

Un Modelo para la Conversación Multimodal Hombre-Máquina: Integración del Habla y la Acción

Luis Villaseñor, Manuel Montes

Laboratorio de Tecnologías del Lenguaje
Instituto Nacional de Astrofísica, Óptica y
Electrónica
Apdo. Postal 51 y 216, Puebla, Pue.
México
{villasen, mmontes}@inaoep.mx

Jean Caelen

Laboratoire de Communication Langagière et
Interaction Personne Système / Fédération IMAG
Campus Universitaire,
BP 53, 38041 Grenoble Cedex 9
France
jean.caelen@imag.fr

RESUMEN

Uno de los objetivos primordiales de las interfaces hombre-máquina es la simplificación de la tarea de comunicación propiamente dicha entre el hombre y la máquina. Hasta el momento el hombre ha estado condicionado al lenguaje de la computadora para comunicarse con ella. Con el advenimiento de las interfaces gráficas y, más recientemente, las interfaces multimodales –donde la interacción es cada vez más simple y *natural*– el esfuerzo en el uso mismo de la computadora ha disminuido considerablemente. Sin embargo, esta dificultad sigue presente y el usuario debe seguir esforzándose en comunicarse mientras busca la solución a su problema a través de la computadora. Nuestro trabajo se enfoca a mejorar la forma de comunicación entre el hombre y la máquina. Este artículo presenta las bases de un modelo para el diálogo multimodal basado en una lógica modal que integra el habla y la acción bajo un mismo esquema.

Palabras Clave

Diálogo cooperativo, actos del habla, comunicación hombre-máquina, interacción multimodal.

ABSTRACT

One major goal of human computer interfaces is to simplify the communication task. Traditionally, users have been restricted to the language of computers for this task. With the emerging of the graphical and multimodal interfaces – where the interaction is day by day easier and simpler– the effort required for working with a computer is decreasing. However, the problem of communication is still present, and users continue caring about the communication task when they deal with a computer. Our work focuses on improving

the communication between the human and the computer. This paper presents the foundations of a multimodal dialog model based on a modal logic, which integrates the speech and the action under the same framework.

Keywords

Cooperative dialogue, speech acts, human-computer communication, multimodal interaction.

INTRODUCCIÓN

Actualmente podemos distinguir dos grandes líneas de desarrollo en las interfaces hombre-máquina [8,12]. Por un lado, las interfaces de manipulación *directa* que adoptan una metáfora donde la máquina es considerada una herramienta virtual que el usuario emplea para realizar su tarea. Y por otro lado, las interfaces de manipulación *indirecta* que integran el lenguaje de manera que el usuario participa hablando con la máquina. En el primer caso, el usuario tiene la sensación de ser el responsable directo de la manipulación de los objetos, dando la impresión que la intermediación del sistema desaparece, al dar toda la libertad de acción posible al usuario. En el segundo caso, el usuario se transforma en interlocutor que habla de su tarea y la delega a la máquina, la cual adopta un papel de asistente. Por supuesto este asistente debe estar dotado de capacidades de comprensión del lenguaje natural y, a la vez, debe conocer las reglas del diálogo o de la conversación¹. Nuestro trabajo integra estos dos tipos de interfaces tomando como metáfora principal el diálogo cooperativo orientado por la tarea. Para esto, hemos desarrollado una lógica de diálogo a partir de los *actos del habla*. Los actos del habla pueden compararse a acciones indirectas *hacer-hacer* (de ahora en adelante *ff*) mientras que las operaciones directas a acciones simples *hacer* (*f*). De esta manera, la aportación principal de este modelo es integrar bajo un mismo esquema acciones solicitadas de manera verbal y las acciones realizadas con ayuda de un dispositivo de

¹ Para una revisión de los diferentes propuestas que existen para el control del diálogo véase [7].

designación directa. Esta teoría unifica estos dos tipos de acción a través de una lógica para el diálogo multimodal hombre-máquina.

El Lenguaje, la Acción y el Diálogo

« El lenguaje se construye por la acción » : éste es uno de los principales resultados de Piaget [9]. Para él, el niño construye el lenguaje como resultado de la asimilación de los efectos que sus acciones tienen sobre el mundo. Llamar a una persona « mamá » no tiene sentido a menos que ella acuda al grito de « mamá ».

Recíprocamente « El lenguaje construye la acción ». Esta es la tesis principal de la escuela anglosajona en filosofía del lenguaje [1,11]. En esta teoría, hablar es actuar, producir actos; comunicar, es actuar sobre el interlocutor. Esta teoría permite considerar al lenguaje como una forma de acción y, por generalización, considerar al diálogo como una secuencia de acciones planificadas que tienen por objetivo una *meta específica sustentada por una intención*. Esta concepción presupone que existe un equilibrio racional entre el conocimiento, las acciones y las intenciones del interlocutor. Por ejemplo, se espera que la serie de acciones que el usuario está haciendo coincida con la realización de sus intenciones y, por supuesto, que el usuario no adopte intenciones imposibles de llevar a cabo.

« El diálogo es una interacción: el lenguaje remite a la acción y recíprocamente ». El diálogo es una serie de acciones (incluyendo o no actos del habla) coordinadas para alcanzar una meta [17]. Esta meta debe ser a la vez alcanzada y satisfecha para que el diálogo concluya exitosamente [16]. A lo largo del diálogo los interlocutores intentan reducir sus diferencias a través de diferentes formas de demandas y de ofertas: preguntas, respuestas, objeciones, cuestionamientos, etc. Es a lo largo de esta interacción que los interlocutores modifican sus conocimientos y adquieren otros nuevos sobre la situación, su interlocutor o su lenguaje.

Acción o Lenguaje: ¿Cuál escoger?

Nuestra hipótesis es que para favorecer al máximo la interacción entre el hombre y la máquina, es necesario dar a ésta última capacidades de acción a través del lenguaje sin dejar de brindar la oportunidad al usuario de actuar por medio de la manipulación directa. Consideramos al diálogo como el medio ideal para la integración de estos dos tipos de interacción. Además, el diálogo nos permite tanto la conducción interactiva de la tarea como la adquisición de conocimientos relativos al lenguaje útiles a la acción. Así,

(a) de una parte, la máquina debe adquirir los conceptos manipulados a través del lenguaje y que son a menudo « naturales » (y por lo tanto implícitos) para el usuario, y

(b) por otro lado, ella debe aprender nuevos planes de acción en un contexto dado y de manera lo suficientemente genérica para ser reutilizables.

En efecto, nosotros no tenemos que distinguir, desde el punto de vista del usuario, una acción elemental que él realiza de manera directa y una acción compleja que él hace cuando usa la palabra y que podría llevarnos a una serie de acciones directas. El nivel de granularidad de una acción directa es por lo general más fino que aquél de un acto del habla. Por ejemplo para dibujar una casa, bastará con ordenar « dibuja una casa » mientras que un dibujo a través de una manipulación directa necesitaría una serie de acciones diversas para trazar el cuerpo de la casa, después el techo, las ventanas, etc. Así nosotros llegamos a distinguir dos niveles de granularidad en las acciones: las instrucciones elementales y los planes elaborados (o tareas). La orden « dibuja una casa » será asociada a un plan de acción, el cual estará compuesto de una serie de acciones elementales. De esta manera, la máquina podrá hacer nuevas acciones (una forma de macroinstrucciones) a partir de este conjunto de instrucciones elementales (una forma de instrucciones de base de un lenguaje de programación) asociándolas a nuevos conceptos como « dibuja », « casa », etc., sin embargo, nosotros no deseamos considerar el lenguaje natural como un tipo de super-lenguaje de programación ya que este enfoque reduciría el diálogo a una entrada interactiva de programas, y perdería su espontaneidad y su aspecto cooperativo (y adoptaría la estructura de los lenguajes artificiales).

Posición Teórica del Problema

Nuestro modelo se apoya sobre la « teoría de actos del habla » [1,11] que, entre los diferentes paradigmas sobre la comunicación, consideramos el más apropiado para un modelado del diálogo hombre-máquina. La hipótesis, sobre la que se basa nuestro modelo, sostiene que el diálogo está dirigido por los *estados mentales* que sustentan las creencias, las acciones y las intenciones del usuario. Este modelo no intenta atribuir una conducta de tipo humano a la máquina sino únicamente proporcionarle los elementos lógicos que fundamenten su comportamiento [2].

La Acción Inmediata y la Acción Diferida

Lo que distingue esencialmente una acción directa de una indirecta es que la primera, en la mayoría de las ocasiones, es inmediata (al oprimir el interruptor, la luz se enciende) y que sus efectos no son siempre reversibles. Por el contrario, una acción que involucra el habla pierde este carácter inmediato puesto que el acto de habla debe ser mediatizado por un agente activo. En este caso los efectos son reversibles durante un breve instante mientras el receptor todavía no ha reaccionado. Sin embargo, los efectos no son siempre predecibles ya que dependen del receptor en su decisión de actuar y en la selección de la acción a ejecutar: este agente puede no entender, realizar una maniobra equivocada, etc. En resumen, los factores que distinguen la acción directa de la acción hablada son:

- el tiempo de reacción entre la emisión de la orden y la obtención del resultado y,
- la pertinencia de los efectos obtenidos. La pertinencia tiene únicamente sentido con relación a las intenciones originales del emisor de la orden.

Así durante el diálogo se debe controlar la evolución de la meta. Entre el momento donde ésta es planteada y el momento donde es satisfecha pueden presentarse infinidad de eventos.

Una Lógica para el Diálogo

La lógica sobre la que dirigimos nuestros esfuerzos, intenta modelar los intercambios de palabra entre el usuario y la computadora usando el concepto de *actos de diálogo*. La idea principal es la integración del acto de diálogo bajo un marco lógico basado en la acción que permita explicar el desarrollo de un diálogo hombre-máquina. Este trabajo se inspira en los trabajos propuestos en [2, 4, 6, 10, 13, 14, 15]. Él contiene elementos de una lógica epistémica –para representar el conocimiento–, de una lógica dinámica –para describir la acción y los efectos producidos por ésta–, y de una lógica dialógica –para expresar las obligaciones, las interrupciones e incomprensiones durante los intercambios de palabra en función de las metas del diálogo.

CONCEPTOS DE BASE

El *conocimiento* está representado por el operador s (saber) p. e. la fórmula Usj expresa que el usuario sabe la proposición j (para hacer la distinción entre el usuario y la máquina, los dos posibles agentes de nuestra lógica, se utilizan las letras U para el usuario y M para la máquina). Para representar la *acción*, introducimos la noción de « evento »: un evento $Uf a$ es la realización de la acción a por el usuario (o $Mf a$ en el caso de la máquina) y este evento tiene la proposición j como resultado. Usando la notación de la lógica dinámica tenemos la fórmula siguiente $[Uf a]j$ (después de terminar la ejecución de a por el usuario, j es verdadero). Para representar la *intención* usamos el operador i , únicamente el usuario es capaz de actuar de manera intencionada, así la fórmula $Ui j$ expresa la intención del usuario de hacer j verdadero. Por último, una acción puede ser (i) una acción de base (instrucción elemental) o (ii) una tarea, es decir, una secuencia de acciones (organizada a través de un plan).

El Acto de Diálogo

Un acto de diálogo es una acción que causa un cambio de situación con respecto a la tarea (y/o del conocimiento que la máquina tiene sobre la tarea) y un cambio de la situación del diálogo. Así, un acto de diálogo está definido como un evento $[Uf a](j_a \dot{U} j_d)$ donde el resultado es el conjunto de cambios de situación con respecto a la tarea j_a y con respecto al diálogo j_d . Los cambios con respecto a la tarea son los efectos propios de la acción (j_a). Y los cambios con respecto al diálogo son las consecuencias que muestran el

avance en el diálogo a cada intercambio de palabra. En nuestro caso, la meta detrás de todo diálogo es la tarea que el usuario intenta realizar. De esta manera, cada acto de diálogo tiene como parte de su realización un efecto sobre la meta (j_d). Este efecto, que hace explícita la evolución de la meta durante el desarrollo del diálogo, está definido a través de los siguientes estados:

- ? meta planteada, acción a realizar
- + meta alcanzada, acción realizada sin confirmación, posiblemente correcta
- ++ meta satisfecha, acción terminada correcta
- @ meta abandonada

Los actos de diálogo son expresados a partir de acciones de la manera siguiente:

- $Uf \alpha, Mf \alpha$ el usuario o la máquina *hace* α
- $Uff \alpha, Mff \alpha$ el usuario *hace-hacer* la acción α a la máquina (o viceversa)
- $Ufs \phi, Mfs \phi$ el usuario *hace-saber* ϕ a la máquina (o viceversa)
- $Uffs \phi, Mffs \phi$ el usuario *hace-hacer-saber* ϕ a la máquina (o viceversa)

Donde la acción fs es una abreviatura para la acción elemental *compartir*: $Ufs \phi \equiv Uf \text{compartir } \phi$

DEFINICIÓN DEL LENGUAJE L_d

1. DEFINICIÓN. Si T es un conjunto enumerable de símbolos proposicionales, Ab un conjunto finito de acciones de base, U el símbolo para nombrar al usuario y M para la máquina, el lenguaje L_d está definido de la siguiente manera:

L_d es el conjunto más pequeño de T tal que:

- si $\phi, \psi \in L_d$ entonces $\neg\phi, \phi \vee \psi \in L_d$

- si $\alpha \in Ac$, y $\phi \in L_d$ entonces

$Us \phi, Ms \phi, [Uf \alpha]\phi, [Mf \alpha]\phi, Ui \phi \in L_d$

donde Ac es el conjunto más pequeño de Ab tal que:

- si $\alpha \in Ab$ entonces $\alpha \in Ac$

- si $\phi \in L_d$ entonces $verificar(\phi) \in Ac$

- si $\alpha \in Ac$ y $\beta \in Ac$ entonces $\alpha;\beta \in Ac$

Se utilizan las abreviaciones $\phi \wedge \psi$ para $\neg(\neg\phi \vee \neg\psi)$ y $\phi \supset \psi$ para $\neg(\phi \wedge \neg\psi)$. La abreviación *verdadero* se toma como alguna fórmula válida, p. e. $\phi \vee \neg\phi$, y *falso* como la abreviación de $\neg\text{verdadero}$.

2. DEFINICIÓN. Semántica para el lenguaje L_d . La clase \mathbf{M} de modelos de Kripke contiene todas las tuplas

$M = \langle S, \pi, R_U, R_M, I_U, r_U, r_M \rangle$ tal que

- i) S es un conjunto de mundos posibles, o estados.
- ii) π es una función que asigna valores de verdad a los símbolos proposicionales T en un mundo posible s .
($\pi(s) : T \rightarrow \{1,0\}$ para todo $s \in S$)

- iii) R_U es una relación binaria entre los mundos posibles de S , ésta es la relación de accesibilidad asociada al conocimiento del usuario. ($R_U \subseteq S \times S$)
- iv) R_M es una relación binaria entre los mundos posibles de S , ésta es la relación de accesibilidad asociada al conocimiento de la máquina. ($R_M \subseteq S \times S$)
- v) I_U es una relación binaria entre los mundos posibles de S , ésta es la relación de accesibilidad asociada a las intenciones del usuario. ($I_U \subseteq S \times S$)
- vi) r_U es la relación entre el conjunto de mundos posibles causados por la realización de una acción a hecha por el usuario en un mundo posible s .
($r_U : Ac \times S \rightarrow \wp(S)$)
- vii) r_M es la relación entre el conjunto de mundos posibles causados por la realización de una acción a hecha por la máquina en un mundo posible s .
($r_M : Ac \times S \rightarrow \wp(S)$)

3. DEFINICIÓN. Sea $M = \langle S, \pi, R_U, R_M, I_U, r_U, r_M \rangle$ un modelo de Kripke de la clase \mathbf{M} . El concepto de verdad de una proposición (\models) en un mundo posible s a partir del modelo M está definido inductivamente de la siguiente manera :

$M, s \models \varphi$	sii (si y solamente si) $\pi(s)(\varphi) = 1$ para $\varphi \in T$
$M, s \models \neg\varphi$	sii $M, s \not\models \varphi$
$M, s \models \varphi \vee \psi$	sii $M, s \models \varphi$ o $M, s \models \psi$
$M, s \models [Uf \alpha] \varphi$	sii $\forall s' [s' \in r_U(\alpha, s) \Rightarrow M, s' \models \varphi]$
$M, s \models [Mf \alpha] \varphi$	sii $\forall s' [s' \in r_M(\alpha, s) \Rightarrow M, s' \models \varphi]$
$M, s \models U_s \varphi$	sii $\forall s' [(s, s') \in R_U \Rightarrow M, s' \models \varphi]$
$M, s \models M_s \varphi$	sii $\forall s' [(s, s') \in R_M \Rightarrow M, s' \models \varphi]$
$M, s \models U_i \varphi$	sii $\forall s' [(s, s') \in I_U \Rightarrow M, s' \models \varphi]$

donde r_U y r_M , por brevedad r_A dada su equivalencia, están definidas por

$r_A(\text{verificar}(\varphi), s)$	$= \{s\}$ si $M, s \models \varphi$ $= \emptyset$ en caso contrario
$r_A(\alpha; \beta), s)$	$= r_A(\beta, r_A(\alpha, s))$

DEFINICIÓN DE AXIOMAS

En esta sección se presentan los axiomas de nuestra lógica. La primera parte muestra brevemente los axiomas que caracterizan el conocimiento, la segunda parte introduce los conceptos asociados a la evolución de la meta necesarios para precisar en la tercera parte, los axiomas que caracterizan el diálogo cooperativo.

Caracterización del Conocimiento

Los siguientes axiomas están orientados a caracterizar el conocimiento que la máquina tiene sobre las intenciones del usuario (p. e. la solicitud de ejecución de una tarea) y sobre la descripción misma de la tarea (el *saber-hacer* requerido para llevarla a cabo). Sin embargo, estos axiomas son usados de manera general tanto para especificar el

conocimiento de la máquina como el del usuario². Para $A = \{U, M\}$:

- (A1) $As \varphi \wedge As(\varphi \supset \psi) \supset As \psi$ *axioma K*
- (A2) $As \varphi \supset \varphi$
se conocen únicamente hechos verdaderos
- (A3) $As \varphi \supset As As \varphi$ *introspección positiva*
se conocen los hechos que uno conoce
- (A4) $\neg As \varphi \supset As \neg As \varphi$ *introspección negativa*
se conocen los hechos que no conocemos

El Desarrollo del Diálogo

En nuestro caso la estructura del diálogo está soportada por la intención del usuario. Suponemos que en un momento dado, un acto de diálogo de la parte del usuario expresa su intención. Esta intención es interpretada como la acción deseada a realizar (o el plan de acción a ejecutar) cuyo efecto es el estado por alcanzar. Así, la realización de un acto de diálogo provoca normalmente el acercamiento a la *meta* deseada, excepto si se trata de un acto que causa un cambio brusco en el desarrollo de la acción (abandono, puesta en espera, etc.). Este acercamiento a la meta depende de la situación actual y del acto de diálogo mismo. Los párrafos siguientes presentan esta evolución en detalle.

Meta Planteada por el Usuario

Una meta es *planteada* por el usuario cuando él ordena a la máquina ya sea la realización de una acción *Uff*, ya sea una solicitud de información *Uffs*. Esta meta es el estado inicial del proceso. Ella es colocada en un pila y sufrirá diferentes transformaciones. En el estado final, la meta es *alcanzada*, si la máquina responde *Mfs* o ejecuta la acción solicitada *Mf*. Ella será considerada *satisfecha* (y por consiguiente sacada de la pila) cuando el usuario exprese su opinión, de manera implícita o explícita, sobre la adecuación de la acción o respuesta obtenida. Una meta también puede ser *abandonada* por el usuario, cuando él informa de manera explícita su cambio de parecer. Para el caso donde el usuario solicita la realización de una acción *Uff*, se tienen las siguientes notaciones:

- i) Meta *planteada*, la abreviación $?Mf a$ designa la acción *Mf a* como meta planteada.

$$[Uff \alpha] (Ms Ui [Mf \alpha] \varphi) \equiv [Uff \alpha] (? [Mf \alpha] \varphi)$$

una meta planteada es el efecto de una solicitud y este efecto es expresado como la intención del usuario (de que la máquina realice una acción) integrada al conocimiento de la máquina.

- ii) Meta *alcanzada*, la abreviación $+Mf a$ designa la acción *Mf a* como meta alcanzada.

$$Ms Ui [Mf \alpha] \varphi \wedge [Mf \alpha] \varphi$$

² Para una presentación detallada véase el sistema S5 en [6].

$$\begin{aligned} &\equiv ? [Mf \alpha] \phi \wedge [Mf \alpha] \phi \\ &\equiv + [Mf \alpha] \phi \end{aligned}$$

una meta planteada es alcanzada cuando la acción solicitada por el usuario también es verdadera.

- iii) Meta *satisfecha*, la abreviación $++Mf a$ designa la acción $Mf a$ como meta satisfecha.

$$\begin{aligned} &Ms Ui [Mf \alpha] \phi \wedge ([Ufs U-i [Mf \alpha] \phi] \phi \vee \\ &\quad ([Uff \beta] \gamma \wedge \neg rel(\phi, \beta))) \\ &\equiv ? [Mf \alpha] \phi \wedge ([Ufs U-i [Mf \alpha] \phi] \phi \vee \\ &\quad ([Uff \beta] \gamma \wedge \neg rel(\phi, \beta))) \\ &\equiv ++ [Mf \alpha] \phi \end{aligned}$$

una meta planteada es considerada satisfecha si el usuario acepta la acción como apropiada, esta aceptación puede manifestarse de manera (a) explícita cuando el usuario informa que él no conserva más su intención con respecto a la acción solicitada; o de manera (b) implícita cuando el usuario solicita una nueva acción que no se deriva de la acción que se acaba de realizar, es decir, no existe ninguna relación entre la nueva acción solicitada y los efectos de la acción precedente.

- iv) Meta *abandonada*, la abreviación $@Mf a$ designa la acción $Mf a$ como meta abandonada.

$$\begin{aligned} &Ms Ui [Mf \alpha] \phi \wedge (Ufs Ui [M-f \alpha] \phi) \\ &\equiv ? [Mf \alpha] \phi \wedge (Ufs Ui [M-f \alpha] \phi) \\ &\equiv @ [Mf \alpha] \phi \end{aligned}$$

una meta planteada es abandonada cuando el usuario hace saber que él desea que la máquina no realice la acción en cuestión.

Para el caso cuando el usuario hace una pregunta $Uffs$, se tienen la misma evolución de la meta y se espera una respuesta del tipo Mfs ³. Es necesario recordar que ffs es una abreviación para $ffcompartir$ donde *compartir* es una acción de base. Así, una pregunta del usuario es una meta planteada $?Mfs f$ la cual es consecuencia de la acción $Uffcompartir$:

$$\begin{aligned} &[Uffs \phi] (Ms Ui [Mfs \phi] \phi) \\ &\equiv [Uffcompartir \phi] (Ms Ui [Mfcompartir \phi] \phi) \\ &\equiv [Uffcompartir \phi] (? [Mfcompartir \phi] \phi) \\ &\equiv [Uffs \phi] (? [Mfs \phi] \phi) \end{aligned}$$

Submeta Desencadenada por la Máquina

En nuestro marco propuesto para la cooperación hombre-máquina, la máquina carece de intención, sin embargo, es

³ Siempre es posible considerar un conjunto de categorías más específicas, por ejemplo: $Uffs_{si}$ para cuando el usuario realiza una pregunta si/no, o $Uffs_{ref}$ si el usuario hace una pregunta para conocer el valor de un parámetro en particular, véase [5, 18] para una clasificación completa.

posible que ella tome cierta iniciativa durante el diálogo con respecto a la tarea solicitada. El primer caso es cuando existe una ignorancia parcial de la tarea a realizar la cual desencadenará una o varias preguntas dirigidas al usuario. El segundo caso es cuando el conocimiento de la máquina entra en contradicción con la proposición afirmada por el usuario. Y, por último, el tercer caso es cuando el usuario muestra a la máquina una nueva tarea. En el primer caso es la intención del usuario el origen de la actividad de la máquina mientras que los otros dos son motivados como características indispensables de un diálogo cooperativo.

Pregunta planteada por la máquina: una pregunta es planteada por la máquina cuando la acción a ejecutar no está definida completamente. Esta situación provoca un subdiálogo para solicitar al usuario la información faltante. A diferencia de una meta planteada por el usuario, la evolución de una submeta planteada por la máquina cuenta únicamente con dos etapas. A una pregunta planteada por la máquina corresponde una respuesta, y ésta es inmediatamente aceptada ya que la máquina es incapaz de expresar opinión alguna sobre la adecuación de la respuesta. Así, si una submeta planteada por la máquina es alcanzada, ella es considerada al mismo tiempo como satisfecha. Para describir estos casos tenemos las siguientes abreviaturas:

- i) Submeta *planteada*, la abreviación $?Ufs f$ designa la acción $Ufs f$ como submeta planteada.
- $$[Mffs \phi] (Ms Ui [Ufs \phi] \phi) \equiv [Mffs \phi] (? [Ufs \phi] \phi)$$
- ii) Submeta *alcanzada y satisfecha*, la abreviación $++Ufs f$ designa la acción $Ufs f$ como submeta alcanzada.

$$\begin{aligned} &Ms Ui [Ufs \phi] \phi \wedge [Ufs \phi] \phi \\ &\equiv ? [Ufs \phi] \phi \wedge [Ufs \phi] \phi \\ &\equiv ++ [Ufs \phi] \phi \end{aligned}$$

una submeta planteada es alcanzada y, por la misma razón, satisfecha cuando la respuesta esperada por la máquina también es verdadera.

- iii) Submeta *abandonada*, la abreviación $@ Ufs f$ designa la acción $Ufs f$ como submeta abandonada.

$$\begin{aligned} &Ms Ui [Ufs \phi] \phi \wedge [Ufs [U-fs \phi] \phi] \gamma \\ &\equiv ? [Ufs \phi] \phi \wedge [Ufs [U-fs \phi] \phi] \gamma \\ &\equiv @ [Ufs \phi] \phi \end{aligned}$$

una submeta planteada es abandonada cuando el usuario informa, de manera explícita, que él no responderá a la pregunta planteada.

Meta desplazada en adquisición: otra situación donde la máquina toma la iniciativa y solicita la intervención del usuario es para aprender la tarea. El origen de esta situación se encuentra en la petición de una tarea desconocida por la máquina. Así la meta original es desplazada para dirigir el diálogo bajo un contexto de

adquisición. Para este caso, tenemos las abreviaturas siguientes:

- i) Meta *desplazada en adquisición*, la abreviación Δa designa como meta del diálogo la adquisición de la acción **a**.

$$[Mff \alpha](Ms Ui [Uf \alpha] Msf \alpha) \equiv [Mff \alpha](? \Delta [Uf \alpha] Msf \alpha)$$

ii) Meta *en adquisición satisfecha*, la abreviación $++\Delta a$ designa la adquisición de la acción a como alcanzada y satisfecha.

$$\begin{aligned} Ms Ui [Uf \alpha] Msf \alpha \wedge [Ufs U-i [Uf \alpha] Msf \alpha] \phi \\ \equiv ? \Delta [Uf \alpha] Msf \alpha \wedge [Ufs U-i [Uf \alpha] Msf \alpha] \phi \\ \equiv ++\Delta [Uf \alpha] Msf \alpha \end{aligned}$$

una meta es adquisición es alcanzada y satisfecha cuando el usuario informa que la tarea está terminada.

iii) Meta *en adquisición abandonada*, la abreviación $@\Delta a$ designa una meta en adquisición abandonada.

$$\begin{aligned} Ms Ui [Uf \alpha] Msf \alpha \wedge [Ufs Ui [U-f \alpha] Msf \alpha] \phi \\ \equiv ? \Delta [Uf \alpha] Msf \alpha \wedge [Ufs Ui [U-f \alpha] Msf \alpha] \phi \\ \equiv @\Delta [Uf \alpha] Msf \alpha \end{aligned}$$

La Caracterización de la Cooperación

Como se vio anteriormente, el diálogo cooperativo se desarrolla alrededor de la realización de una tarea donde la máquina es considerada un asistente que colabora al cumplimiento de dicha tarea. Ahora bien, en nuestro caso, el diálogo también es el medio para dar a conocer el plan para hacer una nueva tarea. Así podemos considerar que el usuario tiene una predisposición a mostrar como hacer la tarea y el sistema una predisposición para aprenderla. Los axiomas que se presentan en esta sección muestran el diálogo cooperativo en función de estas predisposiciones.

Obligación de la Máquina hacia la Meta Planteada

Si el usuario ordena la realización de una tarea y la máquina conoce el plan (Msf sabe-hacer) entonces la máquina ejecuta este plan.

$$(A5.i) \quad ?Mf \alpha \wedge Msf \alpha \wedge \alpha \in Ab \supset [Mf \alpha] \phi$$

para una acción de base

$$(A5.ii) \quad ?Mf (\beta_1 ; \beta_2) \wedge Msf \beta_1 \supset [Mf \beta_1] (\phi_1 \wedge [Uff \beta_2] (? [Mf \beta_2] \phi_2))$$

para una tarea compuesta

Pregunta Desencadenada por la Ignorancia de la Máquina

Entre las posibilidades donde la máquina toma la iniciativa en el diálogo existe el caso de la solicitud de información. La tarea solicitada es conocida por la máquina pero faltan algunos datos para cumplirla. En consecuencia, la tarea es diferida hasta la obtención de la información faltante. Un conjunto de acciones suplementarias deberán hacerse antes de la tarea: la pregunta al usuario de la información faltante y la comprobación de la respuesta prevista. Estas acciones son consideradas como acciones complementarias para alcanzar la meta deseada:

$$(A6) \quad ?Mf \alpha \wedge Msf \alpha \wedge M-s \text{ parámetro}(\alpha, \phi) \supset$$

$$\begin{aligned} &?(Mffs \text{ parámetro}(\alpha, \phi) ; \\ &\text{verificar}(++Ufs \text{ parámetro}(\alpha, \phi) ; Mf \alpha) \end{aligned}$$

La Adquisición de la Tarea

Como lo vimos en párrafos anteriores, una meta planteada por el usuario y desconocida por la máquina será desplazada para conducir al diálogo a un contexto de adquisición de la tarea. Este desplazamiento reposa en la suposición de que el usuario esta en disposición de mostrar la tarea en caso de que la máquina la ignore⁴. Es decir, suponemos que el usuario tiene la intención, siempre presente, de enseñar a la máquina.

$$(A7) \quad ?Mf \alpha \wedge M-sf(\alpha) \supset [Mff \alpha](? \Delta [Uf \alpha] Msf \alpha)$$

$$(A8) \quad ? \Delta [Uf \alpha] Msf \alpha \wedge ++Mf \beta \supset Ms(\beta \in \text{secuencia}(\alpha))$$

Toda acción satisfecha β durante un diálogo de adquisición pertenece a la secuencia para hacer la tarea α .

EJEMPLO

El siguiente ejemplo muestra de manera general la estructura de un diálogo donde los intercambios de palabra convergen hacia una meta en un mundo de dibujo, en este sencillo caso, el dibujo de un cuadrado. Supongamos que esta tarea es conocida por la máquina, pero para completarla es necesario solicitar su color (supondremos conocidos todos los otros parámetros de la tarea *dibujar_cuadrado*). Así el conocimiento supuesto de la máquina es el siguiente:

$$(C1) \quad Msf \text{ dibujar_cuadrado} \wedge M-s \text{ parámetro}(\text{dibujar_cuadrado}, \text{color})$$

U: *dibuja un cuadrado*
[Uff dibujar_cuadrado]

$$1. \quad [Uff \text{ dibujar_cuadrado}](?Mf \text{ dibujar_cuadrado})$$

(definición Uff) *meta planteada*

$$2. \quad ?(Mffs \text{ parámetro}(\text{dibujar_cuadrado}, \text{color}) ; \text{verificar}(++Ufs \text{ parámetro}(\text{dibujar_cuadrado}, \text{color}) ; Mf \text{ dibujar_cuadrado})$$

1, C1, A6 *submeta desencadenada*

$$3. \quad [Mffs \text{ parámetro}(\text{dibujar_cuadrado}, \text{color})]$$

(verificar(++Ufs parámetro(dibujar_cuadrado, color) ; Mf dibujar_cuadrado)

2, A5.i

$$[Mffs \text{ parámetro}(\text{dibujar_cuadrado}, \text{color})]$$

M: *de que color?*

$$4. \quad [Mffs \text{ parámetro}(\text{dibujar_cuadrado}, \text{color})]$$

(?Ufs parámetro(dibujar_cuadrado, color))

(definición Mffs) *submeta planteada*

U : *rojo*

$$[Ufs \text{ parámetro}(\text{dibujar_cuadrado}, \text{color: rojo})]$$

$$5. \quad [Ufs \text{ parámetro}(\text{dibujar_cuadrado}, \text{color: rojo})]$$

⁴ Para profundizar sobre la adquisición de una tarea durante el diálogo véase [3, 18].

- (Ms parámetro(dibujar_cuadrado, color: rojo)
(definición Ufs)
6. ++ Ufs parámetro(dibujar_cuadrado, color)
4, 5 *submeta satisfecha*
7. [Mf dibujar_cuadrado](forma(obj1,cuadrado) ^
tamaño(obj1, 50) ^ posición(obj1,(342, 875)) ^
color(obj1,rojo)
3, 6, (definición dibujar_cuadrado)
- [Mf dibujar_cuadrado]
M: <dibujado del cuadrado>
8. + Mf dibujar_cuadrado
1, 7 *meta alcanzada*

La submeta planteada (en el paso 4) para el intercambio de información se debe a la ignorancia de la máquina, la realización de esta submeta incluye la ejecución de un conjunto de acciones previas –incluyendo actos de diálogo– a la tarea solicitada inicialmente.

CONCLUSIONES

El trabajo presentado establece las ideas principales de una lógica para el diálogo cooperativo. Hemos caracterizado la actividad del usuario en función de las metas por alcanzar que, al mismo tiempo, estructuran el diálogo. Esta lógica considera la acción como el elemento central, donde una intervención que incorpora el habla es considerada como una forma más de acción. Así, el diálogo conduce la realización de una acción y, a la vez, la acción suscita el diálogo; el usuario muestra las acciones y la máquina aprende las secuencias mostradas; se expresa la meta a lograr a través de la acción.

Nuestra lógica permite describir los intercambios de palabra, dentro del marco de la cooperación hombre-máquina, a través del desarrollo de la meta planteada por el usuario. Este desarrollo es una secuencia de acciones (de habla o no) determinada por las intenciones del usuario y el conocimiento de la máquina. Es posible tratar a lo largo del mismo diálogo la falta de información relativa a la tarea en cuestión. Además, esta lógica introduce un mecanismo de predicción –p. e. el caso de una pregunta realizada por la máquina– que permite la recuperación en caso de error o incompreensión.

AGRADECIMIENTOS

El desarrollo de este trabajo fue parcialmente financiado por el CONACYT (No. Ref. 31128A) y el Laboratorio Franco-Mexicano de Informática (LAFMI) bajo el marco del proyecto “Interacción oral hombre-máquina”. Los autores también agradecen el apoyo del Laboratorio de Tecnologías del Lenguaje del INAOE.

REFERENCIAS

1. Austin, J.L. Quand dire, c'est faire. Edition du Seuil, París, 1970.
2. Caelen, J. Vers une logique dialogique. *Séminaire International de Pragmatique*, Jerusalem, 1995.
3. Caelen, J. & Villaseñor-Pineda, L. Dialogue Homme-Machine et Apprentissage. *Apprentissage par l'interaction*. (ed.) Khaldoun Zreik, p 83-117. Europa Productions. París, 1997.
4. Cohen, P. R. & Levesque, H.J. Persistence, Intention and Commitment. *Intentions in Communication*. (eds.) P. R. Cohen, J. Morgan & M. E. Pollack. The MIT Press. 1990.
5. Colineau, N. *Etude des marqueurs discursifs dans le dialogue finalisé*. Thèse Sciences Cognitives, Université Joseph Fourier, Grenoble, 1997.
6. Halpern, J.Y. & Moses, Y. A guide to completeness and complexity for modal logics of knowledge and belief. *Artificial Intelligence* 54, 319-379. Elsevier Science Publishers. 1992.
7. McTear, M. Spoken Dialogue Technology: Enabling the Conversational User Interface. *ACM Computing Surveys*, Vol.34, No. 1, March 2002, pp 90-169.
8. Maybury, M.T. *Intelligent Multimedia Interfaces*. AAAI/MIT Press, USA. 1993
9. Piaget, J. *Development and Learning. Piaget Rediscovered*, Ripple R.E. and Roccastle V.N. ed., School of Education, Cornell University, Ithaca, New York, 1964.
10. Prendinger, H. & Schurz, G. Reasoning about Action and Change. A Dynamic Logic Approach. *Journal of Logic, Language, and Information*, 5:209-245, 1996.
11. Searle, J.R. *Actes de Langage*, Hermann ed., París, 1972.
12. Stein, A. & Maier E. Structuring Collaborative Information-Seeking Dialogues. *Knowledge-Based Systems*, Special Issue on Human-Computer Collaboration. Vol.8 Marzo 1995. Elsevier Science B.V.
13. van der Hoek, W., van Linder, B. & Meyer, Ch. J.-J. A Logic of Capabilities (extended abstract). *Proceedings of the Third International Symposium of Logic Foundations of Computer Science*. LFCS, 94. (eds.) A. Nerode & Yu. V. Matiyasevich. San. Petersburgo, Rusia, Julio 11-14, Springer-Verlag, 1994.
14. van der Hoek, W., van Linder, B. & Meyer, Ch. J.-J. 1994. Unraveling Nondeterminism : On having the Ability to Choose (extended abstract). *Proceedings of the sixth International Conference on Artificial Intelligence : Methodology, Systems, Applications AIMS'94*. (eds.) P. Jorrand & V. Sgurev. Sofia, Bulgaria, Septiembre 21-24, 1994. World Scientific. 1994.
15. van Linder, B., van der Hoek, W. & Meyer, Ch. J.-J. Communicating Rational Agents. *Proceedings of the 18th German Annual Conference on Artificial Intelligence KI-94 : Advances in Artificial Intelligence*. (eds.) B. Nebel & L. Dreschler-Fischer. Saarbrücken, Alemania, Septiembre 18-23, 1994. Springer-Verlag 1994.
16. Vanderveken, D. *La logique illocutoire*. Mandarga éd. Bruselas, 1990.
17. Vernant, D. Modèle projectif et structure actionnelle du dialogue informatif. *Du dialogue, Recherches sur la*

philosophie du langage, Vrin éd., Paris, n°14, p. 295-314, 1992.

18. Villaseñor-Pineda, L. *Contribution à l'apprentissage dans le dialogue homme-machine*. Tesis en Ciencias Computacionales, Université Joseph Fourier, Grenoble. Febrero 1999.

