# Negociación mediante argumentación en sistemas multi-agente

Carles Sierra[†], Nick R. Jennings[‡], Pablo Noriega[†], Simon Parsons[‡]

[†]Institut d'Investigació en Intel·ligència Artificial, IIIA.
Consell Superior d'Investigacions Científiques, CSIC.
Campus UAB, 08193 Bellaterra, Catalunya, España.
{sierra, pablo}@iiia.csic.es

[‡]Department of Electronic Engineering,
Queen Mary and Westfield College,
University of London, London E1 4NS, UK.
{S.D.Parsons,N.R.Jennings}@qmw.ac.uk

26 de mayo de 1998

**Resumen**

Una gran cantidad de agentes autónomos operan en campos donde la cooperación de sus compañeros agentes no se puede garantizar. En tales situaciones la negociación será esencial para persuadir a los demás del valor de la cooperación. Este artículo describe un marco general de negociación en el que los agentes intercambian propuestas respaldados por argumentos que resumen las razones por las que las propuestas deberían ser aceptadas.

## 1 Introducción

La negociación es una forma clave de interaccción en sistemas compuestos por múltiples agentes autónomos. En tales entornos, los agentes no siempre tienen un control prestablecido los unos sobre los otros, de manera que la única forma posible que un agente tiene de influir en el comportamiento de los demás es a través de la persuasion. En algunos casos el persuadido puede necesitar poca, o ninguna persuasión para actuar tal y como desea el persuasor, por ejemplo porque la acción propuesta es consistente con sus planes. En otros casos el persuadido pude resistirse a aceptar inicialmente la propuesta y debe ser persuadido para cambiar sus creencias, sus objetivos o sus preferencias de manera que la propuesta, o alguna de sus variantes, sea aceptada. En cualquier caso el requerimiento mínimo de negociación es que los agentes puedan intercambiarse propuestas. Estas propuestas pueden ser aceptadas o rechazadas como por ejemplo en el caso del protocolo "Contract net" [17]. Otro nivel de sofisticación ocurre cuendo los receptores no tienen únicamente la opción de aceptar o rechazar las propuestas, sino que pueden hacer contraofertas para alterar aspectos de la propuesta que les parecen insatisfactorios [15]. Una forma aun más elaborada de negociacion —basada en la argumentación— es aquella en la que las partes pueden mandar justificaciones o argumentos junto a las (contra) propuestas, indicando por que deberían ser aceptadas [11, 13, 19]. Argumentos tales como: "ésta es mi oferta final, tómala o déjala", "la última vez este trabajo costó 10euros, no voy a pagar 20euros ahora", y "el trabajo tardará más en ser terminado porque un trabajador está enfermo" pueden ser necesarias para cambiar los objetivos o las preferencias del persuadido.

Este artículo trata de la negociación basada en la argumentación. Dado que este es un tema de investigación muy amplio [9, 20], nos limitaremos aquí a presentar las ideas básicas de nuestra área de investigación en argumentación entre agentes computacionales donde un persuasor intenta convencer a un persuadido para que realice una tarea particular (servi-

cio) en su nombre. Presentaremos los componentes de un modelo formal para los procesos de negociación basados en argumentación que puede ser utilizado, en última instancia, para construir agentes negociadores en aplicaciones reales. Aunque nos inspiramos en anteriores trabajos nuestros [15] en esta área, en este artículo cambiamos nuestro foco de atención de los mecanismos que sirven para generar contra-propuestas y para generar e interpretar argumentos hacia los aspectos sociales de la negociación. Además, aprovechamos los trabajos sobre Marcos Dialógicos presentados en [12] para definir los aspectos estáticos de los procesos de negociacion: ontología compartida, relaciones sociales, lenguaje de comunicación y protocolo. Definimos una noción de *estado* de un agente que reproduce el carácter evolutivo de la negociación —permitiendo que el modelo resultante reconozca distintos tipos de argumentos que pueden aportar los agentes para apoyar sus propuestas. Finalmente indicamos como los agentes pueden generar e interpretar dichos argumentos.

En este artículo tratamos tres tipos de ilocuciones: (i) *amenazas* —el no aceptar la propuesta implica que algo negativo le ocurrirá al agente; (ii) *recompensa* —la aceptación de la propuesta significa que algo positivo le va a ocurrir al agente; y (iii) *apelaciones* —v. g. el agente debería preferir esta opción a aquella alternativa por esta razón. Sabemos que se trata de un subconjunto de las ilocuciones involucradas en negociacion persuasiva, (ver [9] para una lista basada en resultados de investigación en psicología), pero nuestro interés se centra en la descripción de un marco en el cual se puedan describir los componentes fundamentales de la argumentación, no en formalizar exhaustivamente todos los tipos de argumento que se pueden encontrar en la bibliografía. La contribución principal de este trabajo es, por lo tanto, presentar un marco formal en el que los agentes puedan realizar una negociación persuasiva para cambiar las creencias y preferencias de los demás, utilizando un lenguage de comunicacion expresivo. Además, el marco conceptual es neutral con respecto a la arquitectura interna de los agentes e impone pocas restricciones en sus recursos formales.

# 2 Modelo de negociación

Nuestro modelo describe el proceso de negociación de un único encuentro entre múltiples agentes. Los acuerdos se realizan siempre entre dos agentes, aunque un agente puede encontrarse simultaneamente negociando con varios agentes. La negociación se consigue a través de un intercambio de ilocuciones en un lenguage de comunicación compartido $CL$. El intercambio real de ilocuciones viene dirigido por las necesidades y los objetivos *individuales* de los agentes —algo que no forma parte del modelo de negociación. De todas formas, éste intercambio está sujeto a algunas *convenciones mínimas compartidas*:

1. Los elementos relevantes para la negociación de un punto del acuerdo —en forma de *características* y *valores* que pueden evolucionar a lo largo del proceso de negociación,
2. La racionalidad de los agentes participantes. Expresada como relaciones de preferencia o funciones de utilidad que permitan a los agentes la evaluación y comparación de diferentes propuestas,
3. La capacidad deliberativa de los agentes. En forma de *estados* internos en los que el agente representa la historia de la negociación, y
4. El significado mínimo compartido de las ilocuciones. Expresado en cómo las ilocuciones *recibidas* por un agente deben ser interpretadas, y haciendo explícitas las condiciones bajo las cuales un agente puede 'generar' una ilocución.

El conjunto mínimo de conceptos necesario para representar los componentes estáticos en negociación automática se presentan en la sección 2.1. Los componentes dinámicos —el *hilo de negociación* y el estado de negociación— se presentan en la sección 2.2. Los aspectos sociales relevantes para los argumentos persuasivos se trabajan en la sección 2.3, y, finalmente, el proceso de interpretar y generar ilocuciones se ejemplifica en la sección 2.4.

Finalmente, señalar que este artículo resume fundamentalmente un artículo presentado en ATAL97 [16].

## 2.1 Una ontología básica de negociación

Una negociación requiere que los agentes participantes en ella se comuniquen, y para que la comunicación sea no ambigua, cada agenter debe poseer un único identificador. Denotaremos el conjunto de identificadores de los agentes participando en una negociación como *Agentes*. Los agentes implicados en una negociación mantienen entre ellos un repertorio de relaciones sociales. Estas relaciones tienen un impacto notable en los procesos de persuasión y argumentación. Por ejemplo, los conferenciantes de prestigio poseen una capacidad de persuasión mayor y los colegas pueden ser persuadidos más fácilmente que aquellos que no lo son [9]. De cara a modelizar estas relaciones, asumimos que se define una relación social general y compartida entre los agentes sobre un conjunto de roles sociales denotados por *Roles*. Finalmente, asumimos que los agentes al negociar intercambian ilocuciones en un lenguaje común $CL$ definido sobre un conjunto de partículas ilocutorias y cuyo contenido proposicional está expresado en un lenguaje lógico compartido $L$[1]. La naturaleza concreta de $L$ no es relevante en nuestro modelo (v.g. podría ser un lenguaje proposicional o un lenguaje modal) aunque debe contar como mínimo con los elementos siguientes:

1. *Variables.* Para representar las características, o puntos, de los acuerdos. Necesitaremos variables ya que las características tomarán valores diferentes en las diferentes propuestas a lo largo de la negociación.

2. *Constantes.* Para representar los valores de las carcterísticas. Tambien necesitaremos una constante especial '?' para representar la ausencia de valor, y permitir propuestas parciales, es decir propuestas que no fijen valor para alguna de sus variables.

3. *Igualdad.* La relación que se debe establecer entre variables y constantes para definir un acuerdo.

4. *Conjunción.* Para definir un acuerdo como un conjunto de igualdades entre variables y constantes.

Un ejemplo de acuerdo sería por los tanto:

$$(Precio = 10euros) \wedge (Calidad = Alta) \wedge$$

$$\wedge (Penalización = ?)$$

donde '$Precio$', '$Calidad$', y '$Penalización$' son las características del acuerdo a negociar y por lo tanto se representan como variables; '$10euros$', '$Alta$', y '?' son valores para tales características y por lo tanto representados como constantes; '=' denota la relación de igualdad; y '$\wedge$' denota la conjunción. A pesar de que estos elementos puedan parecer suficientes, no permiten describir todo lo necesario en un proceso de negociación. En particular, para 'razonar' y 'argumentar' sobre las ofertas es necesario, como mínimo, tener algún mecanismo de definir preferencias entre las ofertas. Las ofertas son fórmulas en $L$, por lo tanto la manera más obvia de representar preferencias entre fórmulas es como una relación de segundo orden en $L$. Esto significaría que $L$ sería una lógica de orden superior, con los problemas computacionales asociados [7]. Por ello, preferimos expresar las preferencias como un meta-lenguaje con los siguientes requerimientos mínimos:

1. *Funciones de codificación.* Para representar las formulas de $L$ como términos en $ML$.

2. *Preferencias.* Un meta-predicado para expresar las preferencias entre formulas de $L$.

Por ejemplo, dadas las fórmulas $Precio = 10euros$, y $Precio = 20euros$ en $L$, podemos expresar la preferencia de la primera sobre la segunda de la manera siguiente:

$$Pref(igual(\lceil Precio \rceil, \lceil 10euros \rceil),$$

$$igual(\lceil Precio \rceil, \lceil 20euros \rceil))$$

donde '$igual$' es la codificación en $ML$ del predicado '=' en $L$, y '$Pref$' es el meta-predicado de preferencias. En el resto del artículo, en lugar de escribir la codificación de las expresiones como: $equal(\lceil Precio \rceil, \lceil 10euros \rceil)$ usaremos una representación más compacta: $\lceil Precio = 10euros \rceil$.

[1] En la práctica, los agentes necesitan a menudo lenguajes diferentes y para interoperar deben utilizar alguna técnica de traducción entre los lenguajes [6, 8]. En este trabajo asumimos el caso simple de que todos compartan un único lenguaje.

El lenguaje de comunicación común, $CL$, contiene el conjunto de partículas ilocutorias necesarias para modelizar el conjunto de actos ilocutorios entre agentes que negocian y argumentan. Los partículas se pueden dividir, por lo tanto, en dos conjuntos, $I_{nego}$ para las partículas de negociación (es decir, las necesarias para hacer ofertas y contraofertas) y $I_{pers}$ para las partículas persuasivas (las usadas en argumentación). En nuestro modelo estos conjuntos son: $I_{nego}$ = {`ofrecer`, `solicitar`, `aceptar`, `rechazar`, `retirar`}, $I_{pers}$ = {`apelación`, `amenaza`, `premio`}. Aunque podríamos pensar en introducir otras partículas en $CL$, el conjunto presentado aquí es suficiente para nuestro objetivo en este artículo.

Un dialogo de negociación entre dos agentes consiste en una secuencia de ofertas y contraofertas conteniendo valores para los puntos del acuerdo. Estas ofertas y contraofertas pueden ser simplemente conjunciones de pares '*característica* = *valor*' contenidas dentro de un acto de habla cuya partícula ilocutoria es `ofrecer`) o pueden venir acompañadas de argumentos persuasivos (`amenaza`, `premio`, `apelación`). 'Persuasión' es un término general que engloba todos los actos de habla por medio de los cuales los agentes intentan cambiar los objetivos y creencias de los otros agentes. La selección de tres partículas ilocutorias en el conjunto $I_{pers}$ es el resultado del análisis de algún dominio real [16], así como del estudio de la bibliografía de persuasión [9, 19]. `apelación` es una partícula con un significado muy amplio, ya que hay muchos tipos diferentes de apelación. Por ejemplo, un agente puede apelar a la autoridad, a la práctica habitual o al interés del agente [19]. La estructura del acto de habla es $\texttt{apelación}(a, b, \xi, [not]\varphi, t)$, donde $\varphi$ es el argumento —una fórmula en $L$ o $ML$, o bien una ilocución en $CL$— que el agente $a$ comunica a $b$ para apoyar la fórmula $\xi$ (que, a su vez, es una fórmula en $L$ o $ML$). Todos los tipos de apelación siguen esta estructura. Las diferentes categorías de apelación se consiguen variando $\varphi$ en $L$ o $ML$, o variando $[not]\varphi$ en $CL$ —*not* $\varphi$ se interpreta como que la acción $\varphi$ no ocurre. `amenaza` y `premio` son más simples ya que tienen un conjunto de interpretaciones posibles más reducido. Su estructura, $\texttt{amenaza}(a, b, [not]\psi_1, [not]\psi_2, t)$ y $\texttt{premio}(a, b, [not]\psi_1, [not]\psi_2, t)$ es recursiva ya que las fórmulas $\psi_1$ y $\psi_2$ pueden ser a su vez ilocuciones en $CL$. Esta definición recursiva permite un conjunto rico de acciones ilocutorias en apoyo de la persuasión. Por ejemplo, el agente $a$ puede amenazar a $b$ con comunicar su incompetencia a su jefe, $c$.

$\texttt{amenaza}(a, b,$
$\quad not\ \texttt{aceptar}(b, a, tiempo = 24h, t_2),$
$\quad \texttt{apelación}(a, c, b = incompetente,$
$\quad\quad not\ \texttt{aceptar}(b, a, tiempo = 24h, t_2), t_3), t_1)$

Una vez introducidos todos los componentes, podemos describir ahora nuestro marco dialógico para negociación persuasiva.

**Definition 1** *Un* Marco dialógico *es una tupla* $DF = \langle Agentes, Roles, R, L, ML, CL, Tiempo\rangle$, *donde*

1. *Agentes es un conjunto de identificadores de agente.*
2. *Roles es un conjunto de identificadores de rol social.*
3. $R : Agentes \times Agentes \rightarrow Roles$, *asigna un rol social a cada par de agentes. La estructura social de un conjunto de agentes puede verse, por lo tanto, como un grafo dirigido etiquetado.*
4. *L es un lenguaje lógico*[2] *que satisface los requerimientos mencionados anteriormente.* $Acuerdos(L)$ *denota el conjunto de todas las posibles fórmulas conjuntivas en L sobre igualdades entre características y valores, es decir* $x_1 = v_1 \wedge ... \wedge x_n = v_n$. $Acuerdos_{?-libre}(L) \subset Acuerdos(L)$ *excluye '?' como valor aceptable en un acuerdo.*
5. *ML es un meta-lenguaje sobre L que satisface los requerimientos mencionados anteriormente.*
6. *CL es el lenguaje para la comunicación entre agentes. Dados* $a, b \in Agentes$ *y* $t \in Tiempo$, *CL se define como:*
   (a) *Si* $\delta \in Acuerdos(L)$ *entonces* $\texttt{solicitar}(a, b, \delta, t) \in CL$.

[2]Dado que queremos ser neutrales respecto a una arquitectura de agente en particular, no nos comprometemos con ningún lenguaje formal específico. $L$ puede ser tan simple como un lenguaje proposicional, o tan elaborado como una lógica multi-modal BDI [10, 14].

(b) *Si* $\delta \in Acuerdos_{?-libre}(L)$ *entonces* `ofrecer`$(a, b, \delta, t)$, `aceptar`$(a, b, \delta, t)$, `rechazar`$(a, b, \delta, t) \in CL$.

(c) `retirar`$(a, b, t) \in CL$.

(d) *Si* $\psi_1, \psi_2 \in CL$, $\xi \in L \cup ML$, *y* $\varphi \in L \cup ML \cup CL$ *entonces* `amenaza`$(a, b, [not]\psi_1, [not]\psi_2, t)$, `premio`$(a, b, [not]\psi_1, [not]\psi_2, t)$, `apelación`$(a, b, \xi, [not]\varphi, t) \in CL$.

7. *Tiempo es un conjunto de instantes discreto y totalmente ordenado.*

La marca de tiempo que aparece como el último argumento en todas las ilocuciones será omitido cuando no haya ambigüedad.

Los agentes usan las ilocuciones en $CL$ conforme al siguiente protocolo de negociación (ver figura 1):

1. Una negociación siempre empieza con una *propuesta de acuerdo*, en otras palabras con un `ofrecer` o `solicitar`. En ilocuciones de `solicitar` la constante especial '?' puede aparecer. Se interpreta esto como una petición del agente emisor al receptor de realizar una propuesta detallada substituyendo los '?' con valores concretos.

2. Esto es seguido por un intercambio de diversas contrapropuestas (que los agentes pueden `rechazar`) y muchas ilocuciones persuasivas.

3. Finalmente, se emite una ilocución *de cierre* que puede ser `acceptar` o `retirar`.

## 2.2 Agentes negociadores

El marco dialógico descrito en la sección anterior representa los componentes estáticos del modelo de negociación —aquellos que son fijos para todas las negociaciones. Esta sección presenta los elementos dinámicos —aquellos que cambian conforme avanza la negociación. A pesar de que nuestro modelo pretende ser neutral respecto de la arquitectura de los agentes, para capturar los aspectos esenciales de la persuasión es necesario asumir que los agentes tengan memoria y sean deliberativos. La memoria de *estados de negociación* se representa mediante un *hilo de negociación* [12] que contiene la historia del dialogo entre un par de agentes.

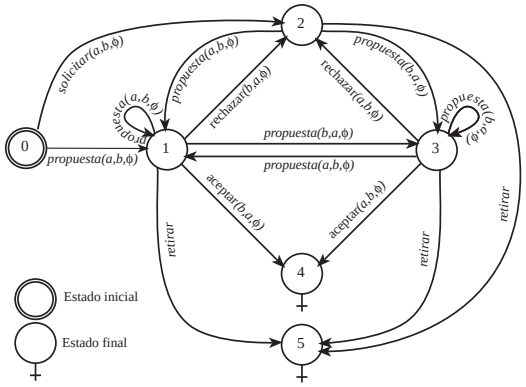


Figura 1: Protocolo de negociación. En las ilocuciones de `acceptar`$(x, y, \varphi)$ y `rechazar`$(x, y, \varphi)$ $\varphi$ se refiere siempre a la última propuesta recibida. $Propuesta(x, y)$ se refiere a cualquier ilocución construida con una de las partículas siguientes: `ofrecer`, `amenaza`, `premio`, `apelar`, y entre los agentes $x$ y $y$. Omitimos la marca de tiempo de las ilocuciones.

**Definition 2** *Un* **Hilo de Negociación** *entre los agentes* $a, b \in Agents$, *en el tiempo* $t_n \in Tiempo$, *denotado como* $X^{t_n}_{a \leftrightarrow b}$, *es cualquier secuencia finita de longitud* $n$ *de la forma* $(x^{t_1}_{a \to b}, x^{t_2}_{b \to a}, x^{t_3}_{a \to b}, \ldots)$ *con* $t_1, t_2 \ldots \leq t_n$, *donde:*

1. $t_{i+1} > t_i$, *la secuencia está ordenada en el tiempo,*

2. *Para cada característica* $j$, $x^i_{a \to b}[j] \in [min^a_j, max^a_j], x^{i+1}_{b \to a}[j] \in [min^b_j, max^b_j]$ *con* $i = 1, 3, 5, \ldots$, *y opcionalmente el último elemento de la secuencia es una de las partículas* $\{aceptar, retirar\}$.

*Decimos que un hilo de negociación está* **activo**[3] *si* $última(X^{t_n}_{a \leftrightarrow b}) \notin \{aceptar, retirar\}$, *donde última es una función que devuelve el último elemento de una secuencia.*

Queremos modelizar la idea, por otro lado muy intuitiva y realista, de que los puntos de negociación puedan variar a lo largo de la negociación. Esto es necesario porque consideramos

[3] Asumimos que una oferta es válida (es decir, que el agente que la emitió está comprometido a mantenerla) hasta que se recibe una contraoferta. Si el tiempo de respuesta es relevante, es possible incluirlo en el conjunto de puntos bajo negociación.

que una de las formas más claras en las que un agente puede persuadir a otro acerca de la aceptación de una propuesta concreta es introduciendo nuevos puntos en la propuesta que la hagan más atractiva. Esto implica que necesitamos una representación explícita del conjunto $\Omega$ de características conocidas por un agente. Las preferencias también evolucionan, bien porque $\Omega$ evoluciona o porque el agente es persuadido para cambiar sus preferencias. Por lo tanto, la teoría interna del agente, $T$, que incorpora sus preferencias como fórmulas en $ML$, y un conjunto de otras formulas en $L$ para modelizar el dominio, debe ser representada como un componente del estado del agente. En nuestro modelo no imponemos ningun requerimiento específico sobre $T$. Así pues tenemos la siguiente definición:

**Definition 3** *Un* **Estado de negociación** *de un agente $a$ en el instante $t$ es una tupla $s = \langle \Omega, T, H \rangle$, donde*

- *$\Omega$ es un conjunto finito de puntos negociables.*
- *$T \subseteq L \cup ML$, es una teoría en el lenguaje común.*
- *$H$, la historia de la negociación, es el conjunto de todos los hilos de negociación del agente $a$. Es decir, $H = \{X_{i \leftrightarrow a} | i \in Agentes\}$.*

Denotaremos por $S_a$ a todos los estados de negociación posibles de un agente $a$.

## 2.3 Agentes persuasivos

$CL$ permite la construcción de actos de habla, a partir de las partículas de $I_{pers}$, que contengan argumentos en favor de un acuerdo. El bloque básico de construcción de argumentos es $\texttt{apelación}(a, b, \xi, [not]\varphi, t)$ donde $a, b \in Agentes$, $\xi \in L \cup ML$, y $\varphi \in L \cup ML \cup CL$. La lectura de tal acto de habla es: "el agente $a$ quiere que el agente $b$ añada $\xi$ a su teoría actual; el argumento de soporte para ello es $[not]\varphi$". Los otros actos de habla persuasivos, $\texttt{amenaza}(a, b, [not]\psi_1, [not]\psi_2, t)$ y $\texttt{premio}(a, b, [not]\psi_1, [not]\psi_2, t)$ con $\psi_1, \psi_2 \in CL$, pueden contener argumentos ya que $\psi_1$ y/o $\psi_2$ son apelaciones, o pueden, recursivamente contener apelaciones.

La interpretación de un argumento persuasivo determina si el agente que lo recibe cambia o no su teoría. Para tomar una decisión, el agente considera los argumentos (eventualmente conflictivos) provenientes de otros agentes y de sí mismo —es decir, generados desde su propia teoría. Suscribimos la opinión de otros autores en sistemas multiagente [2] de que el rol social es un factor determinante en la toma de decisiones sobre qué argumento preferir. Por lo tanto proponemos derivar una relación de autoridad a partir de los roles sociales entre los agentes, para ser utilizada como mecanismo de comparación de argumentos. Los roles sociales que establecen una determinan una relación de poder dependen del dominio concreto. Para construir un grafo dirigido que represente la autoridad que los agentes ejercen los unos sobre los otros, tomamos el grafo asociado a la relación $R$ y eliminamos de él los arcos etiquetados con roles no considerados de poder. Así, tenemos la definición siguiente

**Definition 4** *Dado un marco dialógico $DF = \langle Agentes, Roles, R, L, ML, CL, Tiempo \rangle$ y un conjunto de roles de autoridad $Poder \subseteq Roles$, definimos el* grafo de autoridad, *$AG \subseteq Agentes \times Agentes$, para $DF$ como:*

1. *Si $R(a, b) \in Poder$ entonces $(a, b) \in AG$*
2. *Si $(a, b), (b, c) \in AG$ entonces $(a, c) \in AG$*

*Decimos que un grafo de autoridad está bien definido si es acíclico.*

El grafo de autoridad codifica la relación de autoridad —o falta de ella, dado que en general un $AG$ no está totalmente conectado— entre los agentes.

Nuestra postura en este trabajo es que el 'poder' de un argumento depende *exclusivamente* de la autoridad de los agentes que aportan formulas a su construcción. Para ello es necesario extender la noción de relación de autoridad entre agentes a una noción de autoridad entre conjuntos de agentes. Hay dos maneras obvias de hacerlo. Decimos que un conjunto de agentes $A$ tiene *menor autoridad mínima* que $B$,

$A \sqsubset_{\min} B$, si y sólo si para todos los $b \in B$ existe un $a \in A$ tal que $(b, a) \in AG$. Y que $A$ tiene *menor autoridad máxima* que $B$, $A \sqsubset_{\max} B$, si y sólo si para todos los $a \in A$ existe un $b \in B$ tal que $(b, a) \in AG$. Así, intuitivamente, el orden $\sqsubset_{\min}$ indica que si alguna fórmula utilizada en la construcción del argumento fué propuesta por alguien situado en un punto bajo del grafo de autoridad, el argumento es débil, mientras que $\sqsubset_{\max}$ indica que si alguna fórmula en un argumento es propuesta por alguien con mucha autoridad, el argumento es fuerte. Otras relaciones de autoridad son concebibles. Nos referiremos de manera genérica a cualquiera de ellas con el símbolo $\sqsubset$.

En su forma más general, un argumento es la prueba de una fórmula [1]. Aquí asumimos que los agentes poseen el mismo sistema deductivo para $L$ ($\vdash_L$) y $ML$ ($\vdash_{ML}$); ver [4] para una aproximación en la que esto no es así. Así pues, en este contexto restringido, una prueba puede ser representada como la conjunción de todas las formulas utilizadas en ella ya que la prueba puede ser reconstruida por el agente que recibe la conjunción. Un argumento es, por lo tanto, una fórmula $\varphi \in L \cup ML \cup CL$ construida a partir de fórmulas atómicas presentes en la teoría inicial del agente u obtenidas en encuentros de negociación previos con otros agentes. Suponiendo definida una función $Soporte : L \cup ML \cup CL \rightarrow 2^{Agentes}$ que devuelve los agentes cuyas fórmulas han contribuido en la construcción de un argumento, o el agente que emite la ilocución cuando $\varphi \in CL$, podemos fácilmente usar el rol social de esos agentes para decidir sobre la fuerza de los argumentos. La última noción a tener en cuenta para poder construir un sistema de decisión es la de que un argumento puede atacar a otro [3]. Representaremos el hecho de que una argumento $Arg$ apoya la fórmula $\varphi$ como un par $(Arg, \varphi)$, y el hecho de que la pareja $(Arg_1, \varphi_1)$ ataca a $(Arg_2, \varphi_2)$ como $Ataca((Arg_1, \varphi_1), (Arg_2, \varphi_2))$. El significado preciso de *Ataca* depende, esecialmente, de los lenguajes $L$ y $ML$ concretos que se usen. Aquí seguimos a Dung [3] en suponer que se trata de una noción primitiva, dado que nuestro interés reside en resolver el efecto de un *Ataca* independientemente de como éste sea definido.

**Definition 5** *Dados dos pares de argumentos* $(Arg_1, \varphi_1)$ *y* $(Arg_2, \varphi_2)$ *tales que* $Ataca((Arg_1, \varphi_1), (Arg_2, \varphi_2))$ *diremos que* $(Arg_1, \varphi_1)$ *es preferido a* $(Arg_2, \varphi_2)$*, notado como* $(Arg_2, \varphi_2) \prec (Arg_1, \varphi_1)$*, si y sólo si* $Soporte(Arg_2) \sqsubset Soporte(Arg_1)$*. Cuando* $(Arg_2, \varphi_2) \not\prec (Arg_1, \varphi_1)$ *y* $(Arg_1, \varphi_1) \not\prec (Arg_2, \varphi_2)$ *decimos que el agente es indiferente respecto a los argumentos,* $(Arg_1, \varphi_1) \sim (Arg_2, \varphi_2)$.

Los agentes usan la técnica de argumentación para decidir cómo interpretar las ilocuciones que reciben y para decidir qué ilocuciones generar. Cuando se recibe un par $(Arg_1, \varphi_1)$ que no es atacado por ningún argumento construible a partir de la teoría que en ese momento posee el agente, un agente *bonachón* podría simplemente añadir el argumento $Arg_1$ y la fórmula $\varphi_1$ a su teoría. En cambio, un agente más *conservador* podría no acceptar una proposición a no ser que proviniera de una autoridad superior. En general la idea es que cuando sea cierto $Ataca((Arg_1, \varphi_1), (Arg_2, \varphi_2))$ el agente conserve el par preferido (en el sentido definido arriba). Si $(Arg_1, \varphi_1) \sim (Arg_2, \varphi_2)$ se debe utilizar algún criterio adicional, como por ejemplo el 'epistemic entrenchment' [5].

## 2.4 Interpretación y generación de ilocuciones

Un agente realiza dos operaciones básicas sobre las ilocuciones. Una es interpretar las ilocuciones dirigidas a él con objeto de actualizar su estado mental. La otra es la generación de ilocuciones dirigidas a otros agentes del sistema multi-agente en el que participa. Presentamos en esta sección ambas operaciones como dos funciones $I$ y $G$.

En el caso que nos ocupa, negociación por argumentación, cualquier ilocución puede introducir nuevos puntos en la negociación, mientras que las apelaciones pueden, además, modificar las relaciones de preferencia y las teorías del agente. En cualquier caso, la interpretación de una ilocución depende estrictamente del agente que se diseñe y del dominio en que este se encuentre. No es posible tener una aproximación normativa al diseño de la función $I$. Podemos, eso sí, dar un ejemplo de cómo se podría construir tal función en el caso de un agente 'bonachón':

**Ejemplo** *Interpretación bonachona.* La función de interpretación $I$ de un agente bonachón se basa en las siguientes intuiciones:

- Cada ilocución extiende el hilo de negociación correspondiente. De esta manera, los agentes conservan una memoria completa de la negociación. Se podrían modelizar agentes 'olvidadizos' eliminando parte del hilo de negociación.
- Todas las ilocuciones pueden introducir nuevos puntos en la negociación.
- Las apelaciones pueden cambiar las relaciones de preferencia de un agente. Así mismo, pueden cambiar la teoría, añadiendole las fórmulas del argumento usado en la apelación, en caso de que la teoría actual del agente no pueda construir agumentos que ataquen la apelación.

Dado un lenguaje de comunicación $CL$, un marco dialógico $DF$ y un conjunto de estados de negociación possibles $S_b$ para un agente $b$, la función de interpretación de un agente 'bonachón' se define en la figura 2 como: $I : CL \times S_b \times DF \to S_b$ (siendo $s = (\Omega, T, H)$, $H = \{X_{i \leftrightarrow b} | i \in Agentes\}$, y '^' representando la concatenación[4]). □

Finalmente, la especificación de un agente $a$ debe incluir una manera de calcular la siguiente ilocución a emitir en un hilo de negociación. Es decir, una función $G : S_a \times DF \to CL$. Esta función debe respetar las restricciones impuestas por el protocolo presentado en la figura 1. Una manera cómoda de representar tal función es mediante un conjunto de reglas *condición-acción*, donde la acción corresponde, claro está, a una acción ilocutoria.

La manera en que un agente elige qué ilocución emitir depende de muchos factores: de la historia de la negociación, de los objetivos actuales del agente, o de su teoría, y también, claro está, de la manera en que el agente ha interpretado las ilocuciones recibidas. De la misma manera que en el caso de la interpretación, no podemos ser normativos, ya que cada diseño de agente contiene una 'personalidad' que determinará cómo realizar tales decisiones. Véase en la figura 3 un pequeño ejemplo de función $G$ para un agente 'obediente'. Un agente 'obediente' es aquél que siempre que le es posible realiza aquello que se le pide. En las reglas de decisión de la figura 3, 'self' representa al agente interpretando la ilocución.

[4] Una forma alternativa de mirar la interpretación de ilocuciones es como programas que transformas un estado en otro. Un formalismo natural para ello es el de la Lógica Dinámica [12].

# 3 Trabajos relacionados

Gran parte de los trabajos existentes en negociación basada en agentes se fundamenta en teoría de juegos, v. g. [18]. A pesar de que esta aproximación ha producido resultados significativos y ha resultado existosa en muchas áreas de negociación, encarna un número de supuestos acerca de los conocimientos y funciones de utilidad de los agentes que limitan su uso en muchas otras áreas. Incluso cuando la aproximación es extendida, como en [11], para tener en cuenta condiciones que cambian a lo largo del tiempo, no ataca el problema de cómo estos cambios pueden ser el resultado de la influencia mútua entre agentes, ni el problema de la introducción de nuevos puntos a lo largo de las negociaciones. El cambio de preferencias a través de persuasión, en sistemas multi-agentes, fué atacado en el trabajo de Sycara sobre negociaciones laborales [19], trabajo que fué extendido y formalizado por Kraus *et al.* [10]. Este trabajo se sitúa en el contexto de una arquitectura de agente particular, supone una teoría de dominio fija y compartida y trata de cinco tipos de argumento (amenazas, premios, apelación a precedentes, a práctica habitual, y al interés propio). Además, Kraus *et al.* no tratan la introducción de nuevos puntos ni racionalidad imperfecta. En cambio, nuestro modelo permite poseer conocimiento parcial, racionalidad imperfecta y la introducción de nuevos puntos en la negociación —que son propiedades relevantes en muchas áreas de aplicación— mientras que sólo impone requerimientos mínimos en los estados de los agentes, utilizando un lenguaje retórico general.

También queremos comentar las diferencias entre nuestro trabajo con aquellos en los que el

1 $I(\iota(a,b,\delta,t),s,df) = (\Omega \cup issues(\delta), T, H - X_{b\leftrightarrow a} + X'_{b\leftrightarrow a})$
**con** $\iota \in I_{nego}$; $X'_{b\leftrightarrow a} = X_{b\leftrightarrow a}\hat{}\iota(a,b,\delta,t)$

2 $I(\texttt{amenaza}(a,b,[not]\psi_1,[not]\psi_2,t),s,df) = (\Omega \cup issues(\psi_1) \cup issues(\psi_2), T, H - X_{b\leftrightarrow a} + X'_{b\leftrightarrow a})$
**con** $X'_{b\leftrightarrow a} = X_{b\leftrightarrow a}\hat{}\texttt{amenaza}(a,b,[not]\psi_1,[not]\psi_2,t)$

3 $I(\texttt{premio}(a,b,[not]\psi_1,[not]\psi_2,t),s,df) = (\Omega \cup issues(\psi_1) \cup issues(\psi_2), T, H - X_{b\leftrightarrow a} + X'_{b\leftrightarrow a})$
**con** $X'_{b\leftrightarrow a} = X_{b\leftrightarrow a}\hat{}\texttt{premio}(a,b,[not]\psi_1,[not]\psi_2,t)$

4 $I(\texttt{apelación}(a,b,\xi,[not]\varphi,t),s,df) = (\Omega', T', H - X_{b\leftrightarrow a} + X'_{b\leftrightarrow a})$
**con** $X'_{b\leftrightarrow a} = X_{b\leftrightarrow a}\hat{}\texttt{apelación}(a,b,\xi,[not]\varphi,t)$;
**si** no $(Arg,\psi)$ construido a partir de $T$ tal que $Attacks(([not]\varphi,\xi),(Arg,\psi))$
**entonces** $\Omega' = \Omega \cup issues(\xi) \cup issues(\varphi)$;
**si** $\varphi \in L \cup ML$ **entonces** $T' = T + \xi + \varphi$ **sino** $T' = T + \xi$
**sino** $\Omega' = \Omega$; $T' = T$

Figura 2: Función de interpetación de un agente bonachón

R1 **si** $última(X_{x\leftrightarrow self}) = \texttt{amenaza}(x, self, no\ \texttt{aceptar}(self, x, \delta), \psi_2)$ y $(x, self) \in AG$
**y** $puedo\_hacer(\delta)$ **entonces** $\texttt{aceptar}(self, x, \delta)$

R2 **if** $última(X_{x\leftrightarrow self}) = \texttt{amenaza}(x, self, no\ \texttt{aceptar}(self, x, \delta), \psi_2)$ **y** $(x, self) \in AG$
**y no** $puedo\_hacer(\delta)$ **entonces** $\delta' = calcular\_contra\_oferta(s, DF)$;$\texttt{ofrecer}(self, x, \delta')$

R3 **si** $última(X_{x\leftrightarrow self}) = \texttt{apelación}(x, self, \xi, \varphi)$ **y** $\psi \rightarrow \neg\varphi \in T$ **entonces** $\texttt{apelación}(self, x, \neg\varphi, \psi)$

Figura 3: Ejemplo de generación de ilocuciones de un agente obediente.

uso de la argumentación sirve para explicar cómo razona un agente. En éstos últimos, un agente argumenta consigo mismo para establecer sus creencias. En nuestro trabajo los argumentos se usan por un agente para cambiar las creencias y determinar las acciones de otros agentes. La otra diferencia importante es que el mecanismo de resolución de conflictos cuando los agentes argumentan consigo mismos esta cableado en el lenguaje lógico en el que se construyen los argumentos y se basa en nociones intuitivas acerca de qué es correcto en el mundo. En cambio, nosotros mantenemos este mecanismo en el meta-nivel y lo basamos en conocimientos acerca del dominio. Esto tiene la ventaja dual de asegurar que los conflictos se resuelven de una manera adecuada al dominio en concreto, pudiendo definir mecanismos de resolución que se implementan fácilmente como meta-teorías en $ML$.

# 4 Conclusión

Este artículo ha presentado un marco para describir negociaciones persuasivas entre agentes autónomos. Este marco proporciona una base sólida para construir agentes artificiales específicos instanciando los componentes genéricos $L$, $ML$ y $T$. El marco ha sido influido por nuestra experiencia en el desarrollo de aplicaciones y ello nos hacer creer que captura las necesidades reales de un amplio abanico de dominios. Hay, a pesar de ello, un conjunto de lineas de trabajo que nos gustaría seguir explorando. Primero está el punto de cuan expresivo debe ser $CL$. Por ejemplo, con el formalismo actual, un agente sólo puede hacer promesas o amenazas sobre acciones ilocutorias, es decir, acerca de 'decir' algo a alguien. Sería deseable que acciones no-ilocutorias fueran también la consecuencia de una amenaza o promesa. De manera similar, mientras que las apelaciones pueden ser utilizadas para modelizar un amplio rango de ilocuciones, sería interesante caracterizar tipos sutílmente diferentes de ilocución a través de una pormenorización de las funciones de interpretación y generación. Segundo, hemos expresado las preferencias de los agentes y los cambios en dichas preferencias simplemente como fórmulas y cambios en la teoría $T$. Se requiere más trabajo para ligar estas preferencias con nociones de racionalidad, en particular con ideas estándar de utilidad esperada. Finalmente, realizamos la simplificación de considerar que los agentes tienen un mecanismo de deducción común. Esto puede ser inadecuado para algunos dominios, en cuyo caso sería necesario que los agentes pudieran ser capaces de discutir sobre qué reglas de inferencia son las apropiadas.

## Referencias


[1] S. Benferhat, D. Dubois, and H. Prade. Argumentative inference in uncertain and inconsistent knowledge bases. In *Proc 9th Conf on Uncertainty in AI*, pages 411–419, Washington, USA, 1993.

[2] C. Castelfranchi. Social Power: A Point missed in Multi-Agent, DAI and HCI. In Y. Demazeau and J. P. Müller, editors, *Decentralised AI*, pages 49–62. Elsevier, 1990.

[3] P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77:321–357, 1995.

[4] P. Faratin, C. Sierra, and N. R. Jennings. Negotiation decision functions for autonomous agents. *Robotics and Autonomous Systems*, page (in press).

[5] P. Gärdenfors. *Knowledge in Flux*. MIT Press, Cambridge, MA, 1987.

[6] F. Giunchiglia and L. Serafini. Multilanguage hierarchical logics (or: How we can do without modal logics). *Artificial Intelligence*, 65:29–70, 1994.

[7] W. D. Goldfarb. The undecidability of the second-order unification problem. *Theoretical Computer Science*, 13:225–230, 1981.

[8] T. R. Gruber. The role of common ontology in achieving sharable, reusable knowledge bases. In J. A. Allen, R. Fikes, and E. Sandewall, editors, *Proc. of the Second Int. Conf. on Principles of Knowledge Representation and Reasoning*, San Mateo, CA, 1991. Morgan Kaufman.

[9] M. Karlins and H. I. Abelson. *Persuasion*. Crosby Lockwood & Son, London, UK, 1970.

[10] S. Kraus, M. Nirkhe, and K. Sycara. Reaching agreements through argumentation: a logical model (preliminary report). In *DAI Workshop'93*, pages 233–247, Pensylvania, USA, 1993.

[11] S. Kraus, J. Wilkenfeld, and G. Zlotkin. Multiagent negotiation under time constraints. *Artificial Intelligence*, 75:297–345, 1995.

[12] P. Noriega and C. Sierra. Towards layered dialogical agents. In *Proceedings of the ECAI'96 Workshop Agents Theories, Architectures and Languages, ATAL'96*, number 1193 in LNAI, pages 157–171. Springer, 1996.

[13] S. Parsons and N. R. Jennings. Negotiation through argumentation—a preliminary report. In *Proc. Second Int. Conf. on Multi-Agent Systems, ICMAS'96*, pages 267–274, Kyoto, Japan, 1996.

[14] A. S. Rao and M. P. Georgeff. BDI agents: From Theory to Practice. In *Proc 1st Int Conf on Multi-Agent Systems*, pages 312–319, San Francisco, USA, 1995.

[15] C. Sierra, P. Faratin, and N. R. Jennings. A service-oriented negotiation model between autonomous agents. In *MAAMAW'97*, number 1237 in LNAI, pages 17–35, Ronneby, Sweden, 1997.

[16] C. Sierra, N. R. Jennings, P. Noriega, and S. Parsons. A framework for argumentation-based negotiation. In M. P. Singh, A. Rao, and M. J. Wooldridge, editors, *Intelligent Agents IV*, number 1365 in LNAI, pages 177–192. Springer, 1997.

[17] R. G. Smith and R. Davis. Frameworks for cooperation in distributed problem solving. *IEEE Trans on Systems, Man and Cybernetics*, 11(1):61–70, 1981.

[18] J. S.Rosenschein and G. Zlotkin. *Rules of Encounter*. The MIT Press, Cambridge, USA, 1994.

[19] K. P. Sycara. Persuasive argumentation in negotiation. *Theory and Decision*, 28:203–242, 1990.

[20] D. N. Walton. *Informal Logic*. Cambridge University Press, Cambridge, UK, 1989.