estado del tráfico en una autopista, donde cada automóvil informe acerca de su ubicación actual.

La última categoría, agentes heterogéneos con comunicación, abarca a los sistemas con mayor potencial, pero que a su vez presentan una mayor complejidad. En este escenario, se manifiestan ciertos inconvenientes que deben ser atendidos para hacer efectiva la comunicación entre agentes, como fijar el protocolo de comunicación y su respectivo lenguaje. Sin duda, el lidiar con agentes que presentan diferencias sensibles entre sí hace de esta tarea una empresa no trivial. Una faceta interesante que ha sido explorado en este contexto consiste en interpretar a la misma transmisión de información entre agentes como un tipo especial de acción. Esta concepción permite establecer precondiciones y efectos para cada acto de comunicación, resultando en un tratamiento uniforme de la interacción entre los agentes, y ha dado a lugar a la la teoría de *actos de habla* desarrollada por los autores Cohen y Levesque [Cohen y Levesque, 1995].

3.5. Aspectos a resolver en la etapa de diseño

Hasta el momento hemos analizado las generalidades de los sistemas multiagente, detallando sus principales aplicaciones y enfocándonos a su vez en las distintas categorizaciones ensayadas en la literatura. Para culminar esta introducción a los SMA hemos elegido reseñar un conjunto de aspectos a resolver en la etapa de diseño de todo sistema, considerados por primera vez en [Bond y Gasser, 1988], si bien recién Gasser en [Gasser, 1991] los identificó como problemas inherentes al diseño e implementación de todo sistema multiagente.

A continuación se exponen cada uno de los desafíos de acuerdo a lo planteado en su formulación más reciente [Gasser, 1991], citando en donde corresponda las principales propuestas desarrolladas tendientes a resolver cada uno de ellos.

- ¿Cómo brindar a los agentes la capacidad de descomponer sus objetivos y tareas, entregar subtareas a otros agentes y sintetizar resultados parciales o soluciones?

  En la mayor parte de las propuestas concretas la responsabilidad de establecer los criterios para descomponer y asignar las tareas recae en el diseñador del sistema. Entre los sistemas automatizados que intentan resolver este desafío podemos mencionar al protocolo de redes de contratos,[5] cuyo principal propósito es llevar adelante la descomposición y asignación de tareas. No obstante, como su especificación inicial [Davis y Smith, 1983] no especifica un modelo

[5]concepto considerado en detalle en la sección 3.6, pág. 170.

formal para llevar a cabo el anuncio y adjudicación de tareas, esta responsabilidad (por cierto, nada trivial), queda en manos de cada implementación en particular. En cada uno de los sistemas en concreto se han propuesto diversos criterios para la descomposición, como por caso la funcionalidad, los niveles de abstracción o las dependencias de control.

- ¿Cómo otorgar a los agentes la capacidad de comunicarse entre sí? ¿Qué protocolos y lenguajes de comunicación emplear?

  Los aportes más significativos en este sentido han sido los lenguajes de comunicación entre agentes y los protocolos de interacción introducidos en la literatura. Hoy en día los lenguajes de comunicación entre agentes más utilizados son FIPA-ACL y KQML, los que serán posteriormente abordados en detalle en la sección 3.6.1. En cuanto a los protocolos de interacción, existen diferentes modelos, dependiendo esencialmente de la situación en la que deban aplicarse (*e.g.*, escenarios con agentes cooperativos, antagónicos, etc). No obstante, todavía persisten importantes limitaciones, por lo que estos modelos no resultan aplicables a toda situación. Más adelante, la sección 3.6 considera las principales virtudes y limitaciones de las propuestas más relevantes en este sentido.

- ¿Cómo permitir a los agentes que representen sus acciones, planes y conocimiento, a fin de razonar y poder así interactuar de forma racional?

  Este desafío constituye uno de los puntos cruciales y de mayor complejidad en el desarrollo de un SMA. En general, las teorías desarrolladas se han basado en atribuirle a los agentes actitudes mentales, como creencias, deseos e intenciones. Adaptando esta idea, se han presentado varias *teorías de agencia*, es decir, teorías que explican como los distintos elementos, tales como creencias, metas, acciones e interacciones, son combinadas para generar el comportamiento del agente.

  Las dos propuestas más exitosos en este terreno se basan en una lógica modal temporal: la teoría de intención de Cohen y Levesque [Cohen y Levesque, 1990] y el modelo usualmente conocido bajo la sigla BDI (por su denominación en inglés "*beliefs, desires, and intentions*"), introducido en [Rao y Georgeff, 1991, Rao y Georgeff, 1995]. El modelo de Cohen y Levesque sólo toma como primitivas las creencias y las metas, donde las otras actitudes (por caso, la intención) se construyen a partir de éstas. En franco contraste, Rao y Georgeff argumentan en su propuesta que las intenciones no pueden ser derivadas en base a las restantes actitudes, por lo que deben ser consideradas como primitivas.

Si bien estas propuestas han tenido una gran aceptación, aún deben resolver el problema central de este tipo de teoría de agencia basadas en lógicas modales: su alto costo computacional. En consecuencia, gran parte de los sistemas implementados siguiendo este enfoque simplifican de forma significativa la teoría adoptada, resultando en sistemas simplistas que guardan una tenue relación entre la teoría y la práctica.

- ¿Cómo otorgar a los agentes la capacidad de modelar el estado actual del proceso de interacción y de razonar sobre el mismo?

  Este aspecto depende en gran medida del protocolo de interacción adoptado por los agentes. La mayoría de las propuestas concretas apelan a la teoría de juegos, adecuando la teoría de utilidades para modelar la elección racional de cada agente. En este contexto, el propio protocolo de interacción es quien debe ocuparse de definir los distintos criterios, tales como el de iniciación o el de terminación, los que a su vez permiten razonar sobre el proceso de interacción que se está llevando a cabo.

  Una novel variante surgida recientemente concibe la interacción entre los agentes del sistema como un mero diálogo, capitalizando los resultados ya obtenidos en el seno de las diversas teorías argumentativas de caracter dialéctico introducidas en la literatura. En esta oportunidad, el estado del proceso de interacción se asocia al modelo dialéctico adoptado por la teoría argumentativa.

- ¿Cómo permitir a los agentes que reconcilien sus puntos de vista conflictivos o dispares?

  La alternativa más evidente para atacar este desafío se inspira en el proceso empleado por las propias personas para reconciliar puntos de vista dispares: la negociación. En el marco de los sistema multiagente, la negociación denota al mecanismo de coordinación que permite que dos o más agentes, con objetivos posiblemente encontrados, alcancen una desición conjunta en relación al tópico que generó la confrontación. En la literatura se han esayado diversas alternativas para modelar este proceso, de las que destacan las propuestas de Sycara y Kraus [Kraus et al., 1998] y de Zlotkin y Rosenschein [Zlotkin y Rosenschein, 1996] entre otros. Un modelo de negociación que merece una mención especial es el propuesto por Simon Parsons en [Parsons et al., 1998], que sigue el enfoque de modelar la negociación entre agentes como si fuera un proceso argumentativo. La sección 3.6.4 resume estas propuestas, considerando sus virtudes y desventajas.

- ¿Cómo diseñar e implementar sistemas multiagente de utilidad práctica, descubriendo en el proceso, de ser necesario, nuevas técnicas de desarrollo para los SMA?

  En los últimos años se han diseñado diversos sistemas multiagente de utilidad práctica, aplicados satisfactoriamente en los dominios más dispararares. No obstante, aún no existe una metodología de diseño estándar, por lo que cada proyecto tiende a desarrollar sus propias técnicas de diseño de forma un tanto *ad-hoc*. En ocasiones, las técnicas propuestas para un sistema en concreto se han transformado en modelos genéricos aplicados satisfactoriamente a otros dominios, o bien en marcos de referencia con los cuales contrastar nuevas técnicas de desarrollo, tal como fue el caso del *Distributed Vehicle Monitoring Testbed* comentado en la sección 3.3, o incluso el mismo protocolo de redes de contratos [Davis y Smith, 1983].

En este sentido, la programación orientada a agentes ha surgido como una metodología para el desarrollo de sistemas de alta complejidad, en los que la interacción entre los distintos componentes resulte trascendental. En esta nueva área se investiga como especificar, implementar y verificar esta variante de sistemas, desarrollando las herramientas adeacuadas a tal fin.

Estos aspectos a resolver durante la etapa de diseño resumen la problemática enfrentada por la comunidad de sistemas multiagente. Less Gasser señala en [Gasser, 1991] que las solución adoptada para cada uno se encuentra entrelazada con las restantes. Por caso, la elección de un cierto conjunto de procedimientos para la comunicación e interacción tendrá una fuerte influencia en el tipo de coordinación entre agentes que pueda desarrollarse.

Una consecuencia directa del análisis de estos aspectos es que la interacción entre los agentes se constituye en uno de los componenetes preponderantes de todo sistemas multiagente. Elaborando en este sentido, Michael Wooldridge señala en [Wooldridge y Ciancarini, 1999] que es la interacción la propiedad que distingue a los sistemas de software de gran complejidad, donde el acercamiento tradicional no suele dar los resultado esperados y resulta más natural apelar al enfoque de agentes múltiples. Nick Jennings, en [Jennings, 2000], se atreven a dar un paso más allá al afirmar que a futuro la computación en sí será entendida como mero un proceso de interacción entre componentes.

3.6. La interacción entre los agentes

Se podría afirmar sin temor a equivocarse que prácticamente la totalidad de las virtudes que presentan los sistemas multiagente dependen en algún grado de la interacción entre sus agentes. Como comentáramos anteriormente, es precisamente la interacción en el seno de un sistema multiagente la que permite que el sistema como un todo pueda alcanzar objetivos que se encuentren incluso más allá de las capacidades de cada uno de sus integrantes.

En lo que resta de esta sección analizaremos los requerimientos necesarios para que la tan mentada interacción pueda ser llevada a cabo. En primer lugar, consideraremos las distintas alternativas para lograr una comunicación efectiva entre los agentes, estudiando los acercamientos de mayor relevancia, considerando a continuación los principales protocolos de interacción. Estos protocolos establecen el marco en el cual los agentes podrán intercambiar información, distinguiéndose tres tipos principales de interacciones posibles: la coordinación, la cooperación y la negociación.

3.6.1. Lenguajes de comunicación entre agentes

Desde un punto de vista histórico, el surgimiento de la comunicación entre los seres humanos posibilitó el desarrollo de la sociedad y la evolución de la especie, haciendo posible la interacción compleja entre los distintos miembros de la comunidad. De una manera análoga, la comunicación entre agentes permite que éstos coordinen su comportamiento, lo cual redunda en sistemas mucho más coherentes y menos caóticos.

De acuerdo a lo expuesto en [Huhns y Stephens, 1999], en el estudio formal y acabado de la comunicación entre agentes se distinguen tres aspectos principales:

- La sintaxis.
- La semántica.
- La pragmática.

La *sintaxis* establece cuál es el conjunto de las reglas de estructuración de los símbolos de comunicación. La *semántica* dota de significado a cada una de las estructuras de símbolos consideradas válidas de acuerdo a la sintaxis en curso. Por último, la *pragmática* señala qué uso se le puede dar a estos símbolos y estructuras.

Otro aspecto instructivo es considerar con qué propósito se comunican los agentes: en general se comunican a fin de comprender a los otros agentes, y a su vez de lograr ser entendidos por éstos. No es sólo el contenido del mensaje lo que resulta importante, sino también el conjunto de circunstancias que lo rodean, de las cuales dependerá el significado de lo comunicado.

Si consideramos cualquier conjunto de entidades que intercambian mensajes (como por ejemplo humanos, primates o delfines), observaremos que la comunicación se realiza por medio de un lenguaje en común, conocido y entendido por todos los participantes. En el caso de los SMA, la contrapartida son los lenguajes de comunicación entre agentes (usualmente denotados como ACL, por su sigla en inglés *Agent Communication Language*), que surgieron como un medio para brindar y requerir información a otros agentes. Aún cuando se han ensayado otros medios de comunicación en las distintas aplicaciones, tales como la técnica de llamado remoto a procedimientos o el pasaje de mensajes para la comunicación entre procesos, o el propio estandar CORBA para la comunicación entre objetos, los ACL se distinguen por brindar un nivel de abstracción superior y contar con una semántica especificada formalmente. Yannis Labrou enfatiza esta distinción en [Labrou et al., 1999], señalando que los ACL se caracterizan por estar específicamente definidos para:

- Manejar proposiciones, reglas y acciones, esto es, elementos con una semántica asociada, en vez de operar sobre simples objetos sin un significado preestablecido.
- Contar con mensajes declarativos, esto es, mensajes que definen el estado que se desea alcanzar, y no los procedimientos o métodos que se deban ejecutar.

Desde ya, el solo hecho de contar con un ACL correctamente definido no soluciona el problema de modelar la interacción en un sistemas multiagente. Los agentes deben ser capaces de intercambiar conceptos que estructuren mayor cantidad de información, tales como planes o metas, respetando el protocolo de coordinación, cooperación o negociación adoptado. Por otra parte, al usar un ACL los mensajes se transmiten a través de un medio físico (tal como una red de computadoras), apelando a protocolos de un nivel inferior, como puede ser el caso de `TCP/IP` o `HTTP`. En consecuencia, el lenguaje de comunicación entre agentes se ubica conceptualmente entre estas dos capas, como se ilustra en la figura 3.3.

La primer contribución a lenguajes de comunicación entre agentes es la teoría de los *actos de habla*, que fuera introducida en [Austin, 1962, Searle, 1970] para estudiar

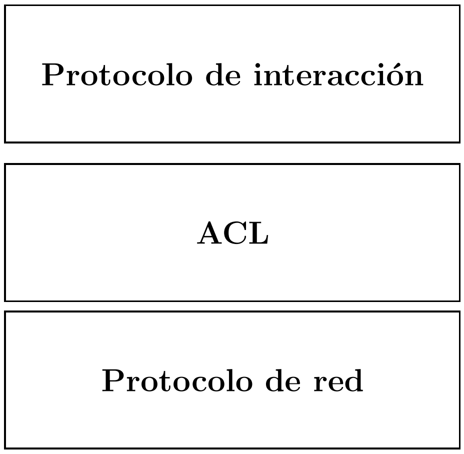


Figura 3.3: Rol de los ACL en la interacción entre agentes

la conversación coloquial entre seres humanos. Esta propuesta ha sido empleada como marco de referencia para el desarrollo de distintos mecanismos de comunicación entre agentes.

La teoría de los actos de habla se inició con un grupo de filósofos interesados en la *prágmatica* del lenguaje natural. La pragmática concibe al lenguaje natural simplemente como un conjunto de *manifestaciones*, los que consisten en descripciones de un estado deseado del mundo, junto con otros aspectos sobre el contexto de producción e interpretación del mensaje. La idea básica consiste en interpretar a las manifestaciones no sólo como meras descripciones, sino como acciones que pueden cambiar el estado del mundo, modificando por caso el estado mental de los participantes. De acuerdo a lo expuesto en [Austin, 1962], existen distintos tipos de acciones llevadas adelante al hablar:

1. **Locución:** la locución resume el acto de producir el mensaje. Abarca un *acto fonético* (emitir ciertos sonidos), un *acto gestual* (produciendo palabras que pertenecen a un cierto vocabulario en concordancia con una gramática) y un *acto retórico*, que capitaliza el producto del acto gestual con un propósito en particular.

2. **Ilocución:** los actos ilocutorios son aquellos realizados al manifestar algo, como por ejemplo al formular o responder una pregunta. Están compuestos de una *fuerza ilocutoria*, que especifica el tipo de acción (*e.g.* requerir, implorar, exigir, etc.), y un *contenido proposicional*, que explicita los detalles de la acción, por caso qué se le solicita al destinatario o cuál es la respuesta a una pregunta recibida anteriormente.

3. **Perlocución:** Los actos perlocutorios consisten de acciones que alcanzan resultados que no dependen del mensaje en sí, sino del contexto y de la situación en particular en que se produce la manifestación. Algunos ejemplos incluyen la persuasión, la sorpresa o la amenaza. Si bien quien produce el mensaje puede desear causar alguno de estos efectos, requiere más que la manifestación para alcanzarlos.

Por citar un ejemplo concreto, supongamos que un cliente le quiere comunicar al vendedor *"deseo comprar un tipo específico de caramelos"*. Para hacerlo, posiblemente emitirá la frase "quiero estos caramelos" mientras apunta con su dedo índice a los caramelos deseados. Notar que tanto la frase como el gesto por separado carecen de significado, alcanzandose el resultado deseado (*i.e.*, comunicar el mensaje correcto), cuando ambos tienen a lugar en simultáneo.

Ya que la teoría filosófica de los actos de habla se ha utilizado para modelar al lenguaje natural, resulta razonable que pueda modelar lenguajes de menor complejidad, tales como los lenguajes de comunicación entre agentes. Los ACL más difundidos en la actualidad son KQML y FIPA-ACL, así como las diversas variantes surgidas de cada uno de ellos. Todos estos lenguajes se encuentran basados en la teoría de los actos de habla. En lo que resta de la presente sección abordaremos los conceptos básicos de éstas propuestas, tomando como punto de partida el análisis expuesto en [Labrou et al., 1999, Labrou, 2001].

KQML

El lenguaje para manipulación y consulta de conocimiento conocido por su sigla como KQML,[6] introducido inicialmente en [Finin et al., 1992], es un protocolo para intercambiar información tanto entre agentes como entre programas de aplicación. KQML ha sido implementado en forma independiente por varios grupos de investigación, y se ha usado exitosamente para desarrollar varios sistemas de información con diferentes arquitecturas de software. Se trata de un lenguaje de comunicación de alto nivel orientado a mensajes, en el cual la interacción tiene a lugar en múltiples niveles. En particular, es independiente del protocolo de transporte, si bien usualmente se suele usar `SMTP`, o directamente `TCP/IP`. Por otra parte, KQML es independiente del lenguaje adoptado para codificar el contenido del mensaje (por caso, Prolog, DeLP, etc.), así como de la ontología empleada por el mismo.

Yannis Labrou señala en [Labrou et al., 1999] que conceptualmente se pueden identificar tres niveles en un mensaje KQML, a saber:

- **El nivel de contenido:** Es el nivel que opera con el mensaje codificado de acuerdo a la representación elegida por la aplicación. KQML puede utilizar cualquier tipo de codificación en este nivel, ya que por definición toda implementación debe ignorar estos aspectos.

- **El nivel de comunicación:** Este nivel incluye un conjunto de características del mensaje que describen los parámetros de comunicación de un nivel más bajo, tales como la identidad del emisor y del receptor.

- **El nivel de mensaje:** Es el nivel que encierra al contenido que una entidad desea transferir a otra, en esencial el núcleo de KQML. La función principal de este nivel es identificar el protocolo de red con el cual se enviará el mensaje, que incluye al acto de habla o *performativa* que el emisor desea transmitir. Este acto de habla indica si el mensaje es una afirmación, una consulta o cualquier otra de las performativas válidas.

[6] del inglés *Knowledge Query and Manipulation Language*.

```
(ask_one
          :sender      george walker
          :content     (QUOTE EUR ?euro)
          :receiver    xe_server
          :language    LPROLOG
          :ontology    XE )
```

Figura 3.4: Un mensaje expresado en KQML

La figura 3.4 ejemplifica la estructura básica de un mensaje en KQML, donde se puede apreciar a la entidad George Walker consultando la cotización actual del euro respecto al dolar. Obsérvese que KQML dota al mensaje a ser comunicado de un envoltorio que puede ser comprendida por cualquiera de los agentes del sistema, si bien para comprender el contenido del mensaje, el receptor debe entender el mismo lenguaje y debe tener acceso a la ontología utilizada. Por esta razón, dado que el contenido del mensaje no está accesible, el nivel de mensaje puede incluir un conjunto de atributos opcionales que describan el lenguaje de contenido, la ontología que éste asume, así como algún tipo de descripción del contenido del mensaje. Estos parámetros hacen posible que algunas implementaciones de KQML analicen, encaucen y hagan llegar a destino mensajes cuyo contenido desconozcan o no puedan comprender.

La sintaxis de KQML está fuertemente inspirada en la sintaxis del lenguaje de programación Lisp. Los argumentos son identificados por su respectiva palabra clave, precedida por el símbolo dos puntos, y admiten ser estructurados en cualquier orden. Las performativas adoptadas por KQML fueron tomadas del conjunto de performativas estudiadas en la teoría de actos de habla, por lo que en consecuencia poseen una semántica independiente del dominio de aplicación. Concretamente, la semántica del mensaje se define en los campos :content (el mensaje a comuni-

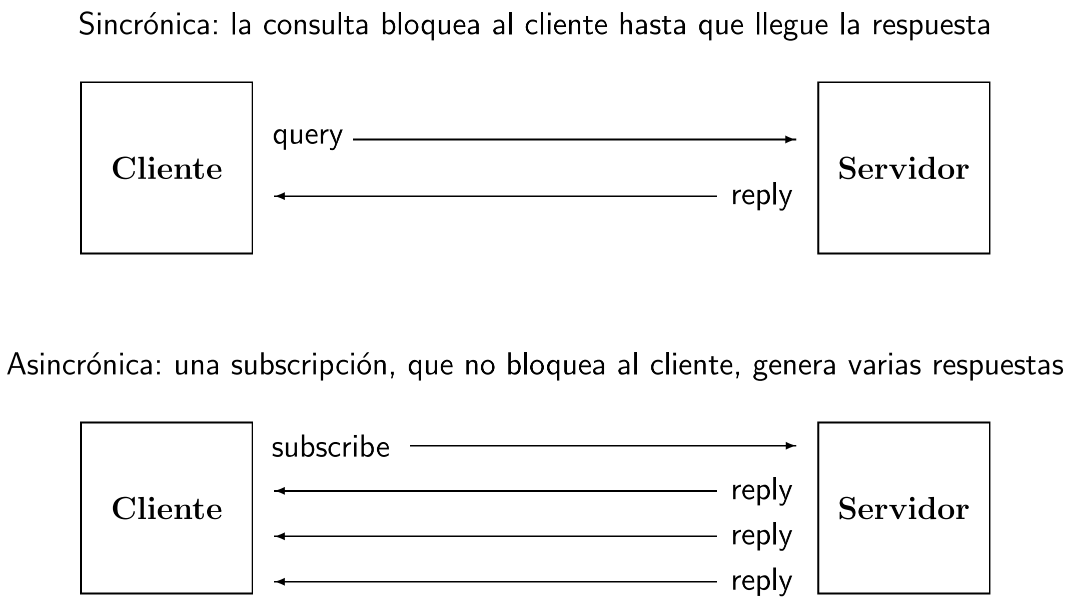


Figura 3.5: Comunicación entre agentes sincrónica y asincrónica via KQML

car), :language (el lenguaje en el cual se expresa el mensaje) y :ontology (el marco ontológico en el cual se compone el mensaje).

Los agentes que utilizan KQML se visualizan mutuamente como clientes y servidores de mensajes. Las comunicaciones que tienen lugar entre ellos pueden ser *sincrónicas* o *asincrónicas*, tal como lo ilustra la figura 3.4, adaptada de un ejemplo expuesto en [Huhns y Stephens, 1999]. Para establecer una comunicación sincrónica el agente que ha emitido el mensaje queda a la espera de la respuesta. En contraste, para establecer una comunicación asincrónica el agente emite el mensaje, pero continúa con sus tareas, que serán interrumpidas recién cuando llegue la respuesta respectiva.

De acuerdo a lo expuesto en [Huhns y Stephens, 1999], el conjunto predefinido de performativas en KQML se puede clasificar en siete categorías básicas, a saber:

- Performativas de consultas simples (formulación de preguntas).
- Performativas de consultas con múltiples respuestas (por ejemplo, solicitar el envío periódico de una cotización).

- Performativas de respuestas (atención de un pedido).
- Performativas informacionales genéricas (por caso, para comunicar algún logro efectuado por el agente o informar de alguna situación).
- Performativas para generadores (tales como permanecer en espera, generar el próximo valor, descansar).
- Performativas para definir capacidades (*e.g.*, publicitar capacidades, subscribir servicios, monitorear prestaciones).
- Performativas para redes de comunicaciones (reenvío de un mensaje).

El conjunto de performativas disponibles no es fijo, sino extensible. Cada implementación en particular puede optar por ampliar o restringir este conjunto, si bien si se incorpora alguna de las performativas predefinidas, se debe respetar su semántica esperada.

En relación a la semántica de KQML como lenguaje de comunicación entre agentes, cabe acotar que en los primeros años de uso del lenguaje sólo existió una descripción informal y parcial de ésta. La primeras críticas hicieron hincapié precisamente en esta falta de formalidad, por lo que posteriormente Labrou y Finnin introdujeron en [Labrou y Finin, 1998] una formalización más acabada. Esta propuesta se basa en asignar *pre-condiciones*, *post-condiciones* y *condiciones de completitud* a cada performativa. Por caso, suponiendo que **A** sea el emisor y **B** el receptor, entonces las pre-condiciones indican los estados necesarios para que un agente envie una performativa (denotado como $Pre(\mathbf{A})$), y para que el receptor la acepte y procese con éxito (denotado como $Pre(\mathbf{B})$). En contraste, si las pre-condiciones no fueran satisfechas, se genera un mensaje de error. Las post-condiciones describen el estado del emisor después de que realizó una performativa en forma exitosa, así como el estado del receptor después de procesar el mensaje recibido, pero antes de responder al mismo. Se asume que las post-condiciones $Post(\mathbf{A})$ y $Post(\mathbf{B})$ se verifican a menos que sea retornado un mensaje de error como respuesta. Por último, la condición de completitud para una performativa indica el estado final, que se alcanza luego de que haya acabado la interacción, cuando ya se cumplió la performativa que la motivó.

Esta caracterización de la semántica de KQML es una de las más difundidas, si bien se han explorado otros acercamientos alternativos, tal como el introducido por Smith y Cohen en [Smith y Cohen, 1995], erigido en torno al concepto de *agencia racional*. En este modelo, cuando dos o más agentes racionales desean comunicarse conforman un equipo de trabajo, donde los actos comunicativos básicos (denominación adoptada en este contexto para las performativas), se reinterpretan como una intención en común del equipo.

En síntesis, KQML fue el primer lenguaje de comunicación entre agentes que se constituyó en un estandar indiscutido. Su modelo de desarrollo está basado en la experimentación y la retroalimentación continua por parte de la comunidad y los usuarios, lo que permitió que el lenguaje se adecue a los distintos aspectos prácticos necesarios para el desarrollo de sistemas multiagentes de envergadura.

Entre las objeciones que pueden señalarse, KQML carece de una especificación formal que cuente con el concenso de la comunidad. Esta falta de uniformidad impide que los distintos sistemas que han adoptado este modelo de comunicación puedan interoperar entre sí, objetivo especialmente atractivo en el marco de los sistemas abiertos. Con respecto a la semántica de KQML, si bien es interesante contar con formalizaciones alternativas, lamentablemente la comunidad no ha alcanzado el concenso acerca de qué modelo semántico utilizar.

FIPA

La asociación FIPA, conocida bajo esta denominación por ser la sigla de su afiliación completa en inglés (*Foundation for Intelligent Physical Agents*), se formó en la década del '90 con el objeto de promover el desarrollo de aplicaciones y servicios basados en sociedades de agentes. Una de las principales metas de esta organización es estandarizar aquellos conceptos más utilizados en el desarrollo de sistemas multiagente. En este sentido, especificaron formalmente un lenguaje de comunicación entre agentes altamente flexible, hoy en día denominado FIPA-ACL. Al igual que KQML, FIPA-ACL fue definido a partir de la teoría de actos de habla. En el contexto de FIPA-ACL, los mensajes entre agentes se interpretan como acciones o como actos comunicativos, ya que el mero hecho de haber sido enviados permite suponer que un cierto agente pretende realizar alguna acción en particular. Su especificación,[7] consiste de un conjunto de tipos de mensaje, y de una descripción de su pragmática, esto es, de los efectos que tienen en las actitudes mentales del emisor y del receptor. Estos actos comunicativos están descriptos mediante una prosa acompañada de una semántica formal, la cual está expresada en la lógica modal conocida como SL.

En líneas generales, el lenguaje FIPA-ACL es análogo a KQML, ya que su sintaxis es similar, salvo ciertas primitivas que han sido denominadas de diferente forma. Se adotpa la misma política que en KQML de separación del lenguaje interno de los mensajes de su lenguaje externo. El lenguaje externo define el significado pretendido del mensaje, mientras que el lenguaje interno, también conocido como contenido, denota la expresión a la cual se aplican los deseos, intenciones o creencias del interlocutor. Aún cuando el documento de especificación de FIPA-ACL manifiesta que éste no está ligado a ningún lenguaje de contenido en particular, el procesar y comprender algunas de las primitivas especificadas requiere que los agentes puedan interpretar SL. Este lenguaje está definido en base a una lógica modal con múltiples cuantificadores, que cuenta con operadores modales para las creencias ($\mathbb{B}$), los deseos ($\mathbb{D}$), las creencias inciertas ($\mathbb{U}$) y para las metas persistentes ($\mathbb{PG}$).

En FIPA-ACL cada acto comunicativo se especifica mediante un conjunto de fórmulas SL que describen las *precondiciones de factibilidad* así como el *efecto racional* del acto comunicativo. Para un acto comunicativo a, las precondiciones de factibilidad $\mathbf{FP}(a)$ especifican las condiciones necesarias para poder llevar adelante ese acto comunicativo, en tanto que el efecto racional $\mathbf{RE}(a)$ representa el efecto que un agente puede esperar obtener como resultado de realizar la acción denotada

[7]siempre disponible en su versión más actulizada en `www.fipa.org`.

[i, inform(j,ϕ)]

FP: $\mathbb{B}_i(\phi) \wedge \neg\mathbb{B}_i(\mathbb{B}_i f_j(\phi) \vee \mathbb{U}_i f_j(\phi))$

RE: $\mathbb{B}_j(\phi)$

Figura 3.6: Un acto comunicativo expresado en el lenguaje FIPA-ACL

en a. No obstante, el agente que reciba un cierto mensaje no necesariamente asegura que el efecto esperado tendrá a lugar; de hecho, es concebible que el efecto deseado sea incluso imposible. En otras palabras, los distintos agentes pueden utilizar el conocimiento acerca de los efectos racionales de cada acto de habla para enriquecer su proceso de toma de decisiones, si bien no pueden asumir que el comunicar un acto a asegure el cumplimiento de sus efectos racionales $\mathbf{RE}(a)$.

La figura 3.6 bosqueja la estructura usual de un acto comunicativo de FIPA-ACL, en este caso informar, a través del cual el agente i informa al agente j del contenido ϕ. El contenido de este informe es una proposición, y su significado es que el emisor comunica al receptor que esa proposición es verdadera. De acuerdo a la semántica de este mensaje, el emisor:

- sostiene que la proposición $\mathbb{B}_i(f)$ es verdadera,
- no considera que el receptor tenga algún conocimiento previo acerca de la proposición $\mathbb{B}_i(\mathbb{B}if_j(f) \vee \mathbb{U}if_j(f))$, y además
- intenta que el agente receptor llegue a considerar que esta proposición es verdadera (debido al efecto racional $\mathbb{B}_j(f)$).

Entre las virtudes de lenguaje FIPA-ACL se debe resaltar que provee no sólo una especificación de sus aspectos sintácticos sino que abarca también la semántica. Otra característica atractiva, producto de haber contado con el respaldo de la organización homónima, es que fue desarrollado en forma ordenada y prolija, respetando fielmente sus objetivos de diseño, lo que permitió obtener un ACL con una semántica clara y bien definida.

Una incipiente vertiente de investigación, la cual está acaparando gran parte de la atención de la comunidad de FIPA-ACL, explora integrar este lenguaje con la *world wide web*. Por lo general, los lenguajes de comunicación entre agentes se han mantenido al margen de las tecnologías y estándares desarrollados para internet. Por citar un ejemplo, Grosof y Labrou introducen en [Grosof y Labrou, 2000] un mecanismo para recodificar los mensajes FIPA-ACL en términos del lenguaje XML. Cabe señalar que esta recodificación no afecta la semántica de los actos comunicativos, la cual permanece invariable ya que por definición no depende de la sintaxis adoptada. Este entrecruzamiento de tecnologías en principio facilita las tareas de ingeniería de software involucradas en el desarrollo de los agentes, permitiendo aprovechar en simultáneo todos los servicios brindados por internet, tales como el uso de certificados o el propio protocolo `SSL` para poder comprobar fehacientemente la identidad de los agentes.

Finalmente, un aspecto pragmático que aún no ha sido atendido de forma satisfactoria por FIPA-ACL es el tratamiento de las primitivas de registración y de selección. Estas primitivas se encargan de tareas tales como tomar nota que un cierto agente cuenta con una habilidad en particular, o como la busqueda y selección de agentes que sean capaces de llevar adelante una cierta tarea. Si bien KQML cuenta con primitivas específicas para realizar estas tareas, como FIPA-ACL en contraste pretende ser un lenguaje puro, no introduce primitivas de este tipo. Dado el importante rol que estas tareas desempeñan en el marco de un sistema multiagente, la ausencia de estas primitivas ha sido una de las principales críticas para con este lenguaje.

3.6.2. Técnicas de coordinación

Los lenguajes de comunicación entre agentes considerados en la sección anterior constituyen efectivos mecanismos para comunicar *un* mensaje. El siguiente nivel de abstracción en la interacción entre agentes abarca el diálogo resultante del intercambio de una *secuencia* de mensajes individuales. El concepto resultante de diálogo permite focalizar la atención en formas más elaboradas de interacción, tales como la *coordinación*, la *cooperación* o la *negociación* entre agentes. Estos patrones de interacción serán abordados en detalle en lo que resta del presente capítulo.

En lo que respecta a la coordinación, resulta un concepto recurrente en la mayoría de los sistemas multiagente, debido principalmente a dos razones: en primer lugar cada agente cuenta sólo con una visión parcial del estado de las cosas, y por otra parte es posible que sus propias acciones interfieran con las acciones de los otros agentes. El no contar con alguna forma de coordinación entre los agentes del sistema puede dar a lugar a comportamientos caóticos, donde las acciones de uno pueden interferir las de otro.

En la literatura, Nick Jennings [Jennings, 1996] define a la coordinación como *el proceso mediante el cual un agente razona sobre sus acciones locales y sobre las acciones anticipadas de otros agentes, con el objeto de lograr que la comunidad actúe de forma coherente*. Lesser y Corkill proponen otra forma de caracterizar a esta noción, en esta ocasión en base a sus efectos, señalando que todo sistema que presenta un comportamiento coordinado manifiesta las siguientes características:

1. Cada una de las subtareas requeridas para solucionar el problema original pueden ser llevadas adelante por alguno de los agentes.

2. Las soluciones parciales de cada agente pueden integrarse para formar la solución global del problema.

3. Los agentes tienen un propósito conciso.

4. Todas estas metas son alcanzables dentro de las limitaciones computacionales y de recursos existentes en el sistema.

La forma más evidente de que el sistema presente estas características sería designar un agente que actue como controlador central, recopilando información de todos los agentes del sistema, formulando los planes y asignando las subtareas. Sin embargo, en la mayoría de los escenarios concretos no resulta factible implementar este modelo de toma de decisiones centralizado. En cierto sentido, las mismas razones expuestas en desmedro de los sistemas centralizados se pueden aplicar a los SMA que centralizan excesivamente el control o la toma de decisiones. Esta objeción fue abordada inicialmente por Thomas Malone en [Malone, 1990], donde plantea que es posible construir sistemas coordinados que no dependan de un controlador central. Esta propuesta gira en torno a tres pricipios fundamentales, que se exponen a continuación.

- **Ajuste mutuo:** es una forma de coordinaciño que ocurre cuando dos agentes comparten recursos para alcanzar un objetivo común.
- **Supervisión directa:** se distingue porque uno o mas agentes establecen una relación en la cual un individuo tiene control sobre los demás, administrando además los recursos compartidos.
- **Estandarización:** este tipo de coordinación puede verse en realidad como una forma especial de control que ejerce un agente supervisor, estableciendo un conjunto de procedimientos genéricos que los subordinados deben seguir ante determinadas situaciones.

En base a estos mecanismos de coordinación, Malone argumenta que se pueden capturar dos de los modelos de coordinación de comportamientos más conocidos: las *jerarquías* y los *mercados*.

Las jerarquías están basadas en procesos de supervisión directa. Esta modalidad es particularmente útil en los sistemas con un gran número de agentes, puesto que aquellos mecanismos más sencillos tales como el ajuste mutuo no resultan viables, pues el volumen de información comunicada crece exponencialmente con el número de participantes. En una jerarquía se divide al conjunto de agentes en pequeños grupos de tarea, de manera tal que la transferencia de información tenga a lugar mayormente dentro de cada grupo. Los eventuales intercambios de información fuera del ámbito del grupo de tareas se realizan a través del agente que asuma el rol de supervisor, estableciendo de esta forma una jerarquía dentro de cada grupo. De igual forma, la coordinación dentro de cada grupo de tareas se puede llevar adelante mediante ajuste mutuo o aplicando nuevamente una jerarquía.

Los mercados están basados en torno al ajuste mutuo. En este modelo existe un conjunto de agentes que no se encuentra relacionado jerárquicamente entre sí, donde cada agente controla un grupo escaso de recursos (*e.g.*, materias primas, mano de obra, etc.). Los distintos agentes arreglan compartir algunos de los recursos que administran con el objeto de poder alcanzar una determinada meta: dos o más agentes ceden sus recursos temporariamente y un agente se convierte en supervisor de la tarea que se esté desarrollando.

Si bien este análisis en relación a cómo dotar a los sistemas de un comportamiento coordinado apelando a jerarquías y mercados resultó influyente en desarrollos posteriores, su autor no sugiere qué conjunto de técnicas o principios se deben aplicar en la construcción de sistemas que presenten estas características. En la última década, la atención de los investigadores de este área se focalizó en el desarrollo de técnicas para distribuir el control y la gestión de la información, ambas tendientes a coordinar el comportamiento de los agentes del sistema. A continuación exploraremos las propuestas más difundidas, siguiendo la presentación delineada en [Jennings, 1996].

Estructuras de búsqueda AND/OR

Victor Lesser propuso en [Lesser, 1991] una de las soluciones más intuitivas para resolver el problema de identificar cuáles son las actividades que requieren la intervención coordinada de varios agentes. Su propuesta esencialmente consiste en representar las acciones de los agentes mediante una búsqueda en un grafo de objetivos AND-OR. Este grafo de objetivos modela las dependencias entre las metas (nodos internos) y los recursos necesarios para alcanzarlas (nodos hoja). El formular los objetivos del sistema multiagente de esta forma permite que las actividades que requieran coordinación resulten evidentes, ya que las interdependecias entre las distintas metas y recursos quedan explicitadas.

Una propiedad interesante de este grafo es que no hace falta computarlo completamente antes de comenzar a construir la solución, sino que se lo puede ir completando a medida que se progresa en la búsqueda. Por otra parte, la principal objeción para con esta propuesta es que la coordinación es analizada sólo en base a los comportamientos que permiten alcanzar los objetivos deseados, sin hacer referencia a una teoría de interacción en concreto. Por caso, Jennings señala que al estudiar el sistema desde una perspectiva puramente basada en el comportamiento no permite distinguir si los agentes coordinaron sus acciones o no.

Para sustentar esta crítica apela un elocuente ejemplo donde un grupo de personas están sentada en una cierta plaza, y que al desencadenarse una lluvia torrencial corren a buscar refugio bajo un gran árbol, el único disponible en esa plaza. En este escenario, aún cuando las personas aparenten actuar de forma coordinada al llevar adelante un conjunto análago de acciones, en realidad cada una en forma independiente está intentando satisfacer su propio objetivo de no empaparse con la lluvia. En contraste, el grupo de bailarinas de ballet en el escenario de un teatro, todas ejecutando un complejo conjunto de acciones, presentan un comportamiento coorinado, respondiendo posiblemente a las exigencias del coreógrafo de la obra.

Compromisos y convenciones

Jennings síntetiza su argumento a favor del cambio de énfasis en el estudio de la coordinación entre agentes, favoreciendo un análisis más abarcativo de este fenómeno, en su propuesta [Jennings, 1996]. Pragmáticamente, constituye un refinamiento del modelo anterior basado en un grafo AND-OR, incorporando dos aspectos que el autor considera cruciales del proceso de coordinación: los *compromisos* y las *convenciones*. La hipótesis de trabajo es que *todos los mecanismos de coordinación pueden reducirse a un mero conjunto de compromisos junto con sus convenciones asociadas.*

Los compromisos se interpretan como promesas acerca de acciones o creencias. Por caso, un agente se puede comprometer comprar un cierto producto en una determinado fecha. La promesa de realizar esta actividad desencadena el compromiso asociado con respecto a los recursos necesarios para llevar a cabo esa acción. En el ejemplo anterior, el agente debe asegurarse de contar con el capital necesario para cumplir lo acordado en la fecha indicada. Se asume que cuando un agente se compromete a realizar una acción en particular, entonces, si las circunstancias no cambian, debe honrar su palabra. En consecuencia, las decisiones del agente con respecto a realizar nuevas acciones se verán limitadas, dado que los recursos—usualmente acotados—deben apartarse para poder cumplir con los compromisos anteriores.

La existencia de compromisos brinda un grado de predecibilidad al sistema, puesto que los agentes podrán tener en consideración las actividades futuras de sus pares para resolver los potenciales conflictos de utilización de recursos. Si bien cuanto mayor sea el nivel de responsabilidad con el que los agente atienden sus compromisos, tanto más predecible se torna el comportamiento del sistema, también es conveniente que los compromisos permitan un cierto grado de libertad, como para asegurar que los agente dispongan de la flexibilidad requerida para poder alcanzar sus metas incluso en un ambiente cambiante. La clave está en encontrar un adecuado balance entre estos conceptos contrapuestos.

En este sentido, las convenciones proveen un mecanismo para gestionar compromisos ante circunstancias cambiantes. Como los agentes suelen estar situados en ambientes cambiantes, es factible que estos cambios en el ambiente provoquen actualizaciones en las creencias y metas del agente incluso antes de que un compromiso anteiormente acordado sea cumplido. Es posible que el cambio en ambiente haya invalidado las circunstancias que rodeaban la toma del compromiso, tornando ahora imposible su cumplimiento. Volviendo al ejemplo anterior, el agente pudo haber tenido la desgracia de que todo su capital quedara atrapado en el "corralito", por lo que ahora le es imposible comprar el producto que había acordado. En consecuencia, los agentes deben evaluar periódicamente si los compromisos preexistentes continúan siendo válidos, apelando al conjunto de convenciones en curso, que son las que restringen las condiciones bajo las cuales un compromiso puede ser revalidado,

Revisión de Compromisos:

Si

el compromiso está satifsfecho, o bien

las motivaciones que originaron el compromiso ya no existen, o bien

el compromiso es ahora inalcanzable,

entonces

dejar de lado ese compromiso.

Figura 3.7: Ejemplo de una convención asociada a la revisión de compromisos

rectificado o descartado. La figura 3.7 ilustra un sencillo ejemplo de este tipo de convenciones, adaptado de [Jennings, 1996].

Las convenciones asisten a los agentes en las gestión de sus compromisos, si bien no especifican como comunicar a otros agentes cuando un compromiso es alterado o descartado. Para el caso de metas interdependientes, donde toda modificación que afecte a una meta anterior influirá en las metas posteriores que de ésta dependan, Jennings introduce el concepto *convenciones sociales*, que especifican cómo comportarse ante los distintos integrantes de la comunidad cuando se altere un compromiso. Por citar un ejemplo, la convención social puede indicar que el agente que rompa un compromiso estará obligado a informar a los restantes agentes involucrados las razones de tal ruptura.

Para concluir podemos resaltar que la propuesta de este autor tiene como principal virtud atacar el problema de la coordinación desde una persepctiva más integral, teniendo en consideración no sólo el comportamiento de los agentes sino también a su estado interno, creencias, intenciones, etc. La hipótesis en torno a la cual se desarrolla el modelo propuesto es que todos los tipos de coordinación puede ser caracterizados en términos de un conjunto de compromisos y de convenciones, respaldando esta afirmación al identificar cuáles son los compromisos y convenciones involucrados en las diferentes técnicas tradicionales de coordinación en los sistemas distribuidos.

3.6.3. Mecanismos de cooperación

La *cooperación* entre agentes es esencialmente una forma refinada de *coordinación*, que tiene lugar entre individuos con objetivos en común. Naturalmente, el desarrollar sistemas eficientes en los cuales sus agentes puedan cooperar en pos de una cierta meta es uno de los objetivos centrales de la investigación en SMA. Moulin y Chaib-draa sintetizan de forma acabada en [Moulin y Chaib-draa, 1996] qué podemos esperar al dotar a un sistema de un adecuado mecanismo de cooperación:

- incrementar la cantidad de tareas resueltas (a través del incremento del paralelismo de tareas),
- incrementar el conjunto de tareas que se pueden realizar (al combinar capacidades individuales, sumar experiencias, etc.),
- incrementar la probabilidad de que una tarea se lleve adelante (asignando la misma tarea a más de un agente), y además
- decrementar la interferencia entre las distintas tareas, evitando las interacciones problemáticas.

Cuando un conjunto de agentes con la intención de cooperar entre sí coordinan sus comportamientos, se torna esencial preservar la coherencia del sistema, cuidando de no avasallar la autonomía de sus agentes. A tal efecto, durante la etapa de diseño se debe hacer incapié en aquellas metas comunes a varios agentes, identificando cuáles son las tareas que deben ser llevadas a cabo en forma grupal. Toda vez que el dominio de aplicación así lo permita, los agentes deben maximizar la información compartida con sus pares, evitando de esta forma conflictos no relacionados con cuestiones de fondo.

En general, la estrategía básica aplicada por la mayoría de los protocolos de cooperación consiste en descomponer y luego distribuir estas subtareas entre los distintos agentes. Este esquema de *divide et impera* permite reducir la complejidad de las tarea al obtener subtareas más sencillas que posiblemente puedan ser llevadas adelante por agentes con capacidades más modestas, insumiendo posiblemente una menor cantidad de recursos. A su vez, en virtud de que la cooperación es en esencia un caso especial de la coordinación, las técnicas genéricas consideradas en la sección anterior pueden también ser aplicadas para modelar este tipo de interacción. En lo que resta repasaremos las propuestas más relevantes introducidas en la literatura para modelar la cooperación en el ámbito de los SMA.

Sistemas de pizarrón

Los *sistemas de pizarrón* fueron desarrollados en la década del '70 específicamente para complejos problemas de interpretación de señales. El objetivo inicial consistió en integrar módulos de software *cooperativos* para alcanzar la flexibilidad que posee un grupo de expertos humanos trabajando en un problema al aplicar la técnica de "tormenta de ideas"[8] El aspecto más interesante de esta técnica es que permite al grupo de expertos resolver problemas complejos, que posiblemente ninguno de ellos hubiera podido resolver por cuenta propia.

Daniel Corkill motiva a los sistemas de pizarrón apelando a la siguiente analogía, tomada de [Corkill, 1991]:

> *"Imaginemos un grupo de especialistas, ya sean humanos o agentes artificiales, ubicados junto a un gran pizarrón. Los especialistas trabajan en forma cooperativa para resolver un cierto problema, empleando el pizarrón como espacio de trabajo en el cual van desarrollando la solución. El proceso de resolución comienza cuando el problema y los datos iniciales se presentan en el pizarrón, momento en el cual los especialistas observan el problema, buscando la oportunidad de aportar su conocimiento*

[8] usualmente referida por su denominación en inglés, *brainstorming*.

> *y pericia. Cuando un especialista encuentra suficiente información como para realizar una contribución, la escribe directamente sobre el pizarrón. A su vez, esta nueva información puede permitir que otro de los expertos aplique su conocimiento. Luego, este proceso continúa hasta que el problema haya sido completamente resuelto."*

A continuación repasaremos los aspectos básicos este popular modelo, siguiendo los lineamientos de su formulación original [Corkill, 1991].

Los sistemas de pizarrón fueron diseñados para atacar problemas relativamente difíciles, que presenten una estructuración poco evidente. La comunicación entre los distintos componentes del sistema siempre se realiza de forma indirecta, a través de un pizarrón. El pizarrón puede abstraerse simplemente como un repositorio de datos, cuyo contenido es público y accesible a todos. De esta forma, no es necesario que exista una jerarquía de control preestablecida entre los agentes del sistema. Un sistema de pizarrón puede descomponerse en tres componentes esenciales:

- **Las fuentes de conocimiento:** son módulos independientes que contienen el conocimiento necesario para resolver el problema. Pueden poseer distintas representaciones, así como contar con mecanismos de inferencia de diversa índole.

- **El pizarrón:** es una base de datos global que puede albergar información inicial, soluciones parciales, o cualquier otro dato que sirva de ayuda en los distintos estadíos de la resolución del problema.

- **El componente de control:** es quien toma las decisiones sobre el rumbo que sigue el proceso de resolución del problema, administrando los recursos disponibles.

Las fuentes de conocimiento (FC) deben ser independientes entre sí, esto es, deben poder aplicar sus capacidades a la resolución del problema sin depender de que existan o participen las restantes fuentes de conociemiento. Sólo se asume que cada FC puede tener acceso al estado actual de la resolución en curso, y puede interpretar toda información relevante codificada en el pizarrón. A su vez, cada FC debe ser capaz reconocer en qué circunstancias puede contribuir a la solución global del problema. Las condiciones que caracterizan a cada una de esas circunstancias en las que la FC puede aportar su experiencia se denominan *condiciones de activación*.

Los agentes activos del sistema de pizarrón no son las FC, sino sus *activaciones*, también denominadas *instancias*, quienes competirán por los recursos disponibles. Estas activaciones son una combinación del conocimiento específico de la FC junto con una condición de activación en particular. La distinción estre FCy activación resulta especialmente relevante en las aplicaciones donde numeroso eventos puedan activar a la misma fuente de conocimiento. En estos casos, las decisiones de control involucran optar entre distintas instanciaciones de la misma FC, donde se debe seleccionar cuál es el contexto más adecuado, en vez de tener que elegir entre FC alternativas, donde la decisión resultaría más demandante, teniendo que elegir entre distintos tipos de conocimiento. En resumen, las FC hacen las veces de repositorios estáticos de información, mientras que las activaciones representan los procesos dinámicos que accesan la información contenida en los repositorios.

El segundo elemento de este tipo de sistemas, el pizarrón, es una estructura global disponible a todas las fuentes de conocimiento, cuyo propósito es:

- constituir una memoria comunitaria para los datos de entrada del problema, las soluciones parciales y la información de control,
- servir como medio de comunicación, y finalmente
- permitir la activación de las fuentes de conocimiento.

Por lo general, las aplicaciones basadas en sistemas de pizarrón cuentan con elaboradas estructuras para representar a este componente, involucrando incluso múltiples niveles de abstracción.

Por último, el componente de control del sistema monitorea el proceso de resolución del problema en forma explícita, permitiendo que las FC respondan de forma oportuna a los cambios que se produzcan en el pizarrón. El razonamiento en en estos sistemas sigue un esquema incremental, construyéndose la solución paso a paso. En todo momento, el componente de control opta entre:

- permitir que la ejecuón de alguna de las FC que hayan sido activadas avance, o bien
- elegir un nuevo foco de atención basándose en las necesidades corrientes de la solución en construcción.

Bajo el esquema de control típico, el avance en la ejecución de la instanciación que esté activa en ese momento va generando eventos a medida que modifica el contenido del pizarrón. Estos eventos se almacenan hasta que termina de ejecutar esta activación, momento en el cual el componente de control elige la próxima que resulte más apropiada para ser ejecutada a continuación. Este ciclo usualmente continúa hasta que se completa la solución buscada. La figura 3.8 esquematiza los principales aspectos de la interacción entre los componentes de un sistema de pizarrón.

Esta introducción a los principales componentes de los sistemas de pizarrón tiene por objeto capturar la esencia de este tipo de sistema. Este tipo de arquitectura resulta particularmente adecuada para integrar sistemas heterogeneos de extracción dispar, ya que los sistemas de pizarrón permiten que las distintas fuentes de conocimiento pueden conservar la representación de información que le resulta más adecuada.

Por otra parte, la comunidad de investigadores en el ́1area de los sistemas multiagente está actualmente revisitando este paradigma, reinterpretando su formulación inicial, concebida como un modelo para la resolución distribuida de problemas, desde una perpespectiva centralizada en agentes [Craig, 1998, Corkill, 2003]. La idea es flexibilizar la metáfora original, otorgando una mayor autonomía a los expertos e incrementando la sofisticación del pizarrón. Este nuevo modelo ha sido aplicado en distintos contextos, abarcando dominios de aplicación que incluyen procesos de control industrial, planificación, diagnóstico e incluso razonamiento basado en casos.

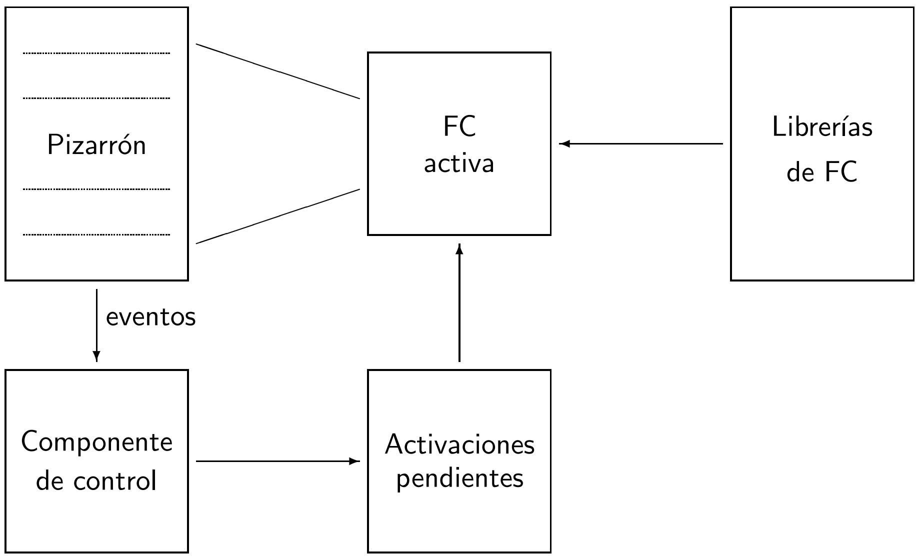


Figura 3.8: Interacción entre los componentes de un sistema de pizarrón

Redes de contratos

Una de las propuestas tal vez más estudiadas para modelar formalmente la interacción en los sistemas multiagente es el modelo de redes de contratos, propuesto por Davis y Smith en [Davis y Smith, 1983]. Este modelo es esencialmente un protocolo de interacción para resolver problemas de forma distribuida. Si bien algunos autores [Moulin y Chaib-draa, 1996] consideran a las redes de contratos como una forma especial de negociación, puesto que los agentes involucrados buscan alcanzar la solución de un problema en común, resulta más natural considerarlo como un mecanismo de cooperación. Seguidamente abordaremos las principales características de este protocolo, siguiendo la formulación presentada en los artículos [Bond y Gasser, 1988, Huhns y Stephens, 1999].

El protocolo de redes de contratos está profundamente influenciados por el modelo de contratación de bienes y servicios empleado en los mercados tradicionales. La arquitectura básica requerida por este protocolo consiste de un conjunto de nodos, donde cada nodo tiene asignado un rol: puede comportarse como *supervisor*, o bien como *contratista*. Esta asignación de roles depende de cada tarea en particular. Por caso, un mismo nodo puede ser en simultáneo supervisor de una cierta tarea y contratista de alguna otra.

Las tareas que cada nodo debe realizar dependen del rol que esté desempeñando. En primer lugar, el supervisor de una cierta tarea debe atender las siguientes responsabilidades:

- Descomponer la tarea encomendada en subtareas. Es posible que la propia descripción de la terea sugiera la descomposición más natural.
- Anunciar las subtareas que se deban realizar. Estos anuncios se pueden comunicar a todos los nodos de la red, o bien sólo al conjunto de especialistas elegidos por el supervisor.
- Evaluar y eventualmente aceptar las ofertas de los potenciales contratistas.
- Consignar el contrato de cada subtarea a cada uno de los postulantes mejores calificados.
- Monitorear el progreso del contrato, solicitando informes periódicos. El supervisor tiene la faculta de reasignar una subtarea si el contrato no se está cumpliendo en forma adecuada.
- Recibir y sintetizar los resultados generados por los distintos contratistas a los que se les delegaron las subtareas.

En contraste, los contratistas debe enfrentar las siguientes responsabilidades:

- Recibir los anuncios de las distintas tareas.
- Evaluar su capacidad individual para realizar la tarea anunciada y tomar un curso de acción acorde, ofertando o bien absteniéndose de participar.
- Llevar adelante la tarea en cuestión, si ésta le resultara encomendada, enviando, de ser necesario, reportes periódicos y notificando por último los resultados obtenidos.

Un aspecto distintivo de este modelo es que el contratista al que se le encomiende una cierta tarea puede a su vez convertirse en supervisor y solicitar otros contratistas que lo asistan con la tarea asignada. Estos vínculos dinámicos entre supervisores y contratistas establecen una jerarquía de control para la subdivisión de tareas y la posterior síntesis de resultados.

La estructura empleada para representar un anuncio incluye campos para denotar a los *agentes contactados*, al *criterio de eligibilidad*, un *resumen de la tarea a realizar*, la *especificación de la oferta*, así como el *tiempo de expiración* de la misma. En primer lugar, los agentes contactados pueden ser uno o varios contratistas en potencia, quienes deben satisfacer el criterio de eligibilidad estipulado. El resumen de la tarea es una breve descripción del objetivo que se desea alcanzar, que los contratantes inspeccionan en todo nuevo anuncio para determinar si se encuentra dentro del alcance de sus capacidades, en cuyo caso formularán una oferta siguiendo los lineamientos que el supervisor indicó en el campo de la especificación de la oferta. Este campo resume qué datos deben suministrar los contratistas que cuenten con las capacidades requeridas al momento de realizar la oferta. Por último, el campo que contiene el tiempo de expiración de la oferta señala por qué intervalo de tiempo el supervisor está dispuesto a aceptar ofertas. Una vez expirado el lapso acordado, el supervisor descartará toda nueva oferta, puesto que se encontrará inmerso en el proceso de evaluar las ofertas ya recibidas.

Cuando el dominio de aplicación así lo requiera, es posible apelar a la *contratación directa*, dirigiendo la propuesta del contrato a un contratista en concreto. Este contrato, como es usual, puede ser aceptado o rechazado por el contratista, si bien en el caso de que acepte, la oferta que formule será necesariamente elegida por el supervisor, por ser la única en consideración. Esta variante de contrato simplifica el protocolo, mejorando el tiempo de respuesta del sistema ante determinadas tareas críticas.

Del análisis de las diversas propuestas introducidas en la literatura se puede concluir que las redes de contratos constituyen uno de los protocolos de distribución de tareas más difundidos e influyentes, posiblemente en virtud de haber sido uno de los pioneros en proporcionar un proceso automatizado de asignación de tareas entre supervisores y contratantes. En contraste, los acercamientos ensayados anteriormente giraban en torno a la asignación realizada por un único supervisor, encargado resolver por su cuenta la distribución de todas las tareas. Otra característica atractiva es que el control no está especificado de forma procedural, sino que es producto de los contratos que dinámicamente van suscribiendo los agentes de forma independiente.

Entre los aspectos cuestionables, el protocolo de redes de contratos no especificaba en su presentación inicial qué esquema seguir en la realización de los anuncios y la adjudicación de tareas, dejando esta importante responsabilidad en manos de cada una de las implementaciones. Otra limitación del protocolo de redes de contratos—señalada inicialmente en [Huhns y Stephens, 1999]—consiste en que si el contratista mejor calificado para una cierta tarea se encuentra circunstancialemente ocupado con otra tarea, es posible que la tarea inicial termine siendo asignada a un contratista con capacidades inferiores.

Por último, se puede objetar que la aplicación de este protocolo no asegura que el sistema presente un comportamiento coherente de un punto de vista global, en función de que el comportamiento del sistema como un todo surge de los distintos criterios de elegibilidad aplicados, naturalmente locales a cada agente. En consecuencia, la responsabilidad para que un sistema que implemente el protocolo de redes de contratos opere de la manera esperada recae un vez más en cada una de las implementaciones en concreto.

3.6.4. Protocolos de negociación

El término "negociación" ha sido empleado para expresar una gran cantidad de conceptos. Por caso, para algunos investigadores tales como Moulin y Chaib-draa [Moulin y Chaib-draa, 1996] o Bond y Gasser [Bond y Gasser, 1988], se define como un mecanismo para asignar tareas o recursos a diferentes agentes, con el objeto último de decidir cómo resolver un determinado problema. En esta concepción, los distintos agentes persiguen un objetivo común que beneficia al sistema en su conjunto.

En contraste, en esta sección consideraremos a la negociación como una refinada forma de coordinación entre agentes que persiguen objetivos posiblemente encontrados. Específifamente, por negociación denotaremos al proceso mediante el cual dos o más agentes alcanzan una desición conjunta en relación a un cierto tópico, pero sin tener que comprometer sus metas individuales al hacerlo. Los agentes comienzan por manifiestar sus tesituras iniciales, las cuales pueden resultar potencialmente en conflicto, para luego intentar arribar a una posición de común acuerdo a través de pequeñas concesiones sucesivas, o bien elaborando propuestas alternativas que minimicen las diferencias entre las partes. Este escenario en el cual la negociación se torna imprescindibel surge naturalmente al aceptar como objetivo de diseño que resulta de interés permitir que agentes heterogéneos, posiblemente diseñados por diferentes autores, puedan interactuar de forma constructiva.

Un primer paso en este sentido lo constituye el contar con protocolos de negociación para dominios específicos, para una vez comprendida la esencia de este proceso formular modelos más abarcativos. Implementando esta política, Zlotkin y Rosenschein [Zlotkin y Rosenschein, 1996] identificaron a dos de los elementos que gobiernan toda negociación: el *protocolo* respetado y la *estrategia* aplicada. El protocolo establece las reglas particulares del dominio de aplicación, especificando tanto las ofertas y contraofertas que se pueden formular como la clase de acuerdos a los que se pueden arribar.

Una vez establecido el protocolo, los agentes deben determinar qué estrategia resulta más apropiada. En cierto sentido, si el protocolo establece los aspectos estáticos de la negociación, la estrategia delinea sus comportamiento dinámico. La estrategia aplicada gobierna la forma en la que el agente se comportará a lo largo de la negociación. Cabe acotar que el tipo de acuerdo que un agente propondrá a lo largo de una negociación depende exclusivamente de la estrategia que esté aplicando, la cual fue codificada previamente por su diseñador.

Diversos grupos de investigación han concebido un gran número de acercamientos alternativos que intentan modelar el proceso de negociación entre agentes. A continuación reseñaremos algunas de las propuestas más significativas presentes en la literatura.

El acercamiento de Zlotkin y Rosenschein

La propuesta de Zlotkin y Rosenschein [Zlotkin y Rosenschein, 1996] modela al proceso de negociación entre agentes apelando al andamiaje provisto por la *teoría de juegos* introducida por John Von Neuman. Como punto de partida, se asume que los agentes son económicamente racionales, en el sentido de que constantemente intetan maximizar su propia—pero a veces arbitraria—función de desempeño.

Para aplicar este modelo de negociación se debe contar con un acotado grupo de agentes, que posean un lenguaje compatible, y que la abstracción del problema que se desea resolver sea común a todos. En adición a estas restricciones iniciales, comunes a todo protocolo basado en la teoría de juegos, los autores señalan un conjunto de suposiciones necesarias para asegurar el correcto funcionamiento de este protocolo:

- Los agentes deben están diseñados para maximizar la utilidad esperada.
- Cada negociación es independiente: un agente no se puede comprometer como parte de una negociación a un determinado comportamiento en otra negociación futura.
- Debe ser posible contrastar y comparar las utilidades de los diversos agentes.
- Todos los agentes deben ser capaces de realizar el mismo conjunto de acciones sobre el entorno.
- Los agentes deben expresar en forma explícita sus compromisos públicos.
- No es posible transferir utilidades en forma explícita entre los agentes.

Finalmente, este modelo de negociación fue concebido para ser aplicado exclusivamente en *dominios orientados a estados*, también conocidos como SOD.[9] En un SOD la descripción del entorne se encuentra capturada a través de un estado global, junto con un conjunto de operadores que causan que el entorno pase de un estado global a otro. En este marco, el propósito de los agentes es lograr que el estado actual del mundo satisfaga sus metas. La negociación se utilizar para que los agentes involucrados formulen un *acuerdo*, el que representa un plan de acción común a todos ellos, de forma tal que la aplicación de este plan transforme al entorno de su estado inicial a un estado final en el cual se satisfagan todos los objetivos que en primer lugar motivaron la negociación.

Los agentes al ejecutar este plan común adoptarán diversos roles. Una de las restricciones para que este modelo sea aplicables es que se debe poder determinar para cada agente cuál es el costo asociado a cada rol. En consecuencia, los agentes de entre todos los planes que le permitan alcanzar sus metas, preferirán aquellos que le asignen un rol con menor costo. Formalmente, cada posible acuerdo brinda una cierta *utilidad* a cada uno de los agentes involucrados, la cual se define como la diferencia entre el costo de alcanzar su objetivo por cuenta propia con el costo de desempeñar el rol que el plan común le asigne. Ya que cada agente desea maximizar su utilidad (este es otro de los requisitos para poder aplicar este modelo de negociación), los agentes al discutir exploran un *espacio de negociación* que abarca a todos los acuerdos que posean una utilidad positiva para cada uno de los agentes participantes.

El protocolo de negociación es quien permite que los agentes elijan que plan conjunto desean llevar adelante. No obstante, existen casos en los cuales no es factible encontrar este plan. Los autores identifican las siguientes situaciones como problemáticas:

- Puede no existir un estado que satisfaga las metas de todos los agentes (dos o más metas son contradictorias entre sí).

- Aún si el estado que satisface las metas de todos los agentes exista, quizás ese

[9] de su sigla en inglés, *state oriented domains*.

estado no pueda ser alcanzado en base a las operaciones disponibles en ese SOD.

- Finalmente, aún cuando el estado que satisface las metas de todos los agentes exista y resulte alcanzable en base a las operaciones dispnibles, quizás resulte demasiado costoso arribar al mismo y los agentes pueden no estar dispuestos a insumir esa cantidad de recursos.

Finalmente, si bien el enfoque adoptado por estos autores permitió comenzar a explorar los conceptos fundacionales presentes en toda negociación (por caso, la importancia del protocolo implementado o el rol de las estrategias), el tipo de dominio para el cual fue concebido resulta en la práctica un tanto restrictivo, puesto que todos los agentes deben contar con un conocimiento perfecto acerca de las metas y de los beneficios que los restantes agentes obtienen de éstas. Actualmente se está explorando la posibilidad de dotar a los agentes de la capacidad de que vayan aprendiendo las preferencias de los restantes agentes al observar su comportamiento, sin depender de que esa información esté disponible de forma previa.

La argumentación como medio para alcanzar consenso

El modelo de negociación desarrollado en forma conjunta por Sarit Kraus y Katia Sycara [Kraus et al., 1998], adopta una teoría argumentativa como medio a través del cual los agentes involucrados intercambian propuestas y contrapropuestas, con el objeto último de alcanzar entre todos una solución unánimemente aceptada. En este contexto, los argumentos son manifestaciones cuyo propósito es modificar las expectativas del interlocutor de turno, influenciando en consecuencia sus acciones.

Este modelo nació en base a Persuader, un sistema de negociación diseñado para gestionar contratos colectivos de trabajo entre empleados y empleadores. Por esta razón, Persuader sólo modelaba tres agentes: la *compañía*, el *sindicato* y el *mediador*. Uno de los objetivos de diseño del nuevo formalismo es brindar una solución más general, que no limite el número de agentes que toman parte de la negociación.

En principio, los autores asumen que los agentes que intervienen en la negociación cuentan con las siguientes capacidades:

- pueden representar y mantener un modelo de sus propias creencias, deseos, metas e itenciones (apelando a un modelo basado en un conjunto de operadores modales introducidos a tal efecto para las distintas actitudes mentales);
- son capaces de razonar en base a las creencias, deseos, metas e intenciones de otros agentes;
- tienen como influenciar el comportamiento de los otros agentes.

En este escenario, un agente usa su modelo del agente a quien desea persuadir para generar argumentos que resulten efectivos en esa situación concreta y para ese tipo en particular de agente. A lo largo de los distintos encuentros con sus pares, cada agente puede analizar los patrones de comportamiento y en base a éstos generar distintas *reputaciones* para cada agente (otorgándoles, por ejemplo, distintos grados de credibilidad). La reputación de un agente puede influenciar la evaluación de los argumentos que construya. Por caso, si un cierto agente recibe una propuesta interesante de parte de otro agente cuya reputación le resulta cuestionable (por ejemplo, pues ha faltado a la verdad en reiteradas ocasiones), es posible que de todas formas desestime esa propuesta.

Este modelo de negociación introduce una teoría lógica formal para asegurar que la generación y el intercambio de argumentos respete los objetivos de diseño del formalismo. Ya que esta teoría cuenta con diferentes modalidades (creencias, intenciones, deseos, etc.), su semántica asocia sendas relaciones de accesibilidad para cada una de estas modalidades, induciendo en definitiva una semántica de mundos posibles.

Los autores consideran una taxonomía de las diferentes formas en que un agente puede conducir una negociación con el objeto de intentar persuadir a otro agente. Se identifican seis formas de conducir la negociación, cuya viabilidad ya ha sido comprobada en las negociaciones entre seres humanos.

- Apelando a *amenazas*, con el propósito de lograr que el agente que está siendo persuadido adopte o abandone una cierta meta.
- Haciendo promesas de una *recompensa futura*.
- Trayendo a colación a un *beneficio otorgado en el pasado*.
- *Resaltando precedentes* que sugieran que existe una contradicción entre el accionar pasado del agente y sus intenciones actuales.
- Citando *experiencias pasadas de otros agentes*, para argumentar a favor o en contra de un determinado curso de acción.
- Empleando *el interés del agente por sí mismo*, mostrando como realizar una determinada acción satisfará una cierta meta de gran importancia para él.

A partir de esta taxonomía, los autores formulan un conjunto de lineamientos para guiar la argumentación, señalando la estrategia que resulta más adecuada cuando se desea aplicar cada una de estas formas de conducir la negociación.

Finalmente, la negociación entre los agentes propiamente dicha se lleva adelante mediante el intercambio de mensajes, que pueden denotar pedidos, respuestas (aceptaciones o rechazos), o sencillas declaraciones. En ocasiones, los mensajes intercambiados contienen argumentos como justificaciones de los mismos, los que deben ser elaborados siguiendo los lineamientos anteriormente citados. El rol de estos argumentos en el modelo de negociación es crucial, ya que la mera recepción de un mensaje no constituye una razón suficiente para alterar el estado mental de ese agente. En general, los agentes evaluarán los mensajes recibidos, decidiendo de qué manera corresponde modificar su estado mental, analizando y considerando de ser el caso la justificación en forma de argumento incluída con el mensaje.

Si bien este formalismo constituye un atractivo banco de prueba para ensayar las diversas modalidades bajo las que se pueden encarar una negociación, su formalización deja algunos aspectos sin elaborar (tales como qué modelo de argumentación emplear, cómo llevar adelante la toma de decisiones o la planificación en cada unos de los agentes, etc.), por lo que no resulta trivial adecuar el modelo propuesto a un dominio de aplicación en concreto. No obstante, se debe reconocer que constituye una de las primeras propuestas en explorar un enfoque lógico formal para llevar adelante la negociación teniendo en consideración los estados mentales de los agente participantes.

La propuesta de Parsons, Sierra y Jennings

El primer formalismo en caracterizar de forma acabada un modelo de negociación elaborado en torno a una teoría argumentativa fue introducida en el inspirador artículo [Parsons et al., 1998]. Este trabajo delinea un marco de negociación en el cual un conjunto de agentes autónomos intentan alcanzar sus respectivas metas, contando con la posibilidad de influenciar el comportamiento de los restantes agentes mediante oportunas negociaciones. En este modelo, los agentes son caracterizados a través de sistemas multicontexto, un formalismo que como particularidad permite capturar las distintas modalidades que componen el estado mental de un agente sin tener que apelar a una lógica modal. Por caso, uno de los ejemplos desarrollados en este artículo reseña cómo modelar agentes del tipo BDI mediante sistemas multicontexto.

En este marco, el término *negociación* se interpreta como el diálogo entre dos o más agentes que les permite alcanzar un acuerdo común acerca del tópico en debate, a través del contínuo intercambio de *propuestas*, *críticas*, *contrapropuestas*, *explicaciones* y *meta-información*. En primer lugar, una propuesta es alguna clase de solución al problema que el agente enfrenta, como por ejemplo el requerimiento de algún recurso o servicio que el agente necesite. El agente que reciba una propuesta cuenta con dos posibles cursos de acción: puede elaborar una crítica, o bien puede concebir una contrapropuesta. Si opta por formular una crítica, debe a su vez indicar si la propuesta inicial fue aceptada o rechazada. La crítica puede también incluir una aclaración sobre qué parte de la propuesta inicial le resultó atractiva o controversial. La otra alternativa para responder a una propuesta consiste formular una contrapropuesta que extienda o modifique a la propuesta original.

Tanto propuestas, contrapropuestas como críticas son básicamente expresiones de los deseos del agente que las fomula. En este sentido, los autores sugieren que el incorporar una explicación al manifestar lo que se desea puede colaborar en la formación de consenso. Por esta razón, se contempla el concepto de explicación, que consiste en la justificación que los agentes pueden incluir para respaldar sus expresiones. Es en este inusual contexto que surge la noción de *argumento* como el medio ideal para conjugar una cierta propuesta con su respectiva fundamentación.

Un último aspecto considerado en este modelo es el intercambio de meta-información a lo largo de una negociación. La meta-información constituye un medio a través del cual focalizar la búsqueda local de cada agente, facilitando información adicional no necesariamente relacionada con el tópico en consideración. Por caso, el proveer a un agente con información adicional de por qué fue rechazada una anterior propuesta puede facilitar la elaboración de una nueva propuesta con más posibilidades de que resulte aceptada. Si bien las explicaciones permiten en cierto sentido el intercambio de meta-información, la noción de meta-información que incorporan los autores resulta más abarcativa que el concepto de explicación, puesto que las explicaciones sólo tienen sentido en el marco de una propuesta, crítica o contrapropuesta, mientras que la meta-información puede ser intercambiada en cualquier circunstancia. No queda claro qué dominio de aplicación tenían en mente los autores al incorporar este concepto, pero el suministrar información adicional no relevante a la discusión en curso no necesariamente facilita encontrar el punto de equilibrio, puede incluso llegar a entorpecer el normal desarrollo de la negociación o ser explotado como un mecanismo de distracción por algún agente con pocos escrúpulos.

En sítesis, el proceso negociación comienza con un agente formulando una propuesta inicial. Esta propuesta posiblemente abaque aquellos recursos o servicios que el agente requiera para alcanzar sus objetivos. En base a esta propuesta, los agentes involucrados pueden elaborar críticas o contrapropuestas a la misma, o bien suministrar meta-información. Acto seguido, el primer agente puede realizar una nueva propuesta, responder a través de meta-información que clarifique la propuesta original, o considerar alguna de las contrapropuestas recibidas, esto es, introducir un nuevo nivel de críticas o contrapropuestas a la negiciación. Este proceso continúa hasta que todos los agentes involucrados alcazan un estado de equilibrio (por caso, todos se encuentra conformes con la última propuesta esbozada), o bien se decide que no es posible alcanzar este estado de equilibrio, en cuyo caso la negociación es abandonada.

El aspecto posiblemente más llamativo de este acercamiento es que el grupo de agentes apela a un formalismo de razonamiento argumentativo tanto para formar sus propias opiniones como para llevar adelante las negociaciones. Si bien el sistema argumentativo adoptado por los autores [Krause et al., 1995] resulta un tanto pobre en prestaciones (por caso, el análisis dialéctico del mismo no permite la restauración de argumentos), el diseño altamente modular de esta propuesta permite subsanar este inconveniente con suma facilidad, pudiéndose adaptar prácticamente cualquier otra teoría de argumentación.

3.7. Conclusiones

La diversidad de dominios de aplicación en los cuales se ha ensayado aplicar el enfoque de agentes múltiples excede a la imaginación. A lo largo del presente capítulo se han sintetizado los aspectos comunes y las diferencias más significativas presentes en las principales propuestas introducidas en la literatura. Si bien es casi imposible encontrar un conjunto de características recurrentes a la totalidad de las propuestas, usualmente los diversos sistemas multiagente suelen presentar como aspectos en común que sus agentes cuentan con conocimiento imperfecto y capacidades limitadas, donde a su vez el control del sistema suele estar distribuido y la información disponible no es centralizable. Por otra parte, otro aspecto recurrente es que la computación en el seno del sistema suele ser realizada de forma asíncrona.

Del análisis de las diversas propuestas consideradas en la sección 3.3, se puede concluir que la mayoría de los sistemas parecen focalizar su atención en torno a aspectos un tanto *ad-hoc*. Por ejemplo, en ciertos casos la estructura interna de los agentes desempeña el rol principal en el sistema. En contraste, en otras propuestas la atención se centra en torno a la delegación de responsabilidades (por ejemplo, la asignación de tareas a los agentes o el seguimiento de las tareas delegadas). Postulamos que la falta de consenso en este sentido se debe principalmente a la combinación de dos factores, los que intervienen simultáneamente en estos tipos de sistema.

Por un lado, el primer factor que contribuye a esta falta de uniformidad es la heterogeneidad de los dominios de aplicación en los que se han ensayado soluciones basadas en sistemas multiagente. Por otro lado, el segundo factor es la propia dificultad de los desafíos surgidos al migrar de un ambiente en el cual tanto el conocimiento como la toma de decisiones se encuentra centralizada a uno en donde múltiples agentes deben sincronizar su comportamiento para poder alcanzar sus objetivos. La conclusión anterior, lejos de ser una mera crítica, constituye un franco reconocimiento al tezón de los pioneros del área, quienes eligieron seguir adelante aún teniendo que afrontar tales contrariedades. A manera de ilustración de este aspecto, la sección 3.4 recopiló los diversos criterios de clasificación ensayados en la literatura, si bien por las razones ya expuestas no es posible señalar a uno de ellos como el más recomendable.

Una de las pocas excepciones a esta regla lo constituye la manifiesta preponderancia que la interacción entre los agentes desempeña en todo sistema multiagente, incluso independientemente de la citada diversidad. Sin duda un sistema multiagente en el cual la interacción entre sus miembros sea muy baja o nula será más bien una mera colección de sistemas centralizados. De hecho, los diversos paradigmas de comunicación entre agentes reseñados en la sección 3.6 no hacen más que evidenciar el incipiente interés en la exploración de los principios y nociones involucrados en la interacción, posiblemente en virtud de que hoy en día nadie está dispuesto a resignar los potenciales beneficios que se pueden obtener al aplicar esta concepción.

Capítulo 4

Un Modelo Dialéctico para la Deliberación Multiagente

A lo largo del capítulo anterior se enfatizó uno de los aspectos más significativos de todo sistema multiagente: *la interacción entre sus agentes*. Sin duda, el contar con modelos adecuados, que permitan no solo describir el comportamiento pasado de un sistema sino también simular su accionar a futuro, constituye una de las vertientes predominantes de esta disciplina. En este sentido, disponer de un modelo formal que provea el marco de referencia para la interacción entre los diversos agentes permitiría estandarizar su tratamiento, posibilitando entre otras cosas adecuar las soluciones ya existentes a nuevas situaciones. Este capítulo presenta el aporte principal de esta tesis, que consiste en la caracterización formal de una de las formas de interacción más refinadas, conocida como *deliberación*.

4.1. Introducción

En base al análisis y la discusión presentados en el capítulo anterior, es posible afirmar que la interacción entre los agentes constituye la característica primordial de este tipo de sistemas. Más aún, es esta interacción la principal responsable del incremento de potencial percibido al migrar de un entorno de agente único a uno en donde existan múltiples agentes. En consecuencia, resulta razonable suponer que el desarrollo de un modelo que describa de forma precisa todas y cada una de las facetas que presenta esta problemática debe constituir una compleja empresa. Al igual que en todo proyecto de gran envergadura, el contar con una estrategia adecuada simplifica en gran medida la tarea por delante, permitiendo aislar e identificar los distintos aspectos que componen el problema a ser resuelto. A continuación detallaremos la estrategia adoptada para llevar a cabo el principal aporte de esta tesis.

Expresándolo en forma sintética, nuestra intención es aplicar en un sentido práctico la equivalencia en forma y estructura identificada por Rescher entre los razonamientos de base dialéctica unilateral y multilateral, uno de los puntos expuestos en [Rescher, 1977], oportunamente considerado en la sección 2.1.2. La idea central es capitalizar el amplio volumen de investigación conducido en el ámbito del razonamiento dialéctico unilateral, para luego, en virtud de este isomorfismo, extrapolar los resultados allí obtenidos al contexto del razonamiento dialéctico multilateral. Considerando que nuestro objetivo es obtener una teoría para modelar la deliberación entre agentes bajo los más diversos escenarios, la teoría deberá poder ser instanciada de acuerdo a las necesidades de cada caso, esto es, las decisiones de diseño tomadas al elaborar el modelo no deben comprometer sus potenciales dominios de aplicación. Por consiguiente, en virtud de la estrategia a seguir debemos partir de un modelo para el razonamiento dialéctico que sea de carácter abstracto. La citada estrategia se puede sintetizar en las siguientes etapas:

1. Formular una teoría abstracta que capture la esencia del razonamiento dialéctico unilateral.

2. Reinterpretar la teoría abstracta obtenida en la etapa anterior como si fuera el marco que gobierna la interacción multilateral de un grupo de agentes.

Lo que resta de este capítulo será destinado a desarrollar estas etapas. En primer lugar la sección 4.2 repasa otros acercamientos a esta problemática que también han seguido la estrategia propuesta. La sección 4.3 cubre la primera mitad de esta estrategia, hasta alcanzar el modelo abstracto para el razonamiento de carácter dialéctico. La otra mitad de la estrategia, el reinterpretar este modelo como si fuera un modelo para la deliberación entre agentes, se desarrolla a lo largo de la sección 4.4. Posteriormente, la sección 4.5 presenta algunas instanciaciones concretas del modelo obtenido, con el objeto de ejemplificar una de sus posibles aplicaciones. Finalmente, la sección 4.6 concluye el presente capítulo sintetizando las principales conclusiones formadas como producto de este estudio.

4.2. Primeros acercamientos

En esta sección desarrollaremos un breve repaso de otros acercamientos que aplicaron una estrategia similar a la adoptada en la presente tesis. En primer lugar, abordaremos en la sección 4.2.1 el modelo propuesto por Ron Loui en [Loui, 1998], hoy en día la referencia obligada de todo trabajo que elabore a lo largo de esta vertiente de investigación. Posteriormente, en la sección 4.2.2 consideraremos en detalle la propuesta de Hendrik Prakken [Prakken, 2001], la cual en cierto sentido instancia de forma concreta al modelo introducido por Loui, lo que permite un análisis exhaustivo de los principales aspectos que caracterizan tanto al razonamiento de base dialéctica como a la interacción entre agentes.

La presente tesis, al transitar un camino similar al recorrido por estos autores, apunta tanto a capitalizar sus principales virtudes como a subsanar, de ser posible, sus aspectos controversiales.

4.2.1. La propuesta de Ron Loui

El trabajo titulado *"Process and Policy: Resource-Bounded Nondemonstrative Reasoning"* [Loui, 1998] describe de forma acabada la perspectiva que su autor, a partir de su vasta experiencia en el campo del razonamiento rebatible, ha adoptado en relación a las formas no demostrativas de razonamiento. Loui señala que los inconvenientes encontrados en este área son producto, generalmente, de haber tomado por sentado al concepto de *default*. Por lo general, la forma en la que se modelará al concepto de default suele estar establecida *a priori* por el diseñador del sistema (por ejemplo, a través de reglas rebatibles o por defecto). El autor enfatiza que la elección de una ontología incorrecta puede causar que el sistema resultante manifieste una semántica plagada de singularidades. Si bien resulta llamativo, la comunidad académica parece no haber reconocido el mérito de este argumento, ni la elegancia de la solución propuesta. En este sentido, Loui sostiene que un default en esencia no es más que un plan o una estrategia para alcanzar una cierta conclusión, la cual posiblemente está más allá del alcance de los métodos clásicos. Bajo la óptica del autor, las reglas rebatibles se asemejan más a una *política* que describe como obtener nuevas conclusiones a partir de la información disponible que a una versión debilitada de la implicación material.

Otro de los aspectos considerados en este artículo, destacado por Loui como apenas explorado en la literatura existente, es el consumo de recursos en que incurre el propio proceso de razonamiento. De hecho, existen diversas propuestas que sólo manifiestan un comportamiento racional cuando tienen acceso a maquinarias de inferencia de una complejidad exorbitante (por ejemplo, teorías de segundo orden, lógicas modales no proposicionales, etc). Si bien estos acercamientos constituyen desafíos intelectuales válidos, donde incluso es posible alcanzar resultados de interés, resulta poco probable que los mismos puedan llegar a ser aplicados a situaciones concretas. El autor desarrolla como caso de estudio un protocolo de base dialéctica que permite la incorporación de nuevas políticas a lo largo del proceso de razonamiento, mostrando cómo este simple protocolo es lo suficientemente sofisticado como para modelar diversas teorías de razonamiento rebatible. Este modelo de disputa, inicialmente introducido para ilustrar los conceptos anteriores, ha resultado más popular que la propia solución propuesta para el tratamiento de los defaults. Considerando que el citado protocolo permite modelar tanto el razonamiento de carácter introspectivo de un agente como la interacción entre dos o más de éstos, dedicaremos el resto de esta sección a introducir sus principales conceptos.

Loui comienza por identificar el siguiente conjunto de características, propias a toda disputa:

Equipos y Contendientes: toda disputa abarca naturalmente múltiples equipos, cada uno compuesto por uno o más contendientes. En general, existen dos equipos, uno proponiendo la tesis a debatir y el otro intentando refutarla, si bien en ocasiones puede existir un tercer equipo compuesto por un contendiente que desempeña el rol de adjudicador o juez de la disputa. Entre los contendientes es posible distinguir los que no cuentan con una postura inicial de aquellos que desde el comienzo ya estén a favor o en contra de la tesis a debatir. Los diversos contendientes de un mismo equipo pueden incluso cooperar entre sí en la formación de justificaciones o contraejemplos.

Protocolo: al repasar las propuestas existentes en relación al estudio de las disputas en el marco del razonamiento no monótono se puede apreciar que los protocolos propuestos suelen resultar un tanto ingenuos. Esto se debe principalmente a que los protocolos tienden a ignorar completamente toda cota en los recursos disponibles. En este sentido, las estrategias de búsqueda resultan importantes precisamente cuando los recursos disponibles son insuficientes como para explorar todas las vertientes del tópico en cuestión, o bien cuando existan una cantidad infinita de éstos.

Todo protocolo presenta dos aspectos primordiales: la *locución* y el *control de la locución*. En el contexto del razonamiento no monótono, la locución suele ser de carácter declarativa. Por caso, las lógicas dialógicas adoptan una clase extensa de actos de habla. Por otra parte, el componente de control de la locución asigna obligaciones a los contendientes, tales como fundamentar una suposición previa cuya validez hubiera sido cuestionada o encontrar nuevas razones que sustenten una cierta tesis cuando las restantes fundamentaciones hayan sido refutadas. Uno de los aspectos más interesantes dentro del control de la locución es la *terminación*, la cual a continuación es considerada en detalle.

Terminación: El concepto de terminación ha sido considerado en profundida en el seno de la problemática de la búsqueda heurística. A diferencia de lo que sucede en esta disciplina, la evaluación de las reglas de terminación en un protocolo no cuentan con la posibilidad de hacer referencia a la respuesta correcta, ya que no existe una solución aceptada a priori por todas las partes. En este contexto, es el proceso mismo quien determina cuál es la aplicación correcta de las políticas existentes, así como la adjudicación más apropiada de los conflictos.

Es posible que en algunas situaciones la disputa termina de forma espontánea, por caso cuando expira el tiempo disponible acordado previamente. Por otra parte, también puede suceder que se llegue a alguna clara condición de terminación, como por ejemplo que alguno de los participantes alcance una posición netamente ganadora. La combinación del protocolo de disputa con el paso del tiempo da a lugar a la noción de *estado*. Esta noción puede ser capturada de diversas formas, desde almacenar tan sólo la opinión general actual, hasta conservar el conjunto de argumentos sin derrotar junto con la clase de respuesta necesaria para motivar un cambio en la opinión general, o incluso llevar un registro detallado de la totalidad de los argumentos formulados a lo largo de la disputa. En síntesis, la adopción de un lenguaje formal que resulte adecuado para describir el estado de la disputa es uno de los principales prerequisitos para establecer un protocolo de disputa.

Estrategia: las diversas teorías propuestas para modelar razonamiento no monótono usualmente minimizan el rol de la estrategia, ya que no se dispone de la información necesaria para evaluar las diferentes tesituras que pueden surgir al considerar una proposición en particular. En contraste, Loui señala que de contar con este tipo de información la estrategia pasaría a desempeñar un papel mucho más trascendental. Sin duda, no resulta sencillo diseñar este tipo de estrategia, pero es posible cosechar beneficios significativos al considerar protocolos acotados en los recursos. Por caso, si los puntos de controversia en un determinado argumento fueran ordenados de acuerdo a su prospecto de éxito, la estrategia dominante consistiría en atacar primero al punto con mayores posibilidades de éxito.

Del análsis de estos aspectos, junto con la exploración de las diversas alternativas posibles para cada una de las decisiones de diseño, el autor deriva su modelo formal para la interacción entre agentes. A continuación repasaremos sus principales componentes.

Definición 4.1. **[Loui, 1998]**
(locución)
Denominaremos *locución* a toda terna $l = \langle p, s, r \rangle$, donde p es un contendiente o conjunto de contendientes, s es una fórmula bien formada en un cierto lenguaje $\mathcal{L}$, o bien es una estructura que verifica ciertas restricciones, como por caso ser un argumento para una formula de $\mathcal{L}$, y cada r es una descripción de recursos, tal como un vector que denote los recursos consumidos. ■

Las locuciones constituyen el comportamiento más elemental reconocido por el protocolo. A su vez, las diversas locuciones se estructuran naturalmente de la siguiente forma:

Definición 4.2. **[Loui, 1998]**
(registro de locución)
Denominaremos *registro de locución* a toda secuencia finita de locuciones de la forma $\Gamma = \langle l_1, \ldots, l_k \rangle$. ■

El concepto de registro de locución da a lugar a un registro de jugadas. En éste, es posible mantener las jugadas consecutivas realizadas por el mismo contendiente.

Definición 4.3. **[Loui, 1998]**
(registro de jugadas)
Sea $\Gamma = \langle l_1, \ldots, l_k \rangle$ un registro de locución. Denominaremos *registro de jugadas* a toda secuencia $M = \langle m_1, \ldots, m_n \rangle$, donde cada m_i, $0 \leq i \leq n$, es la secuencia de locuciones consecutivas en Γ realizadas por alguno de los contendientes. ■

En general, se asume que dos locuciones $l_u = \langle p_u, s_u, r_u \rangle$ y $l_v = \langle p_v, s_v, r_v \rangle$ son adyacentes en algún registro de jugadas M cuando para algún $m_i \in M$, se verifica que $p_u, p_v \in m_i$, o bien cuando para algún par de movidas consecutivas $m_s, m_t \in M$, se verifica que l_u es la última locución de m_s y l_v es la primer locución de m_t. Finalmente, un *registro de disputa* D es una estructura a partir de la cual es posible derivar un registro de locución $\Gamma(D)$ y su respectivo registro de jugadas $M(D)$.

Ejemplo 4.1. Para modelar el protocolo de disputa introducido por Rescher (abordado en la sección 2.1.2), sólo se debe modelar a dos contendientes, el *proponente* de la tesis debatida, y su único *oponente*. Rescher distingue las sentencias de estos contendientes a través de un sencillo lenguaje modal proposicional, que cuenta con el operador "!" para denotar a las afirmaciones categóricas y "†" para las afirmaciones cautas, donde sólo el proponente puede hacer uso de las afirmaciones categóricas y sólo el oponente puede apelar a las afirmaciones cautas. En este contexto, los recursos consumidos no son tenidos en cuenta, ya que Rescher se concentró en modelar los aspectos más evidentes de la disputa. Por ende, el vector de recursos consumidos asociado a toda movida será nulo.
A manera de ejemplo, el registro de locución de una disputa básica que respete el modelo de Rescher puede adoptar la siguiente estructura:

jugada	**locución**	**contendiente**	**sentencia**	**recurso**
m_1	l_1	*proponente*	P	$\langle\rangle$
m_2	l_2	*oponente*	$\sim P/Q$	$\langle\rangle$
m_2	l_3	*oponente*	Q	$\langle\rangle$
m_3	l_4	*proponente*	$P/(Q \wedge R)$	$\langle\rangle$
m_3	l_5	*proponente*	$(Q \wedge R)$	$\langle\rangle$

Finalmente, Loui caracteriza el protocolo de interración entre contendientes en base a las distintas nociones hasta aquí introducidas.

Definición 4.4. **[Loui, 1998]**
(protocolo de disputa)
Sea D un registro de disputa, donde r es el último $r_i \in D$. En este contexto, el par $\sigma = \langle D, r \rangle$ denota una de las configuraciones posibles de la disputa. Diremos que un *protocolo de disputa* queda caracterizado por el siguiente conjunto de funciones potencialmente parciales:[1]

1. *current.opinion*(σ), que determina qué contendiente prevalece actualmente en la disputa en curso. Existe la posibilidad de que el debate no favorezca a jugador alguno, en cuyo caso esta función retornará algún valor preestablecido que denote esta situación.

2. *party.to.move*(σ), que determina qué contendiente debe realizar la próxima movida, a menos que ya se haya alcanzado la finalización natural de la disputa.

3. *move.options*(σ), que retorna el conjunto de movidas permitidas por la configuración actual, esto es, devuelve el conjunto de secuencias de locuciones que podrían ser jugadas a continuación.

4. *player.information*(σ), que devuelve una estructura que contiene todas las subsecuencias de $\Gamma(D)$ y $M(D)$ que satisfacen una determinada condición. Por otra parte, puede además contener información sobre la plausibilidad de las razones expuestas, así como toda información adicional de interés en relación los resultados de las búsquedas realizadas.

■

El protocolo introducido por el autor establece los lineamientos a para la disputa entre contendientes. Si bien este conjunto de funciones parecen no restringir la aplicabilidad del modelo de disputa, Loui observa que de todas formas ya se ha comprometido una decisión con respecto al no determinismo. Por caso, para un

[1] resulta natural suponer que estas funciones sólo se encuentran definidas sobre aquellas configuraciones que el mismo protocolo induzca, es decir que se puede tratar de funciones parciales.

escenario donde existan varios contendientes, el orden de participación quedará establecido por la función *party.to.move*. No obstante, el no determinismo puede ser recuperado permitiendo que los contendientes emitan locuciones nulas.

Un aspecto central en el análisis de Loui es impulsar el tratamiento de las eventuales cotas en los recursos disponibles. Lamentablemente, el conjunto de funciones propuesto no toma en consideración estas limitaciones, puesto que alguno de los recursos puede agotarse incluso durante el cómputo de una de éstas. Loui observa que es necesario extender este modelo para tener en cuenta los recursos de forma tal que el protocolo resultante sea sensible a los recursos consumidos por cada contendiente.

La función *move.options* sintetiza tanto las obligaciones como las restricciones de locución. Las obligaciones se cuantifican en forma existencial sobre toda secuencia de locuciones presente en *move.options*, mientras que, en contraste, las restricciones se cuantifican en forma universal. Por ejemplo, un protocolo puede prohibir la introducción de argumentos repetidos requiriendo que dos locuciones $\langle p, s, r_1 \rangle$ y $\langle p, s, r_2 \rangle$ no puedan formar parte del mismo registro de locución. Claro está, esta restricción tiene sentido si se combina con otra restricción que establezca que un argumento no puede aparecer en una locución si el registro ya contiene un derrotador propio del mismo. Cabe acotar que la mayoría de las obligaciones y restricciones de locución son complejas y suelen hacer referencia tanto a la teoría de argumentación en la cual se basa la disputa como al criterio para adjudicar preferencias modelado por *current.opinion*.

El autor señala que las funciones introducidas deben presentar un cierto grado de consistencia entre ellas. Destaca especialmente que para todo estado posible, si *move.options* dispone de más de una movida, entonces la función *party.to.move* debe retornar por lo menos un contendiente. Esto se debe a que todo estado en donde *move.options* sea vacío denota que la terminación natural de la disputa ha sido alcanzada. En esta situación, resulta lógico suponer que aquel contendiente señalado por *party.to.move* es precisamente quien ha perdido. Sin embargo, Loui no caracteriza a la victoria o derrota sólo en términos del resultado de la función *party.to.move* al finalizar la disputa, sino que introduce una función adicional (*current.opinion*) para determinar si la tesis en consideración debe ser aceptada. Aún cuando es concebible que una disputa termine en un empate, siempre es factible asignar el peso de la prueba de forma tal que la existencia del ganador esté asegurada.

Loui propone una clasificación de los distintos protocolos que se pueden describir en base al modelo propuesto, en función de la información con la cual cuenten los contendientes a la hora de realizar sus movidas. Por caso, el protocolo se dice *dialéctico* cuando, en todos los estados posibles, *player.information* equivale al registro de disputa completo. En particular, el protocolo se dice *dialéctico de respuesta inmediata* cuando se trata de un protocolo dialéctico que a su vez, para todo estado posible σ, se verifica que $party.to.move(\sigma)$ no coincide con $current.opinion(\sigma)$.

4.2.2. El modelo de Hendrik Prakken

Luego de la publicación de la propuesta Ron Loui, Hendrik Prakken adopta un curso de acción análogo al proponer un marco de carácter dialéctico que luego será reinterpretado como base de un protocolo de interacción entre agentes [Prakken, 2001]. La característica distintiva de este acercamiento es permitir que la base de conocimiento a partir de la cual cada agente razona pueda ser modificada dinámicamente mientras se produce la interacción.

La estrategia general coincide a grandes rasgos con la adoptada aquí. Esto es, capturar los aspectos principales de las teorías de prueba de carácter dialéctico más difundidas, tales como aquellas consideradas a lo largo del capítulo 2, para luego reinterpretar el modelo así obtenido como si en realidad se tratara del marco conceptual de un protocolo de disputa. A continuación introduciremos los principales conceptos que conforman la propuesta de Prakken, siguiendo el estilo de presentación adoptado por su autor.

En primer lugar, se repasan los componentes comúnes a las teorías de prueba dialéctica, con el propósito de brindar un panorama general, para luego desarrollar su tratamiento formal. Prakken sostiene que una teoría de prueba dialéctica cuenta en mayor o menor medida con los siguientes componentes:

- Un par de *contrincantes*, por un lado el proponente de la tesis en disputa, y por el otro su oponente.
- Un conjunto *Args* de argumentos considerados como válidos, de acuerdo a una determinada noción de consecuencia monótona.
- Una *relación de fuerza relativa* entre argumentos. La fuerza de un argumento es una de las características a partir de las cuales se determinará la legalidad de una cierta movida.
- Un conjunto de *movidas bien formadas*, las que están compuestas esencialmente por un jugador, un argumento o razón, y una referencia a una movida previa la cual está siendo cuestiona.
- Una función *PlayerToMove*, la cual en cada estado del diálogo determina que jugador debe realizar la próxima movida. En particular, en todo diálogo el proponente está obligado a realizar la primer movida.
- Una función *Legal*, que determina el conjunto de movidas que se considerán aceptables en un cierto estado de la disputa. La legalidad de una movida estará determinada por el desarrollo del diálogo hasta ese momento.
- Una noción de *diálogo*, en principio una serie de movidas legales realizadas por los jugadores que respeta el orden inducido por *PlayerToMove*.
- Un criterio de adjudicación para determinar quién *prevalece* en la disputa. Prakken adopta como criterio básico que un jugador vence a su contrincante cuando logra que éste no pueda llevar a cabo nuevas movidas.
- Un criterio para determinar bajo qué condiciones un argumento resulta aceptable. En este contexto, un argumento $\mathcal{A}$ se considerará aceptable si el proponente del mismo cuenta con una estrategia ganadora para el diálogo que comienza con $\mathcal{A}$.

Las intuiciones reflejadas en este bosquejo de carácter informal quedan capturadas de forma más acabada en la noción de marco dialéctico, comentada a continuación.

Marco dialéctico

Según Prakken, todo sistema argumentativo puede ser caracterizado de forma sucinta a través de tres de sus principales componentes: el conjunto de argumentos considerados válidos, la relación de fuerza relativa entre éstos y la teoría de prueba dialéctica adoptada. El autor hace abstracción de las peculiaridades de los diversos sistemas argumentativos introduciendo la siguiente noción:

Definición 4.5. **[Prakken, 2001]**
(marco dialéctico)
Denominaremos *marco dialéctico* a toda tupla $(Args, \preceq, \mathcal{T})$, donde:

- $Args$ es un conjunto de argumentos.
- $\preceq\ \subseteq Args \times Args$ es una relación de fuerza relativa entre argumentos. Por $\mathcal{A} \preceq \mathcal{B}$ se entiende que $\mathcal{B}$ no es más débil que $\mathcal{A}$, y por $\mathcal{A} \prec \mathcal{B}$ que $\mathcal{A} \preceq \mathcal{B}$, pero a su vez $\mathcal{B} \not\preceq A$.
- $\mathcal{T}$ es una teoría de prueba dialéctica para $(Args, \preceq)$.

■

Los dos primeros componentes de todo marco dialéctico no requieren mayor elaboración, pues a lo largo del capítulo 2 hemos presentado diversos interpretaciones para la noción de argumento así como varios criterios de comparación entre éstos. No obstante, el concepto de teoría de prueba dialéctica requiere de un tratamiento pormenorizado.

Definición 4.6. **[Prakken, 2001]**
(teoría de prueba dialéctica)

Sea *Args* un conjunto de argumentos y sea $\preceq$ una relación de fuerza relativa entre éstos. Denominaremos *teoría de prueba dialéctica para* (Args, $\preceq$) a toda tupla

$$(\mathit{Players}, \mathit{Moves}, \mathit{PlayerToMove}, \mathit{Legal}, \mathit{Dialogues}, \mathit{Winner})$$

donde:

- $\mathit{Players} = \{P, O\}$. Cuando sea conveniente, el complemento de un jugador *Player* será denotado como $\overline{\mathit{Player}}$.

- *Moves* es el conjunto de todas las movidas bien formadas. Se define recursivamente como el menor conjunto que satisface las siguientes condiciones:

  - Si $\mathit{Arg} \in \mathit{Args}$ y $\mathit{Player} \in \mathit{Players}$, entonces $(\mathit{Player}, \mathit{Arg}) \in \mathit{Moves}$.
  - Si $\mathit{Arg} \in \mathit{Args}$, $\mathit{Player} \in \mathit{Players}$ y $M_i \in \mathit{Moves}$, entonces la movida $(\mathit{Player}, \mathit{Arg}, M_i) \in \mathit{Moves}$.

  Cada movida puede ser una *movida inicial*, o bien una *movida de respuesta*. Una movida se dice inicial cuando tiene la forma $M_1 = (\mathit{Player}, \mathit{Arg})$, y se dice de respuesta cuando tiene la forma $M_i = (\mathit{Player}, \mathit{Arg}, \mathit{Move})$, $i > 1$. Por otra parte, se denotará al primer componente de una movida arbitraria M_i como $\mathit{Player}(M_i)$, al segundo como $\mathit{Arg}(M_i)$, y al tercero como $\mathit{Move}(M_i)$. Cuando $\mathit{Move}(M_i) = M_j$ se dice que M_i es una movida en respuesta a M_j.

- *PlayerToMove* es una función que determina a qué jugador le corresponde participar en cada etapa del diálogo. Si por $\mathit{Pow}^\star(\mathit{Moves})$ se entiende al conjunto de todas las secuencias finitas de subconjuntos de *Moves*, entonces la función *PlayerToMove* se define de $\mathit{Pow}^\star(\mathit{Moves})$ en *Players*, con la restricción de que $\mathit{PlayerToMove}(D) = P$ si, y sólo si, D tiene longitud par, y a su vez $\mathit{PlayerToMove}(D) = O$ si, y sólo si, D tiene longitud impar.

- *Legal* es una función que en cada punto de diálogo determina el conjunto de aquellas movidas que resultan razonable realizar en ese momento. Se caracteriza como un función de $\mathit{Pow}^\star(\mathit{Moves})$ en $\mathit{Pow}(\mathit{Moves})$, en donde:

- $M_i \in \mathit{LegalMove}(\emptyset)$ si, y sólo si, M_i es una jugada inicial.
- Si $M_i \in \mathit{LegalMove}(M_1, \ldots, M_{i-1})$, $i > 1$, entonces M_i responde a M_{i-1}.

- *Dialogues* es el conjunto de todas las posibles secuencias de movidas concebibles $M_1, \ldots, M_n$, tales que para todo i, $1 \leq i \leq n$, se verifica que:

  1. $\mathit{Player}(M_i) = \mathit{PlayerToMove}(M_1, \ldots, M_{i-1})$,
  2. $M_i \in \mathit{LegalMove}(M_1, \ldots, M_{i-1})$

- *Winner* es una función parcial que determina quién ha prevalecido a lo largo de un determinado diálogo. Se caracteriza a través de una función de *Dialogues* en *Players*, que sólo señala al jugador *Player* como quien prevaleció en el diálogo D cuando $\mathit{PlayerToMove}(D) = \overline{\mathit{Player}}$ y $\mathit{LegalMove}(D) = \emptyset$.

■

Repasando las definiciones recién presentadas, tomadas literalmente del artículo citado, se puede apreciar que los únicos componentes que quedan parametrizados (*i.e.*, sin especificar), son el conjunto de argumentos considerados válidos *Args*, la relación $\preceq$ y la función *Legal*. Como consecuencia, los marcos dialécticos sólo pueden diferir en cómo interpretarán estos componentes, más allá de toda otra diferencia de orden pragmático que pueda existir.

La teoría de prueba dialéctica tiene por objeto caracterizar al conjunto de inferencias sancionadas por el sistema. A tal efecto, se introduce la noción de estrategia, en esencia un árbol de diálogos en donde para todas y cada una de las posibles jugadas del oponente se registra una única respuesta por parte del proponente.

Definición 4.7. **[Prakken, 2001]**
(estrategia)
Sea $\mathcal{F}$ un marco dialéctico. Diremos que una *estrategia* para el jugador $Player$ en $\mathcal{F}$ es un árbol de diálogos basado en $\mathcal{F}$ que sólo ramifica luego de las movidas de $Player$, y que contiene todas las posibles respuestas válidas que $\overline{Player}$ pueda contraponer a las movidas jugadas por $Player$. ■

La caracterización de esta noción sugerida por el autor resulta un tanto atípica debido a su evidente asimetría. El proponente no cuenta con la posibilidad de equivocarse a la hora de elegir que movida jugar, siempre debe optar por su mejor movida a riesgo de terminar concediendo un debate en el cual hubiera podido prevalecer de haber tomado un curso de acción diferente. En cierto sentido, este estilo de disputa se asemeja más a los debates característicos de los tribunales de justicia que a los propios de la interacción entre agentes.

Combinando la noción de estrategia con el criterio de adjudicación de disputas antes esbozado surge la noción de *estrategia ganadora*. Una estrategia ganadora para un jugador es simplemente una estrategia en la cual todas sus ramas finalizan en una de sus movidas. En base a este concepto, se puede introducir la semántica asociada a los marcos dialécticos de la siguiente forma:

Definición 4.8. **[Prakken, 2001]**
(inferencia rebatible)
Sea $\mathcal{F}$ un marco dialéctico. Diremos que un argumento $\mathcal{A}$ es *inferible rebatiblemente* en base a $\mathcal{F}$ si, y sólo si, el proponente cuenta con al menos una estrategia ganadora para un debate en base a $\mathcal{F}$ que comience con $\mathcal{A}$. ■

La noción de inferencia rebatible completa la presentación del modelo para el razonamiento argumentativo de carácter dialéctico, restando ahora reinterpretar este modelo como si se tratara de un protocolo de disputa.

Protocolo de disputa

La tarea de reinterpretar un marco dialéctico como si fuera el conjunto de reglas que gobierna la interacción a lo largo de un debate resulta relativamente sencilla. Prakken señala un aspecto que, a su entender, distingue a estos escenarios: por un lado, las teorías de prueba dialéctica abarcan todos los diálogos posibles sobre una base de conocimiento de carácter estática, mientras que, por otra parte, los protocolos de disputa también abarcan diálogos concretos con información posiblemente dinámica.

La formalización adoptada por el autor sólo permite modelar disputas en las cuales los contendientes efectúan a lo sumo una jugada por turno, si bien es factible que esta jugada suministre más de una respuesta. Un aspecto interesante es que se asume que los contendientes pueden cambiar sus bases de conocimiento a lo largo del debate, por caso, refinando el modelo mental de su contrincante de turno en base a los argumentos que este último introduzca. Esta peculiar característica se refleja restringiendo al conjunto de todos los argumentos posibles (*Args*) a aquellos que se encuentren habilitados por la información expuesta en la disputa hasta ese momento.

El protocolo de disputa de jugada única se encuentra parametrizado en función de un marco dialéctico. El nuevo marco, ahora para disputas, hereda todas las nociones introducidas para las teorías de pruebas dialécticas a excepción del concepto de diálogo, el cual será reemplazado por el concepto de disputa. Para modelar el carácter cambiante de la base de información se distingue un conjunto de argumentos iniciales (*InitialArgs*), que capturan la información sobre la cual los participantes están de acuerdo al comenzar la discusión. A su vez, la función *Legal* debe ser adecuada para capturar las alternativas existentes al ir construyendo el árbol de diálogos, y la función *Winner* debe incorporar la naturaleza cambiante de la información disponible para determinar qué contendiente ha prevalecido en la disputa.

Con el objeto de evitar confusiones entre los modelos introducidos, el autor distingue con un subíndice R a aquellas nociones correspondientes al protocolo de disputa y con un subíndice T a aquellas pertenecientes a la teoría de prueba dialéctica. Por otra parte, apela a la siguiente notación auxiliar para identificar los diálogos que forman una disputa: para cualquier secuencia de jugadas $D = M_1, \ldots, M_n$, donde M_1 es la jugada inicial, L_j representará al diálogo $M_1, \ldots, M_j$ contenido en D (de acuerdo a la definición 4.6).

Definición 4.9. **[Prakken, 2001]**
(protocolos de disputa)

Diremos que un *protocolo de disputa* consiste de la tupla

$$(\mathcal{F}, \mathit{InitialArgs}, \mathit{Players}_R, \mathit{Moves}_R, \mathit{PlayerToMove}_R, \mathit{Legal}_R, \mathit{Disputes}, \mathit{Winner}_R)$$

en donde:

- $\mathcal{F} = (\mathit{Args}, \preceq, \mathcal{T})$ es un marco dialéctico.
- *InitialArgs* es un subconjunto de *Args*.
- $\mathit{Players}_R = \mathit{Players}_T$
- $\mathit{Moves}_R = \mathit{Moves}_T$
- $\mathit{PlayerToMove}_R = \mathit{PlayerToMove}_T$
- Legal_R es una función de $\mathit{Pow}^\star(\mathit{Moves}_R)$ en $\mathit{Pow}(\mathit{Moves}_R)$, tal que para toda secuencia de jugadas $D = M_1, \ldots, M_i$ verifica que:

  1. $M_i \in \mathit{Legal}_R(\emptyset)$ si, y sólo si, M_i es una jugada inicial.
  2. Si $M_{i+1} \in \mathit{Legal}_R(D)$, entonces se debe verificar que $\mathit{Move}(M_{i+1}) \in D$ y que $\mathit{Player}(\mathit{Move}(M_{i+1})) = \overline{\mathit{Player}(M_{i+1})}$.
  3. Si $M_{i+1} \in \mathit{Legal}_T(L_i)$, entonces $M_{i+1} \in \mathit{Legal}_R(D)$, pero,al mismo tiempo, si $M_{i+1} \in \mathit{Legal}_R(D)$ y $\mathit{Move}(M_{i+1}) = M_k$, entonces $M_{i+1} \in \mathit{Legal}_T(L_k)$.
  4. Si M_{i+1} y M_j, $j < i$, son ambas respuestas posibles para M_k, con $M_j \in D$ y $M_{i+1} \in \mathit{Legal}_R(M_1, \ldots, M_i)$, entonces $\mathit{Arg}(M_i) \neq \mathit{Arg}(M_j)$.

- *Disputes* es el conjunto de todas las secuencias finitas de movidas $M_1, \ldots, M_n$ tales que para todo i, $1 \leq i \leq n$, se verifica que:

  1. $\mathit{Player}_R(M_i) = \mathit{PlayerToMove}_R(M_1, \ldots, M_{i-1})$, y además
  2. $M_i \in \mathit{Legal}_R(M_1, \ldots, M_{i-1})$

  Los integrantes de *Disputes* se denominan *disputas basadas en R*, o sencillamente *R-disputas.* A su vez, para toda disputa D_i, todo L_j, con $j \leq i$, se denominará una *línea de disputa de D_i*.

- $Winner_R$ es una función posiblemente parcial de *Disputes* en $Players_R$, que determina qué contendiente—si es que alguno—ha prevalecido a la disputa.

■

Uno de los aspectos centrales de la definición anterior es la caracterización de la función $Legal_R$. En primer lugar, se puede apreciar que dos de las condiciones son equivalentes a las requeridas para $Legal_T$, con la esencia de la función $Legal_R$ capturada en la tercer condición, la cual establece un puente entre los protocolos de disputa y de dialéctica. Por consiguiente, toda respuesta a la última jugada de una disputa considerada legal en T, será necesariamente legal de acuerdo a R, así como toda movida considerada legal con respecto a R, debe resultar también legal con respecto a T.

El protocolo de interacción hasta aquí expuesto no se encuentra completamente especificado, pues la función para determinar si una jugada resulta aceptable queda caracterizada sólo en forma parcial. Prakken sostiene que la principal dificultad consistirá en obtener una definición sensata y justa para la función *Winner*—aquella que determina qué contendiente logró prevalecer al debate—toda vez que la función $Legal_R$ sea formalmente especificada.

Para poder avanzar hacia este concepto es necesario introducir una noción auxiliar que capture cuáles son los argumentos que resultan relevantes a la hora de determinar el vencedor de una disputa. En principio, se deben considerar al menos todos los argumentos presentados hasta ese momento: para toda disputa D notaremos como $Args(D)$ al conjunto de todos los argumentos mencionados en D. No obstante, Prakken sugiere que el conjunto $Args(D)$ no resulta suficiente *per se* para capturar todos los argumentos relevantes, situación que motiva la introducción de un operador de clausura sobre los mismos. Este operador tiene por objeto incorporar aquellos argumentos *implícitos* que se derivan a partir del conocimiento de cada participante y de la información introducida durante la disputa.

Definición 4.10. **[Prakken, 2001]**
(clausura)

Sea T un conjunto de argumentos arbitrario. Diremos que la *clausura* de T, notada $\mathit{Cl}(T)$, es el conjunto de argumentos que satisface las siguientes restricciones:

- $T \subseteq \mathit{Cl}(T)$,
- si $T \subseteq T'$, entonces $\mathit{Cl}(T) \subseteq \mathit{Cl}(T')$, y además
- si D es una disputa basada en Args y $T = \mathit{Args}(D)$, entonces $\mathit{Cl}(T) \subseteq \mathit{Args}$.

■

A partir de este concepto, es posible caracterizar formalmente al conjunto de argumentos relevantes a una disputa.

Definición 4.11. **[Prakken, 2001]**
(información base)
Sea D una disputa. Diremos que la *información base de* D, notada como $\mathit{Info}(D)$, es la clausura $\mathit{Cl}(\mathit{Args}(D) \cup \mathit{InitialArgs})$. ■

Si bien resulta interesante la propuesta de caracterizar al protocolo de disputa a través de los conceptos propios de los marcos dialécticos, se debe tener en cuenta que al hacerlo estaremos incorporando al modelo de disputa todos los aspectos ya señalados como controversiales en el contexto de los marcos dialécticos. Por caso, el cuestionamiento que sucitó la evidente asimetría de la definición 4.7 también se puede aplicar a la semántica del modelo de disputa. El autor se ve obligado a introducir dos criterios adicionales que permitan distinguir a los protocolos que manifiesten un comportamiento no deseado. En este sentido, propone las nociones de *sensatez* y *justicia*. Desde una perspectiva intuitiva, un protocolo se dice sensato si toda vez que un contendiente prevalece a una disputa acerca de un determinado argumento inicial, entonces ese argumento también resulta inferible en la teoría de prueba dialéctica asociada al tomar como punto de partida la información base empleada en la disputa. La justicia es la noción complementaria, propia de los protocolos que para todo argumento inicial inferible permite que el proponente disponga de al menos una estrategia que asegure su éxito en toda disputa acerca de ese argumento.

Naturalmente, ya que sólo es posible prevalecer en disputas de extensión finita, la formalización de la noción de consecuencia se restringe sólo a argumentos finitamente inferibles.

Definición 4.12. **[Prakken, 2001]**
(finitamente inferible)
Sea $\mathcal{F}$ un marco dialéctico. Diremos que un argumento A es *finitamente inferible* en $\mathcal{F}$ si, y sólo si, A es inferible en F y, además, el proponente cuenta con una estrategia ganadora para A que abarca a un número finito de diálogos. ■

Apelando a esta noción es posible caracterizar formalmente los criterios de sensatez y justicia de la siguiente manera:

Definición 4.13. **[Prakken, 2001]**
(sensatez, justicia)
Sea R un protocolo de disputa en el seno de un marco dialéctico (Args, T). En este contexto, diremos que:

- R es *sensato* si para toda R-disputa D, se verifica que cuando D es ganado por el proponente, entonces $\mathit{Arg}(M_1)$, con $M_1 \in D$, resulta inferible rebatiblemente en el marco de $(\mathit{Info}(D), T)$

- R es *justo* si para toda R-disputa D_i, tal que $\mathit{Arg}(M_1)$, donde $M_1 \in D_i$, resulta finitamente inferible en el marco de $(\mathit{Info}(D_i), T)$, pero a su vez no lo es en el marco $(\mathit{Info}(D_{i-1}), T)$, entonces el proponente puede ganar cualquier posible continuación de D_i que esté basada en $\mathit{Info}(D_i)$.

■

La combinación de estos criterios con el concepto de argumento relevante permite explorar qué variantes de la función *Winner* aseguran que el protocolo obtenido resulte sensato y justo. La alternativa más evidente consiste en establecer que el proponente ha ganado la disputa cuando el argumentos inicial se torna inferible rebatiblemente a partir de $Info(D)$. Prakken denomina a esta alternativa *criterio trivial*, ya que la sensatez y justicia del protocolo se encuentran aseguradas por definición. La principal objeción señalada en contra de esta alternativa—análoga a la adoptada por Loui en su modelo—es su alto costo computacional. Cabe acotar que juzgar el mérito de una semántica en base a qué tan complejo pueda resultar implementarla constituye una práctica riesgosa. Diversos formalismos implementados hoy en día de forma eficiente fueron en sus comienzos definiciones complejas de carácter netamente semántico (por caso, la lógica default o la circunscripción).

Prakken considera otra alternativa para la función $Winner_R$, consistente en asumir que un contendiente ganará la disputa si ha realizado la última jugada, y al mismo tiempo logra que el otro contendiente no pueda producir una nueva respuesta en base al conocimiento relevante hasta allí introducido. Esta variante se denomina, en contraste, *criterio no trivial*. No obstante, este criterio de adjudicación causa ciertos inconvenientes con las nociones de sensatez y justicia. Existen situaciones en las cuales los argumentos habilitados por la nueva información agregada a lo largo de la disputa deberían poder ser introducidos al debate, pero el protocolo lo impida. Este llamativo comportamiento es producto de haber incorporado información una vez comenzada la disputa. Prakken sugiere lidiar con este inconveniente a través de un nuevo tipo de disputa.

Disputas liberales

El criterio de adjudicación no trivial parece ser el más adecuado, si bien al combinarlo con el protocolo hasta aquí delineado da a lugar al tipo de situación problemática comentada anteriormente. Una posible solución consiste en refinar el protocolo de interacción de forma tal que el criterio no trivial resulte justo y sensato. Considerando que el principal problema del protocolo de respuesta única es que en ciertas circunstancias impide la introducción de jugadas relevantes, el autor sugiere un nuevo protocolo más liberal en el cual el introducir una nueva respuesta siempre resulte aceptable. No obstante, el permitir la incorporación irrestricta de nuevas respuestas daría a lugar más controversia, a menos que se acompañe de alguna forma de política que regule la incorporación de movidas. A tal efecto, el autor introduce el concepto de *relevancia*, el cual permite caracterizar al nuevo tipo de disputa.

Definición 4.14. **[Prakken, 2001]**
(disputas liberales)
Sea P un protocolo de disputa. Diremos que P es un *protocolo para disputas liberales* si, y sólo si, para toda disputa D y toda jugada M, se verifica que $M \in Legal_R(D)$ sólo cuando M satisface las condiciones estipuladas en la definición 4.9, y, además, M resulta relevante en D. ■

Para poder comprender correctamente el concepto de relevancia resulta conveniente familiarizarse en primer lugar con la noción de disputa liberal. En este sentido, las disputas liberales inducen una estructura arbórea formada por las líneas de disputa en las cuales los contendientes realizan sus movidas, puediendo acceder a las distintas ramas incluso sin depender de órden preestablecido alguno. El autor reconoce que esta variante admite disputas menos focalizadas, pero señala que este es el precio que hay que pagar para obtener un protocolo sensato y justo. En realidad, este cuestionamiento apunta más las decisiones de diseño tomadas en el contexto de los marcos dialécticos, las que como efecto colateral no deseado permiten la existencia de protocolos de disputa que no sean sensatos o justos.

El criterio de adjudicación en el contexto de las disputas liberales se define de forma análoga:

Definición 4.15. **[Prakken, 2001]**
(criterio de adjudicación)
Sea D una disputa liberal. En este contexto diremos que $Winner(D) = Player$ toda vez que se verifiquen las siguientes condiciones:

1. $PlayerToMove(D) = \overline{Player}$, y además
2. para todo $M \in Legal_R(D)$ se observa que $Arg(M) \notin Info(D)$.

■

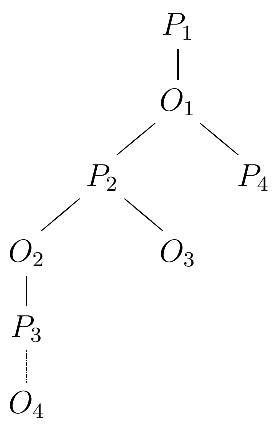


Figura 4.1: Relevancia de las jugadas en una disputa

Este criterio junto con la caracterización del protocolo para conducir disputas liberales concluyen la presentación del modelo, si bien resta aún especificar bajo qué condiciones una determinada movida puede ser considerada como relevante al diálogo en curso. El autor propone analizar esta situación de forma más concreta a través del siguiente ejemplo:

Ejemplo 4.2. Sea D la disputa $\langle P_1, O_1, P_2, O_2, P_3, O_3, P_4 \rangle$, bosquejada en la figura 4.1. En este contexto, observemos que O_4 no es para nada relevante en lo que al resultado de la disputa concierne, puesto que el proponente eventualmente prevalecerá a esta disputa. Incluso suponiendo que $\mathit{Info}(D)$ habilite la movida O_4 como respuesta a P_3, no seá posible cambiar el resultado de la disputa.

Nótese que cuestiones de índole netamente práctico se están introduciendo en la definición teórica del modelo de interacción. De hecho, el concepto de relevancia propuesto no sólo resuelve la falta de sensatez y justicia, sino que también efectiviza una especie de poda del espacio de búsqueda de las movidas disponibles. El autor, ajeno a estos cuestionamientos, prosigue con este curso de acción introduciendo el concepto de *estado disputacional* de cada una de las movidas en una disputa. Esta noción captura el estado que puede ser asociado a una determinada jugada al considerar sólo aquellas jugadas presentes en D, sin tener en cuenta al conjunto completo de las jugadas habilitadas por $\mathit{Info}(D)$.

Definición 4.16. **[Prakken, 2001]**
(estado de controversia de una movida)
Sea D una disputa y M una de las movidas de D. Diremos que el *estado de controversia de la movida M en D* es in si, y sólo si, todos las jugadas de D que respondan a M tienen estado out. En contraste, en toda otra situación M tendrá como estado de controversia en D al estado out. ■

En base a este concepto, una cierta movida se dirá *relevante* sólo si modifica el estado de controversia de la jugada inicial. Este mecanismo de asignación de estados asemeja al proceso ideado en [Simari et al., 1994] para el marcado de árboles dialécticos. En términos formales:

Definición 4.17. **[Prakken, 2001]**
(relevancia)
Sea M una movida y D una disputa. Diremos que *M es relevante en D* si, y sólo si, al incorporar M a D se modifica el estado de controversia de su movida inicial. ■

Prakken concluye la presentación de este modelo de interacción comprobando formalmente que el protocolo de disputas liberales satisface las propiedades de sensatez y justicia. Una de los principales méritos de la propuesta de Prakken, más allá de los cuestionamientos señalados a lo largo de este análisis, es haber logrado instanciar al conjunto de conceptos abstractos que conforman la propuesta de Ron Loui.

4.3. Un modelo abstracto para el razonamiento dialéctico

Esta sección desarrolla la primer mitad de la estrategia propuesta para obtener un modelo para la deliberación entre agentes, que sintéticamente consiste en la formulación de un modelo abstracto para el razonamiento dialéctico. En virtud de que la argumentación en términos dialécticos es el formalismo por excelencia para conducir razonamientos de base dialéctica unilateral, el punto de partida de esta sección consiste en identificar los aspectos primordiales de los exponentes del razonamiento argumentativo reseñados en el capítulo 2, prestando particular atención a sus formulaciones en términos dialécticos, con el objeto de elaborar a partir de éstos un modelo del proceso de razonamiento dialéctico en sí. A tal efecto, la sección 4.3.1 presenta una introducción a los citados aspectos primordiales, motivando la posterior formalización del modelo a lo largo de la sección 4.3.2.

4.3.1. Aspectos primordiales

Consideremos un escenario en el cual un cierto agente se dispone a evaluar la razonabilidad de una tesis, apelando a una disputa dialéctica de carácter introspectiva. Este agente, claramente, debe estar dispuesto a desempeñar al menos dos roles. En ciertas ocasiones deberá consolidar las razones que soportan la tesis en consideración, así como en otras ocasiones deberá adoptar la posición de "abogado del diablo", cuestionando las mismas razones que acaba de exponer. Esta dualidad se encuentra capturada en los diversos formalismos estudiados, donde existe el consenso de denominar *proponente* al rol que asume la defensa de la tesis bajo análisis, y *oponente* al rol que opta por cuestionarla. Como estos roles son desempeñados por el mismo agente, las razones expuestas bajo cualquiera de estos roles tendrán un mismo sustento, esto es, la base de conocimiento del agente que esté llevando a cabo el razonamiento.

A continuación, las nociones más elaboradas serán presentadas apelando a una útil analogía: concebir el proceso llevado adelante en todo razonamiento dialéctico como si se tratara de un juego en donde dos contendientes se alternan para presentar razones a favor y en contra de la tesis en consideración. Esta analogía ha sido empleada en reiteradas ocasiones como medio a través del cual introducir las reformulaciones dialécticas de distintos sistemas argumentativos de una manera más intuitiva (*e.g.*, en los sistemas Simari-Loui o Prator, entre otros).

En este sentido, corresponde preguntarse en qué consiste cada uno de los turnos o movidas que se alternan a lo largo de este juego dialógico. Evidentemente, cada movida debe consolidar o socavar la tesis debatida, en función del rol actual al hacer esa movida. Considerando la naturaleza de los sistemas de razonamiento que estamos intentando abstraer, resulta natural suponer que esta consolidación o socavamiento es llevada a cabo formulando argumentos lógicos. A través del intercambio de argumentos se suele ir explorando una faceta particular del tópico en discusión, posiblemente hasta agotar sus vertientes, para luego seguir explorando alguna otra de sus facetas. Esta noción se encuentra capturada en los diversos formalismos abordados bajo el nombre de *línea de argumentación*. Una línea de argumentación es simplemente una secuencia de movidas relacionadas entre sí—un argumento, seguido de alguno de sus contraargumentos, a su vez seguido de alguno de los contraargumentos del primer contraargumento, etc.

Esta importante noción constituye la piedra basal en torno a la cual se erige nuestro modelo para el razonamiento dialéctico. En contraste, los modelos propuestos por Loui y Prakken, analizados en las secciones anteriores, adoptaban al argumento como punto de partida. Esta decisión refleja el importante rol que las líneas de argumentación desempeñan en este contexto, ya que toda disputa posible puede ser fácilmente descripta a través del conjunto de líneas de argumentación exploradas en la misma. En cierto sentido, se puede afirmar que el conjunto de líneas de argumentación parcialmente exploradas denotan una instantánea del estado actual de la disputa. Una vez introducido el concepto de estado de una disputa, surgen de forma natural los siguientes interrogantes:

- ¿Qué jugador debe mover a continuación?
- ¿Cuál es el conjunto de movidas válidas disponibles?
- ¿Ya finalizó la disputa? De ser así, ¿quién prevaleció?

Al intentar contestar estos interrogantes hay que tener cuidado de no comprometer el modelo propuesto mediante una decisión de diseño que restrinja su aplicabilidad, puesto que una determinada política para responder a cualquiera de estas preguntas puede parecer indicado en un cierto escenario, pero al mismo tiempo resultar completamente inapropiada en un contexto diferente. Para evitar este tipo de inconvenientes, la formalización de estas nociones dejará estos aspectos deliberadamente sin especificar, debiéndoselos parametrizar de acuerdo a la instanciación concreta que se desee obtener. Por caso, cada uno de los sistemas de razonamiento argumentativo considerados en el capítulo 2 establece claramente cuál debe ser la alternancia entre proponente y oponente, cuáles son los argumentos que razonablemente pueden ser formulados en una cierta configuración, o cuál ha sido el resultado del análisis dialéctico recién llevado a cabo.

Un último aspecto que debemos abordar una vez establecido que se entiende por estado de una disputa es cómo se relacionan entre sí los diversos estados que puede transitar un determinado debate. En cierto sentido, las nociones hasta aquí introducidas no hacen más que reflejar el lado estático de una disputa, restando modelar su contrapartida dinámica. Este aspecto admite ser capturado mediante una función de transición que determine cuáles son los efectos de jugar en el marco de una disputa en particular a una cierta movida con el propósito de ahondar una determinada línea de argumentación. Esta función, al relacionar los estados de una disputa antes y después de hacer las movidas que extienden las distintas líneas de argumentación permite capturar la dinámica del juego dialógico.

Este último concepto redondea la introducción de las bases de nuestro modelo. Con el objeto de profundizar el tratamiento de los conceptos introducidos en esta sección debemos llevar adelante su consolidación formal.

4.3.2. Protocolos Dialécticos

Esta sección presenta la terminología y el conjunto de definiciones requeridas para caracterizar al modelo para el razonamiento dialéctico—brevemente comentado en la sección anterior—ahora desde una perspectiva formal.

De acuerdo a lo expuesto, una disputa dialéctica acerca de una cierta tesis $\mathcal{T}$,

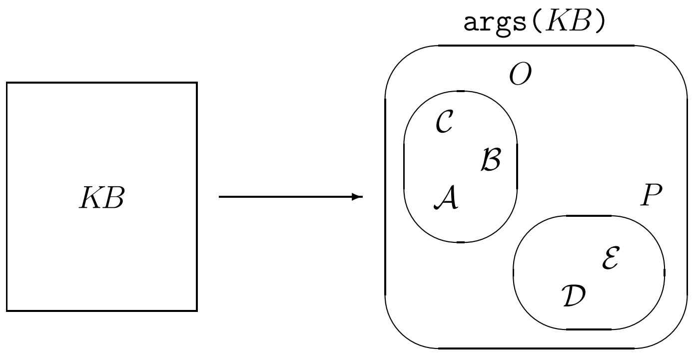


Figura 4.2: Distribución del conocimiento en un contexto de agente único

formulada en un lenguaje de representación de conocimiento $\mathcal{L}$ acerca del cual sólo se estipula que permita instanciar el concepto de argumento de una manera razonable, involucra a dos contendientes: al proponente de la tesis, P, y a su oponente, O. Sea KB la base de conocimiento a disposición de ambos, también codificada en el lenguaje $\mathcal{L}$, acerca del cual sólo se requiere que permita expresar razones o justificaciones, las que serán denominadas en este contexto argumentos, y sea `args` un mapeo que dada una determinada base de conocimiento retorna al conjunto de todos los argumentos que puedan ser construídos a partir esa base de conocimiento. Las condiciones particulares que caracterizan a la noción de argumento adoptada, así como el procedimiento que permite obtener argumentos a partir de la información presente en la base de conocimiento, serán aspectos relevantes en el marco de una instanciación concreta del modelo, pero estos aspectos no forman parte de la formulación inicial de naturaleza abstracta. No obstante, dado que se trata de un razonamiento introspectivo llevado a cabo por un único agente, tanto el conjunto de argumentos a disposición de P como el conjunto de argumentos a disposición de O deben coincidir con `args`(KB).

Los aspectos hasta aquí reseñados constituyen en cierta forma el marco en el cual tendrá a lugar el razonamiento de base dialéctica, situación que motiva la introducción del siguiente concepto:

Definición 4.18.
(contexto)
Denominaremos *contexto* a la tupla $\mathcal{C} = (\mathcal{L}, KB, \texttt{args})$, donde:

- $\mathcal{L}$ es un lenguaje de representación de conocimiento arbitrario, si bien debe ser capaz de expresar argumentos.

- KB es la base de conocimiento, codificada en términos de $\mathcal{L}$, a disposición de todos los contendientes.

- *args* es un mapeo que tomando la base de conocimiento retorna al conjunto de todos los argumentos posibles que se pueden construir a partir del conocimiento allí disponible.

■

En lo que resta de esta sección, asumiremos las existencia en todo momento de un contexto $(\mathcal{L}, KB, \texttt{args})$ arbitrario pero fijo. Retomando la analogía entre el razonamiento de base dialéctica con un juego dialógico de intercambio de argumentos, introduciremos, en primer lugar, qué se entiende formalmente por movida o jugada:

Definición 4.19.
(movida)
Diremos que una *movida* es una tupla $(contendiente, argumento)$, donde *contendiente* es P u O y *argumento* $\in$ $\texttt{args(}KB\texttt{)}$. ■

Cuando resulte conveniente, la movida $m = (contendiente, argumento)$ puede ser descompuesta apelando a la notación auxiliar *player*(m) para denotar a *contendiente* y *arg*(m) para denotar a *argumento*.

De acuerdo a la discusión informal de la sección anterior, las diversas movidas realizadas se estructuran en líneas de argumentación:

Definición 4.20.

(línea de argumentación)

Sea $\mathcal{T}$ una tesis. Diremos que una *línea de argumentación acerca de* $\mathcal{T}$ es una secuencia de movidas $\langle m_0, m_1, \ldots, m_k \rangle$, $k \geq 0$, tal que:

- $m_0 = (P, \mathit{argumento})$, donde *argumento* es un argumento que soporta a $\mathcal{T}$.

■

Esta definición establece que la primer movida de toda línea de argumentación debe ser jugada por P y debe sustentar a la tesis objeto del debate. No obstante, la definición anterior no se comporta de la manera esperada, como se puede apreciar en el siguiente ejemplo:

Ejemplo 4.3. Considerando los siguientes argumentos:

- $\mathcal{A}$ = habría que despedir a Juan porque no viene a trabajar.
- $\mathcal{B}$ = Juan no viene a trabajar porque está de licencia.
- $\mathcal{C}$ = el gato de Juan es de color beige.

En este marco, una posible línea de argumentación acerca de si hay que despedir o no a Juan puede avarcar las movidas $\langle m_0, m_1, m_2 \rangle$, donde $m_0 = (P, \mathcal{A})$, $m_1 = (O, \mathcal{B})$ y $m_2 = (P, \mathcal{C})$.

Como se observa, no toda secuencia posible de movidas representa una línea de argumentación que interece considerar. Por caso, en el ejemplo 4.3, la secuencia de movidas indicada, que hace referencia a argumentos no relacionados entre sí, ciertamente no constituye la exploración de una faceta del debate. Con el objeto de evitar esta situación, la noción de línea de argumentación será refinada a través de una función denominada `legal`, cuyo propósito es determinar el conjunto de movidas que tienen permitido extender a una línea de argumentación particular. Formalmente, si *Lines* es el conjunto de todas las líneas de argumentación posibles y *Moves* es el conjunto de todas las movidas válidas, `legal` denota a una función de *Lines* en $\mathcal{P}(Moves)$. Esta signatura es la única restricción impuesta sobre la función `legal`, ya que su definición pormenorizada es una de las tareas que se deben llevar adelante al instanciar el modelo abstracto.

Apelando a esta función, es posible refinar el concepto de línea de argumentación de forma tal que evite agrupar argumentos no relacionado entre sí

Definición 4.21.
(línea de argumentación revisada)
Sea $\mathcal{T}$ una tesis. Diremos que una *línea de argumentación acerca de* $\mathcal{T}$ es una secuencia de movidas $\langle m_0, m_1, \ldots, m_k \rangle$, $k \geq 0$, tal que:

- $m_0 = (P, argumento)$, donde *argumento* es un argumento que soporta a $\mathcal{T}$, y además
- para cada i, $1 \leq i \leq k$, se verifica que $m_i \in \texttt{legal}(\langle m_0, \ldots, m_{i-1} \rangle)$.

■

El disponer de una definición de línea de argumentación capaz de capturar las situaciones de interés nos permite capturar formalmente el estado de una disputa:

Definición 4.22.
(estado de una disputa)
Sea $\mathcal{T}$ una tesis. El *estado de una disputa acerca de* $\mathcal{T}$ es un conjunto no vacío de líneas de argumentación acerca de $\mathcal{T}$. ■

Aún cuando la definición anterior impone pocas restricciones sobre el concepto de estado de una disputa, no todo estado posible será alcanzable en un sentido práctico. Por ejemplo, aquellos estados que denoten la continuación de un debate ya ganado por alguno de los contendientes posiblemente nunca sean alcanzados por una disputa en concreto, aún tratándose de estados válidos de acuerdo a la definición anterior.

Una vez modelado el concepto de estado de una disputa surgen, una vez más, los mismos interrogantes planteados en la discusión informal de los aspectos primordiales. Resulta natural preguntarse quién debe mover a continuación o cuál de los contenientes se impuso en un determinado estado de la disputa. En este sentido, es importante enfatizar que responder alguno de estos planteos de forma concreta y definitiva no hace más que restringir la aplicabilidad del modelo. No obstante, la toma de decisión que involucra determinar estas respuestas afecta de manera directa al desempeño del modelo, por lo que los mecanismos adoptados para abordar cada uno estos planteos deberán ser tenidos en cuenta en todo intento de formalización. Para reconciliar ambas visiones, el modelo que estamos proponiendo atiende estos cuestionamientos a través de sendas funciones de carácter abstracto, las cuales quedarán sin especificar en el modelo general, pero que deberán ser instanciadas de forma concreta toda vez que se aplique este modelo a un escenario en particular.

A continuación, en la especificación de las citadas funciones, *States* representa el conjunto de todos los posibles estados de una disputa:

- `toMove`: una función de *States* en $\{P, O\}$ que, a partir del estado de una disputa acerca de una tesis, determina cuál de los jugadores debe realizar el próximo movimiento.

- `winner`: una función de *States* en $\{P, O\} \cup \{\texttt{none}\}$ que partiendo del estado de una disputa acerca de una tesis determina cuál de los contendientes—si es que alguno—se ha impuesto en la misma. En el caso que ningún contrincante haya logrado prevalecer, la función debe retornar la constante `none`.

La función `toMove` tiene por objeto modelar al peso de la prueba,[2] el cual va cambiando de lado a lo largo de la disputa. Ya que la función `toMove` permite indagar uno de los aspectos de la jugada que se está a punto de realizar (*i.e.*, quién la realizará), esta función debe presentar un comportamiento compatible con la función `legal`, la cual en cierto sentido también caracteriza a la movida que está por ser

[2]en el sentido dado a la expresión por Rescher.

jugada. Por caso, el siguiente ejemplo ilustra una situación indeseada producto del conflicto entre lo retornado por estas funciones.

Ejemplo 4.4. Considerando a los siguientes argumentos:

- $\mathcal{A}$ = habría que despedir a Juan porque no viene a trabajar.
- $\mathcal{B}$ = Juan no viene a trabajar porque está de licencia.
- $\mathcal{C}$ = Juan no viene a trabajar porque estamos en vacaciones.

Si se adopta una definición para la función `legal` que permita a los dos contendientes mover cualquiera de los argumentos disponibles siempre que no se produzca una reiteración de argumentos, y una definición para la función `toMove` que obligue a los contendientes a alternarse en el uso de la palabra, la línea de argumentación $\langle m_0, m_1, m_2 \rangle$, donde $m_0 = (P, \mathcal{A})$, $m_1 = (O, \mathcal{B})$ y $m_2 = (O, \mathcal{C})$, consitituye una línea de argumentación válida, ya que la primer movida guarda la forma debida y respeta a la función `legal` al no reiterar argumentos (*i.e.*, $m_1 \in \texttt{legal}(\langle m_0 \rangle)$ y $m_2 \in \texttt{legal}(\langle m_0, m_1 \rangle)$). No obstante, esta secuencia de movidas, donde el oponente formula dos razones de forma consecutiva, no respeta el orden de participación establecido por la función `toMove`.

En concreto, para evitar el tipo de conflicto señalado en el ejemplo anterior debemos imponer la restricción adicional de que las funciones `legal` y `toMove` coincidan en lo que retornan, es decir, que el contendiente que introdujo cada una de las movidas a lo largo de una línea de argumentación sea exáctamente aquel que `toMove` hubiera retornado de ser el caso que esa línea de argumentación fuera la única explorada en la disputa. En otras palabras, sólo se consideraran como movidas válidas aquellas realizadas por el contendiente que debía mover en esa situación. Esta restricción adicional puede ser capturada en términos formales estableciendo que toda vez que $(\mathit{contendiente}, \mathit{argumento}) \in \texttt{legal}(\mathit{línea})$, entonces necesariamente $\texttt{toMove}(\{\mathit{línea}\}) = \mathit{contendiente}$.

Para concluir, corresponde caracterizar al componente que aporta el dinamismo al marco conceptual delineado por este conjunto de definiciones: la función de transición `next`. Esta función, definida de $States \times Moves \times Lines$ en *States*, tiene por objeto determinar qué estado resulta cuando una cierta movida cuyo propósito es ahondar una línea de argumentación dada es judaga en el contexto de un estado de disputa en particular. Cabe acotar que esta función es el único componente capaz de alterar el estado de una disputa, es decir, ninguno de los aspectos introducidos previamente cuentan con la habilidad de *crear*, *cambiar* o *eliminar* líneas de argumentación. Nuevamente, especificaremos a esta función sólo de forma abstracta, por las mismas razones antes expuestas: no limitar la aplicabilidad del modelo propuesto. Una vez más, uno de los aspectos más importantes a la hora de instanciar el modelo abstracto consiste precisamente en definir esta función de forma acabada.

Finalmente, estamos en condiciones de estructurar el conjunto de definiciones introducidas junto con los conceptos parcialmente especificados para dar a lugar a la noción de protocolo dialéctico:

Definición 4.23.
(protocolo dialéctico)
Sea $\mathcal{C}$ un contexto. Diremos que un *protocolo dialéctico* para $\mathcal{C}$, notado $\mathbb{PD}_{\mathcal{C}}$, se caracteriza a través de la tupla

$$(Moves, Lines, States, \texttt{legal}, \texttt{toMove}, \texttt{winner}, \texttt{next})$$

donde:

- *Moves* es el conjunto de movidas consideradas válidas (def. 4.19, pág. 214).
- *Lines* es el conjunto de líneas de argumentación consideradas válidas (def. 4.21, pág. 216).
- *States* es el conjunto de todos los posibles estados de disputa (def. 4.22, pág. 216).

- `legal` es una función definida de *Lines* en $\mathcal{P}(Moves)$, que retorna el conjunto de jugadas que tienen permitido extender una determinada línea de argumentación.

- `toMove` es una función definida de *States* en $\{P, O\}$, que retorna qué contrincante debe jugar a continuación.

- `winner` es una función definida de *States* en $\{P, O\} \cup \{\texttt{none}\}$, que retorna qué contendiente prevaleció a un cierto debate, donde la constante `none` denota que todavía no existe un ganador.

- `next` es una función definida de $States \times Moves \times Lines$ en *States*, que captura el efecto de hacer un cierta movida para extender una determinada línea de argumentación en un estado de disputa en particular.

Con el objeto de preservar la consistencia del protocolo dialéctico, se debe verificar que cuando $(\textit{contendiente}, \textit{argumento}) \in \texttt{legal}(\textit{línea})$, entonces necesariamente $\texttt{toMove}(\{\textit{línea}\}) = \textit{contendiente}$. ■

Cuando el contexto al cual hace referencia un protocolo dialéctico sea evidente, notaremos al protocolo $\mathbb{PD}_{\mathcal{C}}$ simplemente como $\mathbb{PD}$.

Toda instancia concreta de un protocolo dialéctico tiene asociado una semántica propia, la cual se deriva de la interacción de sus componentes:

Definición 4.24.
(sustento)
Sea $\mathbb{PD}_{\mathcal{C}} = (Moves, Lines, States, \texttt{legal}, \texttt{toMove}, \texttt{winner}, \texttt{next})$ una instancia particular de un protocolo dialéctico, y sea $\mathcal{T}$ una tesis. Diremos que $\mathcal{T}$ está *sustentada* por $\mathbb{PD}_{\mathcal{C}}$ si, y sólo si, existe una secuencia finita $s_0, s_1, \ldots, s_n$ de estados de disputa acerca de $\mathcal{T}$, tales que:

- $s_0 = \{\langle m_0 \rangle\}$, donde $\langle m_0 \rangle$ es una línea de argumentación acerca de $\mathcal{T}$,

- para cada i, $0 \leq i < n$, es posible encontrar una línea de argumentación $\langle m_0, \ldots, m_k \rangle \in s_i$ en la cual para algún j, $0 \leq j \leq k$, existe una movida $(\mathit{contendiente}, \mathit{argumento}) \in \texttt{legal(}\langle m_0, \ldots, m_j \rangle\texttt{)}$, que verifique que:
    - $\mathit{contendiente} = \texttt{toMove(}s_i\texttt{)}$, y además
    - $\texttt{next(}s_i\texttt{,}(\mathit{contendiente}, \mathit{argumento})\texttt{,}\langle m_0, \ldots, m_j \rangle\texttt{)} = s_{i+1}$.
- para cada i, $0 \leq i < n$, $\texttt{winner(}s_i\texttt{)} = \texttt{none}$, y además
- $\texttt{winner(}s_n\texttt{)} = P$.

■

La semántica asociada a las distintas instancias de protocolos dialécticos permite modelar el comportamiento de diversas teorías dialécticas. En particular, la sección 4.5 aboradará como reformular dos de los formalismos estudiados en el capítulo 2 en términos del modelo propuesto.

4.4. Un modelo abstracto para la deliberación

La sección anterior abarcó el desarrollo de la primera mitad de laestrategia propuesta para obtener un modelo para la deliberación entre agentes. A través del concepto de protocolo dialéctico, junto con su semántica asociada, se puede capturar la dinámica de diversas teorías argumentativas. La presente sección completa el desarrollo de esta estrategia, tomando el modelo obtenido como punto de partida el cual, en virtud de isomorfismo identificado por Rescher, es reinterpretarlo como si fuera un modelo para la interacción entre agentes. En función de la estrategia adoptada, esta segunda etapa no involucra la toma de decisiones de diseño de relevancia, puesto que consiste en una adaptación a un nuevo contexto del modelo obtenido, no así un un completo rediseño.

Rescher destacó en [Rescher, 1977] que los razonamiento de base dialéctica unilateral y multilateral presentaban un estructura análoga, casi isomorfa. El autor, concentrado más en destacar el importante rol de la dialéctica que en ahondar en los aspectos que efectivamente diferencian a estos tipos de razonamientos, adoptó una simplificación que de conservarla haría que el modelo reinterpretado producto de aplicar nuestra estrategia sea demasiado restrictivo. Concretamente, Rescher supuso que tanto en el ámbito unilateral como en el multilateral existe una base de conocimiento única, a disposición de todos los participantes. Esta suposición resulta justificada en el contexto que el autor tenía en mente, esto es, los debates que surgen entre un grupo de científicos al discutir la viabilidad de nuevas ideas y conceptos. No obstante, al reemplazar el grupo de científicos por un conjunto de agentes autónomos, la situación naturalmente cambia. Si bien en el contexto de la dialéctica unilateral es válido asumir una fuente de información única a partir de la cual basar todo razonamiento, en el marco de la dialéctica multilateral resulta mucho más sensato suponer que existirán tantas bases de conocimiento como actores tomando parte en la deliberación. En lo que resta de esta sección, desglosaremos el conjunto de definiciones formales que conforman el modelo para la deliberación que se obtiene de aplicar la estrategia introducida a tal efecto.

Sea $\mathcal{T}$ una tesis acerca de la que se desea deliberar, formulada en un lenguaje de representación de conocimiento $\mathcal{L}$ donde, una vez más, sólo se estipula que permita instanciar el concepto de argumento de una manera razonable. Una disputa dialéctica acerca de $\mathcal{T}$, involucra en este nuevo contexto a varios agentes: $P_1, P_2, \ldots, P_n$ sustentando a $\mathcal{T}$, y $O_1, O_2, \ldots, O_m$ oponiéndose a ésta. En ocasiones, se hará refencia a los agentes sustentan la tesis como $Agents_P$, y a los que están en su contra como $Agents_O$. Como la simplificación de Rescher de que sólo existe una única fuente de información no será mantenida, debemos incorporar al modelo la posibilidad de que los diversos agentes cuentes con sendas bases de conocimiento, la que no necesariamente coincidirán entre sí. Por consiguiente, para cada uno de los agentes Ag, denotaremos por KB_{Ag} a su base de conocimiento particular. La única restricción sobre la base de conocimiento de los agentes es que deben estar codificadas en $\mathcal{L}$.

Sea `args` un mapeo que tomando una base de conocimiento retorna al conjunto

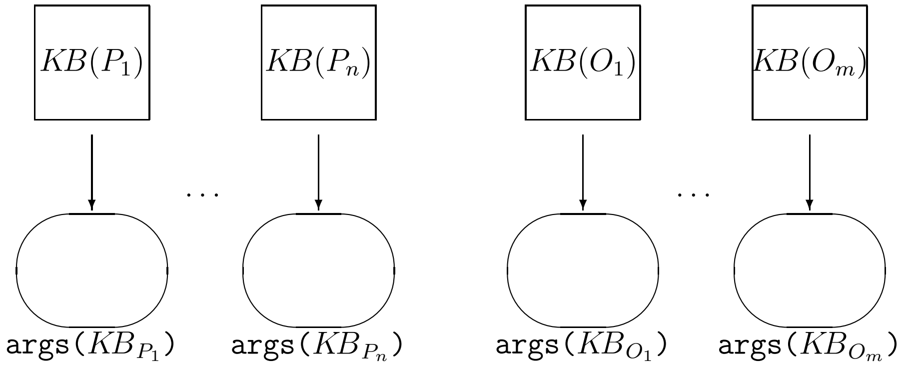


Figura 4.3: Distribución del conocimiento en un contexto de agentes múltiples

de todos los argumentos que puedan ser construídos a partir de ésta. En particular, para cada agente Ag, $\texttt{args}(KB_{Ag})$ denotará al conjunto de todos los argumentos posibles que puedan ser construídos utilizando la información almacenada en su propia base de conocimiento KB_{Ag}. Nuevamente, las condiciones particulares que caracterizan a la noción de argumento adoptada, así como el procedimiento que permite construir argumentos a partir de la información presente en la base de conocimiento, serán aspectos relevantes en el marco de una instanciación concreta del modelo, pero estos aspectos no forman parte de la formulación general.

Los aspectos hasta aquí introducidos constituyen el marco en el cual tendrá a lugar la deliberación entre los agentes:

Definición 4.25.
(contexto)
Denominaremos *contexto* a la tupla $\mathcal{C} = (\mathcal{L}, Agents_P, Agents_O, KBs, \texttt{args})$, donde:

- $\mathcal{L}$ es un lenguaje de representación de conocimiento arbitrario, si bien debe ser capaz de expresar argumentos.

- $Agents_P$ es el conjunto de agentes que sustentan la tesis a ser debatida.

- $Agents_O$ es el conjunto de agentes que se oponen a la tesis a ser debatida.

- KBs es el conjunto de bases de conocimiento asociadas a cada uno de los agentes presentes en $Agents_P \cup Agents_O$, que presentan como característica común el encotrarse codificadas en términos de $\mathcal{L}$.

- *args* es un mapeo que tomando una base de conocimiento retorna al conjunto de todos los argumentos posibles que se pueden construir a partir del conocimiento allí almacenado.

Cuando resulte conveniente, se denotará al conjunto $Agents_P \cup Agents_O$ de todos los agentes que participan de la deliberación simplemente como *Agents*. ■

Apelando a una estrategia simplificación de la notación similar a la adoptada en la sección anterior, en lo que resta asumiremos la existencia en todo momento de un contexto

$$(\mathcal{L}, \{P_1, \ldots, P_n\}, \{O_1, \ldots, O_m\}, \{KB_{P_1}, \ldots, KB_{P_n}, KB_{O_1}, \ldots, KB_{O_m}\}, \texttt{args})$$

arbitrario pero fijo. Bajo el nuevo enfoque de agentes múltiples, la noción de *movida* no requiere mayor actualización:

Definición 4.26.
(movida)
Diremos que una *movida* es una tupla $(agente, argumento)$, donde se verifica que $agente \in Agents$ y $argumento \in \texttt{args}(KB_{agente})$. ■

El concepto anterior de línea de argumentación da a lugar bajo este nuevo enfoque al de *línea de deliberación*:

Definición 4.27.
(línea de deliberación)
Sea $\mathcal{T}$ una tesis. Diremos que una *línea de deliberación acerca de* $\mathcal{T}$ es una secuencia de movidas $\langle m_0, m_1, \ldots, m_k\rangle$, $k \geq 0$, tales que:

- $m_0 = (agente, argumento)$, donde para algún $agente \in Agents_P$ se verifica que *argumento*, uno de los argumentos a disposición de *agente*, soporta a $\mathcal{T}$, y además

- para cada i, $1 \leq i \leq k$, se verifica que $m_i \in \texttt{legal}(\langle m_0, \ldots, m_{i-1}\rangle)$.

■

La función `legal` responde a las mismas motivaciones que antes, esto es, asegurar que los argumentos esgrimidos en las distintas líneas de deliberación presenten algún aspecto que los relacione, por caso, la exploración de una de las facetas de abordadas en la deliberación. En este caso, si *Lines* denota al conjunto de todas las líneas de deliberación posibles, y *Moves* al conjunto de todas las movidas consideradas válidas, la función `legal`, definida de *Lines* en $\mathcal{P}(Moves)$, debe retornar al conjunto de movidas que tienen permitido extender a una determinada línea de deliberación.

Para capturar el estado de una deliberación apelaremos a una estrategia similar a la adoptada para modelar el estado de una disputa. El estado particular en el que se encuentra una cierta deliberación será caracterizado a través del conjunto de líneas de deliberación parcialmente exploradas hasta ese instante de su evolución en el tiempo.

Definición 4.28.
(estado de una deliberación)
Sea $\mathcal{T}$ una tesis. Diremos que el *estado de una deliberation acerca de* $\mathcal{T}$ es un conjunto no vacío de líneas de deliberación acerca de $\mathcal{T}$. ■

Una vez más, recordemos que si bien la definición anterior engloba a una gran cantidad de estados posibles, no todos estos estados serán alcanzables a lo largo de una deliberación. El ejemplo introducido en la sección anterior resulta adecuado para ilustrar este punto: es posible que aquellos estados que denoten la continuación de un deliberación en la que ya se alcanzó el consenso acerca de que hacer con la tesis inicial nunca sean alcanzados por una deliberación en concreto.

Una vez introducido el concepto de estado de una deliberación surgen los mismos interrogantes que se consideraron una vez establecida la noción análoga en el marco del razonamiento dialéctico. Tanto el análisis como la solución adoptada en ese contexto puede ser adecuada al marco del razonamiento multilateral, es decir, se puede apelar a un conjunto de funciones auxiliares que resuelvan los planteos del tipo quién debe realizar la siguiente movida en un determinado estado de deliberación, o conocer si alguna de las partes ha logrado a esta altura imponerse en la deliberación. En este sentido, recordemos que la adopción de una política en particular no hace más que restringir la aplicabilidad del modelo, por lo que, en consecuencia, apelaremos a sendas funciones de carácter abstracto para atender estos planteos. Naturalmente, las funciones quedarán sin especificar en el modelo general, pero deben ser instanciadas de forma concreta toda vez que se aplique el modelo a un escenario en particular.

En la especificación de estas funciones, *States* representa en este contexto al conjunto de todos los posibles estados de deliberación:

- `toMove`, una función de *States* en $\mathcal{P}(Agents)$ que, a partir de un estado de deliberación acerca de una tesis, determina el conjunto de los jugadores que tienen permitido realizar el próximo movimiento.
- `winner`, una función de *States* en $\mathcal{P}(Agents) \cup \{\texttt{none}\}$ que, a partir de un estado de deliberación acerca de una tesis, determina que conjunto de contendientes han prevalecido en la misma. En el caso que ningún conjunto de contrincantes haya logrado imponerse, la función debe retornar la constante `none`.

En base a las consideraciones respecto a la consistencia del modelo motivadas por el ejemplo 4.4, sólo se impondrá la restricción en relación al comportamiento de estas funciones de que `toMove` y `legal` deben alcanzar la misma conclusión acerca de quiénes pueden realizar la próxima movida. Concretamente, cada uno de los contendientes que en virtud de la función `legal` estén en condiciones de hacer una cierta movida en el marco de una línea de argumentación, deben, a su vez, estar contemplados en el conjunto de contendientes que de acuerdo a la función `toMove` hubieran podido hacer la siguiente movida en el contexto de una deliberación que sólo haya explorado esa línea. Formalmente, la concordancia deseada se alcanza requiriendo que toda vez que $(agente, argumento) \in \texttt{legal}(línea)$, entonces necesariamente $agente \in \texttt{toMove}(\{línea\})$.

Siguiendo el espíritu del modelo para el razonamiento dialéctico, una función denominada `next` tendrá por objeto capturar la dinámica de una deliberación. Recordemos que las nociones hasta aquí introducidas sólo caracterizan los aspectos estáticos de la deliberación, esto es, su situación actual. Si *States* denota al conjunto de todos los posibles estados de una deliberación, la función `next`, definida de $\mathit{States} \times \mathit{Moves} \times \mathit{Lines}$ en *States*, determina cuál será el estado sucesor que resulte de hacer una determinada movida con el objeto de extender una línea de deliberación en el marco del estado actual de la deliberación. Nuevamente, su especificación completa será una de las tareas más importantes a la hora de instanciar el modelo abstracto, ya que la única restricción impuesta es su signatura.

Finalmente, es posible recopilar el conjunto de definiciones introducidas hasta este punto para dar a lugar al concepto de *protocolo deliberativo*:

Definición 4.29.
(protocolo deliberativo)
Sea $\mathcal{C}$ un contexto. Diremos que un *protocolo deliberativo* para $\mathcal{C}$, notado $\mathbb{PD}_{\mathcal{C}}$, se caracteriza a través de la tupla

$$(\mathit{Moves}, \mathit{Lines}, \mathit{States}, \mathtt{legal}, \mathtt{toMove}, \mathtt{winner}, \mathtt{next})$$

donde:

- *Moves* es el conjunto de movidas consideradas válidas (def. 4.26, pág. 224).
- *Lines* es el conjunto de líneas de deliberación consideradas válidas (def. 4.27, pág. 225).
- *States* es el conjunto de todos los posibles estados que una deliberación puede atravezar (def. 4.28, pág. 225).
- `legal` es una función definida de *Lines* en $\mathcal{P}(\mathit{Moves})$, que retorna el conjunto de jugadas que tienen permitido a extender una línea de deliberación.
- `toMove` es una función definida de *States* en $\mathcal{P}(\mathit{Agents})$, que retorna qué contrincantes pueden jugar a continuación.

- `winner` es una función definida de *States* en $\mathcal{P}(Agents) \cup \{\texttt{none}\}$, que retorna qué contendientes prevalecieron en una cierta deliberación, donde la constante `none` denota que todavía nadie se ha impuesto.

- `next` es una función definida de *States* $\times$ *Moves* $\times$ *Lines* en *States*, que captura el efecto de hacer un cierta movida para extender una determinada línea de deliberación en el marco de un estado de deliberación en particular.

Con el objeto de preservar la consistencia del protocolo deliberativo, se debe asegurar que toda vez que $(agente, argumento) \in \texttt{legal(}línea\texttt{)}$, entonces necesariamente $agente \in \texttt{toMove(}\{línea\}\texttt{)}$. ■

A manera de simplificación de la notación, cuando el contexto al cual hace referencia un protocolo deliberativo resulte evidente, notaremos al protocolo $\mathbb{PD}_{\mathcal{C}}$ simplemente como $\mathbb{PD}$. Finalmente, sobre la base provista por el concepto de protocolo deliberativo es posible caracterizar una semántica inspirada en aquella definida para el escenario dialéctico.

Definición 4.30.
(sustento)
Sea $\mathbb{PD}_{\mathcal{C}} = (Moves, Lines, States, \texttt{legal}, \texttt{toMove}, \texttt{winner}, \texttt{next})$ una instancia particular de un protocolo deliberativo, y sea $\mathcal{T}$ una tesis. Diremos que $\mathcal{T}$ está *sustentada* por $\mathbb{PD}_{\mathcal{C}}$ si, y sólo si, existe una secuencia finita $s_0, s_1, \ldots, s_n$ de estados de deliberación acerca de $\mathcal{T}$, tales que:

- $s_0 = \{\langle m_0 \rangle\}$, donde $\langle m_0 \rangle$ es una línea de deliberación acerca de $\mathcal{T}$,

- para cada i, $0 \leq i < n$, es posible identificar una línea de deliberación $\langle m_0, \ldots, m_k \rangle \in s_i$ en la cual para algún j, $0 \leq j \leq k$, existe una movida $(agente, argumento) \in \texttt{legal(}\langle m_0, \ldots, m_j \rangle\texttt{)}$, que verifique que:

    - $agente \in \texttt{toMove(}s_i\texttt{)}$, y además

    - $\texttt{next(}s_i\texttt{,}(agente, argumento)\texttt{,}\langle m_0, \ldots, m_j \rangle\texttt{)} = s_{i+1}$.

- para cada i, $0 \leq i < n$, $\texttt{winner(}s_i\texttt{)} = \texttt{none}$, y además
- $\texttt{winner(}s_n\texttt{)} \cap Agents_P \neq \emptyset$, pero a su vez $\texttt{winner(}s_n\texttt{)} \cap Agents_O = \emptyset$.

■

Al contrastar el modelo para el razonamiento de base dialéctica de la sección anterior con el modelo para la deliberación entre agentes recién introducido se puede apreciar que las diferencias más evidentes se circunscriben a la formalización del concepto de contexto. Esto es, la dinámica propia de los modelos resulta altamente análoga, lo que asegura que las decisiones de diseño tomadas para el modelo inicial, las cuales contaban con el respaldo del cuerpo de desarrollos en el ámbito de la argumentación dialéctica, constituyan decisiones de diseño acertadas en el contexto del nuevo modelo.

Por otra parte, resulta de particular interés enfatizar que la estrategia desarrollada con el objeto de obtener este modelo para la deliberación multiagente induce a su vez un modelo abstracto para argumentar de forma distribuída, en el cual cada agente tiene la libertad de aportar a la argumentación desde su propia posición lo que opina acerca de la tesis que motivó la discusión, o en relación a algún argumento previo esgrimido por otro agente. Más aún, en aquellos agentes dentro del mismo equipo surge de forma natural otro tipo de interacción ya comentado en el capítulo 3, conocido como *colaboración*, puesto que todos ellos contribuyen, en la medida de sus posibilidades individuales, hacia el objetivo en común, sea éste sustentar o atacar la tesis bajo consideración. Por caso, un cierto agente puede asistir a uno de sus compañeros al atacar un argumento que este último no era capaz de refutar por sí mismo.

Para elaborar un poco más acerca de las posibles aplicaciones que brinda el modelo propuesto debemos en primer lugar familiarizarnos con su dinámica a través de sus instanciaciones concretas.

4.5. Ejemplos de instanciaciones concretas

Esta sección tiene por objeto ilustrar las posibilidades que brinda el modelo desarrollado en las secciones anteriores. En este sentido, consideraremos cuáles son los protocolos de interacción que resulten de tomar como punto de partida a dos prominentes teorías para la argumentación rebatible. En primer lugar, en la sección 4.5.1 analizaremos qué protocolo resulta de tomar como punto de partida a la Programación en Lógica Rebatible, fomalismo argumentativo abordado en detalle en la sección 2.2.6. Por último, en la sección 4.5.2 discutiremos como modelar prioridades entre las reglas tomando como punto de partida al sistema propuesto por Prakken y Sartor, abordado minuciosamente en la sección 2.2.4.

4.5.1. Modelando la Programación en Lógica Rebatible

La Programación en Lógica Rebatible (DeLP) constituye uno de los formalismos de representación de conocimiento y razonamiento más versátiles entre los introducidos en la literatura. El ilustrar como capturar la dinámica de este formalismo en términos de un protocolo dialéctico, con el objeto último de obtener su protocolo deliberativo asociado, no sólo permite ejemplificar una de las principales aplicaciones del modelo propuesto, sino que también nos proveerá de un modelo de deliberación con atractivas propiedades. De hecho, el análisis del protocolo de interacción resultante constituye el tópico central del artículo [Stankevicius y Simari, 2000]. Recordemos que, en principio, para obtener el protocolo dialéctico inducido por la DeLP debemos describir formalmente en qué consiste su contexto, así como definir acabadamente cada uno de los aspectos que caracterizan al protocolo en sí.

El contexto de un protocolo dialéctico engloba tres componentes: el lenguaje de representación de conocimiento adoptado, la estructura interna de la base de conocimiento, y el mapeo que permite obtener el conjunto de todos los argumentos construibles a partir de esa base de conocimiento. En DeLP el conocimiento se codifica a través de reglas estrictas y reglas rebatibles, expresadas en un lenguaje similar al adoptado en la programación en lógica, cuya definición precisa puede consultarse en la sección 2.2.6. La base de conocimiento se representa a través de un programa lógico rebatible $\mathcal{P} = (\Pi, \Delta)$, donde Π es un conjunto de hechos y reglas estrictas y Δ es un conjunto de reglas rebatibles. Finalmente, $\texttt{args}(KB) = \texttt{args}(\mathcal{P})$, donde $\texttt{args}(\mathcal{P})$ denota al conjunto de todas las estructuras de argumento construíbles a partir de $\mathcal{P}$. Naturalmente, en este caso estamos haciendo referencia a la noción de estructura de argumento propia de DeLP (def. 2.88, pág. 106). Denotando como $\mathcal{L}$ al lenguaje de representación de conocimiento a partir del cual se contruyen las reglas estrictas y rebatibles en DeLP, $\mathcal{C} = (\mathcal{L}, \mathcal{P}, \texttt{args})$ es el contexto en el cual se desempeñará el protocolo dialéctico caracterizado a continuación.

Toda estructura de argumento válida puede constituir una potencial movida. Sin embargo, para poder establecer qué entendemos por una línea de argumentación en el contexto de este protocolo dialéctico debemos caracterizar formalmente el comportamiento de uno de sus principales componentes: la función `legal`. En esencia, debemos determinar cuáles son las estructuras de argumentos que constituyen una extensión plausible para una línea de argumentación. En principio, todo argumento que derrote al ultimo argumento expuesto en la línea de argumentación considerada debería ser tenido en cuenta. No obstante, no todos estos derrotadores constituyen una extensión aceptable, puesto que DeLP incorpora ciertas restricciones adicionales que permiten descartar las líneas de argumentación que conduzcan a conclusiones incorrectas, producto de haber conducido un razonamiento argumentativo de tipo falaz. Concretamente, no se acepta reintroducir una estructura de argumento que haya formado parte de otra estructura ya presente en esa línea de argumentación, ni incorporar argumentos que estén en conflicto con otros argumentos propuestos anteriormente por el contendiente que va a realizar la siguiente movida. Tampoco debemos olvidar la última condición en relación a la noción de línea de argumentación válida (def. 2.99, pág. 114), la cual límita la derrota por bloqueo sólo a aquellos escenarios en los que haya existido una derrota propia en el paso inmediato anterior.

En términos formales, para toda línea de argumentación $\langle m_0, \ldots, m_n \rangle$, el conjunto `legal(`$\langle m_0, \ldots, m_n \rangle$`)` contendrá a aquellas movidas (*contendiente*, $\langle \mathcal{A}, h \rangle$) que verifiquen las siguientes restricciones:

1. $\langle \mathcal{A}, h \rangle$ debe derrotar a la estructura de argumento $arg(m_n)$, si bien cuando $n > 0$, sólo se aceptará que $\langle \mathcal{A}, h \rangle$ derrote por bloqueo a $arg(m_n)$ si $arg(m_n)$ previamente derrotó de forma propia a $arg(m_{n-1})$.

2. Cuando *contendiente* $= P$, el conjunto $\{\langle \mathcal{A}, h \rangle\} \cup \{arg(m_i) \mid 0 \leq i \leq n,\ i = 2k\}$ debe resultar concordante. Caso contrario, cuando *contendiente* $= O$, el conjunto $\{\langle \mathcal{A}, h \rangle\} \cup \{arg(m_i) \mid 0 \leq i \leq n, i = 2k+1\}$ debe resultar concordante.

3. $\langle \mathcal{A}, h \rangle$ no aparece como subargumento de ninguna de las estructuras de argumento $arg(m_i)$, $0 \leq i \leq n$.

Contando con una definición acabada de los conceptos de movida y línea de argumentación queda especificado a su vez el concepto de estado de una disputa y, por consiguiente, los conjuntos *Moves*, *Lines* y *States*.

El próximo componente del protocolo dialéctico que debe ser caracterizado es la función `toMove`, la cual determina cuál de los participantes debe realizar la siguiente movida. En el contexto de DeLP, esta función presenta un comportamiento altamente predecible, puesto que manifiesta una alternancia estricta (*i.e.*, no es posible que el mismo contendiente realice dos movidas consecutivas). Si bien existen diversas formas de alcanzar este comportamiento, en esta ocasión apelaremos al árbol de dialéctica unívocamente asociado a todo estado de disputa. En particular, el árbol de dialéctica asociado a un conjunto *estado* de líneas de argumentación puede ser obtenido a través de la siguiente caracterización recursiva: *estado*

1. Si *estado* $= \emptyset$, el árbol dialéctico asociado es el árbol nulo.

2. Si *línea* es una de las líneas de argumentación contenidas en *estado*, el árbol dialéctico asociado se construye actualizando al árbol dialéctico asociado a *estado* $- \{$*línea*$\}$ de forma tal que, si *línea* $= \langle m_0, \ldots, m_n \rangle$, entonces el nodo

etiquetado $arg(m_i)$ cuenta con el nodo etiquetado $arg(m_{i+1})$ como uno de sus sucesores, para todo $0 \leq i < n$.

En base a este concepto, es posible caracterizar el comportamiento de la función `toMove`. Concretamente, si *estado* es un conjunto arbitrario de líneas de argumentación denotando el estado actal de una disputa, `toMove(`*estado*`)` debe retornar P si el árbol dialéctico asociado a *estado*, obtenido mediante el procedimiento recursivo recién introducido, es tal que su raíz resulta etiquetada "D" al aplicar el marcado correspondiente (def. 2.101, pág. 117), retornándose O en caso contrario, es decir, cuando en este mismo árbol su nodo raíz resulte etiquetado "U".

Con respecto a la consistencia que debe existir entre lo retornado por las funciones `legal` y `toMove`, considerando que el comportamiento de estas funciones se deriva a partir de un sistema argumentativo en funcionamiento, el requisito de concorancia impuesto para el protocolo deliberativo se satisface de forma directa. En otras palabras, toda vez que $(\mathit{contendiente}, \langle \mathcal{A}, h \rangle) \in$ `legal(`*línea*`)`, entonces necesariamente `toMove({`*línea*`})` $= \mathit{contendiente}$.

El próximo componente que debe ser definido en este protocolo dialéctico es la función `winner`, la cual tiene por objeto determinar a partir de un cierto estado de disputa cuál de los contendientes se ha impuesto a lo largo del debate. Uno de los aspectos más díficiles de capturar de esta función es su capacidad de determinar si el diálogo actual ha alcanzado su terminación, ya que mientras un debate no se haya agotado, esta función debe retornar consistentemente la constante `none`. Para lograr este comportamiento apelaremos nuevamente al concepto de árbol de dialectica asociado a un conjunto *estado* de líneas de argumentación. En este contexto, la función `winner` puede caracterizarse formalmente de la siguiente manera:

- Toda vez que exista alguna línea de argumentación $\mathit{línea} \in \mathit{estado}$, para la cual alguna de las movidas $\mathit{movida} \in$ `legal(`*línea*`)` sea tal que verifica que `toMove(`*estado*`)` $= \mathit{player}(\mathit{movida})$, y, además, esa movida resulte relevante en el debate actual, esto es, `next(`*estado*`,`*movida*`,`*línea*`)` $\neq \mathit{estado}$, entonces se retornará `none`.

l_1	l_2	l_3
$m_0^1 = (P,\mathcal{A})$	$m_0^2 = (P,\mathcal{A})$	$m_0^3 = (P,\mathcal{A})$
$m_1^1 = (O,\mathcal{C})$	$m_1^2 = (O,\mathcal{B})$	$m_1^3 = (O,\mathcal{B})$
	$m_2^2 = (P,\mathcal{D})$	$m_2^3 = (P,\mathcal{E})$

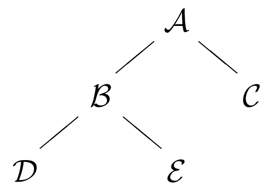


Figura 4.4: Arbol dialéctico asociado a un estado de disputa

- En caso contrario, cuando no queden movidas relevantes por ser jugadas, se retornará P si el árbol dialéctico asociado a *estado*, obtenido mediante el procedimiento recursivo indicado anteriormente, es tal que su raíz resulta etiquetada como "U" al aplicar el marcado correspondiente, retornándose O cuando en este mismo árbol su nodo raíz resulte etiquetado como "D".

La primer condición detecta si aún no se ha alcanzado la terminación natural del debate, en cuyo caso se debe retornar consistentemente `none`. La terminación natural de la disputa habrá sido alcanzada toda vez que no sea posible encontrar nuevo argumentos que puedan extender alguna de las líneas de argumentación exploradas hasta ese momento. Esta situación es capturada mediante la propia función `next`, definida formalmente más adelante. Cuando esta primer condición no se satisface, la segunda condición apela nuevamente al marcado del árbol de dialéctica asociado para determinar quién ha logrado prevalecer a lo largo de la disputa, la cual tenemos la certeza ya ha culminado.

Finalmente, para poder instanciar completamente al protocolo dialéctico en consideración se debe caracterizar formalmente la función `next`, el componente del modelo que tiene por objeto capturar la dinámica del formalismo analizado. Este componente, si bien desempeña un rol crucial, cuenta con una definición accesible, producto de una observación acabada de la evolución del análisis dialéctico que se lleva adelante en el marco de DeLP. Formalmente, suponiendo que *estado* sea un conjunto arbitrario de líneas de argumentación, $movida = (contendiente, argumento)$ una cierta movida que se desea poner en juego para extender a la línea de argumentación $línea = \langle m_1, \ldots, m_n \rangle$, entonces `next(`*estado*`,`*movida*`,`*línea*`)` debe retornar el siguiente conjunto de líneas de argumentación:

- Si la línea de argumentación $\langle m_1, \ldots, m_n, movida \rangle$ figura como prefijo de alguna de las líneas ya contenidas en *estado*, se debe retornar el mismo conjunto *estado*.

- En caso contrario, se debe retornar el siguiente conjunto de líneas de argumentación:
$$(estado - \{línea\}) \cup \{\langle m_1, \ldots, m_n, movida \rangle\}$$

La primer condición en este caso detecta si el efecto del argumento que se está a punto de considerar ya ha sido tenido en cuenta, en cuyo caso el estado actual no se modifica. De no ser así, la segunda condición refleja el impacto que ese nuevo argumento tiene sobre el conjunto de líneas de argumentación que venían siendo consideradas: si se está atacando el último argumento de la línea de argumentación, la línea de argumentación extendida con el nuevo argumento debe reemplazar a la anterior, pero si se está atacando un argumento que no aparece en la última posición, la línea anterior debe ser conservada, pues el nuevo argumento puesto en juego denota la exploración de una línea de argumentación colateral.

Finalmente, el protocolo dialéctico que describe el comportamiento de DeLP es el que se obtiene al instanciar cada uno de sus componentes de la manera indicada. La semántica asociada a este protocolo dialéctico es la inducida por el conjunto de las tesis sustentadas por aquellas estructuras de argumento que el proponente logre establecer respetando el protocolo. En este caso, la noción de tesis sustentada se corresponde a la de literal garantizado, propia de la programación en lógica rebatible.

Una vez llevada adelante la instaciación de todos los componentes del protocolo dialéctico, su posterior reinterpretación en términos de un marco de agentes múltiples resulta esencialmente directa, ya que será posible extrapolar la especificación de la mayoría de los componentes ya introducidos sin mayores modificaciones. En principio, la diferencia más evidente se focaliza en la definición del contexto, donde si bien hace uso del mismo lenguaje de representación de conocimiento, ahora abarca un mayor número de contendientes, organizados en dos grupos: por un lado $Agents_P = \{P_1, \ldots, P_n\}$, el conjunto de aquellos a favor de la tesis a ser debatida, y por otra parte $Agents_O = \{O_1, \ldots, O_m\}$, el conjunto de los que están en su contra.

Como lo sugieren los aspectos que conforman el contexto de un protocolo deliberativo, en vez de contar con una base de conocimiento única, KBs se compone de tantas bases de conocimiento como contendientes participen en la disputa, siendo cada una de ellas el programa lógico rebatible que modela el estado epistémico de los agentes. De forma análoga, el mapeo `args` retorna para cada uno de los agentes Ag que toman parte en la deliberación, al conjunto $\mathtt{args}(KB_{Ag})$ de todas las estructuras de argumento construíbles a partir de KB_{Ag}. Considerando que este lenguaje de representación de conocimiento permite expresar información estricta, se debe imponer la restricción adicional de que la suma de la información estricta de todos los agentes intervinientes resulte consistente (de no serlo, la interacción carecerá de una base en común a partir de la cual deliberar).

Las nociones de movida y de línea de deliberación para el contexto multiagente son idénticas a las ya consideradas, si bien la función `legal` debe ser adecuada al nuevo escenario: toda referencia a P, denotará ahora a alguno de los contendientes en $Agents_P$, y toda referencia a O, a alguno de los contendientes en $Agents_O$. Debemos recordar que la definición formal del concepto de movida nos obliga a construir la estructura de argumento que está siendo puesta en juego apelando sólo a la información contenida en la base de conocimiento del contendiente que poniendo en juego esa movida. Sin duda, es factible extender el protocolo obtenido en base a la estrategia aquí sugerida con el objeto de incorporar toda otra característica adicional que se desee. Por caso, se puede explorar la posibilidad de permitir que la tarea de proponer nuevas estructuras de argumento se lleve a cabo de forma colaborativa, apelando al conocimiento de múltiples agentes. Por último, recordemos que el concepto de estado de una deliberación queda a su vez completamente caracterizado una vez especificado qué se entiende por línea de deliberación.

En relación al siguiente componente, debemos señalar que la interacción gobernada por el protocolo que estamos describiendo presentará la misma característica que su contraparte dialéctica: la alternancia estricta en la realización de las movidas. En este sentido, la función `toMove` se define de una manera análoga a la anterior, con la salvedad de que cuando antes retornaba P, ahora debe retornar el conjunto $Agents_P$, y, caso contrario, cuando antes retornaba O, debe ahora retornar el conjunto $Agents_O$. En lo que respecta a la función `winner`, su comportamiento es similar al manifestado por la función introducida en el escenario dialéctico. La diferencia más evidente es que cada vez que se retornaba P, ahora se debe retornar $Agents_P$, y que cuando antes se retornaba O, ahora se debe retornar $Agents_O$. Finalmente, el comportamiento de la función `next` no necesita ser actualizado, pues su dinámica se encuentra definida a partir de otras nociones ya adecuadas al marco multiagente.

El protocolo deliberativo que resulta de instanciar los distintos componentes de la forma indicada, al igual que su semántica asociada, no resultan directamente implementables, ya que existen varias tareas de vital importancia que deben ser llevadas a cabo antes y después de conducir la interacción propiamente dicha. Por ejemplo, en base al marco de interacción propuesto en [Stankevicius y Simari, 2000], se deben llevar adelante por lo menos las siguientes tareas:

1. Un cierto agente decide que necesita deliberar con otro agente, en relación a un determinado tema.

2. El agente propone al contrincante elegido el debatir sobre ese tema. En caso de que la propuesta sea aceptada, la deliberación se encuentra lista para comenzar, pero, en caso contrario, el agente debe proponer un nuevo asunto,

elegir otro contrincante, o directamente abandonar el deseo de llevar a cabo la deliberación.

3. Se desarrolla la deliberación propiamente dicha.

4. El resultado de la deliberación es procesado según corresponda.

En este contexto, cabe enfatizar que el protocolo deliberativo introducido tiene por objeto dictar las reglas que gobernarán la tercer etapa.

4.5.2. Modelando prioridades entre las reglas

El sistema argumentativo propuesto por G. Sartor y H. Prakken es una de las pocas teorías que han incorporado un tratamiento formal de la comparación de la fuerza relativa entre las reglas empleadas para representar el conocimiento por defecto. El formalismo resultante presenta un comportamiento único, donde es factible incluso cuestionar dinámicamente el propio ordenamiento entre las reglas. En lo que resta de esta sección, ilustraremos cómo modelar este comportamiento en el seno de un protocolo dialéctico.

La tarea por delante es similar a la realizada en la sección anterior, esto es, caracterizar de forma acabada tanto al contexto como a cada uno de los componentes que integran el protocolo dialéctico, para luego, reinterpretación mediante, derivar el protocolo deliberativo asociado a este formalismo argumentativo. A los efectos de simplificar la presentación, consideraremos la primer versión del sistema, aquella en la que las prioridades entre las reglas son fijas, si bien apelando a una estrategia similar se puede obtener una instanciación afín a su segunda variante.

En primer lugar, debemos caracterizar en qué consiste el contexto del protocolo dialéctico en este escenario. El lenguaje de representación de conocimiento es el adoptado por los autores para formular las reglas con las cuales se codifica el conocimiento, esto es, el lenguaje de la programación en lógica extendida complementado mediante una relación de orden estricto parcial. Por consiguiente, la base de conocimiento KB esta compuesta por la tupla $(S, D, <)$, donde S y D son conjuntos de reglas, ninguna regla de S contiene literales default, y '$<$' es un orden estricto parcial sobre los elementos de D. Finalmente, el último componente que caracteriza el contexto del protocolo dialéctico es el mapeo `args`. En este caso, `args(`KB`)` denotará al conjunto de todos los posibles argumentos que se puedan construir a partir de KB. Naturalmente, en esta oportunidad estamos haciendo referencia al concepto de argumento correspondiente al sistema Prator (def. 2.38, pág. 65).

Para comenzar a caracterizar el protocolo dialéctico debemos capturar la noción que suele describir la mayor parte del funcionamiento del sistema que está siendo modelado, esto es, la función `legal`. Para hacerlo, debemos determinar cuál es el conjunto de argumentos que constituyen una extensión razonable para una cierta línea de argumentación. Si bien todo argumento que derrote estrictamente o no a al último argumentos formulado representa una réplica plausible, debemos tener en cuenta las restricciones adicionales que los autores estipularon a la par del concepto de diálogo (def. 2.49, pág. 73). En otras palabras, no se aceptarán aquellas movidas que reintroduzcan un argumento ya presente en la línea de argumentación en consideración, y, además, se requerirá que P siempre derrote a los argumentos de O de forma estricta, empleando argumentos minimales.

Formalmente, para una línea de argumentación arbitraria $\langle m_0, \ldots, m_n \rangle$, la función `legal` retorna el conjunto de todas las movidas (*contendiente*, A) que verifiquen el siguiente conjunto de restricciones:

1. A derrota al argumento $arg(m_n)$.

2. Si $n = 2k + 1$, entonces *contendiente* debe ser P y A debe ser es el menor argumento (en términos de inclusión de conjuntos[3]) que derrota estrictamente a $arg(m_n)$. En caso contrario, cuando $n = 2k$, *contendiente* debe ser O.

3. A difiere de todos los argumentos previamente expuestos por *contendiente* a lo largo de la línea de argumentación $\langle m_0, \ldots, m_n \rangle$. En otras palabras, cuando

[3]un argumento es este contexto es una secuencia de reglas y no un conjunto. Para poder respetar el espíritu de esta restricción debemos aplicar la minimalidad al conjunto formado por aquellas reglas presentes en la secuencia.

contendiente $= P$, se debe cumplir que $A \notin \{arg(m_i) \mid 0 \leq i \leq n, i = 2k+1\}$. De manera análoga, cuando *contendiente* $= O$, entonces se debe verificar que $A \notin \{arg(m_i) \mid 0 \leq i \leq n, i = 2k\}$.

Una vez especificada esta función, los conceptos de movida y de línea de argumentación quedan completamente definidos, ya que sólo dependen de la combinación de la función recién caracterizada con el concepto de contexto. Por consiguiente, tanto el concepto de estado de una disputa como los conjuntos *Moves*, *Lines* y *States* quedan totalmente especificados.

Para el siguiente componente, la función `toMove`, apelaremos a la misma propiedad explotada en protocolo dialéctico introducido en la sección anterior: la estricta alternancia que debe ser observada al introducir cada nuevo argumento. A tal efecto, haremos uso de la noción de árbol de diálogo asociado a un conjunto de líneas de argumentación. Formalmente, para todo el un conjunto *estado* de líneas de argumentación, representando el estado de una disputa, el árbol de diálogo asociado se caracteriza recursivamente de la siguiente manera:

1. Si *estado* $= \emptyset$, el árbol de diálogos asociado será el árbol nulo.

2. Si *línea* es una de las líneas de argumentación contenidas en *estado*, el árbol de diálogos asociado se construye actualizando al árbol de diálogos asociado a *estado* $-$ $\{$*línea*$\}$ de forma tal que, si *línea* $= \langle m_0, \ldots, m_n \rangle$, entonces el nodo conteniendo al argumento $arg(m_i)$ figura como padre del nodo que contiene a $arg(m_{i+1})$, para todo $0 \leq i < n$.

En base a esta noción, es posible caracterizar de forma acabada al comportamiento de la función `toMove`. Concretamente, si *estado* es un conjunto arbitrario de líneas de argumentación denotando el estado actal de una disputa, `toMove(`*estado*`)` debe retornar P u O en función de quién se haya impuesto en el árbol de diálogos asociado a *estado*, obtenido de acuerdo al procedimiento recursivo anterior. La función así obtenida, en conjunción con la primer restricción en la especificación de la función `legal`, asegura la concordancia requerida por los protocolos dialécticos entre estas funciones, ya que cuando $(contendiente, A) \in$ `legal(`*línea*`)`, entonces necesariamente `toMove({`*línea*`})` $=$ *contendiente*.

Para especificar el próximo componente del protocolo dialéctico debemos atender dos cuestiones: bajo qué condiciones una disputa alcanza su terminación, y cuál de los contendientes se impuso en ésta una vez finalizada. Prakken y Sartor señalan un claro criterio de terminación: un debate finaliza cuando se agotan las movidas a disposición del contendiente que tenga que realizar la próxima movida. Para contestar al segundo interrogante, apelaremos nuevamente al criterio de adjudicación utilizado en la propia teoría de prueba dialéctica del sistema Prator. Formalmente, el resultado de aplicar la función `winner` a un conjunto *estado* de líneas de argumentación, representando el estado actual de una disputa, se obtiene siguiendo el siguiente procedimiento:

- Toda vez que exista alguna línea de argumentación *línea* $\in$ *estado*, para la cual alguna de las movidas *movida* $\in$ `legal(`*línea*`)` sea tal que verifica que `toMove(`*estado*`)` $=$ *player*(*movida*), y, además, esa movida resulte relevante en el debate actual, esto es, `next(`*estado*`,`*movida*`,`*línea*`)` $\neq$ *estado*, entonces se retornará `none`.

- En caso contrario, cuando no queden movidas relevantes por ser jugadas, se retornará P u O, en función de quién se haya impuesto en el árbol de diálogos asociado a *estado*, el cual se obtiene mediante el procedimiento recursivo antes indicado.

El componente final que completa la instanciación del protocolo dialéctico es la función `next`. Emplearemos una estrategia análoga a la introducida en la instanciación de la sección anterior, ya que la dinámica de ambos sistemas es similar en lo que a la acumulación de argumentos formulados respecta. Siendo *estado* un conjunto de líneas de argumentación, y *movida* $= (contendiente, A)$ una movida que se desea poner en juego para extender a la líneade argumentación *línea* $= \langle m_0, \ldots, m_n \rangle$, entonces `next(`*estado*`,`*movida*`,`*línea*`)` debe retornar el siguiente conjunto de líneas de argumentación:

- Si la línea de argumentación $\langle m_0, \ldots, m_n, movida\rangle$ aparece como prefijo de alguna otra línea ya presente en *estado*, se retorna el mismo conjunto *estado*.
- En caso contrario, se retorna el conjunto compuesto de las siguientes líneas de argumentación:

$$(estado - \{línea\}) \cup \{\langle m_0, \ldots, m_n, movida\rangle\}$$

Contando con una instanciación acabada tanto de los componentes del protocolo dialéctico como de los de su contexto, la semántica asociada, aquella inducida por el concepto de tesis sustentada, surge de forma directa y natural. En esta ocasión, el conjunto de las tesis sancionadas por la semántica coincide con la noción de literal justificado, propia del sistema Prator.

En lo que resta de esta sección, repasaremos los principales aspectos a tener en cuenta para reinterpretar el protocolo dialéctico hasta aquí obtenido como si se tratara de un protocolo deliberativo. Nuevamente, la principal diferencia se encuentra en la formalización del contexto, ya que si bien se adopta el mismo lenguaje de representación de conocimiento que antes, ahora la existencia de múltiples agentes obliga a contemplar un mayor número de contendientes. Como es usual, los agentes interesados en tomar parte en la deliberación se encuentra organizados en dos grupos: por un lado los que están a favor de la tesis por ser debatida $Agents_P = \{P_1, \ldots, P_n\}$, y por otro lado los que están en su contra $Agents_O = \{O_1, \ldots, O_n\}$. Cada uno de estos agentes tiene a su disposición una base de conocimiento codificada al igual que antes como una tupla $(S, D, <)$, donde S y D son conjuntos de reglas, ninguna regla de S contiene literales default, y '$<$' es un orden estricto parcial sobre los elementos de D. A su vez, el conjunto KBs contendrá tantas bases de conocimiento como agentes intervengan en la deliberación. Cada agente empleará su propia base de conocimiento para construir argumentos, los cuales serán utilizados a lo largo de la deliberación para tratar de solidificar la posición de su grupo y al mismo tiempo atacar las razones expuestas por sus contrincantes. Una vez más, como el lenguaje de la programación en lógica extendidida permite expresar información estricta, debemos asegurar que la suma de la información estricta almacenada en las distintas bases de conocimiento resulte consistente. Finalmente, en relación al mapeo `args`, se puede adoptar el mismo mapeo utilizado en para el protocolo dialéctico, pues tanto la política de representación de conocimiento como el mecanismo de contrucción de argumentos no han cambiado.

Adoptaremos un orden de presentación similar al empleado para el protocolo dialéctico recién introducido, comenzado por la caracterización de la función `legal`. Esta función puede ser adaptada sencillamente al nuevo escenario, interpretando que toda referencia a P en realidad denota a alguno de los contendientes en $Agents_P$, y que toda referencia a O en realidad denota a alguno de los contendientes en $Agents_O$. Por consiguiente, las nociones de movida y de línea de deliberación no requieren mayor elaboración, quedando adaptadas al nuevo entorno pues sólo dependen de conceptos ya reinterpretados. Una situación análoga sucede con el concepto de estado de deliberación, el cual sólo depende de la noción de línea de deliberación.

El protocolo deliberativo que estamos introduciendo presentará las mismas características que el protocolo dialéctico de partida. Entre otras cosas, la interacción resultante manifiesta el mismo patrón de alternancia estricta entre los contendientes. En consecuencia, la función `toMove` se define de una manera análoga, con la salvedad de que cuando antes retornaba P, ahora debe retornar al conjunto $Agents_P$, y, en caso contrario, cuando antes retornaba O, ahora debe retornar al conjunto $Agents_O$. En cuanto al criterio de adjudicación inducido por la función `winner`, también presenta un comportamiento similar al anterior, si bien la principal diferencia consiste en retornar a $Agents_P$ cuando antes retornaba P, y de retornar a $Agents_O$ cuando antes retornaba O. Por último, la función `next` se define de manera idéntica a la propuesta en el marco del protocolo dialéctico, ya que, una vez más, su comportamiento se encuentra caracterizado en términos de nociones ya extrapoladas al marco de agentes múltiples.

Resulta interesante contrastar el protocolo deliberativo resultante con aquel obtenido en la sección anterior, ya que a pesar de las evidentes diferencias a nivel de las teorías de argumentación rebatible tomadas como punto de partida, los modelos de deliberación resultantes presentaron varios aspectos en común (por caso, el comportamiento de las funciones `toMove` o `next`). No obstante, al contrastar la función `legal` introducida en cada escenario se puede apreciar que los protocolos obtenidos no presentan un comportamiento exáctamente igual, producto de que es esta función justamente la que embebe las peculiaridades de los sistemas modelados.

4.6. Conclusiones

Motivados por la necesidad de contar con un modelo formal que permita caracterizar los distintos tipos de interacción entre agentes repasados en el capítulo 3, se diseñó en primer lugar una estrategia que permite derivar un modelo para la interacción entre agentes a partir de otro para el razonamiento de base dialéctica. En esencia, la estrategia propuesta descansa sobre una apreciación anterior del filósofo N. Rescher: el aparente isomorfismo entre el razonamiento dialéctico unilateral y aquel de carácter multilateral. En este sentido, se analizaron en detalle las propuestas presentes en la literaturas que en líneas generales hayan adoptado un curso de acción compatible con el antes expuesto, con la intención de cotejar tanto los resultados obtenidos como los inconvenientes encontrados.

Luego de considerar las virtudes y las falencias de estas propuestas, se desarrolló el concepto de protocolo diaéctico, el cual, a partir de factorizar las principales características de los sistemas argumentativos considerados a lo largo del capítulo 2, resultó adecuado para modelar una amplia gama de razonamientos de base dialéctica. El concepto de protocolo dialéctico fue posteriormente tomado como punto de partida para desarrollar un segundo modelo, el cual, si bien estructuralmente análogo al primero, tenía por objeto capturar uno de los principales patrones de interacción entre agentes: *la deliberación.* El modelo obtenido aplicando la estrategia propuesta, sintentizado en el concepto de protocolo deliberativo, heredó las principales características del modelo inicial, entre las que se destaca el hecho de estar especificado de forma abstracta. En particular, uno de los aspectos más importantes a la hora de instanciar el modelo por este medio obtenido consiste en especificar de forma acabada cada uno de sus componentes.

Por último, con el objeto de ilustrar el proceso de instanciación, se recorrieron paso a paso las etapas requeridas para poder modelar a dos de las principales teorías de argumentación rebatible, analizando, por otra parte, los respectivos protocolos deliberativos resultantes de tomar como punto de partida a cada una de estas teorías. En primer lugar, en la sección 4.5.1, se consideró cómo capturar la dinámica de la programación en lógica rebatible en términos de un protocolo dialéctico, analizando posteriormente su reinterpretación como si fuera un protocolo deliberativo. Posteriormente, se contrastó cómo modelar un sistema radicalmente distinto, que admita prioridades entre las propias reglas empleadas para construir los argumentos. A tal efecto, en la sección 4.5.2 se ilustró cómo capturar el comportamiento del sistema argumentativo ideado por H. Prakken y G. Sartor, abordando también la reinterpretación del protocolo dialéctico resultante como si fuera un protocolo deliberativo.

El modelo obtenido presenta una gran versatilidad, evidenciada principalmente en la facilidad con la cual se puede adecuar a teorías tan dispares como las consideradas anteriormente. De hecho, de la experimentación con este modelo abstracto así como con sus instanciaciones concretas surgió el interrogante de determinar si es posible relacionar el comportamiento de sus diversos componentes con las propiedades manifestadas por sus instanciaciones. Por caso, qué condiciones deben imponerse sobre el comportamiento de la función `next` de forma tal que el protocolo de deliberación resultante sea de carácter progresivo, o bien determinar qućondiciones permiten asegurar que todos los proponentes mantienen una cierta coherencia a nivel de las creencias sancionadas por sus bases de conocimiento. Este vínculo, de existir, permitiría establecer propiedades a nivel de familias de protocolos, en vez de tener que reiterar este análisis caso por caso.

Capítulo 5

Conclusiones y Resultados Obtenidos

Este capítulo tiene por objeto recopilar las principales conclusiones elaboradas a lo largo de los estudios reseñados en la presente tesis, sintetizando a su vez los resultados más importantes.

En relación al acabado análisis de los diversos sistemas argumentativos desarrollado en la sección 2.2, se identificaron aquellos aspectos comunes a la mayoría de los sistemas argumentativos, aislando, a su vez, las características particulares que diferencian a cada uno de estos acercamientos de los restantes. Ciertamente, todos los sistemas argumentativos considerados compartes las generalidades comentadas en la sección 2.1.3, puesto que todos apelan a alguna forma de lógica subyacente en términos de la cual caracterizan al concepto de argumento, donde algunos de estos argumentos entrarán en conflicto entre sí, relación usualmente simétrica, la cual a través de la noción de derrota permite determinar la direccionalidad del ataque implícito en el conflicto. A su vez, todos estos formalismos basan su semántica en el estado final de los argumentos construíbles, concepto intrínsecamente ligado al de derrota, puesto que, en general, la semántica del sistema se asocia al conjunto de argumentos que presentan como característica no tener derrotadores, o bien cuyos derrotadores estén a su vez derrotados.

Otro aspecto sobre el cual se aprecia un evidente consenso entre las diversas propuestas es en la metodología de representación de conocimiento por defecto, donde usualmente se adopta alguna forma de regla rebatible. Esta característica—la representación de conocimiento por defecto apelando a construcciones que asemejan una implicación material con menor fuerza conclusiva—de por sí constituye un rico tópico de investigación. No obstante, en esta ocasión nos limitaremos simplemente a identificar la solución adoptada por la mayoría de los sistemas argumentativos sin ahondar en las justificaciones filosóficas o pragmáticas que sustentan tal elección.

Por último, uno de los patrones recurrentes más importantes en las teorías que modelan la argumentación rebatible, característica crucial sobre la cual descansa gran parte de los desarrollos presentados en esta tesis, es que la mayoría de las propuestas consideradas admiten reformular sus semánticas mediante alguna forma de caracterización dialéctica. La existencia de este aspecto común a la mayoría de los sistemas argumentativos estudiados fue adelantada en la sección 2.1.2, al considerar los argumentos expuestos en ese sentido por el filósofo N. Rescher en su estudio del rol de dialéctica en la propia investigación científica.

En relación a las peculiaridades distintivas de cada acercamiento, debe resaltarse una vez más el profundo impacto que tuvo el sistema argumentativo introducido por J. Pollock (analizado en la sección 2.2.2), influenciando de forma marcada a los diversos desarrollos posteriores del área de la argumentación rebatible. Pollock, a través de su arquitectura Oscar, enfatizó la importancia de mantener acotada la distancia entre la teoría y su respectiva implementación computacional, principio altamente encomiable. En este sentido, propone un ciclo de evolución donde la experimentación con la implementación de una teoría permite depurar sus falencias, posibilitando la introducción de nueva teoría refinada, la cual, a su vez, dará a lugar a una nueva implementación que permita continuar así sucesivamente. Coincidimos plenamente con la posición del autor acerca de que si la distancia entre un desarrollo teórico y su respectiva implementación resulta excesiva, la propuesta en cuestión carecerá de sentido práctico.

Por su parte, el sistema argumentativo propuesto por H. Prakken y G. Sartor (abordado en la sección 2.2.4), se caracteriza por ser actualmente la única propuesta que incorpora la posibilidad de razonar rebatiblemente acerca de la relación de comparación entre argumentos a ser aplicada en la obtención de las conclusiones sancionadas por una base de conocimiento. Ciertamente, se debe reconocer que este aspecto se encuentra presente en las disputas entre humanos. No obstante, las restantes propuestas existentes en la literatura han optado por ignorar esta faceta, posiblemente debido a la complejidad inherente al proceso de argumentar de forma rebatible.

Una característica relevante pero a la vez largamente ignorada en el marco de las teorías argumentativas, es el adecuado tratamiento de la denominada argumentación falaz. La reformulación dialéctica del sistema Simari-Loui (considerada en la sección 2.2.5), consituye el primer sistema argumentativo en incorporar los controles necesarios para que este tipo no deseado de argumentación, bajo la cual es posible obtener conclusiones incorrectas no debidamente sustentadas en la base de conocimiento en curso, no termine perjudicando a la semántica del sistema en su conjunto. Por otra parte, se debe destacar otra característica que distingue esta reformulación dialéctica: el haber reconocido tempranamente las virtudes que devienen de adoptar una caracterización dialéctica para describir la semántica de un sistema argumentativo. Si bien hoy en día la mayoría de los restantes acercamientos cuentan con alguna variante dialéctica de sus semánticas, la reformulación dialéctica del sistema Simari-Loui puede ser sindicada como la pionera en adoptar esta postura en lo que a sistemas argumentativos respecta.

Finalmente en relación al paradigma de la Programación en Lógica Rebatible (repasado en la sección 2.2.6), se destacó su loable objetivo de diseño de reconciliar las ideas desarrolladas en el marco del sistema Simari-Loui con el principio postulado por J. Pollock, obteniéndose así un sistema argumentativo que presenta un adecuado balance entre poder expresivo e implementabilidad. En este sentido, por un lado incorpora el poder expresivo del sistema Simari-Loui, tratamiento de la argumentación falaz incluído, y por otro lado, hereda el razonamiento guiado por la metas de la programación en lógica, el cual admite implementaciones más eficientes que aquellos que no lo sean. Para lograr esta fina mixtura, las definiciones del sistema fueron cuidadosamente adecuadas de forma tal de preservar el citado carácter de ser guiado por las metas. Si bien hoy en día no existe una debida batería casos de prueba como para poder establecer fehacientemente un ordenamiento de los distintos sistemas argumentativos en función de su desempeño, el hecho de que los restantes acercamientos no hayan tenido en cuenta este aspecto evidencia que sus respectivas implementaciones deberán incurrir en costos computacionales adicionales.

Con respecto al compendio de propuestas alternativas ensayadas en el ámbito de los sistemas multiagente, reseñadas a lo largo del capítulo 3, la primer y más evidente conclusión es que la diversidad de dominios de aplicación en los cuales se ha experimentado aplicar este tipo de sistema excede a la imaginación. Si bien es casi imposible encontrar un conjunto de características recurrentes a la totalidad de las propuestas, usualmente los diversos sistemas multiagente suelen presentar como aspectos en común que sus agentes cuentan con conocimiento imperfecto y capacidades limitadas, donde a su vez el control del sistema suele estar distribuido y la información disponible no es centralizable. Por otra parte, otro aspecto recurrente es que la computación en el seno del sistema multiagente suele ser llevada adelante de forma asíncrona.

Del análisis de las diversas propuestas consideradas en la sección 3.3, se puede concluir que la mayoría de los sistemas parecen focalizar su atención en torno a aspectos un tanto *ad-hoc*. Postulamos que la falta de consenso en este sentido se debe principalmente a la combinación de dos factores, los que intervienen simultáneamente en estos tipos de sistema. Por un lado, el primer factor que contribuye a esta falta de uniformidad es la heterogeneidad de los dominios de aplicación en los que se han ensayado soluciones basadas en sistemas multiagente. Por otro lado, el segundo factor es la propia dificultad de los desafíos surgidos al migrar de un ambiente en el cual tanto el conocimiento como la toma de decisiones se encuentra centralizada a uno en donde múltiples agentes deben sincronizar su comportamiento para poder alcanzar sus objetivos. La conclusión anterior, lejos de ser una mera crítica, constituye un franco reconocimiento al tezón de los pioneros del área, quienes eligieron seguir adelante aún teniendo que afrontar tales contrariedades.

Una de las pocas excepciones a esta regla lo constituye la manifiesta preponderancia que la interacción entre los agentes desempeña en todo sistema multiagente, incluso independientemente de la citada diversidad. Sin duda, un sistema multiagente en el cual la interacción entre sus miembros sea muy baja o nula se asemeja más bien una mera colección de sistemas centralizados. De hecho, los diversos paradigmas de comunicación entre agentes reseñados en la sección 3.6 no hacen más que evidenciar el incipiente interés en la exploración de los principios y nociones involucrados en la interacción.

En consecuencia, motivados por la necesidad de contar con un modelo formal que permita caracterizar los distintos tipos de interacción entre agentes considerados, se diseño en primer lugar una estrategia que permite derivar un modelo para la interacción entre agentes a partir de otro para el razonamiento de base dialéctica. En esencia, la estrategia propuesta descansa sobre una apreciación anterior del filósofo N. Rescher: el aparente isomorfismo entre el razonamiento dialéctico unilateral y aquel de carácter multilateral. En este sentido, se analizaron en detalle las propuestas presentes en la literaturas que en líneas generales hayan adoptado un curso de acción compatible con el antes expuesto, con la intención de cotejar tanto los resultados obtenidos como los inconvenientes encontrados.

Como producto de haber considerado las virtudes y las falencias de estas propuestas, se desarrolló el concepto de protocolo diaéctico, el cual, a partir de factorizar las principales características de los sistemas argumentativos considerados a lo largo del capítulo 2, resultó adecuado para modelar una amplia gama de razonamientos de base dialéctica. El concepto de protocolo dialéctico fue posteriormente tomado como punto de partida para desarrollar un segundo modelo, el cual, si bien estructuralmente análogo al primero, tenía por objeto capturar uno de los principales patrones de interacción entre agentes: *la deliberación*. El modelo obtenido aplicando la estrategia propuesta, sintentizado en el concepto de protocolo deliberativo, heredó las principales características del modelo inicial, entre las que se destaca el hecho de estar especificado de forma abstracta. En particular, uno de los aspectos más importantes a la hora de instanciar el modelo por este medio obtenido consiste en especificar de forma acabada cada uno de sus componentes.

Por último, con el objeto de ilustrar el proceso de instanciación, se recorrieron paso a paso las etapas requeridas para poder modelar a dos de las principales teorías de argumentación rebatible, analizando, por otra parte, los respectivos protocolos deliberativos resultantes de tomar como punto de partida a cada una de estas teorías. En primer lugar, se consideró cómo capturar la dinámica de la programación en lógica rebatible en términos de un protocolo dialéctico, contrastándose posteriormente cómo modelar un sistema radicalmente opuesto, que admite prioridades entre las mismas reglas empleadas para construir argumentos.

El modelo obtenido como principal aporte de esta tesis presenta una gran versatilidad, evidenciada principalmente en la facilidad con la cual se puede adecuar a teorías tan dispares como las consideradas anteriormente. De hecho, de la experimentación con este modelo abstracto así como con sus instanciaciones concretas surgió el interrogante de determinar si es posible relacionar el comportamiento de sus diversos componentes con las propiedades manifestadas por sus instanciaciones. Por caso, qué condiciones deben imponerse sobre el comportamiento de la función `next` de forma tal que el protocolo de deliberación resultante sea de carácter progresivo, o bien determinar qúcondiciones permiten asegurar que todos los proponentes mantienen una cierta coherencia a nivel de las creencias sancionadas por sus bases de conocimiento. Este vínculo, de existir, permitiría establecer propiedades a nivel de familias de protocolos, en vez de tener que reiterar este análisis caso por caso.

Glosario

Esta Tesis involucra un conjunto de términos técnicos del idioma inglés, provenientes en su mayoría del area de la argumentación rebatible, que han sido traducidos al castellano. Con el objeto de evitar confusiones producto de estas traducciones, a continuación se incluye como referencia el listado de términos en inglés y los correspondientes vocablos adoptados en castellano.

Término en inglés	**Vocablo adoptado en castellano**
abstract argumentation system	sistema argumentativo abstracto
agent communication language	lenguaje de comunicación entre agentes
agent oriented programming	programación orientada a agentes
argumentation framework	marco argumentativo
autoepistemic logic	lógica autoepistémica
assumption attack	ataque a las suposiciones
backing	fundamentos
blackboard system	sistema de pizarrón
blocking defeater	derrotador de bloqueo
broadcast	difusión en masa
burden of proof	peso de la prueba
categorical assertion	afirmación categórica
categorical counterassertion	contra-afirmación categórica
cautious assertion	afirmación cauta
cautious denial	refutación cauta

Término en inglés	**Vocablo adoptado en castellano**
circunscription	circunscripción
commonsense reasoning	razonamiento de sentido común
commitment	compromiso
complete extension	extensión completa
complete semantic	semántica completa
conclusion	conclusión
conclusive reason	razón conclusiva
contract net	redes de contratos
convention	convención
datum	evidencia
default logic	lógica default
default reasoning	razonamiento por defecto
defeasible reasoning	razonamiento rebatible
defeasible rule	regla rebatible
defensible argument	argumento defensible
defensible literal	literal defensible
defeated outright	totalmente derrotado
dialectical framework	marco dialéctico
dialectical proof theory	teoría de prueba dialéctica
dispute protocol	protocolo de disputa
distributed problem solving	resolución distribuída de problemas
distributed artificial intelligence	inteligencia artificial distribuida
enablement	habilitamiento
grounded extension	extensión fija
grounded semantic	semántica fija
justified argument	argumento justificado
justified literal	literal justificado
knowledfe base	base de conocimiento

Término en inglés	**Vocablo adoptado en castellano**
knowledge representation	representación de conocimiento
manufacturing system	sistema de producción
multiagent system	sistema multiagente
nonmonotonic reasoning	razonamiento no monótono
nonmonotonic logic	lógica no monótona
ordered theory	teoría ordenada
overruled argument	argumento denegado
overruled literal	literal denegado
performative	performativa
preferred extension	extensión preferida
preferred semantic	semántica preferida
prima facie reason	razón prima facie
proper defeater	derrotador propio
provisoed assertion	afirmación condicionada
provisoed counterassertion	contra-afirmación condicionada
provisoed denial	refutación condicionada
provisionally defeated	provisionalmente derrotado
rebuttal attack	ataque por rebatimiento
rebuttal defeater	derrotador por rebatimiento
reinstate	restaurar
sensory network	red de sensado
social abilities	capacidades sociales
specificity criterion	criterio de especificidad
speech act	acto de habla
stable extensión	extensión estable
stable semantic	semántica estable
state oriented domain	dominio orientado a estados
strict rule	regla estricta

Término en inglés	**Vocablo adoptado en castellano**
stricly more specific	estrictamente más específico
strong distinction	distinción fuerte
undefered	no derrotado
undercut attack	ataque por socavamiento
undercut defeat	derrota por socavamiento
utterance	manifestación
warrant	normativa
weak distinction	distinción débil

Bibliografía

[Alferes et al., 1998] Alferes, J. J., Pereira, L. M., y Przymusinski, T. C. (1998). "Classical" Negation in Non-monotonic Reasoning and Logic Programming. *Journal of Automated Reasoning*, 20(1):107–142.

Este artículo analiza el rol desempeñado por los diversos tipos de negación en los principales sistemas de razonamiento no monótono, con el objeto de establecer una clasificación de éstos. Se identifican las propiedades satisfechas por los dos tipos de negación ya conocidos: la negación clásica (*i.e.*, la negación presente en la lógica clásica), y la negación default (*i.e.*, la negación cuyo comportamiento asemeja a la negación por falla de la programación en lógica). Se fundamenta por qué un sólo tipo de negación no es suficiente para satisfacer las demandas surgidas al representar conocimiento y se proponen dos nuevos tipos de negación, la negación fuerte y la negación explícita. Con el objeto de lograr caracterizaciones lo suficientemente generales, las nuevas definiciones son introducidas en el contexto de una lógica autoepistémica de conocimiento, formalismo capaz de modelar a la mayoría de los sistemas de razonamiento conocidos.

[Antoniu, 1996] Antoniu, G. (1996). *Nonmonotonic Reasoning*. MIT Press.

Este libro indaga la problemática del razonamiento no monótono. Se analizan diversos formalismos para modelar este tipo de razonamiento, tales como la lógica default [Reiter, 1980], la lógica autoepistémica [Moore, 1985], la circunscripción [McCarthy, 1980] y la programación en lógica [Lloyd, 1987]. La lógica default es estudiada en profundidad, presentando sus diversas variantes e introduciendo una novedosa semántica procedural. También se elabora un ingenioso algoritmo para calcular expansiones en la lógica autoepistémica y se explora la interrelación entre el razonamiento no monótono y la programación en lógica bajo las semánticas estable y bien fundada.

[Austin, 1962] Austin, J. A. (1962). *How to Do Things with Words*. Harvard university press.

Este libro presenta la teoría de actos de habla basada en la noción de lenguaje performativo, de acuerdo a la cual decir algo es hacer algo. El autor propone que quien realiza un acto de habla está en realidad creando una nueva realidad social en un determinado contexto. En particular, distingue tres clases de actos o sentencias: actos locutorios (equivalentes a producir una sentencia con un cierto significado en el sentido usual), actos ilocutorios (como informar, ordenar, persuadir, etc) y actos perlocutorios (el resultado que logramos al decir algo como una persuación). Austin sostiene que lo importante no es estudiar la sentencia sino el acto en el cual se expresa la sentencia. Esto resume su creencia de que el estudio de las palabras o sentencias (actos locutorios) fuera de su contexto social nos dice poco sobre la comunicación (actos ilocutorios) o su efecto en la audiencia (actos perlocutorios).

[Bond y Gasser, 1988] Bond, A. H. y Gasser, L. (1988). An Analysis of Problems and Research in DAI. En Bond, A. H. y Gasser, L., editores, *Readings in Distributed Artificial Intelligence*, págs. 3–35. Morgan Kaufmann.

Este artículo presenta una revisión de los distintos trabajos realizandos en el área de Inteligencia Artificial Distribuida (DAI). Los autores definen en primer lugar qué se entiende por DAI, señalando a qué área circunscribirán su análisis. Se realiza una reseña histórica de los principales sistemas desarrollados en este terreno, identificando los diferentes desafíos básicos que se deben enfrentar al diseñar un sistema distribuido. El resto del artículo se organiza en torno a los problemas identificados, considerando cada uno de ellos en mayor detalle, repasandos herramientas y lenguajes introducidos para modelar sistemas distribuidos.

[Bondarenko et al., 1997] Bondarenko, A., Dung, P. M., Kowalski, R. A., y Toni, F. (1997). An abstract, argumentation-theoretic approach to default reasoning. *Artificial Intelligence*, 93(1–2):63–101.

Este artículo presenta una teoría abstracta basada en la argumentación rebatible cuyo principal propósito es unificar bajo un mismo enfoque diversos acercamientos al razonamiento no monótono. El formalismo descripto extiende al sistema THEORIST [Poole, 1988] incorporando las diversas semánticas argumentativas desarrolladas en [Dung, 1995]. En esta ocasión la noción de argumento es asociada con el conjunto de hipótesis necesarias para derivar una determinada conclusión, caracterizando el conflicto entre argumentos a través de la refutación de hipótesis. Como semántica del sistema se extrapolan aquellas definidas para los frameworks argumentativos. Cabe destacar que se alcanza con creces el objetivo estipulado, logrando modelar dentro de este formalismo a diversas teorías de razonamiento no monótono.

[Chaib-draa, 1995] Chaib-draa, B. (1995). Industrial Applications of Distributed AI. *Communications of the ACM*, 38(11):49–53.

Este artículo reseña una serie de aplicaciones de la Inteligencia Artificial Distribuida a diversos escenarios industriales. La presentación se divide en dos partes, las aplicaciones que han tenido un éxito comprobado y las que están en desarrollo pero poseen un futuro prometedor. Entre las primeras se mencionan ejemplos en distintas áreas: telecomunicaciones, bases de datos, asignación de recursos en una fábrica, aplicaciones en tiempo real. Entre las restantes, se encuentra el manejo de tráfico urbano y un sistema para el desarrollo concurrente de software. Basándose en los ejemplos considerados, el autor argumenta que el desarrollo de sistemas distribuidos resulta sumamente útil para lidiar con la complejidad de los sistemas industriales y especula que su uso está en franco crecimiento.

[Chesñevar et al., 2000] Chesñevar, C. I., Maguitman, A., y Loui, R. P. (2000). Logical Models of Argument. *ACM Computing Surveys*, 32(4):337–383.

Este artículo de repaso del estado del arte analiza las principales ideas que caracterizan cada uno de los diversos modelos de argumentación existentes. En este sentido, junto con [Prakken y Vreeswijk, 2002] constituyen los primeros artículos en llevar a cabo esta tarea. Se escudriña el origen de este acercamiento al razonamiento no monótono, cuando simplemente constituía una alternativa a las propuestas basada en lógicas clásicas, hasta sus más recientes manifestaciones en la actualidad, donde ya constituye un area muy activa de investigación en sí misma. Por otra parte, también se estudian aquellos dominios de aplicación en los cuales este tipo de sistema han manifestado un desempeño atractivo, como ser el razonamiento basado en casos, la toma de decisiones ante la presencia de incertidumbre o la negociación entre agentes inteligentes.

[Cohen y Levesque, 1990] Cohen, P. R. y Levesque, H. J. (1990). Intention is Choice with Commitment. *Artificial Intelligence*, 42(2-3):213–261.

Este artículo explora una teoría de agencia que intenta balancear las creencias, metas, compromisos, acciones e intenciones del agente. Para esto proponen una lógica que sirva al mismo tiempo para describir y para razonar sobre los estados mentales del agente. Cabe destacar que en este modelo las intenciones juegan un rol preponderante y no pueden ser reducidas a las restantes actitudes. Los autores analizan las propiedades de la lógica propuesta y detallan las implicaciones que esta teoría de agencia acarrerá sobre las potenciales arquitectura que la implementen.

[Cohen y Levesque, 1995] Cohen, P. R. y Levesque, H. J. (1995). Communicative Actions for Artificial Agents. En *Proceedings of the First International Conference on Multi-Agent Systems (ICMAS'95)*, págs. 65–72, Menlo Park, CA, USA.

Este artículo analiza el lenguaje KQML en general, prestando especial atención a la caracterización de su semántica. Se detallan los inconvenientes asociados a la adopción de este lenguaje de comunicación entre agentes, considerando como poder solventarlos. Luego, los autores estudian la clase de semántica que resulta adecuada para los lenguajes de comunicación basados en la teoría de actos de habla. Se identifica en particular una propiedad que deben satisfacer las teorías de actos de habla, denominada composicionalidad. Por medio de esta propiedad se establece un conjunto de condiciones que debe satisfacer toda semántica para resultar viable.

[Corkill, 1991] Corkill, D. D. (1991). Blackboard Systems. *Journal of AI Expert*, 6(9):40–47.

Este trabajo describe tanto las características como el potencial de los sistemas de pizarrón, desde el punto de vista de su creador. Se define en forma precisa en qué consisten este tipo de sistemas, repasando sus principales componentes. El autor identifica cuándo resulta conveniente usar un sistema de pizarron para una aplicación determinada, destacando como candidatos a los sistemas en los cuales no se conoce un marco de trabajo para representar el conocimiento, o bien donde la aplicación debe tomar desiciones de control dinámicas y el producto terminado debe poder integrarse como parte de un sistema más grande. Finalmente, se presentan una serie de lineamientos y ejemplos concretos tendientes a auxiliar al diseñador de un sistema de pizarrón.

[Corkill, 2003] Corkill, D. D. (2003). Collaborating Software: Blackboard and Multi-Agent Systems & the future. En *Proceedings of the International Lisp Conference*, New York, USA.

Este artículo analiza los sistemas de pizarrón desde una perspectiva histórica, como un primer acercamiento para la creación de software colaborativo. En este terreno se realiza un análisis comparativo de los sistemas multiagente orientados a la colaboración con respecto a los sistemas de pizarrón, considerando en qué casos resulta más adecuado uno u otro sistema, planteando incluso la posibilidad de integrar ambos acercamientos.

[Craig, 1998] Craig, I. D. (1998). From Blackboards to Agents. En *Online Proceedings of the VIM Project Spring Workshop on Collaboration Between Human and Artificial Societies*, Lanjarón, España.

Este artículo reinterpreta la metafora de los sistemas de pizarrón propuesta originalmente por Corkill para el contexto de los sistemas multiagente [Corkill, 1991]. El autor argumenta que la interpretación original crea algunos problemas en lo concerniente al desempeño, acceso concurrente y autonomía de los expertos. La reinterpretación propuesta resulta en una arquitectura basada en agentes, en la cual los expertos no estan restringidos a comunicarse a través del pizarrón, sino que pueden interactuar unos con otros, incluso actuando en forma concurrente. Por otra parte, los expertos deciden por sí mismos en qué aspectos del problema trabajarán y cómo realizarán sus tareas. En este contexto el pizarrón se convierte en un medio para el intercambio de mensajes, ya que los expertos son autónomos, es decir, el pizarrón ya no resume el estado global del sistema.

[Dautenhahn, 1995] Dautenhahn, K. (1995). Getting to know each other – Artificial social intelligence for autonomous robots. *Robotics and Autonomous Systems*, 1(16):333–356.

Este trabajo propone el fomento de la inteligencia social artificial, como una forma de obtener entidades autónomas inteligentes. Este acercamiento está inspirado en la hipótesis de la inteligencia social, la cual establece que la inteligencia se originó en los primates como una forma de resolver problemas sociales, para luego extenderse a la resolución de otra clase de situaciones. El autor sugiere aplicar este principio de evolución a la propia Inteligencia Artificial. Para sustentar este teoría se describen en primer lugar escenarios de aplicación en los cuales se necesitan habilidades sociales para interactuar entre entidades artificiales, o bien con usuarios humanos. Se discuten estudios en sociedades naturales que afirman que la adquisición de inteligencia social es un prerequisito para el desarrollo de otras habilidades de inteligencia no sociales. Finalmente, se consideran mecanismos que permiten estudiar la inteligencia social en entidades artificiales.

[Davis y Smith, 1983] Davis, R. y Smith, R. G. (1983). Negotiation as a Metaphor for Distributed Problem Solving. *Artificial Intelligence*, 20(1):63–109.

Este artículo introduce el mecanismo de delegación de tareas conocido como redes de contratos. La particularidad de este modelo es que la distribución de tareas para la resolución distribuida de problemas se realiza en forma interactiva: un nodo anuncia una tarea a realizar y un grupo de nodos con la capacidad de ejecutarla formulan distintas ofertas. Los autores enfatizan que la negociación, esto es, la interacción entre los componentes, es una extensión adecuada de los mecanismos de transferencia de control usados en los primeros sistemas. Finalmente, se presenta una aplicación de las redes de contrato a un problema de monitoreo concreto, conceptualmente análogo al propio del control del tráfico aereo.

[Davis, 1989] Davis, R. E. (1989). *Truth, Deduction and Computation: logic and semantics for computer science*. Computer Science Press.

Este libro presenta un acabado análisis de las teorías formales más conocidas (*i.e.*, la lógica proposicional y el cálculo de predicados), con especial atención en distinguir los ámbitos semánticos, sintácticos y de cómputo. Se prueban las metapropiedades de completitud, sensatez, consistencia y decidibilidad para cada una de estas teorías. Se aborda la incompletitud de la teoría numérica elemental, resultado debido a Kurt Göedel, ejemplificando una teoría formal incompleta con innumerables aplicaciones prácticas. Para resaltar la distinción entre deducción y cómputo se presenta una cuarta teoría formal, el cálculo lambda, donde también se estudia la interrelación entre las metapropiedades.

[Dung, 1993a] Dung, P. M. (1993a). An Argumentation Semantics for Logic Programming with Explicit Negation. En Warren, D. S., editor, *Proceedings of the Tenth International Conference on Logic Programming*, págs. 616–630, Budapest, Hungría.

Este artículo describe una semántica basada en argumentos para una extensión de la programación en lógica convencional introducida por [Gelfond y Lifschitz, 1990], que incorpora una clase de negación inspirada en aquella presente en la lógica clásica, denominada negación explícita. Se desarrolla una teoría de argumentación en la cual la noción de argumento es asociada con el conjunto de suposiciones requeridas para derivar un cierto literal. Se consideran dos tipos de conflicto entre argumentos: el ataque a los basamentos y el ataque por reducción al absurdo. Se exploran cuatro alternativas para dotar de significado a este formalismo: las semánticas preferida, estable, básica y completa. También se estudian las propiedades que deben ser satisfechas por un programa lógico para asegurar la existencia de las diversas semánticas.

[Dung, 1993b] Dung, P. M. (1993b). On the Acceptability of Arguments and its Fundamental Role in Nonmonotonic Reasoning and Logic Programming. En *Proceedings of the 13th International Joint Conference on Artificial Intelligence*, págs. 852–857, Chambéry, Francia. Morgan Kaufmann Publishers.

Este artículo se propone investigar los mecanismos fundamentales que gobiernan la argumentación entre personas. A tal efecto se introduce la noción de un framework argumentativo. A diferencia de otros acercamientos a la argumentación rebatible, en un framework argumentativo los argumentos carecen de estructura interna prefijada, alcanzando de esta forma un alto nivel de abstracción. Se introducen las definiciones de cuatro semánticas que pueden asociarse a un framework argumentativo: la admisible, la preferida, la básica y la completa. Tambiém se delinea como reinterpretar diversos paradigmas de razonamiento no monótono (*e.g.*, lógica default [Reiter, 1980], programación en lógica [Lloyd, 1987], etc.) en términos de frameworks argumentativos. Una versión ampliada y revisada de este artículo se encuentra publicada en [Dung, 1995].

[Dung, 1995] Dung, P. M. (1995). On the Acceptability of Arguments and its Fundamental Role in Nonmonotonic Reasoning and Logic Programming and n-Person Games. *Artificial Intelligence*, 77(2):321–357.

Este artículo elabora sobre la base de un artículo anterior [Dung, 1993b], ahondando el análisis allí emprendido de los mecanismos fundamentales que gobiernan la argumentación entre personas. Se revisita el concepto de framework argumentativo junto con sus semánticas asociadas definias en torno a la noción de aceptabilidad. Se aborda la problemática de la indefinición de la semántica, estudiando propiedades que al ser satisfechas por un framework argumentativo aseguran la existencia y/o coincidencia de las diversas semánticas. También se estudia el rol la argumentación rebatible en el razonamiento no monótono, reinterpretando diversos paradigmas en términos de frameworks argumentativos.

[Etherington y Reiter, 1983] Etherington, D. W. y Reiter, R. (1983). On Inheritance Hierarchies With Exceptions. En *Proceedings of the 3rd National Conference on Artificial Intelligence*, págs. 104–108.

Este artículo se compromete a proveer de una semántica adecuada a las jerarquías de herencia que admiten excepciones. Teniendo en cuenta el éxito alcanzado adoptando lógica clásica en la formalización de la semántica de las jerarquías sin excepciones, se postula una semántica alternativa para el caso con excepciones basada en la lógica default [Reiter, 1980], teoría formal que extiende a la lógica clásica. Se presenta un algoritmo de cómputo para la semántica definida, restringido a el caso de redes acíclicas. También se estudia como explotar el potencial paralelismo presente en este algoritmo, sugiriendo que para el caso sin restricciones no será posible formular un algoritmo masivamente paralelo.

[Finin et al., 1992] Finin, T., McKay, D., Fritzson, R., y McEntire, R. (1992). KQML–a language and protocol for knowledge and information exchange. En Fuchi, K. y Yokoi, T., editores, *Knowledge Building and Knowledge Sharing*, págs. 456–463. Ohmsha and IOS Press.

Este artículo constituye la definición inicial del lenguaje de comunicación entre agentes KQML. Este lenguaje fue desarrollado por el grupo KSE (Knowledge Sharing Effort) de la organización ARPA. Este grupo tiene por misión desarrollar un conjunto de convenciones para que resulte más fácil compartir y reusar bases de conocimiento. Los autores desarrollaron el lenguaje KQML para solucionar las dificultades existentes en la comunicación entre agentes, proveyendo un lenguaje común con una sintáxis, semántica y pragmática certeramente definidas. El aporte principal consiste en la definición de protocolo de comunicación basado en este lenguaje.

[Fitting, 1996] Fitting, M. (1996). *First-Order Logic and Automated Theorem Proving*. Springer-Verlag, segunda edición.

Este libro presenta un novedoso tratamiento de la lógica proposicional y la lógica de primer orden, donde se presta especial detenimiento en explorar sistemas de prueba no convencionales. Se estudia la sensatez y completitud de los sitemas de prueba por tableaux semántico y por resolución, tanto para el caso proposicional como para el caso de primer orden. Se contrastan otros métodos de prueba adicionales para el cálculo proposicional, tales como la deducción natural, el cálculo de resultantes y los sistemas axiomáticos. También se analiza un tipo particular de teoría de primer orden que incluye una axiomatización de la relación de igualdad, abarcando desde su sintáxis y semántica hasta la implementación de su sistema de prueba.

[García, 1997] García, A. J. (1997). La Programación en Lógica Rebatible: su definición teórica y computacional. Tesis de Magister, Departamento de Ciencias de la Computación, Universidad Nacional del Sur, Bahía Blanca, Argentina.

Esta Tesis constituye la primer formalización extensiva de la Programación en Lógica Rebatible, teoría formal inspirada en los sistemas Simari-Loui [Simari y Loui, 1992] y su reformulación dialéctica [Simari et al., 1994]. La Programación en Lógica Rebatible hereda las principales virtudes de las disciplinas que combina (*i.e.*, la argumentación rebatible y la programación en lógica). Este formalismo interpreta a las implicaciones materiales como reglas de inferencia. También incorpora la noción de presuposición [Nute, 1987], y se analizan las virtudes de admitir un nuevo tipo de negación, la negación default. Otro aporte de esta Tesis consiste en la definición de una máquina abstracta apropiada para la interpretación eficiente de programas lógicos rebatibles, inspirada en aquella definida por David Warren para la programación en lógica convencional.

[García, 2000] García, A. J. (2000). *Programación en Lógica Rebatible: Lenguaje, Semántica Operacional, y Paralelismo.* Tesis Doctoral, Departamento de Ciencias de la Computación, Universidad Nacional del Sur, Bahía Blanca, Argentina.

Esta Tesis presenta un refinamiento de la Programación en Lógica Rebatible introducida oportunamente en [García, 1997], formalismo de representación de conocimiento y razonamiento que al combinar la argumentación rebatible con la programación en lógica hereda las principales virtudes de estas disciplinas. Se introduce en primer lugar una versión simplificada del sistema sin negación default ni presuposiciones, con el objeto de caracterizar con precisión tanto su sintáxis como su semántica. Cabe acotar que la semántica del sistema es definida con exactitud mediante una especificación operacional. La negación default y las presupociones son incorporadas posteriormente como extensiones al sistema simplificado, adecuando la semántica operacional de manera acorde. También se analiza la explotación del paralelismo, indicando como extrapolar las ideas desarrolladas en el marco de la programación en lógica convencional al contexto de la programación en lógica rebatible.

[García et al., 2000] García, A. J., Gollapally, D., Tarau, P., y Simari, G. R. (2000). Deliberative Stock Markets using Jinni and Defeasible Logic Programming. En *Proceedings of Engineering Societies in the agents' world, ECAI 2000.*

Este artículo describe un escenario para la implementación de sistemas multiagente donde se combinan la Programación en Lógica Rebatible [García, 2000] con el lenguaje Jinni, una plataforma para la construcción de agentes inteligentes autónomos. La aplicación desarrollada permite formular un conjunto de agentes que extraen información del mercado de valores por medio de internet y otro conjunto de agentes deliberativos que representan su conocimiento y razonan en base a la Programación en Lógica Rebatible. Se combina entonces el poder expresivo de este formalismo con los servicios que provee Jinni para el desarrollo de agentes autónomos, aprovechando las ventajas de ambos formalismos.

[Garijo et al., 1996] Garijo, M., Cancer, A., y Sanchez, J. (1996). A Multiagent System for Cooperative Network-Fault Management. En *Proceedings of the First International Conference on the Practical Applications of Intelligent Agents and Multi-agent Technology*, págs. 279–294, Londres, UK.

Este trabajo describe una aplicación de los sistemas multiagente para asistir a los operadores en el manejo de las redes de comunicaciones. Mientras que los acercamientos tradicionales se basaban en sistemas centralizados, que solamente funcionaban para un dominio en particular, en esta propuesta se introduce un marco de trabajo en el cual un grupo heterogéneo de agentes autónomos cooperan para realizar el manejo de las fallas de un sistema de redes. Los agentes usan un lenguaje de interacción de alto nivel para comunicarse con sus pares, y primitivas de un protocolo estandar para obtener información sobre el estado de la red.

[Gasser, 1991] Gasser, L. (1991). Social Conceptions of Knowledge and Actions: DAI Foundations and Open Systems Semantics. *Artificial Intelligence*, 47(1):107–138.

Este artículo brinda una visión panorámica de la investigación en la Inteligencia Artificial Distribuida (DAI). Se discuten un conjunto de problemas básicos en el contexto de la DAI, presentando además las motivaciones que guían al desarrollo de éste tipo de sistemas. Luego, se presenta una serie de principios que fundamentan una concepción social de la acción y el conocimiento en la investigación en este área. Estos principios son luego contrastados con la propuesta realizada por Carl Hewitt sobre la semántica de los sistemas abiertos, mostrando como éstos puede ser fortalecidos y refinados para servir como la base a partir de la cual construir sistemas en el área de la Inteligencia Artificial Distribuida.

[Gelfond y Lifschitz, 1990] Gelfond, M. y Lifschitz, V. (1990). Logic Programs with Classical Negation. En Warren, D. H. D. y Szeredi, P., editores, *Proceedings of the 7th International Conference on Logic Programming*, págs. 579–597.

Este artículo presenta una extensión a la programación en lógica convencional motivada por su inhabilidad para representar información incompleta. A tal efecto, se formaliza la Programación en Lógica Extendida, la cual introduce un nuevo tipo de negación, denominado negación fuerte, con el objeto de capturar el tipo de negación presente en la lógica clásica. La correspondiente semántica se define en términos de un operador de punto fijo que computa los conjuntos de respuestas asociados a un programa, una noción derivada de la semántica estable de la programación en lógica convencional.

[Grosof y Labrou, 2000] Grosof, B.Ñ. y Labrou, Y. (2000). An Approach to using XML and a Rule-based Content Language with an Agent Communication Language. En Dignum, F. y Greaves, M., editores, *Issues in Agent Communication*, págs. 96–117. Springer-Verlag.

Este artículo presenta una codificación en XML del lenguage de comunicación entre agentes FIPA-ACL. Los autores argumentan que esta propuesta facilitará desarrollar y mantener compiladores para integrarlos con la ingeniería de software de internet. Aunque FIPA-ACL es el lenguaje elegido para este trabajo, la codificación podría también llevarse a cabo en KQML. Se propone un nuevo lenguaje de contenido para FIPA-ACL, el cual resulta adecuado para algunas aplicaciones en particular, como es el caso de reglas de negocios para el comercio electrónico. El lenguaje de contenidos desarrollado es una forma extendida de programas lógicos codificada también en XML, y denominado BRML (Business Rules Markup Language).

[Hayton et al., 1998] Hayton, R., Bacon, J., y Moody, K. (1998). OASIS: Access Control in an Open, Distributed Environment. En *Proceedings of IEEE Symposium on Security and Privacy*, págs. 3–14.

Este trabajo describe un acercamiento para realizar el control de acceso en un sistema abierto, que consiste en una arquitectura de servicios de red segura e independiente. Cada uno de los componente del sistema puede nombrar a sus clientes de acuerdo a un conjunto de roles designados, usando una lógica formal para expresar las condiciones necesarias para asumir cada rol. Aquellos clientes que resulten aceptados para un determinado servicio se les otorga un certificado de membresía, el cual está protegido por un protocolo de cifrado. El sistema resultante es inherentemente distribuido y puede adaptarse a diferentes escenarios. Los autores postulan que es posible incorporar esta arquitectura a aplicaciones preexistentes. Como caso concreto, esta propuesta fue utilizada para implementar un controlador de tráfico aereo.

[Huhns y Stephens, 1999] Huhns, M.Ñ. y Stephens, L. M. (1999). Multiagent Systems and Societies of Agents. En Weiss, G., editor, *Multiagent Systems: A Mo-*

dern Approach to Distributed Artificial Intelligence, págs. 79–120. The MIT Press, Cambridge, MA, USA.

Este trabajo presenta un análisis de los principales mecanismos de interacción entre agentes. Se analiza la comunicación entre agentes, reseñando algunos de los acercamientos más relevantes, como la teoría de actos de habla y los lenguajes KQML y KIF. Luego, se estudian los protocolos de interacción entre agentes, distinguiendo entre diferentes categorías: la coordinación, la negociación, los mecanismos de mercado, el mantenimiento de creencias en sistemas multiagente y la cooperación. Los autores proponen analizar el concepto de sociedades de agentes, donde el sistema de agentes no se analiza desde el punto de vista de un agente individual, sino que la visión se centra en el conjunto de los agentes involucrados, los cuales en esencia conforman una sociedad. Los roles y compromisos sociales de cada uno de estos agentes son elementos claves para formular este tipo de sistemas.

[Jennings, 1996] Jennings, N. R. (1996). Coordination Techniques for Distributed Artificial Intelligence. En O'Hare, G. y Jennings, N. R., editores, *Foundations of Distributed Artificial Intelligence*, págs. 187–210. Wiley Interscience.

Este artículo explora el problema de la coordinación entre agentes, analizando los problemas que se presentan usualmente, considerando las posibles soluciones a los mismos. En particular, se presenta una técnica de coordinación basada en el uso de un formalismo de búsqueda distribuida, donde las estructuras claves son los compromisos y las convenciones. Los compromisos se interpretan como promesas para seguir un determinado curso de acción, mientras que las convenciones permiten monitorear los compromisos adquiridos ante un ambiente cambiante. Finalmente, para demostrar la expresividad del modelo propuesto, se reformula un conjunto de mecanismos para la coordinación multiagente en términos de compromisos y convenciones.

[Jennings, 2000] Jennings, N. R. (2000). On agent-based software engineering. *Artificial Intelligence*, 117(1):277–296.

Este artículo analiza el problema de construir software de calidad para problemas del mundo real, considerando aquellos escenarios donde la ingeniería de software aún no provee herramientas adecuadas. El autor sostiene que los acercamientos basados en agentes pueden aumentar la capacidad de construir sistemas de software complejos y distribuidos. Esta afirmación se sustenta en las cualidades que manifiestan estos acercamientos, por caso que la división en agentes es una buena forma de particionar el espacio de problema en los sistemas complejos, o que las abstracciones claves de los marcos orientados a agentes resultan usualmente más naturales. Por otra parte, se señalan cuáles son las desventajas asociadas al uso de un paradigma orientado a agentes y se proponen soluciones para éstos a partir de un análisis basado en el nivel social de los sistemas.

[Kraus et al., 1998] Kraus, S., Sycara, K., y Evenchik, A. (1998). Reaching agreements through argumentation: a logical model and implementation. *Artificial Intelligence*, 1(104):1–69.

Este artículo desarrolla un modelo de negociación en el cual los agentes involucrados intercambian argumentos para lograr arribar a una solución común. En este escenario, los argumentos se interpretan como manifestaciones cuyo propósito es persuadir a otro agente a fin de influenciar su futuro curso de acción. Se introduce una lógica formal para gobernar el intercambio de argumentos entre los participantes, de forma que se respeten los objetivos de diseño del formalismo. Los autores exploran una taxonomía de las diferentes formas en que un agente puede conducir una negociación, identificando seis formas diferentes, cuya eficacia ya ha sido ampliamente probada en las negociaciones entre seres humanos. Finalmente, se presenta la implementación de un sistema de simulación para la creación y uso de agentes que negocian siguiendo el esquema propuesto.

[Krause et al., 1995] Krause, P., Ambler, S., Elvang-Gøransson, M., y Fox, J. (1995). A Logic of Argumentation for Reasoning under Uncertainty. *Computational Intelligence*, 11(1):113–131.

Este artículo presenta una lógica para la argumentación denominada LA, que se basa en un formalismo preexistente denominado lógica minimal. El proceso de inferencia se realiza en dos niveles, uno de ellos se encarga de la construcción de argumentos, que son pruebas en la lógica minimal, mientras que el razonamiento sobre los argumentos que se realiza en un metanivel. Para razonar sobre los argumentos se aumenta la lógica con un conjunto de etiquetas que permiten anotar las proposiciones con los argumentos que las sustentan. Además de presentar la sintaxis de su lógica, los autores describen las líneas generales para el desarrollo de un demostrador de teoremas basado en argumentación. Proponen también una semántica formal para esta lógica basada en una teoría de categorías.

[Labrou, 2001] Labrou, Y. (2001). Standardizing Agent Communication. *Lecture Notes in Computer Science*, 2086:74–97.

Este artículo reconoce a la estadarización de los lenguajes de comunicación como un aspecto clave para promulgar su uso. Se comienza por analizar el lenguaje KQML desde una perspectiva histórica, repasando luego al lenguaje FIPA-ACL, el primer esfuerzo coordinado para desarrollar un lenguaje de comunicación en el área de los Sistemas Multiagente. Posteriormente, se considera qué problemas pueden surgir al utilizar un lenguaje de comunicación de agentes en un sistema concreto. El autor remarca que existen numerosos aspectos pragmáticos que no fueron solucionados por la semántica de estos lenguajes. En consecuencia, cada implementación debe solucionar estos incovenientes según su criterio, pero como no existe un estandar establecido se crean nuevos dialectos de los lenguajes originales. Finalmente, se analiza como solucionar este tipo de situaciones para lograr la estandarización de los lenguajes de comunicación de agentes anteriormente repasados.

[Labrou y Finin, 1998] Labrou, Y. y Finin, T. (1998). Semantics for an Agent Communication Language. En *Agent Theories, Architectures and Languages IV*, Lecture Notes in Artificial Intelligence, págs. 209–214. Springer-Verlag.

Este artículo presenta una propuesta para describir la semántica de KQML basada en ideas tomadas de la teoría de actos de habla. Una de las principales críticas al lenguaje KQML consiste en que su semántica no fue definida formalmente. Esto provocó entre otras cosas que existan diferencias irreconciliables entre las distintas implementaciones. Para otorgar una semántica a KQML los autores proponen interpretar a cada performativa del lenguaje como un acto de habla. Se utiliza entonces un conjunto de expresiones que especifican las precondiciones y postcondiciones asociadas a cada performativa, junto con condiciones que describen el estado final de una ejecución exitosa de la performativa en cuestión. Para esto se valen de un conjunto de operadores que representan las actitudes de los agentes, dotados de un significado reservado, para representar creencias, deseos, conocimiento e intenciones. En base a esta propuesta se describe el conjunto completo de las performativas de KQML. Cabe señalar que este mismo método puede ser extrapolado para describir la semántica de otros lenguajes de comunicación basados en la teoría de actos de habla.

[Labrou et al., 1999] Labrou, Y., Finin, T., y Peng, Y. (1999). Agent Communication Languages: The Current Landscape. *Intelligent Systems*, 14(2):45–52.

Este artículo reseña los lenguajes de comunicación entre agentes existentes, comparando sus caraterísticas primordiales. El autor presenta un conjunto de nociones que a su juicio consituyen los conceptos básicos de los lenguajes de comunicación entre agentes. Estos conceptos se ejemplifican en el contexto del lenguaje KQML, de quien se brinda una descripción detallada, distinguiendo la sintáxis y la semántica del lenguaje. A continuación se considera al lenguaje FIPA-ACL, destacando sus similitudes y diferencias con respecto a KQML.

[Lesser, 1991] Lesser, V. R. (1991). A Retrospective View of FA/C Distributed Problem Solving. *IEEE Transactions on System Man and Cybernetics*, 1(21):1347–1363.

Este trabajo tiene como objetivo extender el modelo FA/C (Functionally Accurate, Cooperative), que fuera desarrollado para realizar distribución de tareas entre procesos en la década de los ‘80. Este modelo nació como un mecanismo para que los agentes (procesos) cooperaran en forma efectiva aún cuando cuenten con información inconsistente o limitada sobre las actividades o criterios de los otros agentes. La definición del modelo en cuestión viene acompañada de los resultados obtenidos en el conjunto de experimentos realizados a fin de establecer el desempeño del nuevo modelo. Finalmente, se comparan y analizan las distintas alternativas para modelar la cooperación en el marco de un conjunto de agentes.

[Lin y Shoham, 1989] Lin, F. y Shoham, Y. (1989). Argument Systems: a uniform basis for nonmonotonic reasoning. En Levesque, H. J., Brachman, R. J., y Reiter, R., editores, *Proceedings del 1st International Conference on Principles of Knowledge Representation and Reasoning*, págs. 245–255, Toronto, Canada. Morgan Kaufmann.

Este artículo describe un formalismo argumentativo capaz de modelar el razonamiento no monótono, cuyo propósito principal es unificar las diversas teorías existentes bajo una misma concepción. En este sistema los argumentos guardan la forma de árboles de derivación basados en reglas de inferencia monótonas y no monótonas. El conjunto de argumentos razonables se captura a través de la noción de estructura de argumento. Estos conjuntos distinguidos constituyen las consecuencias sancionadas por el sistema. También se detalla como reconstruir diversas teorías (*e.g.*, la lógica default [Reiter, 1980], la circunscripción [McCarthy, 1980]) en términos de este formalismo. Como contrapartida, no se discute la comparación de argumentos, por lo que en consecuencia se provee de poca guía para dirimir aquellas situaciones donde múltiples estructuras de argumentos en conflicto resultan aceptables.

[Lloyd, 1987] Lloyd, J. W. (1987). *Foundations of Logic Programming.* Springer-Verlag, segunda edición.

Este libro recorre los fundamentos teóricos que sustentan a la programación en lógica. Se introduce la semántica declarativa heredada de la lógica clásica. Se aborda la semántica operacional, demostrando su sensatez y completitud en relación a la semántica declarativa. Se estudia el limitado poder expresivo de los programas lógicos definidos, destacando su incapacidad para representar información negativa. Extendiendo los programas lógicos definidos a través de una negación por falla se arriba a la noción de programa lógico normal. Se extiende la semántica operacional para proveer de significado a este tipo de negación, comprobando que los resultados de sensatez y completitud aún se verifican para la semántica extendida.

[Loui, 1987] Loui, R. P. (1987). Defeat Among Arguments: A System of Defeasible Inference. *Computational Intelligence*, 3(2):100–106.

Este artículo puede ser sindicado como la primer reconstrucción del concepto de argumento en el seno de la Inteligencia Artificial. Se define una teoría que utiliza como lógica subyacente un lenguaje de primer orden, extendido por medio de un conjunto de reglas de inferencia que permiten representar conocimiento de caracter tentativo. La base de conocimiento, codificada por medio de este lenguaje, permite la construcción de argumento a partir de la información allí presente. Este concepto se asocia al grafo de inferencia recorrido para alcanzar una cierta meta. Por otra parte, se introducen las relaciones de contraargumentación y de derrota entre los argumentos del sistema, estableciendo la semántica del sistema en términos argumentativos.

[Loui, 1998] Loui, R. P. (1998). Process and Policy: Resource-Bounded Nondemonstrative Reasoning. *Computational Intelligence*, 14:1–38.

Este artículo explica el punto de vista de su autor en relación a las formas no demostrativas de razonamiento. En particular, se destaca que los inconvenientes encontrados en este área son producto, generalmente, de haber tomado por sentado al concepto de default. En la mayoría de los casos el concepto de default se modela mediante una forma de implicación de naturaleza más débil que la implicación material. El autor enfatiza que la elección de una ontología incorrecta puede causar que el sistema resultante manifieste una semántica plagada de singularidades. Sostiene que un default en esencia no es más que un plan o una estrategia para alcanzar una cierta conclusión, que se asemeja más a una política que describe como obtener nuevas conclusiones a partir de la información disponible. Finalmente, el autor explora la problemática del consumo de recursos que genera el proceso de razonamiento. Se desarrolla como caso de estudio un protocolo de base dialéctica que permite la incorporación de nuevas políticas a lo largo del proceso de razonamiento, mostrando cómo este simple protocolo es lo suficientemente sofisticado como para modelar diversas teorías de razonamiento rebatible.

[Malone, 1990] Malone, T. W. (1990). Organizing information processing systems: Parallels between human organizations and computer systems. En Robertson, S. P., Zachary, W., y Black, J. B., editores, *Cognition, computation and cooperation*, págs. 56–83. Ablex Publishing Corporation.

Este artículo concentra en las similitudes que existen entre las estructuras organizacionales de las sociedades humanas y los sistemas computacionales. El objetivo principal que el autor pretende alcanzar es definir una metodología de desarrollo que permita usar a las organizaciones humanas como fuente de inspiración para el desarrollo de sistemas computacionales. Para esto se analiza la descomposición de problemas existente en ambos campos y la importancia de la coordinación como medio para la asignación de recursos y la comunicación de resultados intermedios entre las distintas partes. Se estudian las diferentes formas de coordinacion presentes en la literatura, desde las básicas (ajuste mutuo, supervisión directa y estandarización), hasta formas más elaboradas, como las jerarquías y los mecanismos de mercado, que se construyen en base a las anteriores. Como resultado de este estudio el autor presenta un marco conceptual para estudiar las estructuras organizacionales y la coordinación de distintos sistemas, que integra el trabajo previo realizado por varios especialistas en el área. Este marco conceptual cumple con el objetivo de facilitar la transferencia de ideas entre el campo que analiza las organizaciones humanas y el que contempla los sistemas computacionales.

[Martínez y García, 1999] Martínez, D. C. y García, A. J. (1999). Significancia de las falacias en los sistemas argumentativos. En *Proceedings del 6to Congreso*

Argentino de Ciencias de la Computacin.

Este artículo analiza en profundidad el problema de las falacias en los sistemas argumentativos. Los autores reconocen que en la mayoría de los modelos formales para la argumentación rebatible es posible contruir argumentos falaces, esto es, argumentos que, aún cuando parezcan válidos a simple vista, han sido estructurados en base a supuestos en conflicto. A continuación desarrollan una interesante clasificación de las falacias en la argumentación, distinguiendo las falacias relacionadas con apectos subjetivos del debate de aquellas intrínsecas a los procesos argumentativos, señalando que estas últimas deberían ser evitadas por el propio sistema. Finalmente, se analiza la situación de los diversos formalismos argumentativos existentes en relación al tratamiento de las falacias.

[McCarthy, 1980] McCarthy, J. L. (1980). Circumscription—A Form of Non-Monotonic Reasoning. *Artificial Intelligence*, 13(1–2):27–39.

Este artículo introduce una teoría de razonamiento no monótono basada en la minimización de modelos denominada circunscripción. La principal intuición detrás de este formalismo consiste en modelar un patrón de razonamiento que parece propio de los seres humanos, el cual en esencia permite asumir que el conjunto de los elementos que presentan una cierta propiedad constituye precisamente el conjunto de todos los elementos que satifacen esa propiedad. Esta sencilla noción permite obtener un diagnóstico correcto frente a los casos de prueba usuales, si bien su formalización presenta una alta complejidad. El autor enfatiza que la circunscripción no constituye una nueva lógica no monótona, sino que en realidad representa una metodología para poder razonar de esta manera a partir de la lógica clásica.

[McDermott y Doyle, 1980] McDermott, D. V. y Doyle, J. (1980). Non-monotonic logic I. *Artificial Intelligence*, 13(1–2):41–72.

Este artículo desarrolla un estudio de las lógicas no monótonas, su historia y las motivaciones que condujeron al desarrollo estos sistemas. Los autores consideran en su análisis cada uno de los formalismos no monótonos existentes, como la circunscripción, los sistemas de mantenimento de verdad y la lógica default. Con el afán de disponer de un formalismo no monótono con una sólida base teórica, los autores introducen a continuación la Lógica No Monótona, cuya estructura asemeja aquella de las lógicas clásicas. Se detallan tanto su sintáxis como su semántica, introduciendo a su vez una teoría de prueba apropiada para la misma.

[Moore, 1985] Moore, R. C. (1985). Semantical considerations in non-monotonic reasoning. *Artificial Intelligence*, 25(1):75–94.

Este artículo analiza en profundidad la lógica no monótona propuesta por McDermott y Doyle [McDermott y Doyle, 1980]. Moore afirma que ese formalismo posee algunas peculiaridades que ponen en evidencia la falla del mismo para cumplir con sus objetivos de diseño. En consecuencia, el autor plantea una reconstrucción de la lógica no monótona que denomina lógica autoepistémica. Se define una semántica para el sistema resultante en base a la cual se establece formalmente que la lógica autoepistémica es sensata y completa. Finalmente, se contrastan ambos formalismos, enfatizando las virtudes del nuevo acercamiento.

[Moulin y Chaib-draa, 1996] Moulin, B. y Chaib-draa, B. (1996). An overview of Distributed Artificial Intelligence. En O'Hare, G. M. y Jennings, N. R., editores, *Foundations of Distributed Artificial Intelligence*, págs. 3–55. John Wiley and Sons.

Este trabajo presenta una síntesis del estado de la Inteligencia Artificial Distribuida, en el cual se detallan las incumbencias del área y las motivaciones para utilizar sistemas distribuidos. Para repasar las contribuciones existentes, los autores proponen en primer lugar una clasificación propia. Se analizan los sistemas desde la perspectiva del agente, esto es, lo que atañe a un agente que participa de un sistema multiagente. Luego, se considera la perspectiva de grupo, esto es, los elementos que caracterizan a un grupo de agentes (como la coordinación, la cooperación o la negociación). Una tercer categoría reune acercamientos específicos que escapan a estas dos visiones, y por último la perspectiva del diseñador, donde se congregan los métodos y técnicas de implementación propuestas para construir sistemas multiagente.

[Nute, 1987] Nute, D. (1987). Defeasible Reasoning. En *Proceedings of the XX Annual Hawaii International Conference on System Sciences*, págs. 470–477.

Este trabajo, junto con el acercamiento de John Pollock [Pollock, 1987], fue uno de los pioneros en formalizar el concepto de razonamiento rebatible. El autor identifica que en numerosas oportunidades conocer nuevos hechos puede conducirnos a descartar algunas de nuestras creencias anteriores, denominando a esta clase de inferencia razonamiento rebatible. Se define la lógica LDR como mecanismo para representar conocimiento y para obtener conclusiones utilizando este tipo de razonamiento. Posteriormente, se desarrolla una implementación computacional de la misma, como una extensión del lenguaje Prolog.

[Parsons et al., 1998] Parsons, S., Sierra, C., y Jennings, N. (1998). Agents that reason and negotiate by arguing. *Journal of Logic and Computation*, 8(3):261–292.

Este artículo delinea un marco de negociación en el cual un conjunto de agentes autónomos intentan alcanzar sus respectivas metas personales, contando con la posibilidad de influenciar el comportamiento de los restantes agentes mediante oportunas negociaciones. En este modelo los agentes son caracterizados a través de sistemas multicontexto, un formalismo que como particularidad permite capturar las distintas modalidades que componen el estado mental de un agente sin tener que depender de una lógica modal. En este contexto, el término negociación se interpreta como el diálogo entre dos o más agentes que les permite alcanzar un acuerdo común acerca del tópico en debate, apelando al contínuo intercambio de argumentos, que sintetizan las propuestas de los agentes, proveyendo además sus respectivas fundamentaciones. Este intercambio de argumentos es gobernado por medio de un modelo formal basado en argumentación, por lo que el aspecto más llamativo de esta propuesta es que el grupo de agentes apela a la argumentación tanto para formar sus propias opiniones como para llevar adelante la negociación.

[Parunak, 1998] Parunak, V. D. (1998). Practical and Industrial Applications of Agent-Based Systems. En O'Hare, G. y Jennings, N. R., editores, *Foundations of Distributed Artificial Intelligence*, págs. 139–164. Wiley Interscience.

Este artículo describe un conjunto de aplicaciones industriales de los sistemas centrados en el concepto de agencia. El autor presenta en primer lugar una taxonomía de agentes a fin de catalogar los distintos tipos de sistemas e introducir la terminología necesaria, para luego describir las aplicaciones en cuestión. Esta taxonomía distingue los distintos sistemas en base a las características de los agentes individuales, así como de los aspectos de la comunidad de agentes. Se analizan las motivaciones que pueden existir en un ambiente industrial para elegir una determinada arquitectura de agentes y describe un amplio conjunto de aplicaciones en concreto, mostrando como éstas pueden guiar las distintas etapas del proceso industrial, desde la especificación del proyecto hasta la implementación del mismo.

[Pereira et al., 1991] Pereira, L. M., Aparício, J.Ñ., y Alferes, J. J. (1991). Nonmonotonic Reasoning with Well Founded Semantic. En Furokawa, K., editor, *Proceedings of the 8th International Conference on Logic Programming*, págs. 475–489.

Este artículo describe minuciosamente una metodología de representación de conocimiento a través de la semántica bien fundada de los programas lógicos. Se discuten políticas para representar reglas estrictas, hechos estrictos, reglas condicionales, excepciones a las reglas condicionales, etc. También se reconstruyen diversos problemas tomados de la literatura mediante la metodología brindada, y se realiza una contrastación con respecto a los principales formalismos de representación de conocimiento de aquella época.

[Pollock, 1987] Pollock, J. L. (1987). Defeasible Reasoning. *Cognitive Science*, 11(4):481–518.

Este artículo postula una acabada teoría epistémica, basada en la argumentación, capaz de modelar el razonamiento de tipo rebatible. Se establece una correlación entre la noción de razonamiento rebatible estudiada en los círculos filosóficos y la noción de razonamiento no monótono analizada en el seno de la Inteligencia Artificial. La teoría propuesta intenta capturar el mismo conjunto de conclusiones que un razonador idealizado estaría dispuesto a aceptar. A tal efecto se intoduce una caracterización inductiva que apela a la idea de argumentos activos por niveles. Cabe señalar que esta caracterización influenció notablemente otros desarrollos en el area [Simari y Loui, 1992, Vreeswijk, 1993b]. También se delinea como implementar esta teoría, abordando los principales desafíos que involucra esta tarea.

[Pollock, 1992] Pollock, J. L. (1992). How to reason defeasibly. *Artificial Intelligence*, 57(1):1–42.

Este artículo describe la construcción de un razonador rebatible de propósito general basado en una actualización de la teoría argumentativa formulada previamente por el autor [Pollock, 1987]. Uno de los principales resultados obtenidos fundamenta la imposibilidad de emplear un demostrador de teoremas tradicional como base para elaborar un razonador rebatible, ya que las conclusiones sancionadas por este tipo de razonador no resultan recursivamente enumerables. Se postula un criterio de correctitud alternativo para evaluar la fidelidad de una dada implementación en relación a la teoría modelada. Por último, se comprueba formalmente que bajo ciertas restricciones razonables el razonador rebatible delineado satisface el nuevo criterio propuesto.

[Pollock, 1994] Pollock, J. L. (1994). Justification and Defeat. *Artificial Intelligence*, 67(2):377–407.

Este artículo estudia el comportamiento de las principales teorías de razonamiento no monótono ante ciertas situaciones paradigmáticas tales como aquellas estipuladas por la paradoja de la lotería o la paradoja del prefacio. Se critica el pobre desempeño obtenido en estas circunstancias por la lógica default [Reiter, 1980], la circunscripción [McCarthy, 1980] e incluso la teoría argumentativa del propio autor [Pollock, 1987].

Se postula que estas paradojas deben recibir diferente tratamiento una de otra, incluso a pesar de presentar estructuras similares. A tal efecto, se describe una modificación en la semántica de su teoría argumentiva con el objeto de lidiar adecuadamente con estas situaciones, abandonando la asignación única de estados a los argumentos en favor de una asignación múltiple.

[Pollock, 1995] Pollock, J. L. (1995). *Cognitive Carpentry: A Blueprint for How to Build a Person*. The MIT Press.

Este libro describe minuciosamente la arquitectura de un razonador rebatible de propósito general denominado OSCAR, desde sus basamentos filosóficos hasta su implementación en un lenguaje de programación, recopilando las principales conclusiones obtenidas por el autor a lo largo de sus investigaciones. Se analiza con especial detenimiento la maquinaria de inferencia de OSCAR, la cual está basada en la última manifestación de la teoría argumentativa elaborada por el autor [Pollock, 1994]. Se enfatiza el rol de la capacidad de planificar en todo agente inteligente que aspire a presentar un comportamiento racional. En este sentido se explora la posibilidad de incorporar un planificador de propósito general a OSCAR, delineando un mecanismo de construcción de planes que apela tan sólo a la maquinaria de inferencia preexistente.

[Poole, 1985] Poole, D. L. (1985). On the Comparison of Theories: Preferring the Most Specific Explanation. En *Proceedings of the 9th International Joint Conference on Artificial Intelligence*, págs. 144–147.

Este artículo explora la idea de modelar al razonamiento por defecto como si fuera una teoría científica, donde las suposiciones por defecto empleados constituyen las hipótesis sobre las que descansa la teoría. En este contexto, para obtener resultados satisfactorios es necesario contar con un criterio de comparación entre teorías, con el objeto de resolver las situaciones donde resulten viables teorías científicas fundamentando tesis opuestas. A tal efecto, se introduce una comparador de teorías que combina formalmente dos intuitivos principios: favorecer las explicaciones que recurran a información más especíca, y preferir los razonamientos más directos. Este comparador de teorías es actualmente conocido como el criterio de especificidad, el cual ha sido bien recibido en el seno de la argumentación rebatible como potencial criterio para contrastar argumentos [Simari y Loui, 1992, García, 1997, García, 2000].

[Poole, 1988] Poole, D. L. (1988). A Logical Framework for Default Reasoning. *Artificial Intelligence*, 36(1):27–47.

Este artículo presenta una formalización del razonamiento rebatible denominada sistema THEORIST, que se caracteriza por adoptar una postura contrapuesta a la tendencia de la época. A diferencia de otras propuestas elaboradas sobre la base de extender la lógica clásica (*e.g.*, la lógica default [Reiter, 1980], las lógicas modales [McDermott y Doyle, 1980, Moore, 1985], etc.), este acercamiento interpreta al razonamiento rebatible como un proceso de formación de teorías sustentadas por conjuntos de hipótesis. El formalismo bosquedajo guarda la forma de aquel introducido en [Poole, 1985], pero, a diferencia de éste, no incorpora el comparador de teorías por especificidad. El argumento principal expuesto a lo largo de este trabajo fundamenta que no es necesario recurrir a una nueva lógica para razonar de manera no monotona, ya que es posible alcanzar el mismo comportamiento a través del razonamiento hipotético.

[Prakken, 1993] Prakken, H. (1993). *Logical Tools for Modelling Legal Argument*. Tesis Doctoral, Vrije Universiteit, Amsterdam, Holanda.

Esta Tesis explora el punto de contacto existente entre la Inteligencia Artificial y el razonamiento legal. Se postula que los métodos de representación de conocimiento y razonamiento desarrollados en el seno de la Inteligencia Artificial pueden constituir las bases de una formalización del razonamiento legal. A tal efecto, se estudian los principios fundacionales detrás de la mayoría de los sistemas de razonamiento no monótono de aquella época, por caso la lógica default [Reiter, 1980], la circunscripción [McCarthy, 1980], el sistema Simari-Loui [Simari y Loui, 1992] y el sistema argumentativo de Lin y Shoham [Lin y Shoham, 1989] entre otros. Sobre la base de este análisis se formula un sistema argumentativo basado en el lenguaje de la lógica default, que en teoría reconcilia la esencia de los sistemas considerados. La semántica del sistema se estipula en términos de un operador de punto fijo. También se estudian diversos criterios de comparación entre argumentos, analizando sus virtudes y defectos, e indicando como incorporar esta posibilidad en el sistema argumentativo propuesto.

[Prakken, 2001] Prakken, H. (2001). Relating Protocols for Dynamic Dispute with Logics for Defeasible Argumentation. *Synthese*, 127(1/2):187–219.

Este artículo presenta una instanciación concreta del modelo de disputa dialéctica introducido por Ron Loui [Loui, 1998]. Se formula un marco de carácter dialéctico el cual luego es reinterpretado como base de un protocolo de interacción entre agentes. La característica distintiva de este acercamiento es permitir que la base de conocimiento a partir de la cual cada agente razona pueda ser modificada dinámicamente mientras se produce la interacción. El autor extrapola las nociones de sensatez y justicia al contexto de éstos protocolos de disputa, con el objeto de constrastarlos con las teorías de pruebas dialécticas. Del análisis de estas propiedades el autor sostiene que determinados protocolos que funcionan adecuadamente en el contexto de una base de conocimiento estática no satisfacen estas propiedades al contemplar una base de conocimiento dinámica. Por último, se introduce un nuevo protocolo de disputa para el cual es posible asegurar las propiedades de sensatez y justicia aún estando en presencia de información cambiante.

[Prakken y Sartor, 1996] Prakken, H. y Sartor, G. (1996). A System for Defeasible Argumentation, with Defeasible Priorities. En Gabbay, D. M. y Ohlbach, H. J., editores, *Proceedings of the International Conference on Formal and Applied Practical Reasoning (FAPR)*, págs. 510–524, Bonn, Alemania. Springer-Verlag.

Este artículo describe un sistema argumentativo que conjuga el lenguaje de la programación extendida [Gelfond y Lifschitz, 1990] con una de las semánticas introducidas por Phan M. Dung [Dung, 1993a]. El lenguaje de representación de conocimiento es altamente fléxible ya que cuenta con dos tipos de negación, la negación estricta y la negación default. La noción de argumento se caracteriza como el encadenamiento de reglas requerido para derivar una conclusión. Se define al conjunto de argumentos justificados en términos de un operador de punto fijo. Cabe señalar que una característica distintiva del sistema es la habilidad de derivar rebatiblemente al propio conjunto de prioridades que debe ser empleado al contrastar argumentos.

[Prakken y Sartor, 1997] Prakken, H. y Sartor, G. (1997). Argument-based extended logic programming with defeasible priorities. *Journal of Applied Non-classical Logics*, 7:25–75.

Este artículo elabora sobre la base de un artículo anterior [Prakken y Sartor, 1996], incorporando nuevas características al sistema argumentativo allí presentado. Este formalismo combina el lenguaje de la programación en lógica extendida [Gelfond y Lifschitz, 1990] con una semánticas basada en un operador de punto fijo [Dung, 1993a]. La característica distintiva de este sistema es su habilidad para representar prioridades entre reglas que permite inducir un criterio de comparación entre argumentos. Como innovación en relación al trabajo anterior, la base de conocimiento distingue dos subconjuntos disjuntos, uno representando el conocimiento estricto y otro representando el conocimiento rebatible. Cabe señalar que esta distinción ya había sido considerada en los sistemas Simari-Loui [Simari y Loui, 1992], así como en su reformulación dialéctica [Simari et al., 1994]. También se incluye un sistema de prueba dialéctico como alternativa a la caracterización semántica basada en un operador de punto fijo.

[Prakken y Vreeswijk, 2002] Prakken, H. y Vreeswijk, G. (2002). Logics for defeasible argumentation. En Gabbay, D. y Guenthner, F., editores, *Handbook of Philosophical Logic*, volumen 4, págs. 218–319. Kluwer Academic Publishers.

Este trabajo presente una minuciosa recorrida por el trabajo realizado hasta la actualidad en la comunidad de razonamiento no monótono, con especial atención al campo de la argumentación rebatible. Se enfatiza la relevancia tanto en la Inteligencia Artificial como en la Filosofía de los principales acercamientos que intentan formalizar el razonamiento no monótono. Luego, se presenta un tratamiento abstracto de las características genéricas de los formalismos de argumentación rebatible. Este tratamiento abstracto luego es instanciado en el contexto de los principales sistemas de argumentación rebatible introducidos en la literatura [Simari y Loui, 1992, Pollock, 1995, Vreeswijk, 1997, Prakken y Sartor, 1997]. Finalmente, se abordan los principales interrogantes abiertos existentes en el área, los cuales de acuerdos a los autores, deberían ser abordados en trabajos futuros.

[Rao y Georgeff, 1991] Rao, A. S. y Georgeff, M. P. (1991). Modeling Agents Within a BDI-Architecture. En *Proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning*, págs. 473–484.

Este artículo presenta una formalización para el modelo BDI, que posee varias diferencias con respecto al acercamiento presentado previamente por Cohen y Levesque [Cohen y Levesque, 1990], por caso que la noción de intención tiene la misma preponderancia que las nociones de creencias y de deseos, y por tanto no puede ser reducida a éstas. Los autores conciben a las intenciones como planes parciales que el agente adopta a fin para alcanzar sus metas. Este modelo filosófico se formaliza mediante una lógica modal con operadores para las creencias, metas e intenciones del agente, junto con su correspondiente semántica de mundos posibles. Este trabajo a su vez explora la definición de un proceso de revisión de intenciones, que permite relacionar las intenciones actuales del agente con sus acciones futuras. Este proceso puede interpretarse intuitivamente como una estrategia de compromiso propia de cada agente, donde distintos agentes poseerán estrategias posiblemente diferentes. Finalmente, se consideran las distintas propiedades exhibidas por la lógica introducida.

[Rao y Georgeff, 1995] Rao, A. S. y Georgeff, M. P. (1995). BDI Agents: from theory to practice. En Lesser, V., editor, *Proceedings of the First International Conference on Multi-Agent Systems*, págs. 312–319, San Francisco, CA. MIT Press.

Este artículo aborda las cuestiones teóricas y prácticas que surgen del modelo BDI presentado en [Rao y Georgeff, 1991]. En primer lugar argumenta sobre las dos principales críticas que sufrió este modelo, en primer lugar el poseer a las creencias, deseos e intenciones a un mismo nivel y en segundo lugar la utilidad de estudiar lógicas multi-modales que no poseen una axiomatización completa y no son eficientemente computables. Cabe destacar que para responder al segundo cuestionamiento se estudia como es posible implementar un agente que siga el modelo BDI desde una perspectiva práctica. La aplicación elegida para en este caso es un sistema para el manejo de tráfico aereo denominada OASIS. Sin embargo, los autores destacan que para emprender esta tarea es necesario realizar algunas suposiciones que simplifiquen el formalismo teórico, sacrificando poder expresivo. Es por tanto cuestionable la relación existente entre el modelo BDI y la aplicación resultante.

[Reiter, 1980] Reiter, R. (1980). A Logic for Default Reasoning. *Artificial Intelligence*, 13(1–2):81–132.

Este artículo constituye la primer formalización de la lógica default. Sorprendentemente, el autor se ve obligado a sustentar el caracter no monótono del razonamiento por defecto, ya que por aquellos tiempos aún se cuestionaba el hecho de resignar la monotonía. La lógica default extiende a un conjunto de fórmula bien formadas de alguna lógica clásica mediante un conjunto de defaults, reglas de inferencia válidas por defecto. La semántica de este sistema es capturada mediante un operador cuyos puntos fijos denotan los conjuntos de creencias sancionados, denominados extensiones. Paradójicamente, es posible formular teorías defaults que no posean extensiones. Este inconveniente es resuelto circunscribiendo la atención a una subclase, las teorías default normales, que entre otras propiedades asegura la existencia de extensiones. No obstante, con posterioridad se comprobó que tales teorías son incapaces de modelar prioridades entre defaults [Antoniu, 1996], situación que desalentó su adopción.

[Rescher, 1977] Rescher, N. (1977). *Dialectics: A Controversy-Oriented Approach to the Theory of Knowledge.* State University of New York Press, Albany, Estados Unidos.

Este libro constituye una alabanza al rol desempeñado por la dialéctica en la investigación científica. Se define un juego dialógico en donde, con el objeto de analizar la aceptabilidad de una cierta tesis, un proponente y un oponente intercambian razones a favor y en detrimento de ésta. El juego dialógico resultante se define en términos de un conjunto de reglas que describen las movidas consideradas válidas para cada situación. Rescher utiliza este modelo de disputación para explorar otros interesantes aspectos como la terminación de las disputas y el concepto de "peso de prueba". Se estudia la noción de presuposición, reconociendo su caracter rebatible y se postula a la plausibilidad como criterio de preferencia entre presuposiciones. Uno de los resultados alcanzados más importantes es la identificación del isomorfismo entre la dialectica multilateral (la formación de consenso entre varias partes) y la dialectica unilateral (la formación de consenso introspectivo).

[Russell, 1997] Russell, S. J. (1997). Rationality and Intelligence. *Artificial Intelligence*, 94(1–2):55–78.

Este artículo desarrolla una caracterización de la racionalidad, concepto usualmente tomado por sentado, aún cuando su formalización precisa presenta serios desafíos. El autor afirma que el campo de la Inteligencia Artificial se vería beneficiado ante la creación de una definición de inteligencia que sea al mismo tiempo precisa y general, que englobe la gran diversidad de entidades que poseen esta propiedad. En este sentido, se analiza la noción de *racionalidad*, identificando cuatro formas diferentes: la perfecta, la calculativa, la del metanivel y la acotada. En particular, una característica común a estas cuatro definiciones es que en todas ellas la inteligencia de una entidad está asociada a su capacidad de alcanzar un comportamiento exitoso.

[Searle, 1970] Searle, J. R. (1970). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press.

Este libro constituye en punto de introducción de la conocida teoría de los actos de habla. El autor manifiesta que el acto ilocutorio [Austin, 1962] es la mínima unidad de la comunicación lingüística humana. Estos actos ilocutorios tienen una intencionalidad asociada, denominada presunción comunicativa. Un acto ilocutorio se considera exitoso si la intención de quien habla es reconocida por quien escucha. El autor destaca que los actos ilocutorios generalmente tiene como intención primaria lograr un efecto perlocutorio. Finalmente, se clasifican los tipos de intenciones en cinco categorías diferentes, que resumen qué tipo de tareas pueden realizarse a través del lenguaje.

[Simari et al., 1994] Simari, G. R., Chesñevar, C. I., y García, A. J. (1994). The Role of Dialectics in Defeasible Argumentation. En *Proceedings of the XIV Conferencia Internacional de la Sociedad Chilena para Ciencias de la Computación*, págs. 111–121, Concepción, Chile. Universidad de Concepción.

Este artículo presenta una extensión al sistema Simari-Loui [Simari y Loui, 1992], incorporando el tratamiento de dos novedosos aspectos: una reformulación dialéctica del proceso de justificación, y la identificación de un tipo de razonamiento no deseado, basado en argumentación falaz, con el objeto de evitar la sanción de conclusiones sin sustento. El proceso de búsqueda de justificaciones reemplaza la noción original de niveles de argumentos activos por una estructura mucho más intuitiva denominada árbol de dialéctica. Por otra parte, el razonamiento basado en argumentación falaz es evitado mediante la imposición de restricciones sobre las líneas de argumentación (cada uno de los caminos que componen un árbol de dialéctica) que serán aceptadas como válidas.

[Simari y Loui, 1992] Simari, G. R. y Loui, R. P. (1992). A mathematical treatment of defeasible reasoning and its implementation. *Artificial Intelligence*, 53(2–3):125–157.

Este artículo reseña los principales resultados alcanzados en la Tesis Doctoral de G. Simari, presentando un acabado análisis del razonamiento rebatible desde una óptica matemática, basada en la confrontación de argumentos. Constituye una de las primeras formalizaciones de la argumentación rebatible. Se presenta un formalismo capaz de expresar tanto información concreta como información tentativa, que combina las principales ideas de J. Pollock [Pollock, 1987] y D. Poole [Poole, 1985]. Esta presentación, a diferencia de otros acercamientos dejados de informalidad, prueba formalmente diversas propiedades satisfechas por el álgebra de argumentos inducido por el sistema.

[Smith y Cohen, 1995] Smith, I. A. y Cohen, P. R. (1995). Toward a Semantics for a Speech Act Based Agent Communication Language. En *Proceedings of the CIKM Workshop on Intelligent Information Agents*, págs. 24–31, New York, USA.

Este artículo propone una semántica para un lenguaje de comunicación entre agentes basada en las habilidades de los agentes para interactuar entre sí. En este modelo cuando dos o más agentes racionales desean comunicarse forman un equipo de trabajo. Luego, los actos comunicativos básicos (esto es, las performativas que caracterizan a la teoría de actos de habla), se reformulan como la creación de una intención en común para ese equipo. La teoría propuesta adopta un lenguaje particular de comunicación, para el cual se describe la semántica de cada acto comunicativo en base al estado mental del agente que lo formula, señalando como estos actos se pueden usar para crear y disolver grupos de agentes.

[Stankevicius, 2001] Stankevicius, A. G. (2001). Modelling Multiagent Deliberation from an Abstract Standpoint. En *Proceedings del 8vo Workshop sobre Aspectos Teóricos de la Inteligencia Artificial (ATIA), 3er Workshop de Investigadores en Ciencias de la Computación (WICC)*, págs. 74–76, San Luis, Argentina. Universidad Nacional de San Luis.

Este resumen extendido delinea una estrategia que permite sacar partido del conjunto de desarrollos teóricos ya conducidos en el seno de la argumentación rebatible con el objeto último de obtener un modelo abstracto formal para la forma de deliberación que se sucita en los sistemas multiagente. En particular se rescata la reciente tendencia de reintroducir caracterizaciones netamente dialécticas para las teorías argumentativas, incluso para aquellas que inicialmente contaron con otro tipo de semántica basado ya sea en puntos fijos o en especificaciones recursivas. Por otra parte se señalaron otras propuestas presentes en la literatura que han ensayado un camino análogo, tales como [Loui, 1998] o [Prakken, 2001].

[Stankevicius y García, 1999] Stankevicius, A. G. y García, A. J. (1999). Modelling Negotiation Protocols in a Dialectical Framework. En *Proceedings del 6to Workshop sobre Aspectos Teóricos de la Inteligencia Artificial (ATIA), 1er Workshop de Investigadores en Ciencias de la Computación (WICC)*, págs. 69–76, San Juan, Argentina. Universidad Nacional de San Juan.

Este trabajo introduce la noción de apelar a un modelo dialéctico de argumentación rebatible como marco conceptual para guiar la evolución de una negociación entre dos partes. El principal aporte consiste en la definición de un protocolo para gobernar el intercambio de información entre las partes involucradas, elaborado tomando como punto de partida la caracterización dialéctica del sistema de Simari y Loui presentada en [Simari et al., 1994]. Se identifican los principales desafíos que deben ser atendidos a la hora considerar un dominio de aplicación en concreto, por caso la selección de un adecuado criterio de comparación entre las razones expuestas, el grado de coherencia entre las creencias propias y las ajenas o incluso los diversos grados de confianza entre las partes involucradas. Por otra parte también se explora el abanico de alternativas existentes para lidiar con cada uno de estos desafíos, señalando tanto sus virtudes como sus defectos.

[Stankevicius y García, 2000] Stankevicius, A. G. y García, A. J. (2000). Could negotiation among agents be regarded as an argumentative process? En *Proceedings del 7mo Workshop sobre Aspectos Teóricos de la Inteligencia Artificial (ATIA), 2do Workshop de Investigadores en Ciencias de la Computación (WICC)*, págs. 22–25, La Plata, Argentina. Universidad Nacional de La Plata.

Este resumen extendido fundamenta las razones que sugieren que la negociación entre agentes inteligentes puede ser reinterpretada como si fuera simplemente un proceso argumentativo. La intuición expuesta consiste en reconocer el caracter dialógico de toda negociación, que en esencia abarca el intercambio de razones a favor de la posición propia y en desmedro de la posición del interlocutor de turno. Bajo el enfoque descripto, la negociación ciertamente se asemeja a la mayoría de los procesos argumentativos de base dialéctica. A su vez también se revisitan otras propuestas presentes en la literatura que siguen este mismo espíritu.

[Stankevicius y Simari, 2000] Stankevicius, A. G. y Simari, G. R. (2000). A Framework for Multiagent Deliberation Based on Dialectical Argumentation. En *Proceedings del 1er Workshop en Agentes y Sistemas Inteligentes (WASI), 6to Congreso Argentino de Ciencias de la Computación (CACiC)*, págs. 1469–1480, Ushuaia, Argentina. Universidad Nacional de la Patagonia San Juan Bosco.

Este artículo elabora sobre la base delineada en trabajos anteriores tales como [Stankevicius y García, 1999] y [Stankevicius y García, 2000], donde el interés inicial en modelar el proceso de negociación se troca por una forma más sutil de interacción entre agentes: *la deliberación*. Entre otras cosas se caracteriza formalmente el marco de conceptual en cual se conduce la deliberación, así como el respectivo protocolo de intercambio de información necesario para llevarla adelante, presentando las definiciones introducidas un grado suficiente de especificación como para poder ser puestas en práctica con un mínimo esfuerzo adicional. Por otra parte, a los efectos de conducir una deliberación se identifican cuatro etapas principales: la toma de la decisión de deliberar con otro agente, el contacto inicial con el contendiente de turno, el desarrollo de la deliberación en sí misma y finalmente la incorporación a la base de conocimiento del resultado obtenido. Cabe acotar que del análisis de estas etapas se desprende que contrariamente a lo esperado es la última de éstas la que supone un desafío de consideración, en particular para el contendiente que no logre imponerse en la deliberación, ya que posiblemente para aceptar el resultado de la misma deba apelar a alguna forma de revisión de creencias.

[Stankevicius y Simari, 2001] Stankevicius, A. G. y Simari, G. R. (2001). An Abstract Model for the Process of Deliberation within Multiagent Systems. En *Proceedings del 2er Workshop en Agentes y Sistemas Inteligentes (WASI), 7mo Congreso Argentino de Ciencias de la Computación (CACiC)*, págs. 1017–1026, El Calafate, Argentina. Universidad Nacional de la Patagonia Austral.

Este artículo en principio desarrolla la estrategia tendiente a obtener un modelo abstracto para el proceso de deliberación entre agentes inicialmente bosquejada en [Stankevicius, 2001]. En este sentido a partir de un análsis de los principales exponentes de la argumentación rebatible se postula un modelo que factoriza las principales características presentes en las teorías consideradas. El modelo resultante, de carácter abstracto, permite ser instanciado de forma tal de describir fielmente el comportamiento de las principales teorías para la argumentación rebatible. Se introducen las definiciones necesarias para brindar el andamiaje teórico que permite establecer el concepto más importante: el conjunto de tesis sustentadas por una dada instanciación del modelo abtracto. Este modelo del proceso de deliberación en presencia de un agente único es posteriormente reinterpretado como un modelo para el proceso de deliberación ahora en presencia de múltiples agentes. Finalmente, al instanciar el marco conceptual resultante se pueden obtener diversos protocolos de deliberación entre agentes, los cuales cubren un amplio espectro de escenarios.

[Stone y Veloso, 2000] Stone, P. y Veloso, M. M. (2000). Multiagent Systems: A Survey from a Machine Learning Perspective. *Autonomous Robots*, 8(3):345–383.

Este artículo presenta un analísis de los sistemas multiagente existentes, reseñando fundamentalmente las técnicas para el desarrollo de los mismos desde la perspectiva del área de aprendizaje automático. La presentación se organiza como una serie de categorías o escenarios propicios para el uso de sistemas multiagente. Existen diversas formas de clasificar los sistemas multiagente, en este caso los autores eligen realizarlo en función de dos dimensiones principales: el grado de heterogeneidad y el volumen de la comunicación entre los agentes. Para cada escenario resultante de la combinación de estos dos atributos se describe una instanciación apropiada a fin de ilustrar las características del mismo, reseñando los acercamientos previos introducidos previamente en el área.

[Sycara et al., 1991] Sycara, K. P., Roth, S. F., Sadeh, N., y Fox, M. (1991). Resource Allocation in Distributed Factory Scheduling. *IEEE Expert 6*, 1(6):29–40.

Este artículo introduce un sistema de control de producción multiagente heterogeneo y descentralizado, donde las distintas áreas de manufactura se modelan como agentes que realizan la planificación distribuida de las tareas de producción. Estos agentes se ejecutan en distintas estaciones de trabajo y se comunican entre sí mediante un mecanismo de pasaje de mensajes. Para formalizar el algoritmo de planificación distribuida, se identificaron un conjunto de heurísticas utilizadas en los algortimo de búsqueda con restricciones de los sistema centralizados. Estas heurísticas fueron posteriormente adaptadas a un escenario distribuido, en el cual el proceso de planificación es realizado por el grupo de agentes, los cuales cuentan con un conocimiento limitado del ambiente, así como de las restricciones e intenciones de los otros agentes. Cabe destacar que en este sistema ningún agente posee una visión global del problema. Las soluciones se construyen intercalando los computos locales con la comunicación con otros agentes. Finalmente, los autores citan los resultados empíricos obtenidos que aseguran la viabilidad del sistema propuesto.

[Toulmin, 1958] Toulmin, S. (1958). *The Uses of Argument.* Cambridge University Press.

Este libro desarrolla un acabado análsis acerca de los diversos roles que desempeñan los argumentos desde una óptica filosófica. Como principal contribución se indentifican cuatro componentes esenciales a todo argumento: la evidencia, la conclusión, las normativas y los fundamentos. Se postula que la conclusión es alcanzaba a partir de la evidencia por virtud de las normativas y que éstas se encuentra sustentadas en los fundamento. Se estudian los diversos tipos de argumentos en función de la interacción entre sus componentes. Dos tipos de argumentos reciben una mayor atención: los análiticos y los demostrablemente válidos. Se postula que la argumentación únicamente en términos de argumentos análiticos carecerá de completitud, enfatizando la importancia de los argumentos demostrablemente válidos.

[Verheij, 1996] Verheij, H. B. (1996). *Rules, Reasons, Arguments: Formal studies of argumentation and defeat.* Tesis Doctoral, Universiteit Maastricht, Maastricht, Holanda.

Esta Tesis analiza los principales aspectos del proceso de la argumentación, prestanto particular énfasis a su carácter rebatible. Se introduce la lógica basada en razones, una formalización del razonamiento basado en razones característico de los ámbitos legales, en donde es posible declarar con exactitud las razones que fundamentan la preferencia de un cierto patrón de inferencia por sobre otro al formar una conclusión. Se define un sistema argumentativo de espectro más amplio denominado CumulA. Este sistema permite la coordinación de argumentos en adición a la usual subordinación. Este técnica, más conocida como "acopio de argumentos", permite construir argumentos más persuasivos al acumular múltiples razones para una misma conclusión. La semántica del sistema es caracterizada mediante el novel concepto de etapas, donde cada etapa captura el efecto de tener en cuenta un nuevo argumento sobre el estado de los argumentos considerados en etapas anteriores.

[Vreeswijk, 1993a] Vreeswijk, G. A. W. (1993a). Defeasible Dialectics: A Controversy-Oriented Approach towards Defeasible Argumentation. *Journal of Logic and Computation*, 3(3):317–334.

Este artículo postula que la dialéctica rebatible representa un modelo lo suficientemente universal como para cubrir a la argumentación rebatible en su totalidad. Se introduce la noción de debate rebatible, esencialmente una disputa dialéctica entre un proponente y un oponente acerca de una cierta tesis. Se presenta una versión preliminar de los sistemas argumentativos abstractos [Vreeswijk, 1993b] como modelo de la argumentación rebatible en general, adecuando la disputación dialéctica a sus peculiaridades. Como principal resultado se prueba la equivalencia bajo la suposición de finitud entre la semántica tradicional basada en la ecuación de punto fijo y la derivada de prevalecer en un debate rebatible.

[Vreeswijk, 1993b] Vreeswijk, G. A. W. (1993b). *Studies in Defeasible Argumentation.* Tesis Doctoral, Vrije Universiteit, Amsterdam, Holanda.

Esta Tesis estudia en profundidad las características generales de la argumentación rebatible, con el propósito último de identificar aquellos aspectos recurrentes. En aporte central de este estudio consiste en la definición de los *sistemas argumentativos abstractos*, un formalismo argumentativo en el cual su alto grado de abstracción redunda en un marcado incremento de su aplicabilidad práctica. En este contexto, puede apreciarse una cierta similitud entre este acercamiento y aquel previo de Lin y Shoham [Lin y Shoham, 1989], si bien el autor ha extendido a esta propuesta incorporando un mecanismo de contrastación entre argumentos, aspecto notablemente ausente en el formalismo de partida. Se proponen diversas semántica para el sistema desarrollado, incluyendo una con un cariz dialéctico. En partular, a partir de esta última se explora el potencial de los debates como un mecanismo computacionalmente viable para determinar la semántica de un sistema argumentativo abtracto.

[Vreeswijk, 1997] Vreeswijk, G. A. W. (1997). Abstract Argumentation Systems. *Artificial Intelligence*, 90(1–2):225–279.

Este artículo analiza minuciosamente los fundamentos de los sistemas argumentativos abstractos [Vreeswijk, 1993b], un formalismo de razonamiento basado en argumentos cuyo principal objetivo es proveer un marco unificador en donde estudiar los principios que gobiernan la argumentación rebatible en general. Este formalismo se basa en un lenguaje sin especificar, construyendo argumentos como agregados de reglas de inferencia. Los argumentos se comparan en base a un orden de fuerza conclusiva previamente suministrado. El estado final de los argumentos se determina a partir de dos semánticas, una inductiva y la otra basada en una ecuación de punto fijo. Atendiendo al no constructivismo de la última, se explora la noción de secuencia argumentativa, un procedimiento tendiente a caracterizar la semántica basada en la ecuación de punto fijo de manera constructiva.

[Wooldridge et al., 1996] Wooldridge, M., Bussmann, S., y Klosterberg, M. (1996). Production Sequencing as Negotiation. En *Proceedings of the First International*

Conference on the Practical Application of Intelligent Agents and Multi-Agent Technology, págs. 709–726, London, UK.

Este artículo presenta una aplicación de la negociación entre agentes al problema del secuenciamiento de la producción de una fábrica. Este problema consiste en decidir en qué orden procesar los productos a fin de satisfacer los requerimientos y minimizar los costos de elaboración. Los autores proponen utilizar un sistema multiagente en el cual cada unidad de producción está representada por un agente, los cuales negocian sobre las posibles secuencias de productos a ser producidos, a fin de distribuir los costos en forma equitativa y minimizar el costo total de su fabricación. Se describe un mecanismo de negociación adaptado para este escenario, analizando las suposiciones y simplificaciones que resultan necesarias para poder implementarlo. Cabe destacar que la propuesta de negociación presentada resulta interesante desde un punto de vista teórico, pero posee varias limitaciones que dificultan su adaptación a escenarios en concreto.

[Wooldridge y Ciancarini, 1999] Wooldridge, M. y Ciancarini, P. (1999). Agent-oriented Software Engineering. En *Handbook of Software Engineering and Knowledge Engineering*, págs. 1–28. World Scientific Publishing Company.

Este artículo destaca a los agentes inteligentes o racionales como una herramienta adecuada para la construcción de sistemas de software complejos, en los que la interacción desempeñe un rol preponderante. En estos sistemas la noción de agencia adoptada surge como una extensión natural del concepto de objeto, con la salvedad de que los objetos son entidades pasivas, pero los agentes se caracterizan por ser proactivos e interactuar con su entorno. Para sustentar esta idea se abordan las técnicas de especificación orientadas a agentes. Los agentes son modelados como sistemas intencionales, a los cuales se les atribuyen estados mentales tales como creencias, metas e intenciones. Posteriormente, se discute como implementar las especificaciones por este medio obtenidas. Se contemplan dos posibilidades: ejecutar directamente la especificación obtenida, o bien compilarla, esto es, tranformarla en un modelo computacional concreto por medio de un proceso de síntesis automático.

[Zlotkin y Rosenschein, 1996] Zlotkin, G. y Rosenschein, J. (1996). Mechanisms for Automated Negotiation in State Oriented Domains. *Artificial Intelligence Research*, 1(5):163–238.

Este artículo introduce un modelo de negociación dependiente del dominio. Los autores plantean una clasificación de los tipos de interacciones por medio de las cuales sería posible elegir, dado un determinado dominio, qué estratégias y mecanismos resultan apropiados. Se define también un protocolo de negociación unificado, que pueda ser utilizado en cualquier tipo de encuentro. El acercamiento brindado exige que los agentes deben ser económicamente racionales, en el sentido de que siempre intentan maximizar su propia función de desempeño. En este escenario, resultan propicias las herramientas provistas por la teoría de juegos. Finalmente, se considera una de las principales críticas a las propuestas que siguen una visión económicamente racional, esto es, el hecho de que cada agente deba poseer total conocimiento sobre las metas y utilidades de los agentes involucrados, considerńdose algunos protocolos alternativos para escenarios con información incompleta sobre las metas o las utilidades de los agentes.