Diálogos Inteligentes Multimodales en Español

Luis Alberto Pineda Cortés

Departamento de Ciencias de la Computación
IIMAS,UNAM, México
luis@leibniz.iimas.unam.mx

Resumen. En este trabajo se presenta un modelo para la administración de diálogos multimodales orientados hacia la solución de problemas en dominios específicos. En este modelo, un dominio lingüístico se define en términos de un conjunto de diálogos y cada diálogo como un conjunto de situaciones conversacionales; asimismo, cada situación codifica tanto los actos del habla como los actos retóricos que el agente computacional puede interpretar y realizar en la misma. Como antecedente de esta investigación se describe un corpus constituido por un conjunto de diálogos orientados hacia el diseño de cocinas; se discuten algunas propiedades de estos diálogos y los problemas que enfrenta su modelación computacional; posteriormente se presenta un modelo para un diálogo de carácter más sencillo en el que un robot da una visita guiada a las instalaciones de nuestro departamento, y se describe su implementación; en las conclusiones se discute la posibilidad de extender el modelo a diálogos multimodales de mayor complejidad, como los que se presentan en nuestro corpus. Asimismo, se discute la necesidad de contar con corpus anotados en diferentes niveles de representación lingüística, incluyendo el prosódico, para dar un fundamento empírico a esta clase de investigación.

1. Introducción

Esta investigación tiene por objetivo desarrollar e implementar modelos de diálogos de lenguaje natural hablados en español coordinados con información proveniente de otras modalidades; en particular, tenemos por objetivo la modelación de diálogos cuya realización vaya a la par de la solución de un problema o la realización de una tarea, como los que ocurren, una oficina de información turística, en un restaurante de comida rápida o los que se dan en algunos contextos de diseño. De manera más general, nuestra investigación tiene por objetivo la creación de modelos de diálogos para contextos conversacionales en los que las intenciones que los participantes puedan expresar o interpretar tengan un carácter concreto, además de que su número sea finito y reducido y, por lo mismo, puedan elucidarse y especificarse a través del análisis; asimismo, asumimos que la estructura de estos diálogos tiene un carácter esquemático, por lo que su modelación se reduce a la identificación de un número reducido de esquemas conversacionales y los modos en que éstos se combinan en el curso de la conversación.

La modelación de estos diálogos es atractiva tanto por sus posibilidad comerciales, ya que es posible encontrar una gran cantidad de situaciones conversacionales cuya automatización sería sumamente útil, como desde el punto de vista científico, ya que la labor implica tomar una postura respecto a la naturaleza y estructura del lenguaje y su relación con la acción en el mundo. Asimismo y desde el punto de vista de la tecnología computacional, la modelación de estos diálogos ofrece también una oportunidad y un reto ya que son lo suficientemente complejos para ameritar el uso del lenguaje natural, y al mismo tiempo, lo suficientemente simples para que su implementación computacional sea factible.

2. El Corpus DIME

Con el fin de dar una base empírica a nuestra investigación se recolectó un corpus de conversaciones orientadas hacia la solución de una tarea de diseño, en este caso el diseño de cocinas (Villaseñor *et al*., 2001). El experimento se llevó a cabo en un escenario del tipo del Mago de Oz (Dalhbäck, *et al*., 1993), excepto que los sujetos sabían que el mago era un ser humano; durante el experimento la información lingüística y visual producida tanto por el sujeto como por el mago fue recopilada de manera automática. El escenario utilizado para el experimento se muestra en la figura 1.

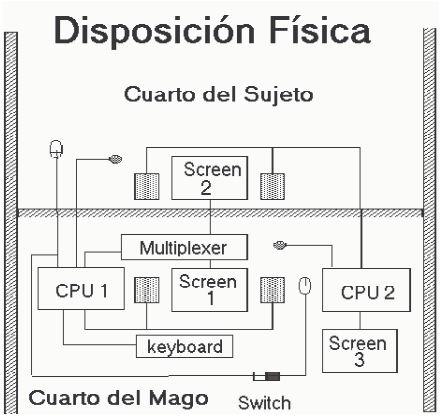

Figura 1. Escenario del Mago de Oz

Como se puede apreciar, el sujeto y el mago estaban en cuartos separados, por lo que la comunicación se realizaba necesariamente a través del micrófono, o por medio del monitor y los dispositivos asociados de interacción gráfica. El cuarto del mago estaba equipado con un procesador (CPU 1) al que se conectaron los monitores y ratones de ambos cuartos; en particular, la imagen de salida de este CPU se desplegaba simultáneamente en los monitores del sujeto y del mago a través de un multiplexor, por lo que ambos interlocutores tenían acceso a la misma imagen visual a lo largo de

toda la conversación; asimismo, ambos ratones estaban asociados a este monitor, por lo que el cursor correspondiente era también el mismo para el sujeto y el mago; el sistema incluía también un interruptor del lado del mago, con el que se habilitaba el control del ratón a uno sólo de los participantes en un momento dado de la conversación. Este CPU controlaba también el micrófono del sujeto y las bocinas del mago, habilitando la comunicación hablada del sujeto al mago, mientras que la vía inversa se habilitaba mediante un segundo procesador (CPU 2) cuyo micrófono estaba en el cuarto del mago y sus bocinas en el del sujeto. Por lo mismo, en este ambiente, la imagen del monitor, con ambas direcciones de entrada gráfica, así como el audio del sujeto se recolectó en un archivo de audio y video, mientras que el audio del mago, se recopiló de manera independiente; en la etapa de post-procesamiento, ambos archivos fueron integrados, y para cada diálogo se cuenta con un archivo con el video, la entrada gráfica y el audio de ambos participantes (.avi), así como un archivo de audio (.wav) con el audio del sujeto y el mago.

Para la definición de la interfaz de diseño se habilitó un producto comercial[1] con el que se definió el espacio de trabajo para las tareas de diseño. En el experimento, este programa era operado directamente por el mago, quien cedía el control del ratón al sujeto sólo en sus turnos conversacionales. La interfaz, con el espacio de trabajo de los experimentos, se muestra en la figura 2.

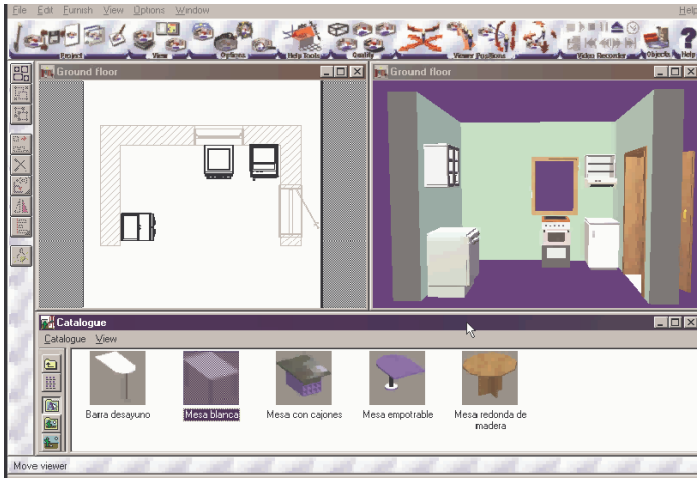

Figura 2. Escenario para el diseño de cocinas

La interfaz tiene tres áreas de trabajo principales: la vista de planta en dos dimensiones del cuarto de cocina, la perspectiva tridimensional del cuarto vista desde la pared contraria al área de trabajo, y un área inferior donde se despliegan los catálogos de muebles que se utilizan durante el proceso de diseño. En el

[1] Complete Home Designer Interiors, User´s Guide 1998. Alpha Software Corporation and Data Becker GmbH & Co KG.

experimento, la tarea consistía en escoger el mobiliario y posicionarlo en su lugar en la cocina; también se exploraron dos tipos de problemas: en el primero se partía de un cuarto de cocina parcialmente amueblado, como la que se muestra en la figura 2, y en el segundo la tarea se iniciaba con un cuarto vacío. Mediante este escenario se recolectaron 26 diálogos útiles, y cada uno de éstos consiste en la solución de un problema simple de diseño propuesto por el mago a través del mismo escenario, y resuelto a través de la conversación, donde el sujeto es el cliente y el mago un experto en cocinas que asiste al sujeto en el proceso de diseño. El corpus se etiquetó primero ortográficamente y actualmente se están realizando varios niveles de transcripción adicionales, como se verá más adelante.

Para la realización de este experimento no se impusieron restricciones en el tipo de lenguaje que se podía utilizar, y como era de esperar, el lenguaje presenta los fenómenos característicos del habla espontánea: interjecciones, tartamudeos, reparaciones, habla simultáneas, reparaciones, pausas y ruidos varios; adicionalmente, hay una gran de elocuciones que no tienen una estructura gramatical completa, como las palabras aisladas que expresan intenciones de confirmación y rechazo; también hay un número significativo de frases hechas que se emplean con el mismo fin (e.g. "ahí está bien"). Estos son sólo algunos ejemplos que ilustran la complejidad de los diálogos de nuestro corpus; con el fin de dar una idea más intuitiva, en la tabla 1 se presenta la transcripción ortográfica de las primeras 24 elocuciones de uno de los diálogos, que en total consta de 117 elocuciones y 69 turnos:

En nuestra investigación es también de interés estudiar el contexto multimodal en el que se desarrollan los diálogos; para este efecto se cuenta con los videos con las imágenes y eventos de interacción que ocurren en todos los diálogos. Un análisis del video asociado al segmento de diálogo en la tabla 1 muestra, por ejemplo, que a lo largo del mismo el mago realiza cinco acciones gráficas: un acto de apuntar simple (utt4), dos actos de apuntar compuestos (utt11 y utt15) donde se señala un objeto y luego la dirección en la que se desea moverlo, y dos movimientos de la estufa propiamente (utt18 y utt22). De hecho a lo largo del todo este diálogo se produce un número muy reducido de tipos acciones sobre el espacio de diseño: apuntar a un objeto, apuntar a un objeto e indicar dirección de movimiento, mostrar un catálogo, mostrar objeto de catálogo, además de mover, agregar y quitar muebles del espacio de diseño. También es posible observar que estos acciones motoras tienen una estructura, la cual está relacionada con las elocuciones de lenguaje natural a las que acompañan.

Turno	Part.	**Expression de lenguaje natural**	**Acción gráfica**
utt1	s:	quieres que desplace o traiga algún objeto a la cocina?	
utt2	u:	<ruido> no	
utt3		¿puedes mover la estufa hacia la izquierda?	
utt4	s:	<ruido> ¿hacia dónde ?	Apuntar objeto (estufa)
utt5	u:	<ruido> hacia <sil> hacia la derecha	
utt6	s:	<no-vocal> hacia la derecha	
utt7		<no-vocal> okey	
utt8		¿que tanto quieres que la desplace ?	
utt9	u:	a la mitad del espacio que hay entre la ventana y la pared	
utt10	s:	okey	
utt11		¿quieres que mueva este objeto <sil> hacia acá ?	Apuntar objeto e indicar. dir. mov. (derecha)
utt12	u:	no	
utt13		hacia el otro lado	
utt14	s:	okey	
utt15		¿este objeto hacia acá ?	Apuntar objeto e indicar. dir. mov. (izquierda)
utt16	u:	sí	
utt17	s:	okey	
utt18		<ruido>	Mueve estufa (izquierda)
utt19		¿ahí está bien ? <no-vocal>	
utt20	u:	un poco menos por favor	
utt21	s:	okey	
utt22		<ruido>	Mueve estufa (derecha)
utt23		¿ahí está bien ?	
utt24	u:	ahí está bien	

Tabla 1. Segmento de un diálogo

Tomando en cuenta tanto el lenguaje empleado como las acciones de interacción gráfica podemos hacer algunas observaciones relevantes para el proceso de modelación; por un lado, hay una frecuencia muy baja de expresiones gramaticalmente completas y la gran mayoría de las elocuciones son fragmentos y frases sueltas. En nuestro segmento hay tan sólo cuatro expresiones completas y las cuatro son preguntas (utt1, utt3, utt8 y utt1); de éstas, la primera es realmente una invitación a iniciar la tarea expresada como pregunta de cortesía (i.e un acto del habla indirecto); la segunda es realmente una directiva o comando expresada también como pregunta de cortesía (i.e. *mueve la estufa hacia la izquierda*); la tercera es una pregunta genuina que solicita información para determinar la magnitud del desplazamiento, y en la cuarta se solicitud una confirmación de que la

intención expresada por el sujeto ha sido comprendida correctamente por el sujeto (una pregunta de verificación o *check*). Podemos notar también que mediante la segunda pregunta se inicia una acción conversacional que tiene por objeto mover un objeto (la estufa) y el resto del segmento de diálogo se enfoca a determinar y confirmar los argumentos de la acción de movimiento solicitada.

Este proceso de determinación de los parámetros de la acción puede ser conceptuado como un proceso de resolución de referencia: es necesario determinar quien es el agente de la acción de mover, del objeto a ser movido (i.e. la estufa) y de la posición a la que ésta se desea desplazar; la resolución del agente es en este caso trivial dado que en la situación conversacional éste es el mago, y el referente de la frase nominal *la estufa* es la estufa que está presente en el espacio de diseño, y puede ser seleccionada directamente en el contexto espacial; sin embargo, la resolución de *hacia la izquierda* es mucho más compleja, ya que se requiere determinar una orientación relativa, y una posición concreta, para poder llevar a cabo la acción de mover. Como se puede apreciar, un número significativo de elocuciones en el segmento consisten en actos del habla que confirman y precisan las hipótesis que se forman para la resolución de este último parámetro de la acción, empezando por la determinación de un marco de referencia espacial común al sujeto y al mago. Asimismo, el lenguaje espacial no basta para resolver este argumento, y es necesario realizar actos ostensivos directos tanto sobre el objeto a desplazar como a la dirección en que se debe realizar este desplazamiento. Es este proceso de determinación de los argumentos de la acción de mover lo que consume la parte substancial del diálogo; adicionalmente, el léxico involucrado en esta parte del diálogo está también orientado hacia la determinación de este argumento (i.e. el uso de la proposición *hacia* y el pronombre *donde*).

Por otro lado, para la realización de las acciones propiamente, una vez que la intención queda determinada, se requiere de planear y ejecutar la acción sobre el dominio espacial. El proceso de planeación consiste en decidir la secuencia movimientos necesarios para satisfacer la intención solicitada; en algunas ocasiones la acción puede llevarse a cabo directamente, pero en otros se requiere de mover otros muebles para abrir el espacio requerido y llevar la acción a su término. Sin embargo, estas acciones de planeación sólo se especifican una vez que los argumentos de la acción (e.g. de mover o poner) han sido completamente determinados, y son hasta cierto punto independientes al flujo conversacional. Si todas las acciones de movimiento que se llevan a cabo en el diálogo de ejemplo estuvieran completamente determinadas en la mente del sujeto, y éste tuviera la forma de expresarlas con la precisión suficiente para que dicha información quedara también determinada en la mente del mago, el sujeto se limitaría a formular y ejecutar planes directamente, y el diálogo tendría el mismo número de elocuciones que de acciones, ya que el mago ni siquiera necesitaría confirmar sus acciones o buscar un acuerdo respecto a las mismas. Este es el caso, por ejemplo, del diseñador que interactúa con un sistema de dibujo realizando acciones perfectamente determinadas, con la ayuda de los dispositivos de interacción gráfica.

Por lo anterior, una manera de conceptuar a estos diálogos en términos de un conjunto de situaciones conversacionales relacionados con el dominio de aplicación, en las que se tiene la intención de satisfacer dicha acción; sin embargo, con el fin de determinar a los argumentos de estas acciones es necesario realizar un sub-diálogo, con intenciones y lenguaje dirigidos a satisfacer cada uno de estos argumentos. Adicionalmente, los procesos asociados a la interpretación de estas situaciones requieren de formular e interpretar acciones comunicativas, como presentar, notificar, confirmar, aceptar, rechazar, etc., proposiciones relacionadas con la solución de la tarea. Adicionalmente, la modelación de estos diálogos requiere tomar en cuenta, además del lenguaje, el contexto de interpretación, y las expectativas acerca de las intenciones, expresadas mediante actos del habla, en cada situación del diálogo, además del lenguaje que el mismo agente computacional puede utilizar en cada situación para atender a las intenciones del sujeto, o para expresar sus propias intenciones. Asociadas a estas intenciones están también los actos motores, como apuntar, o indicar la dirección del movimiento, que tiene por función elucidar, o confirmar, la información necesaria para determinar las acciones de diseño propiamente.

En base a estas intuiciones, en este trabajo proponemos un modelo de diálogo basado en la noción de *situación conversacional*; cada situación está asociada a un conjunto de intenciones y a una modalidad de interacción y el propósito de estar en una situación es satisfacer dichas intenciones. Asimismo, en cada situación hay un conjunto limitados de actos del habla que pueden ser interpretados en relación a la situación, y un conjunto de estructuras retóricas que pueden ser expresadas por el agente cuando se encuentra en dicha situación. Finalmente, concebimos al diálogo como un proceso de navegación en estas situaciones, incluyendo situaciones complejas que pueden a su vez analizarse como sub-diálogos, con lo que se permite el análisis de dominios conversacionales complejos en términos de componentes más simples.

Por supuesto, el análisis de los diálogos colectados en nuestro experimento es sumamente complejo, y antes de proponer un modelo, aterrizamos y formalizamos estas ideas con el estudio de un modelo básico, el cual se expone a continuación.

3. Modelos y administración de diálogos

El modelo se basa en la noción de situación conversacional; cada situación está asociada a un conjunto de intenciones del dominio del diálogo y a una modalidad de interacción; asimismo, concebimos a la situación como un objeto representacional que se interpreta por el agente conversacional, y el producto de este proceso es satisfacer la o las intenciones asociadas a la situación. Al ser un objeto sumamente contextualizado, sólo un pequeño conjunto de actos del habla expresados por el sujeto son significativos en una situación dada; asimismo, como resultado de

interpretar una situación el agente puede expresar un número reducido de conductas lingüísticas o motoras relacionadas con la satisfacción de las intenciones y estas conductas se codifican como actos retóricos asociados a la situación. Un modelo de diálogo se define como una red de situaciones donde cada situación puede ser precedida o seguida de varias situaciones. Finalmente, un domino conversacional se define como un conjunto de modelos de diálogos. Asimismo, asociado a cada modelo de diálogo, se define el conjunto de actos retóricos que el agente puede realizar durante la interpretación de dicho modelo.

En el modelo se define un conjunto de tipos de situaciones y cada instancia de éstas está asociada a un tipo específico. Los tipos definidos en la versión actual del modelo son *inicial*, *final*, *escuchando*, *diciendo*, *motora*, *recursiva* y *error* (*initial*, *final*, *listening*, *telling*, *motor*, *recursive & error*). Los tipos inicial, final y recursivo están asociados al control y todo diálogo tiene una situación inicial y una final; asimismo, el tipo recursivo es una situación en la que el control del diálogo pasa a la situación inicial de un modelo de dialogo subordinado y cuando se llega al estado final de este último el control regresa a la situación correspondiente del diálogo subordinante, como sucede en las redes de transición recursivas. Cada modelo de diálogo puede tener tantas situaciones recursivas como sea necesario, por lo que este mecanismo permite estructurar el dominio conversacional en un diálogo principal y varios sub-dialogos destinados a atender los contextos conversacionales específicos, con lo que se puede modelar la estructura del dominio a un nivel de detalle sumamente fino. Por su parte los tipos "escuchando", "diciendo" y "motora" están asociados a modalidades preceptúales específicas y orientados hacia a la acción; incluso la acción de escuchar que tiene por objetivo la recepción de información se piensa como una conducta activa, orientada a recolectar información relevante para satisfacer las intenciones de la situación, como se elabora más adelante. Del mismo modo, a pesar de que moverse es una acción, los actos motores se conciben como un proceso embebido dentro de una situación motora. Asimismo, con la misma filosofía, se podría definir un tipo situación para otras modalidades preceptúales, como el tacto, si esta modalidad estuviera considerada en el sistema de interacción multimodal.

En el modelo, el formato de todos los tipos de situaciones es exactamente el mismo, y el sistema de administración del diálogo interpreta a todas las situaciones con la misma disciplina; de esta manera, el proceso de interpretación del diálogo integra de manera uniforme la información de entrada y salida de todas las modalidades consideradas en el sistema.

Las transiciones entre situaciones se definen en términos de *pares de entrada y salida*; el par de entrada consiste en el nombre de la situación previa y una transición de forma *i-sa*:*o-rha*, donde *i-sa* es el acto del habla de entrada (*input speech act*) que originó la transición de la situación previa a la actual, y *o-rha* es el acto retórico (*rhetorical act*) realizado por el agente en la misma transición. Asimismo, el par de salida consiste en un par *i-sa*:*o-rha* que específica la transición de la situación actual a la siguiente y el nombre de la situación a la que se llega con dicha transición. En

algunos casos, el *i-sa* del par de salida puede ser vacío, lo que origina una transición incondicional; del mismo modo, el *o-rht* puede también ser vacío, con lo que la transición se realiza sin que se produzca ninguna conducta lingüística. En resumen, una situación es un objeto abstracto *s* con un par de parámetros de entrada y salida $s(i_i{:}o_i, i_o{:}o_o)$. La coherencia del diálogo se logra garantizando que el par de salida de una situación es siempre el par de entrada de la que le sigue; por lo mismo, para todas las situaciones $s_{i\text{-}1}(x_{i\text{-}1}, y_{i\text{-}1})$, $s_i(x_i, y_i)$, $s_{i+1}(x_{i+1}, y_{i+1})$ se cumplen las igualdades $y_{i\text{-}1} = x_i$ & $y_i = x_{i+1}$, donde *x* & *y* son pares de transición *i-sa*:*o-rha*. Para efectos de nuestra notación los pares de entrada y salida para la situación s_i se especifican como $s_{i\text{-}1}$ => $i\text{-}sa_i{:}o\text{-}rha_i$ y $i\text{-}sa_i{:}o\text{-}rha_i$ => s_{i+1} respectivamente.

A continuación ilustramos la notación con a una aplicación muy sencilla. En esta aplicación el sistema de administración de diálogo controla a un robot que da una visita guiada (simplificada) a las instalaciones del departamento de ciencias de la computación de nuestro instituto. En este caso la situación motora consiste en la navegación del robot hacia las áreas del departamento en las que realizan los intercambios lingüísticos entre el robot y el visitante. El modelo del diálogo principal para esta aplicación se muestra en la figura 3.

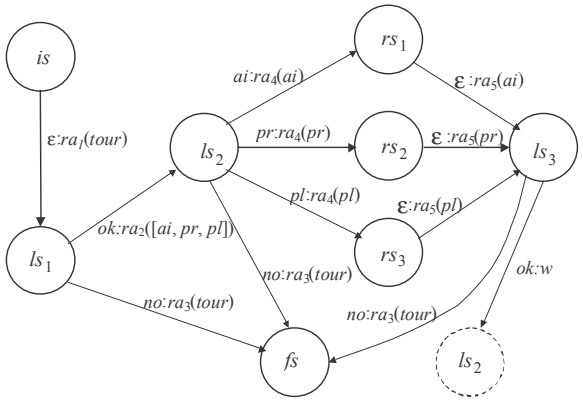

Figure 3. Modelo diálogo principal

En el diagrama los nombres de las situaciones se codifican de acuerdo al tipo: los prefijos *i*, *f*, *m*, *l*, *r* & *t* corresponden a los tipos inicial, final, motora, escuchando (*listening*), recursiva y diciendo (*telling*). Como se puede apreciar, toda situación puede tener más de una arco de entrada y también más de uno de salida. Las etiquetas de las transiciones especifican (i.e *acto-del-habla*:*acto-retórico*) se interpretan directamente por el administrador del diálogo, e identifican directamente a los actos del habla que se interpretan en la situación actual, y a los actos retóricos que se realizan cuando se pasa a la siguiente. Los actos retóricos, por su parte, admiten uno o más parámetros, los cuales determinan la forma lingüística final que el agente realiza. En este diálogo, el acto del habla de la situación inicial se define

como vacío, por lo que el control pasa incondicionalmente a la situación ls_1 que es del tipo escuchando; durante esta transición, el agente realiza el acto retórico ra_1 que tiene como parámetro la constante *tour*; como veremos más adelante este acto del habla está constituido por un saludo, una presentación del "tour" (el parámetro) y una pregunta si/no con la que se pregunta al visitante si desea realizar la visita; en la situación ls_1 el agente espera una respuesta, y tiene la expectativa de dos intenciones posibles por parte del visitante: éste desea ya sea aceptar o declinar la invitación; por lo mismo, independientemente de la conducta lingüística proferida de manera explícita, esta tiene que ser interpretada por el agente computacional en términos de las dos intenciones posibles en la situación.

En nuestro sistema, el sujeto puede usar una amplia variedad de conducta lingüística, y puede decir, por ejemplo, un simple "si" o "no", una interjección "aja", o una oración completa como "si deseo hacer la visita". Para este efecto, hemos desarrollado una técnica muy simple a la que llamamos *Interpretación Directa de los Actos del Habla*; en esta técnica aprovechamos una vez más que en cada situación del diálogo sólo un conjunto de intenciones es relevante, e identificamos empíricamente las diversas formas en que dichas intenciones pueden ser expresadas; finalmente, como las mismas formas explícitas pueden utilizarse para expresar intenciones diferentes en diferentes estados conversacionales, sólo las formas que expresan las intenciones posibles de la situación actual son consideradas en el proceso de interpretación directa. Más adelante se describe está técnica en detalle.

Si el visitante acepta la visita, se hace una transición de ls_1 a ls_2 a través del arco cuyo acto del habla es "*ok*"; para este efecto el administrador del diálogo verifica que el acto del habla reconocido en ls_1 corresponde con el especificado en la transición de salida de dicha situación en el modelo del diálogo actual. Todas las situaciones del diálogo se van recorriendo con esta misma disciplina, hasta llegar a la situación final, donde se ejecuta un acto retórico en el que se agradece y se despide la visita. El transito por las situaciones recursivas se lleva a cabo del mismo modo: cuando se llega a una de éstas, el administrador del diálogo toma el parámetro del acto retórico de entrada, que es el nombre del diálogo subordinado correspondiente, y pasa el control a la situación inicial de dicho diálogo; cuando se llega a la situación final de este último se regresa el control a la situación recursiva del diálogo subordinante, el cual ejecuta su transición de salida de manera estándar.

En el presente modelo es también posible dejar sub-especificado un acto retórico de salida, y este puede ser decido de manera dinámica, con información local a la situación correspondiente; por ejemplo, si se llega a la situación ls_3 después de haber realizado el diálogo de inteligencia artificial, el acto retórico de salida ra_2, mediante el cual se ofrece la siguiente alternativa al visitante, no debería incluir lo ya visitado; por lo mismo es posible definir dinámicamente el acto retórico $ra_2([pr, lp])$, realizarlo en la transición que regresa a ls_2 y ofrecer la visita sólo a las áreas que aún no han sido visitadas, en este caso reconocimiento de patrones (*pr*) y lenguajes de

programación (*lp*), en vez de la oferta que se hace al llegar a esta situación originalmente.

En la situación ls_2 el sistema de administración del diálogo espera que el sujeto responda a la oferta expresada por $ra_2([ai, pr, lp])$, y se interpreta la situación recursiva correspondiente, como ya se ha explicado. Como se puede apreciar, el acto de salida de las situaciones recursivas es vacío, ya que la situación siguiente esta determinada de antemano; asimismo, el acto retórico correspondiente es de un tipo especial al que llamamos *continuación* (i.e. ra_3) que consiste en agradecer al sujeto su visita parcial, y preguntarle si desea continuar la visita. Una instancia de un subdiálogo (*ai*) se ilustra en la figura 4. Los elementos de este modelo son similares al anterior, pero nos permiten ilustrar la funcionalidad de las situaciones de tipo motor y diciendo (*telling*).

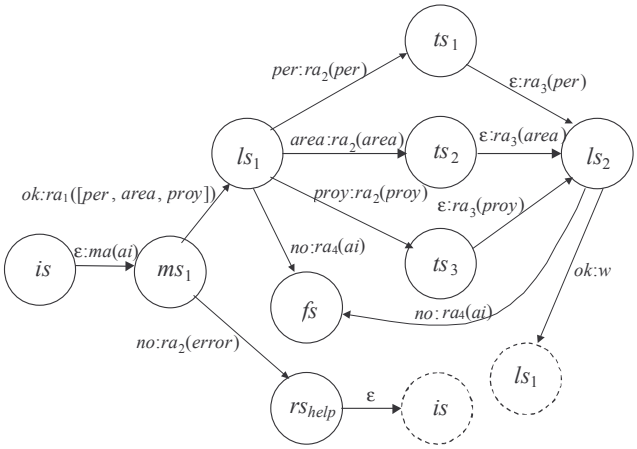

Figura 4. Modelo de diálogo con situaciones *motora* y *diciendo*

En este diálogo, el acto retórico de salida del estado inicial es de tipo motor y tiene como parámetro el nombre del área a la que el robot debe dirigirse. Al pasar el control, se inicia la conducta motora, y la situación espera la conclusión de la misma; en caso de que ésta sea exitosa, el sistema de navegación del robot regresa la confirmación correspondiente mediante un "*ok*"; en caso contrario, recibe un "*no*" y el administrador del diálogo se mueve a una situación recursiva donde se ejecuta un modelo de diálogo para atender la contingencia motora (aunque los detalles de este modelo están todavía en proceso de investigación). Finalmente, la transición de salida de la situación motora se representa e interpreta de la misma forma que el resto de las situaciones

En la figura 4 se ilustra también el tipo de situación diciendo o *telling*; dado que todas las transiciones de una situación a la siguiente realizan un acto del habla, en principio podríamos prescindir de este último tipo; sin embargo, su inclusión nos da un grado adicional de flexibilidad para definir conductas lingüísticas complejas en el

mismo turno conversacional; como se vera más adelante, el acto retórico que se ejecuta en la transición de ls_1 a ts_1 es una presentación y una elaboración, mientras que el acto retórico que se realiza de ts_1 a ls_2 es una continuación y, en principio, estos actos retóricos podrían integrarse en uno sólo con mayor nivel de estructura; sin embargo, la definición del estado del tipo "diciendo" brinda una mayor modularidad en el proceso del diseño del modelo del diálogo, y en la definición de los actos retóricos, por lo que se decidió incluirla.

Adicionalmente a los estados mostrados, el modelo de diálogo incluye también una situación a la cual se llega si el acto del habla no está entre los esperados en una situación situación; esta situación es de tipo error, y en su transición de entrada el habla no está especificado; el acto retórico de entrada, por su parte, es de un tipo *error*, y su realización expresa que el mensaje anterior no se comprendió. Por otra parte, la situación siguiente a la situación de error, es la misma que su situación previa, por lo que una vez comunicada la falta de comprensión, el sistema regresa a la situación en la que falló la comunicación, y el diálogo se reinicia normalmente.

Pasamos ahora a la definición de los actos retóricos; de la misma forma que cada modelo de diálogo se define con la especificación de sus situaciones, cada acto retórico que puede ser realizado en las transiciones debe ser definido de manera explícita; siguiendo los lineamientos de la teoría de los actos retóricos o *Rhetorical Structure Theory*, *RST* (Mann & Thompson, 1988), cada acto retórico incluido en el modelo se define en términos de un conjunto actos retóricos básicos cuya forma está especificada de antemano para todos los modelos. En la tabla 2 se muestra la definición de los actos retóricos asociados al modelo de diálogo principal (figura 1); cada acto retórico tienen un identificador, un tipo, una lista de parámetros y un conjunto de actos básicos; el acto ra_1 del diálogo principal, por ejemplo, se define en términos de un saludo, una presentación y una opción que se realiza como una pregunta absoluta; a su vez, la presentación tiene como parámetro al identificador de lo que se presenta, en este caso el *tour*. En la tabla 3 se muestran los actos retóricos de un diálogo subordinado. Este esquema permite definir actos retóricos estructurados relacionados con el dominio de aplicación de manera flexible, mediante el uso de actos retóricos básicos; por su parte, los actos retóricos básicos son bastante generales, y su número y características son objeto de nuestra investigación empírica. Para el proceso de realización final, se define una plantilla de texto para cada forma retórica básica; en el proceso de interpretación, durante las transiciones de salida de cada situación, los actos retóricos se realizan de manera secuencial; como parte final de este proceso se invoca a la plantilla correspondiente, y de esta forma se genera la conducta lingüística del agente computacional.

Id	Tipo	Pars.	Actos Retóricos Básicos
Ra_1	*Invitación*	*Tour*	*Saludo* *Presentación*(*Tour)* *Opción-s/n*(*hacer, 'una visita guiada'*)
Ra_2	*Oferta*	*Topicos*	*Introducción*(*Topicos,'el departamento','las áreas'*) *Opción-abierta-que*('*visitar'*, *areas*)
Ra_3	*Despedida*		*Agradecimiento*(*visitar, departamento*) *Despedida*
Ra_4	*Confirmación*	*Area*	*Confirmación*(*'vamos a', Area*)
Ra_5	*Continuation*	*Area*	*Agradecimiento*(*visitar, Area*) *Opción-s/n*('*seguir visitando'*, '*el departamento'*)
error	*Error*	*Concepto*	*error*(*Concepto*) *solicitar-repetición*

Tabla 2. Actos Retóricos del modelo de diálogo principal

Id	Tipo	Pars.	Actos Retóricos Básicos
ra_1	*Oferta*	*Topicos*	*Introducción*(*Topics, ai*) *Opción-abierta-como*(*visitar, ai*)
ra_2	*Descripción*	*Concepto*	*Presentación*(*Concepto, ai*) *Elaboración*(*Concepto, ai*)
ra_3	*Continuación*	*Concepto*	*Notificación*(*terminamos*, *Concepto, ai*) *Opción-s/n*('*seguir visitando'*, *ai*)
ra_4	*Conclusión*	*Concepto*	*Notificación*(*terminamos*, *Concepto*)
error	*Error*	*Concepto*	*error*(*Concepto*) *solicitar-repetición*

Tabla 3. Actos Retóricos del modelo de diálogo de inteligencia artificial

Pasamos ahora a la interpretación de las expresiones lingüísticas proferidas por el sujeto; como ya se ha visto, cada situación codifica las intenciones esperadas por el agente computacional durante en la interpretación de la misma; en nuestra notación, el identificador del acto de habla esperado en el par de salida de cada situación es el mismo que el identificador de la intención correspondiente en la definición de la situación; sin embargo, esta intención o acto del habla puede ser expresada o realizado con una gran variedad de conductas lingüísticas; por lo mismo, el sistema de administración de la conversación incluye también la definición de un modelo de lenguaje que asocia los identificadores de las intenciones (o actos del habla) con las formas lingüísticas superficiales mediante las cuales estos se realizan. En particular, los actos del habla interpretables en el diálogo principal de nuestro ejemplo son "*ok*", "*no*", "*ai*", "*pr*" & "*pl*", como se puede apreciar en la figura 3, y las diálogo subordinado "*ok*", "*no*", "*per*", "*area*" & "*proy*", como se puede apreciar en la figura 4. De hecho, en este modelo el usuario puede expresar tan sólo ocho intenciones diferentes, y dos de éstas ("*ok*", "*no*") ocurren en ambos diálogos. El modelo de lenguaje para la interpretación de estas intenciones se muestra en la tabla 4.

Intención	Expresiones
ok	*"si", "okey", "ok", "si, por favor", "por favor", "encantado", "me encantaría", "si, gracias", "aja", "hum, si"*
no	*"no", "no, gracias", "hum, no", "no tengo tiempo", "ahorita no, gracias", "nel", "nel pastel"*
ai	*"ai", "ia", "inteligencia artificial", "a inteligencia artificial", "inteligencia", "a inteligencia", "la de inteligencia", "la de inteligencia artificial"*
pr	*"lenguajes de programación", "lenguajes", "el area de lenguajes", "a lenguajes", "la de lenguajes"*
lp	*"por reconocimiento de patrones", "reconocimiento de patrones", "reconocimiento", "patrones"*
per	*"personal", "por personal", "el personal", "los investigadores", "por investigadores", "a personal"*
proy	*"proyectos", "por proyectos", "los proyectos", "quiero ver los proyectos", "los investigadores"*
area	*"áreas", "por áreas", "por área", "las áreas", "por secciones", "secciones", "las secciones"*

Tabla 4. Modelo del Lenguaje

Con esta información, el mecanismo de interpretación de la entrada lingüística se reduce a verificar en la tabla 3 (i.e. el modelo del lenguaje) si el mensaje recibido en una situación de tipo "escucha" corresponde con la realización de alguna de las intenciones esperadas en dicha situación. Adicionalmente, como la misma conducta lingüística puede expresar intenciones diferentes en diferentes situaciones, es posible que la inspección del modelo del lenguaje arroje más de una intención; sin embargo, de todas estas, sólo las que estén explícitamente definidas en la situación actual son consideradas por el proceso de interpretación, con lo que la ambigüedad pragmática que resulta de la inspección original de la tabla se reduce significativamente. En el contexto de nuestra investigación hemos denominado a este proceso *interpretación directa de los actos del habla*. Con el fin de ampliar la gama de expresiones con las que se pueda expresar una intención estamos considerando actualmente incluir expresiones regulares que incluyan a objetos léxicos designados, precedidos y seguidos con material léxico adicional. Adicionalmente, esta arquitectura abra la posibilidad de explorar modelos de lenguaje más complejos, tomando en cuenta también parámetros prosódicos, como se discute más adelante.

Finalmente, con estos elementos, podemos mostrar el tipo de diálogo producido por el sistema. En la tabla 5 ilustramos un diálogo actual producido por la versión actual del manejador; en la tabla se incluyen los turnos conversacionales, los actos retóricos realizados por el sistema en cada uno de sus turnos, y le texto final, realizado mediante plantillas. Por su parte, en los turnos del usuario se muestra la entrada

lingüística explícita, que se interpreta como una intención posible en la situación, de acuerdo al modelo del lenguaje. Como se puede apreciar el diálogo luce bastante natural, a pesar del limitado número de intenciones de este dominio. Por otra parte, a pesar de su simplicidad, el modelo puede ser aplicado a una gran variedad de dominios con la condición de que el conjunto de intenciones relevantes sea limitado, y pueda ser identificado a través del análisis. Asimismo, la conducta lingüística del agente computacional debe poder ser estructurada en términos de un conjunto de actos retóricos básicos, cuya identificación dependerá de la investigación empírica del lenguaje. Sin embargo, el modelo, tal como se ha presentado, no es suficientemente rico para modelar los diálogos de diseño de cocinas de nuestro experimento. En la siguiente sección hacemos un análisis de estas limitaciones, e investigamos, de manera especulativa, cómo generalizar el modelo; asimismo, exploramos cómo enriquecer el modelo mediante el uso de información prosódica.

Turno	**Part.**	**Actos Retóricos**	**Texto**
1	s:	*Saludo* *Presentación* *Opción-s/n*	*Bienvenido* *Este es el departamento de ciencias de la computación* *quieres una visita guiada?*
2	u:		*Si, por favor*
3	s:	*Introducción* *Opción-abierta-que*	*El departamento tiene las áreas de inteligencia artificial, reconocimiento de patrones y lenguajes de programación* *Qué área quieres visitar*?
4	u:		*Inteligencia*
5	s:	*Confirmación* *Acto motor* *Introducción* *Opción-abierta-como*	*Okey, vamos a inteligencia artificial* El robot se mueve a IA *Esta es el área de inteligencia artificial; podemos hacer la visita por personal académico, área o proyectos* *Cómo quieres hacer la visita?*
6	u:		*Por proys*
7	s:	*Error* *solicitar_repetición*	*Perdón no te entendí* *Podrías repetirlo?*
8	u:		*Por proyectos*
9	s:	*Presentación* *Elaboración* *Notificación*	*En inteligencia artificial se desarrollan los proyectos DIME y GOLEM* *Las tareas principales son la creación de prototipos con manejadores de diálogos inteligentes, con entrada y salida hablada en español, y visión por computadora* *Terminamos con proyectos de inteligencia artificial*

		Opción-s/n	*Quieres seguir visitando inteligencia artificial?*
10	u:		*Okey*
11	s:	*Introducción* *Opción_abierta*	*inteligencia artificial tiene personal académico y área* *Cómo quieres visitar inteligencia artificial?*
12	u:		*Por áreas*
13	s:	*Presentación* *Elaboración* *Notificación* *Opción-s/n*	*En inteligencia artificial se investiga en sistemas multimodales inteligentes, gramáticas computacionales del español, razonamiento gráfico y creatividad* *Y también en visión computacional* *Terminamos con área de inteligencia artificial* *Quieres seguir visitando inteligencia artificial*
14	u:		*No*
15	s:	*Notificación* *Agradecimiento* *Opción-s/n*	*Terminamos con inteligencia artificial* *Gracias por visitar inteligencia artificial* *Quieres seguir visitando el departamento?*
16	u:		*No, gracias*
17	s:	*Agradecimiento* *Despedida*	*Gracias por visitar el departamento de ciencias de la computación* *Hasta luego!*

Tabla 5. Diálogo producido por la versión actual del administrador de diálogos

4. Enriqueciendo el modelo

El modelo presentado en la sección anterior tiene una flexibilidad mayor que los modelos de diálogos basados en autómatas de estados finitos, y su fuerza expresiva puede compararse a aquellos basados en *frames* o *templates*[2]; de hecho, el sistema permite definir modelos de diálogos con bastante facilidad, siempre y cuando las intenciones del dominio puedan ser caracterizadas de antemano mediante el análisis. Sin embargo, el tipo de diálogos orientados a solución de problemas y tareas de diseño, como los que se presentan en nuestro corpus, presenta una complejidad

[2] Para una discusión acerca de los tipos de manejadores de diálogos así como su expresividad y limitaciones ver, por ejemplo, Jurafsky & Martín (2000), Capítulo 19.

adicional, por lo que nuestro modelo debe extenderse en varias dimensiones; por lo mismo, actualmente estamos contemplando revisar la definición de situación conversacional, la caracterización de los actos del habla o intenciones de la situación y el modelo de interpretación directa de los actos del habla. Como un consecuencia de esta revisión se plantea la necesidad de enriquecer el modelo con información prosódica, y se describen las acciones que estamos realizando en esta dirección.

Una limitación del modelo es que las intenciones asociadas a cada situación tienen que estar completamente especificadas, y en contextos conversacionales ricos éste no es siempre el caso. En nuestro modelo es posible, por ejemplo, asociar una situación de tipo "escuchar" a las cuatro acciones posibles en el dominio (mover, poner, quitar y mostrar); adicionalmente, por el contexto conversacional es posible determinar que el agente de dichas acciones es necesariamente el sistema; sin embargo, los argumentos restantes de estas acciones no están determinados y es necesario embarcarse en un sub-diálogo para establecer estas referencias, lo cual queda ya fuera del alcance del modelo.

En este sentido, tenemos contemplado permitir que las intenciones asociadas a las situaciones queden sub-especificadas; esto tendrá dos consecuencias inmediatas: la primera es la necesidad de definir también transiciones sub-especificadas en el acto del habla de salida, y que estas transiciones lleven al sistema hacia modelos de diálogos subordinados (mediante situaciones recursivas) cuyo objetivo sea determinar los argumentos de la acción. En relación a nuestro ejemplo, esto nos llevará a definir un sub-diálogo para determinar la referencia de la estufa (i.e. en utt3) y otro más para determinar la referencia de "hacia la izquierda". Este último sub-diálogo estará especializado en el lenguaje espacial, por lo que sus situaciones e intenciones estarán dirigidas a resolver este problema de posicionamiento; dicho sub-diálogo incluirá también situaciones de tipo motor que serán responsables de la realización de los actos de apuntar, así como de las expresiones lingüísticas que acompañan a estos actos. Una vez que los argumentos de la acción de mover queden determinados, el control regresará a la situación asociada a la acción de movimiento.

La segunda consecuencia es que para satisfacer las intenciones asociadas al dominio de diseño es necesario realizar acciones concretas, como los movimientos de la estufa en utt18 y utt22, además de la acciones meramente comunicativas. Estas acciones sólo pueden llevarse a cabo cuando los argumentos de *mover* están ya completamente determinados, y para su realización se requiere la elaboración y ejecución de un plan, en este caso con razonamiento gráfico, sobre el dominio de diseño. Por lo mismo, la especificación de la situación debe ser también enriquecida con los esquemas que permitan generar estos planes, y con la definición del contexto sobre el cual se llevan a cabo dichas acciones. En la siguiente etapa de nuestra investigación tenemos contemplado agregar estos elementos a la estructura de las situaciones, con el fin de modelar a los diálogos como un proceso de navegación sobre situaciones cuyas intenciones sean la satisfacción de acciones sobre el dominio de diseño.

La segunda línea para la generalización del modelo se centra en la naturaleza de los actos del habla; en la presente versión, el sistema reconoce instancias concretas y no tipos de actos del habla; por ejemplo, nuestro sistema reconoce que la expresión "*si, por favor*" es un "*okey*", pero no reconoce que este acto del habla es la *aceptación de una oferta*. Sin embargo, en diálogos con una conducta lingüística más rica, y tomando en cuenta la sub-especificación de las intenciones, más importante que reconocer el nombre de una intención concreta es reconocer su tipo. Por ejemplo, la utt11 es una pregunta de confirmación o *check*, e inferir este hecho es indispensable para determinar su fuerza ilocusoria de la elocución, es decir, la intención que se expresa (el agente verifica que entendió correctamente la directiva que el sujeto expresó anteriormente). En dicha situación, interpretar la conducta lingüística explícita en términos del tipo de acto del habla es todo lo que se necesita para realizar el acto retórico asociado a la intención, y pasar a la situación siguiente. Por lo mismo, es necesario extender nuestro modelo de interpretación directa de los actos del habla codificando, además de la intención concreta, su tipo. En este sentido, uno de los objetivos actuales de nuestra investigación es determinar de manera empírica los tipos de actos del habla que se dan en el corpus DIME; para este efecto, tomamos como punto de partida el esquema de anotación de actos del habla DAMSL (Allen & Core, 1997) y versiones subsecuentes como la de Stent y Allen (2000). Asimismo, consideramos que los actos del habla producidos por el sujeto e interpretados por el agente computacional deben corresponder con los actos retóricos ejecutados por el agente e interpretados por el sujeto. Esta última restricción nos lleva a establecer una correlación entre las categorías de actos del habla estudiadas en sistemas de diálogo, como en los sistemas TRAINS y TRIPS (Allen *et al*., 2000), y las formas retóricas de RST (Mann & Thompson, 1988). Como resultado de este desarrollado esperamos contar con un inventario de los tipos de actos del habla que se dan en nuestro corpus, así como de los tipos de actos retóricos necesarios para modelar la conducta lingüística del agente computacional.

La identificación de los tipos de actos del habla nos lleva también a la necesidad de incorporar información prosódica al sistema de reconocimiento; intuitivamente, los tipos de los actos del habla están asociados a patrones prosódicos característicos; esto es evidente a nivel de la modalidad declarativa, interrogativa e imperativa de los enunciados, pero existe evidencia que la información prosódica puede estar asociada a tipos más específicos, como las preguntas *check* en oposición a los enunciados declarativos, las preguntas absolutas y las preguntas pronominales (Shriberg *et al*., 1998). En base a estas consideraciones tenemos contemplado utilizar información prosódica para determinar el tipo de acto del habla interpretado. La meta es mantener el esquema de interpretación directa de los actos del habla, pero aumentarlo con el tipo de acto que surja del patrón entonativo, así como con la información léxica relevante para la identificación de la intención.

La identificación de los patrones prosódicos permitirá adicionalmente la generación de la salida no sólo en términos del texto que se genera actualmente mediante las

plantillas asociadas a cada acto retórico, sino también con el patrón entonativo correspondiente, y aumentar de este modo la calidad del salida hablada.

El programa de investigación aquí esbozado requiere de una amplia investigación empírica; con este propósito, actualmente se está llevando a cabo la transcripción de los diálogos del corpus DIME en varios niveles de representación lingüística; en particular, se está etiquetando el corpus a nivel fonético y fonológico, con el fin de estudiar los fenómenos del habla a nivel acústico y utilizar esta información en el proceso de reconocimiento de voz; a nivel prosódico se está transcribiendo la sílaba, tanto fonética como fonológica, la palabra y los índices de separación (*Break Indices*) de acuerdo al sistema de transcripción prosódica ToBI (Beckman *et al*, 2000); además, se está llevando a cabo la transcripción prosódica a nivel fonético utilizando el sistema INTSINT, junto con su algoritmo asociado de estilización de la curva de frecuencia fundamental MOMEL (Baqué & Estruch, 2003). Finalmente, estamos también iniciando la transcripción pragmática para la identificación de los actos del habla y las estructuras retóricas que ocurren en el corpus, como ya se ha comentado.

Adicionalmente se tiene planeado realizar la etiquetación de los actos multimodales, como los actos de apuntar y manipular objetos, como se ve en la tabla 1. Este esfuerzo de etiquetación se está llevando a cabo de manera coordinada, y las transcripciones a los diferentes niveles de representación se están llevando a cabo de manera alineda. En la figura 5 se muestra la trancripción de la utt1, de la figura 1. Los primeros 6 niveles de transcripción del cuadro superior (alófonos, fonemas, sílaba fonética, sílaba fonológica, palabra e índices de separación), así como el nivel pragmático en la última línea, se están llevando a cabo en la herramienta CSLU Tool Kit´s SpeechView[3]; en el séptimo nivel se muestra la transcripción prosódica a nivel fonético de la entonación de acuerdo con el sistema INTSINT; sin embargo, con el fin de alinear esta información con el resto de los niveles de transcripción, se desarrolló un *script* para traducir estas etiquetas al formato del SpeechView y poder compararlas con el resto de los niveles de transcripción; por su parte, la curva estilizada de la frecuencia fundamental generada por el algoritmo MOMEL, se muestra, alineada con el resto de los niveles, en el cuadro inferior de la misma figura.

Concluimos este artículo mencionando que el objetivo a largo de este proyecto es construir un sistema conversacional capaz de llevar a cabo el tipo de diálogos que aparecen en el corpus DIME. En el corto y mediano plazo se pretende terminar la etiquetación del corpus en todos los niveles mencionados, con el fin de tener una base empírica para el estudio de este tipo de diálogos, incluyendo la transcripción de los eventos gráficos. Adicionalmente se espera contar con un inventario de los actos del habla y las estructuras retóricas que ocurren en el dominio, así como la caracterización de su estructura entonativa; se pretende también contar con un

[3] http://cslu.cse.ogi.edu

algoritmo para la interpretación directa de los actos del habla que permita identificar el tipo del acto del habla expresado por el sujeto en base a información prosódica, así como la información léxica necesaria para caracterizar la intención específica expresada por el sujeto.

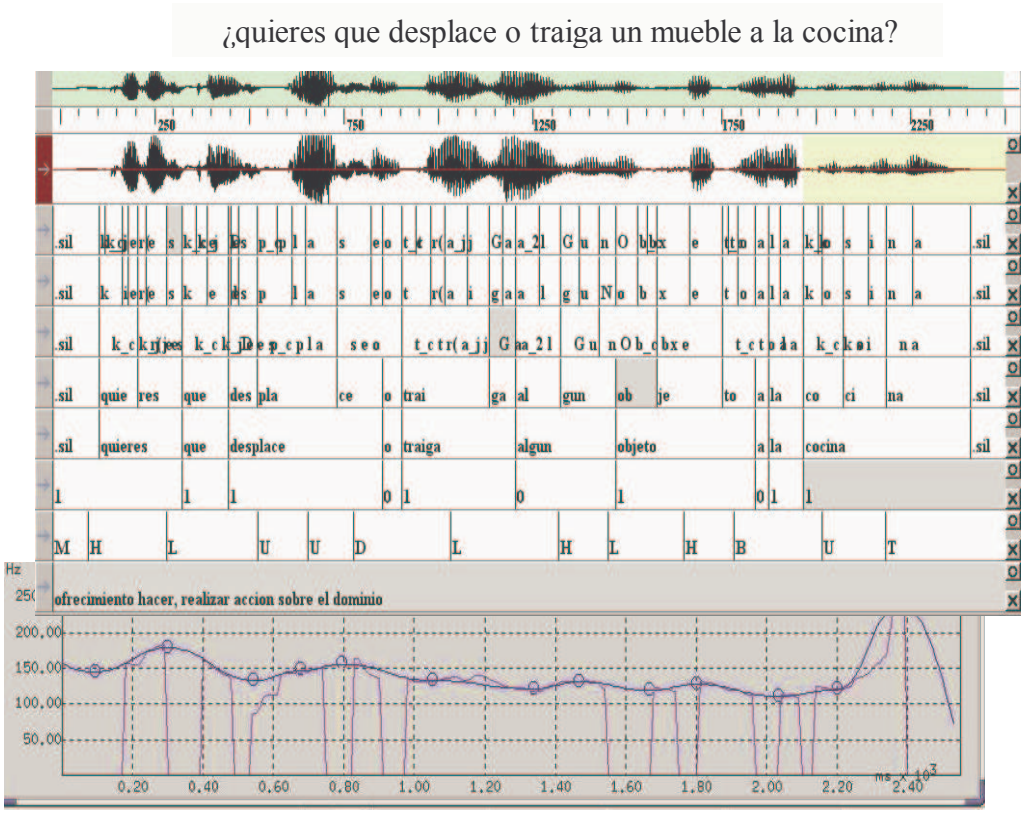

Figura 5. Niveles de etiquetación del Corpus DIME

5. Agradecimientos

Se agradece la entusiasta participación de Ivan Meza, Javier Cuétara, Sergio Coria, Hayde Castellanos, Ivonne López, Fernanda López, Varinia Estrada, Patricia Pérez, Iván Moreno, Isabel López, Israel Vázquez y demás miembros del proyecto DIME; se agradece también la entusiasta participación de Luis Villaseñor en el INAOE, en Tonanzintla, y de Joaquim Llisterri y Montserrat Riera, en la Universidad Autónoma de Barcelona. Se agradece también el apoyo de los proyecto CONACYT 39380-A y PAPIIT-UNAM IN111700.

6. Referencias

[Allen & Core, 1997] Allen, J. and Core. M. (1997). Draft of DAMSL: Dialog Act Markup Several-Layers. http://www.cs.rochester.edu:80/research/trains/annotation/RevisedManual/RevisedManual.html

[Allen *et al*., 2000] Allen, J.F., D.K. Byron, M.O. Dzikovska, G.M. Ferguson, L. Galescu, and A.J. Stent. (2000). "An architecture for a generic dialogue shell," J. Natural Language Engineering 6, 3 (Special Issue on Best Practices in Spoken Language Dialogue Systems Engg.), 1–16.

[Baqué & Estruch, 2003]. Baqué, L. y Estruch, M. (2003). Modelo de Aix-en-Provence, en *Teorías de la Entonación*, Pilar Prieto (Ed.), Editorial Ariel, S. A., Barcelona.

[Beckman, *et al*., 2000] Beckman, M., Diaz Campos, M., Tevis and J, Morgan, T. (2000). Intonation across Spanish, in the Tones and Break Indices framework, *Forbus* (14), pp. 9 – 36.

[Dahlbäck, *et al*., 1993] Dählback, N., Jönsson, A and Ahrenberg, L (1993). Wizard of Oz Studies – Why and How, *Knowledge-based Systems*, 6(4), pp. 258 – 266.

[Jurafsky & Martin, 2000]. Jurafsky, D. and Martín, J. (2000). Speech and Language Processing, Prentice Hall, New Jersey.

[Mann &Thompson, 1988] Mann, W. C. and Thompson, S. A. (1988). Rhetorical Structure Theory: Towards a functional theory of text organization", *Text* 8(3), pp. 243 – 281.

[Shriberg, E., 1998] Shriberg, E., Bates, R., Tylor, P., Stolcke, A., Jurafsky, D., Ries, K., Coccaro, N., Martin, R., Meeter, M. and Ess-Dykema, C. V. (1998). Can Prosody aid the automatic classification of dialog acts in conversational speech? *Language and Speech*, 41(3 – 4), 439 – 487.

[Stent & Allen, 2000] Stent, A.J. and J.F. Allen (2000). Annotating argumentation acts in spoken dialog, TR 740 and TRAINS TN 00-1, Computer Science Dept., U. Rochester.

[Villaseñor *et al*., 2001] Villaseñor, L., Massé, A. & Pineda, L. A. (2001). The DIME Corpus, Memorias 3º. Encuentro Internacional de Ciencias de la Computación ENC01, Tomo II, C. Zozaya, M. Mejía, P. Noriega y A. Sánchez (eds.), SMCC, Aguascalientes, Ags. México, Septiembre, 2001.