

Un modelo de interpretación pragmática basado en la SDRT para el diálogo hombre-máquina

Luis Villaseñor-Pineda¹, Anne Xuereb²,
Jean Caelen² & Manuel Montes-y-Gómez^{1,3}

¹Laboratorio de Tecnologías del Lenguaje
Instituto Nacional de Astrofísica, Óptica y Electrónica,
Apdo. Postal 51 y 216, Puebla, México
{villasen,mmontesg}@inaoep.mx

²Laboratoire CLIPS-IMAG – Université Joseph Fourier
Domaine universitaire, BP 53, 38041 Grenoble Cedex 9
Anne.Xuereb@imag.fr

³ Departamento de Sistemas Informáticos y Computación
Universidad Politécnica de Valencia, España
mmontes@dsic.upv.es

Resumen

En este trabajo presentamos un modelo de interpretación pragmática para un sistema de diálogo hombre-máquina. Este modelo es parte fundamental de un sistema conversacional, y amplía el mecanismo de comprensión semántica para permitir la resolución de las referencias deícticas y anafóricas. En este trabajo se describen los principios de este modelo, el cual integra el histórico del diálogo, el estado de la tarea y una ontología propia del dominio para la interpretación de una locución.

1.- Introducción

El objetivo final de este trabajo persigue la resolución de las referencias deícticas y anafóricas, además de las presuposiciones y las implicaciones conversacionales, en diferentes contextos en la comunicación oral hombre-máquina. En particular este trabajo se orienta a la resolución pragmática en diálogos multimodales orientados por una tarea. Para este estudio se utilizó un corpus multimodal [1]. Este corpus fue recolectado a través de la técnica del mago de Oz. El usuario dialoga en lenguaje natural con un “agente conversacional” simulado para realizar el diseño de una cocina. Las acciones permitidas son sencillas, como colocar objetos, moverlos o eliminarlos. Para precisar la orden el usuario puede expresarse de forma multimodal al usar la voz conjuntamente con gestos indicativos realizados a través del cursor gráfico asociado al ratón. A pesar de lo simple de la tarea se tiene un contexto lingüístico rico tanto en referencias deícticas como anafóricas.

El desarrollo de un sistema conversacional involucra diferentes subsistemas: el subsistema de reconocimiento de habla, el subsistema de comprensión del lenguaje, el subsistema de gestión del diálogo, etc. El presente trabajo se enfoca únicamente al problema de la interpretación pragmática. De esta manera, suponemos que el subsistema de

comprensión del lenguaje¹ provee una representación semántica del enunciado de entrada, la cual puede estar incompleta al no precisar puntualmente los referentes a los cuales hace alusión. Por ejemplo en las intervenciones siguientes:

- (1) *muévelo a la derecha*
- (2) *pon el mueble aquí*

es necesario resolver la anáfora pronominal (*lo* – pronombre clítico) así como las anáforas asociativas (*el mueble*, refiriéndose al último mueble precisado en la conversación). Estos fenómenos lingüísticos son comunes y ampliamente usados por los locutores. Así un proceso de interpretación pragmática es necesario, el cual debe considerar para la interpretación del enunciado el estado de la tarea y el histórico del diálogo. Este proceso de interpretación incluye:

- la resolución de los deícticos a partir del contexto
- la resolución de la elipsis
- la resolución de la anáfora y la catáfora
- el tratamiento de las presuposiciones y las implicaciones conversacionales

En las secciones subsecuentes se discutirá únicamente el caso de la resolución de la anáfora. Sin embargo, este mismo esquema nos brinda los elementos para el tratamiento de las presuposiciones e implicaciones conversacionales. Por otro lado, la resolución de los deícticos gramaticales (por ejemplo, *aquí*) sólo es posible gracias a la información proveída por los gestos. Para la descripción de un posible mecanismo de resolución de ellos véase nuestro estudio en [3].

2.- Fundamentos

Este trabajo se basa en la SDRT (*Segmented Discourse Representation Theory*) [4]. Esta teoría es una formalización

¹ Para una descripción detallada del subsistema de gestión de diálogo y del subsistema de administrador de la tarea véase [2].

de la interpretación dinámica de un enunciado en determinado contexto. Ésta a su vez, extiende la DRT (*Discourse Representation Theory*) [5] completándola con relaciones retóricas al ligar los segmentos del discurso. El presente trabajo reutiliza la SDRT y la aplica a nuestra problemática (para una presentación detallada véase [6]).

El aspecto dinámico de la interpretación de un enunciado es capturado a través de la construcción incremental de la estructura lógica del diálogo (SDRS *Segmented Discourse Representation Structure*). El sistema de comprensión de lenguaje provee una representación lógico-semántica del enunciado bajo la forma de una DRS (*Discourse Representation Structure*).

A partir de un análisis detallado de nuestro corpus se estableció una clasificación específica de relaciones retóricas, además de las reglas para la identificación de dichas relaciones. Con estos elementos se adaptó la SDRT a nuestro dominio de aplicación. A continuación presentamos las bases de la DRS para posteriormente describir la clasificación propuesta.

2.1. La representación lógico-semántica

El subsistema de comprensión del lenguaje provee una representación del enunciado bajo una forma lógica: una DRS. Un enunciado es modelado por una conjunción de actos de lenguaje, cada uno de la forma Fp , donde F es la fuerza ilocutoria y p el contenido proposicional [7].

La fuerza ilocutoria varía dependiendo de la intención del acto, ésta puede ser:

- FA : *hacer* una acción sobre el mundo
- FS : *hacer saber* (afirmar) cierta información
- FFS : *hacer hacer saber* (preguntar) cierta información
- FF : *hacer hacer* (ordenar) una acción
- FD : *hacer deber* (obligar)
- FP : *hacer poder* (proponer una opción)

Esta fuerza es calculada a partir de un tratamiento conjunto entre diferentes módulos del sistema, el cual se basa, por un lado en marcadores pragmáticos del enunciado [8], y por otro lado la predicción del acto esperado más probable [9].

El contenido proposicional es representado bajo forma lógica: ésta comprende una lista de variable, de predicados y de ecuaciones entre las variables. Las variables son tipadas semánticamente (los tipos son definidos en una ontología local), por ejemplo, tenemos *muebles*, *actores*, etc. además de tipos generales para describir, por ejemplo, a los pronombres cuyo tipo es *indefinido*.

A partir de (i) la representación DRS del enunciado en cuestión, de (ii) la SDRS en curso (representando el histórico del diálogo) y (iii) del contexto de la tarea (el conocimiento del dominio), el intérprete actualiza la estructura lógica del diálogo, generando una nueva SDRS. La siguiente tabla da un ejemplo de las representaciones DRS para los enunciados de un segmento de diálogo entre U (el usuario) y M (el agente conversacional):

<i>Intervención y representación DRS</i>	<i>Ontología y diccionario</i>
<p>U: puedes mover la estufa hacia la izquierda?</p> <p>π_1: [FFS ; a1 : actor, obj1 : mueble ; pos1 : posición ; mover(a1, obj1, pos1), izquierda(pos0, pos1), a1.identidad = "agente conversacional", obj1.nombre = estufa]</p>	<p>a1 : actor + propiedades : identidad + funciones : participante, actor</p> <p>obj1 : mueble + propiedades : nombre, color, tamaño, posición pos1 : posición</p> <p>predicado: mover(actor, mueble, posición) <i>modificación del contexto gráfico</i></p>
<p>M: que tanto quieres que la desplace?</p> <p>π_2: [FFS ; a1 : actor, v : indefinido, pos2 : posición ; mover(a1, v, pos2), a1.identidad = "agente conversacional", v = ?, pos2 = ?]</p>	<p>a1 : actor pos2 : posición v : indefinido pronombre = tipo indefinido y co-referencia</p> <p>predicado: mover(actor, mueble, posición) <i>modificación del contexto gráfico</i></p>

La intervención de M contiene la subespecificación ($v = ?$) representando el mueble sobre el cual recae el predicado *mover*. Esta subespecificación se resolverá durante el cálculo de la estructura lógica del diálogo (la SDRS). La resolución de esta variable y de los actos de habla debe conducir a:

π_2 : [FFS ; a1 : actor, obj1 : mueble, pos2 : posición ; mover(a1, obj1, pos2), a1.identidad = "agente conversacional", obj1.nombre = estufa , pos2 = ?]

Lo que significa que el objeto sobre el que discurre el segundo enunciado es la estufa.

2.2. Las relaciones retóricas en el diálogo hombre-máquina

Dado que nos encontramos en el marco de un diálogo hombre-máquina orientado por una tarea es posible suponer: los interlocutores son cooperativos, los enunciados son pertinentes, y el mundo del dominio es cerrado. A partir del análisis detallado del corpus multimodal recolectado se

encontraron las siguientes características: los diálogos están estructurados en ciclos, cada ciclo orientado a la colocación de un mueble, estos subdiálogos son cortos pero complejos. Estos son estructurados por secuencias de orden-ejecución-evaluación. Las intervenciones son concisas a menudo con elipsis. Las referencias anafóricas son frecuentes. La realización completa de la tarea se hace a través de varios ciclos, cada uno de ellos casi completamente independiente del ciclo anterior, a pesar de que en ocasiones se retome la colocación de un mismo mueble.

Por el momento no se considera un análisis sintáctico y semántico fino que permita la identificación de marcadores lingüísticos elaborados. De esta manera, la caracterización de las relaciones retóricas se basa esencialmente en la forma del enunciado (interrogación, declaración, orden), los conceptos introducidos por los predicados, así como, los referentes mismos.

Cabe mencionar que este módulo no modela las intenciones, planes o creencias del locutor. Las metas del locutor son administradas por el controlador del diálogo en función del plan de la tarea.

En este contexto específico nosotros hemos identificado seis relaciones retóricas. A continuación se presentan cada una de estas relaciones junto con sus criterios de identificación ($K\pi$ designa la fórmula SDRS con etiqueta π):

QAP(π_1, π_2) (*Question-Answer Pair*): Par pregunta-respuesta completo.

En la relación QAP, π_2 es la respuesta completa a la pregunta π_1 . $K\pi_1$ es una pregunta, $K\pi_2$ es una afirmación. Los criterios para su identificación son los siguientes:

- $K\pi_1$ es de fuerza FFS o FF; $K\pi_2$ es de fuerza FS o FA,
- $K\pi_1$ y $K\pi_2$ versan sobre el mismo tópico: mismo tema de los predicados y tipos semánticos de referentes,
- $K\pi_2$ tiene un contenido semántico « sí » o « no » o todos los referentes cuestionados de $K\pi_1$ pueden unificarse con un referente definido en $K\pi_2$.

PQAP(π_1, π_2) (*Partial Question-Answer Pair*): Par pregunta-respuesta parcial

$K\pi_2$ es una respuesta incompleta, que reduce el conjunto de respuestas posibles de $K\pi_1$, pero no permite inferir la respuesta completa. Los criterios para su identificación son los siguientes:

- $K\pi_1$ es de fuerza FFS o FF; $K\pi_2$ es de fuerza FS o FA,
- $K\pi_1$ y $K\pi_2$ versan sobre el mismo tópico,
- $K\pi_2$ no permite más que una resolución parcial de $K\pi_1$

IQAP(π_1, π_2) (*Indirect Question-Answer Pair*): Par pregunta-respuesta indirecta

- $K\pi_1$ es de fuerza FFS o FF; $K\pi_2$ es de fuerza FS o FA,
- $K\pi_2$ no resuelve directamente pero provee la información para inferir la resolución de $K\pi_1$.

Q-Elab(π_1, π_2) (*Question-Elaboration*): Elaboración de pregunta

En esta relación Q-Elab, $K\pi_2$ es una pregunta, cuya respuesta es necesaria para elaborar la respuesta a $K\pi_1$ (se

trata por ejemplo, de una pregunta de clarificación originada por una orden).

- $K\pi_1$ es de fuerza FFS o FF; $K\pi_2$ es de fuerza FFS,
- El tópico de $K\pi_2$ es una elaboración del de $K\pi_1$.

Contexto (π_1, π_2)

π_1 constituye el contexto de π_2 . $K\pi_1$ es un estado. El contexto es el conocimiento de dominio en cierto punto del diálogo, por ejemplo, las condiciones iniciales al momento de iniciar una sesión. Esta relación nos permite integrar el conocimiento del estado de la aplicación al inicio de cada subdiálogo.

Coordinación (π_1, π_2)

La coordinación es la relación por omisión cuando ninguna otra relación es identificada. Por ejemplo cuando se presenta un cambio de tópico o una pregunta inmediatamente después de una secuencia QAP. La coordinación es una simple sucesión de segmentos discursivos sin tratar de distinguir una relación estructural paralela o de contraste.

Las primeras cinco relaciones QAP, PQAP, IQAP, Q-Elab, Contexto son subordinantes: $R(\pi_1, \pi_2)$ define una dependencia de π_2 con respecto a π_1 . El segmento subordinado (π_2) no puede ser interpretado sin referencia al segmento subordinante (π_1).

2.3. La noción de tópico

Retomando la idea de [10] nosotros definimos el tópico como *el conjunto de elementos en curso de discusión*. La caracterización automática es un problema complejo y aún es objeto de investigación. En primera instancia, nosotros consideramos que el tópico es caracterizado a la vez por el tema del predicado y los tipos semánticos de los referentes del enunciado. Esta información es capturada por una ontología de conceptos, que contiene la noción de compatibilidad de tópicos. Esta noción de compatibilidad permitirá detectar la elaboración de un tópico, justo a la apertura de la relación Q-Elab, así como la clausura y cambio de tópico en el transcurso del diálogo. Así, para cada intercambio pregunta-respuesta, se identifica un tópico, el cual está representado por una SDRS, marcada con una T , dominando el intercambio. Esta estructura marca el conocimiento común establecido por los participantes en el diálogo: ésta contendrá el resultado de la resolución de la secuencia pregunta-respuesta subyacente, y es el punto de enlace disponible para otro segmento discursivo [11]. QAP y IQAP pueden cerrar o no un tópico, y es posible tener secuencias de QAP coordinadas bajo el mismo tópico.

2.4. Las resoluciones desencadenadas por la estructura lógica del diálogo

La modelización de la estructura del diálogo por la SDRT captura las restricciones de accesibilidad de la resolución de referentes. En efecto, cada relación retórica restringe de manera específica la resolución de pregunta-respuesta y de las subespecificaciones.

En las relaciones QAP, PQAP, las subespecificaciones del segmento de la respuesta son resueltas accediendo al segmento de la pregunta. Si dos SDRS π_1, π_n están ligados por una relación $R(\pi_1, \pi_n)$, entonces π_n y sus DRS contenidas acceden

a los referentes de π_1 (DRS-accesibles). Las referencias anafóricas son resueltas por unificación con un referente definido y accesible del histórico. Las anáforas π_n se resuelven preferentemente en el segmento π_1 . En caso de no tener éxito, se intenta la resolución con los referentes definidos y accesibles al nivel inmediato superior.

Por otro lado, cada relación retórica desencadena la actualización de la base de hechos tomando en cuenta tanto el contenido proposicional de los segmentos ligados, como de la semántica propia de la relación retórica. En el caso de los pares QAP, si la respuesta tiene un valor semántico de “sí” o “no”, entonces la anáfora proposicional del segmento pregunta, y los hechos correspondientes son agregados o retirados del contexto.

3.- Ejemplo

A continuación se muestra el análisis de un segmento de diálogo mostrando el mecanismo de interpretación pragmática propuesto. En los segmentos la U designa al usuario y la M el agente conversacional. Los detalles de la notación fueron descritos en §2.1.

U: puedes poner este mueble en esta esquina debajo del estante ?	π_1
M: <realización de la acción>	π_2
U: un poco más a la derecha	π_3
M: <realización de la acción>	π_4
U: ahora pon un refrigerador aquí	π_5

Tabla 1. Diálogo ejemplo

Enseguida se detallan las etapas de cálculo de las relaciones retóricas y la resolución de referencias. Para efectos de presentación del ejemplo colocamos una estructura π_0 , que representa la estructura final del diálogo anterior.

π_1 : [FFS ; a1 : actor ; ac1 : acción ; obj1, obj2 : mueble ; lug1 : lugar ; pos0, pos1 : posición ; ac1 = poner(a1, obj1, pos1), debajo(pos1, pos0), a1. identidad = "agente conversacional", obj1.nombre = estufa, obj2.nombre = estante, lug1.nombre = "esquina", pos1 = ?, pos0 = (x1,y1)]

π_2 : [FA ; a1 : actor; ac1 : acción ; obj1, obj2 : mueble ; lug1 : lugar ; pos0, pos1 : posición ; ac1 = poner (a1, obj1, pos1), debajo(pos1, pos0), a1. identidad = "agente conversacional", obj1.nombre = estufa, obj2.nombre = estante, lug1.nombre = "esquina", pos1 = ?, pos0 = (x1,y1)]

$QAP(\pi_1, \pi_2)$

Con la primera intervención del usuario un nuevo tópico es identificado T_1 y asociado a π_0 . La secuencia de actos FFS+FA nos orienta a un par QAP, donde el usuario solicita la realización de una acción. Es importante notar que la resolución de “este mueble” en π_1 se realiza a través de un gesto asociado.

La respuesta π_2 involucra un proceso de inferencia espacial que permite proponer una solución para $pos1$ a partir de la

$pos0$. Por otro lado, la acción provoca un cambio en el contexto gráfico lo que a su vez es considerado como parte

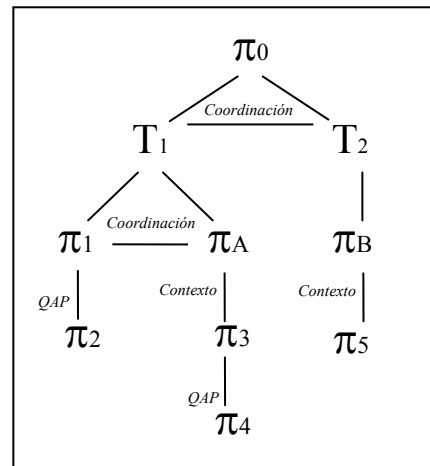


Figura 1. Gráfica de la SDRS del diálogo ejemplo

del diálogo mismo. Esto se captura por medio de una relación de coordinación con el estado inmediato anterior al tópico T_1 . El nuevo subdiálogo se enlaza con el contexto gráfico a través de una relación de *contexto* entre el estado resultante de la acción realizada π_A y la nueva solicitud del usuario π_3 .

π_A : [obj1 : mueble ; pos1 : posición ; obj1.nombre = estufa, obj1.posición = pos1 ; pos1 = (x2, y2)]

π_3 : [FFS ; a1 : actor, ac2 : acción ; pos3, pos4 : posición ; derecha(pos3, pos4), a1. identidad = "agente conversacional", ac2 = ?, pos3 = ?, pos4 = ?]

Después de observar los efectos de la acción el usuario “corrige” el resultado. La nueva intervención π_3 solicita el cambio. En este caso la elipsis recae primero sobre la acción. Así que la acción (ac2 : acción, ac2 = ?) debe ser resuelta por unificación con un referente del mismo tipo presente en el histórico del diálogo. La búsqueda es restringida por el tipo de la variable *ac2* y por la estructura definida por las relaciones retóricas (véase la Fig. 1). Así el primero estado donde tratar de resolver la elipsis es π_A , estado en el cual es imposible resolver la unificación; dado que existe una relación de coordinación con π_1 la unificación es posible con *ac1*. A través de esta unificación resolvemos la elipsis dando por resultado lo siguiente:

π_3 : [FFS ; a1 : actor, ac2 : acción ; obj1 : mueble ; pos3, pos4 : posición ; derecha(pos3, pos4), a1. identidad = "agente conversacional", ac2 = poner (a1, obj1, pos1), obj1.nombre = estufa, pos1 = (x2, y2) , pos3 = ?, pos4 = pos1]

π_4 : [FA ; a1 : actor, ac2 : acción ; obj1 : mueble ; pos3, pos4 : posición ; derecha(pos3, pos4), a1. identidad = "agente conversacional", ac2 = poner (a1, obj1, pos1), obj1.nombre = estufa, pos1 = (x2, y2) , pos3 = ?, pos4 = pos1]

Una nueva intervención π_5 solicita una nueva acción, esta nueva solicitud provoca un cambio de tópico T2 dado que no existe anáfora o elipsis en la elocución. Dado que la acción anterior modificó el contexto gráfico un nuevo estado π_B es insertado en la estructura del diálogo. Una relación de contexto asocia π_B a π_5 .

π_B : [obj1 : mueble ; pos1 : posición ; obj1.nombre = estufa,
obj1.posición = pos1 ; pos1 = (x3, y3)]

π_2 : [FFS ; a1 : actor ; ac2 : acción ; obj3 : mueble ;
pos5 : posición ; ac2 = poner(a1, obj3, pos5),
a1.identidad = "agente conversacional",
obj3.nombre = refrigerador, pos5 = ?]

4.- Conclusiones y perspectivas

Este trabajo propone un modelo derivado de la SRDT en el marco específico del diálogo hombre-máquina orientado por la tarea. Hasta ahora hemos identificado seis relaciones retóricas reconocidas a través de criterios simples. A partir de una estructura — la SDRS global — es posible representar el contexto común a ambos interlocutores (lo dicho así como lo implícito). Hasta ahora la validación del modelo sobre el corpus ha sido manual y hemos comprobado su aplicación, además de verificar la resolución de la anáfora y la elipsis. Por otro lado, estamos trabajando en el tratamiento de las presuposiciones, donde el conocimiento de sentido común o conocimiento externo a la tarea es involucrado.

Actualmente trabajamos sobre un prototipo automático donde las reglas de cálculo de las relaciones están basadas en la tipología de actos del habla y las propiedades semánticas del contenido proposicional. Para ello estamos usando una lógica monótona limitando el razonamiento a una sola intervención. Después de la inserción de la SDRS para cada intervención los sitios de enlace disponibles son calculados así como una posible relación para cada SDRS T no etiquetada. Las inferencias se desencadenan en base a estas hipótesis para intentar su resolución. La hipótesis es entonces aceptada o rechazada según el éxito o fracaso de la resolución. Una hipótesis aceptada no será nuevamente evaluada en el ciclo siguiente.

Agradecimientos

El presente trabajo fue parcialmente financiado por el Laboratorio Franco-Mexicano de Informática (LAFMI). El cuarto autor también agradece a la Secretaría de Estado de Educación y Universidades de España por el apoyo recibido.

Referencias

- [1] Villaseñor, L., Massé, A. & Pineda, L. A. The DIME Corpus. *ENC'01 3er Encuentro Internacional de Ciencias de la Computación*, SMCC-INEGI, Aguascalientes, México, Septiembre 2001.
- [2] Nguyen H., Caelen J. (2003), Generic manager for spoken dialogue systems. *DiaBruck : 7th Workshop on the Semantics and Pragmatics of Dialogue*, ed. Ivana Kruijff-Korbayova and Claudia Kosny, Saarland University, Saarbrücken, Proceedings pp.201-202.
- [3] Villaseñor-Pineda L., Vazquez-Valerdi C. (2004), Estudio sobre el uso de gestos en una interfaz multimodal hombre-máquina simulada. *Taller de Interacción Humano-Computadora: interacción-integración*. (eds.) Miguel Arias & Alexander Gelbukh, 2004, pp 37-46. SMCC ISBN 970-692-170-2.
- [4] Asher N., Lascarides A. (2003), *Logics of Conversation*. Cambridge University Press.
- [5] Kamp H., Reyle U. (1993), *From Discourse to Logic*. Kluwer Academic Publishers.
- [6] Xuereb, A. (2004) *Pragmatique de Dialogue Homme - Machine*. Master Recherche Mathématiques - Informatique, Spécialité Intelligence, Interaction, Information. Université Joseph Fourier.
- [7] Vanderveken D. (1985), *Logique illocutoire*, Bruxelles, Mardaga éd.
- [8] Colineau N., Caelen J. (1995), "Etude de marqueurs dans les actes de dialogue dans un corpus de conception" Actes de 01Design'95, 127-139 : Aspects communicatifs en conception, EuropaIA éd.
- [9] Fouquet Y. (2002), Un modèle de dialogue par les attentes du locuteur, Actes de *TALN'02*, 371-377.
- [10] Muller P., Prévot L. (2002), Conversation sous les topiques, du contenu propositionnel à la structure du dialogue. *Information - Interaction - Intelligence*, Hors série 2002, pp.179-196.
- [11] Prévot L., Muller P., Denis P., Vieu L. (2002), Une approche sémantique et rhétorique du dialogue – Un cas d'étude : l'explication d'un itinéraire. *Traitement Automatique des Langues*, Vol. 43(2), pp.43-70.