



Inteligencia Artificial. Revista Iberoamericana
de Inteligencia Artificial

ISSN: 1137-3601

revista@aepia.org

Asociación Española para la Inteligencia
Artificial
España

López Cózar, R.

Procesamiento de Información Multilingüe en un Sistema de Diálogo

Inteligencia Artificial. Revista Iberoamericana de Inteligencia Artificial, vol. 8, núm. 22, primavera,
2004, pp. 161-174

Asociación Española para la Inteligencia Artificial
Valencia, España

Disponible en: <http://www.redalyc.org/articulo.oa?id=92582212>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

redalyc.org

Sistema de Información Científica

Red de Revistas Científicas de América Latina, el Caribe, España y Portugal

Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto

Procesamiento de Información Multilingüe en un Sistema de Diálogo

R. López-Cózar

Dpto. Lenguajes y Sistemas Informáticos, E.T.S. Ingeniería Informática
18071 Universidad de Granada, Tel.: +34 958 240579, FAX: +34 958 243179
rlopezc@ugr.es

En este artículo se presenta el resultado de nuestro trabajo actual en el campo de los sistemas de diálogo relacionado con el procesamiento multilingüe, cuyo objetivo es permitir que los usuarios de un sistema de diálogo puedan interactuar con el mismo utilizando diversos idiomas, como por ejemplo castellano, inglés, francés, etc. El artículo se inicia con una introducción a la tecnología de los sistemas de diálogo que incluye una visión panorámica de su evolución histórica, así como una descripción de la arquitectura básica de un sistema de diálogo convencional. A continuación se presenta un sistema de diálogo en el que hemos incorporado la capacidad de procesamiento multilingüe, comentándose la finalidad y funcionamiento de sus componentes principales.

Posteriormente se describe la aproximación que hemos seguido para incorporar este tipo de procesamiento, tratando diversos aspectos relacionados con reconocimiento del habla, comprensión del lenguaje, control del diálogo, generación de frases y síntesis de voz. Finalmente se presentan las conclusiones y se indican algunas líneas de trabajo futuro.

Procesamiento de Información Multilingüe en un Sistema de Diálogo

R. López-Cózar

Dpto. Lenguajes y Sistemas Informáticos, E.T.S. Ingeniería Informática
18071 Universidad de Granada, Tel.: +34 958 240579, FAX: +34 958 243179
rlopezc@ugr.es

Resumen

En este artículo se presenta el resultado de nuestro trabajo actual en el campo de los sistemas de diálogo relacionado con el procesamiento multilingüe, cuyo objetivo es permitir que los usuarios de un sistema de diálogo puedan interactuar con el mismo utilizando diversos idiomas, como por ejemplo castellano, inglés, francés, etc. El artículo se inicia con una introducción a la tecnología de los sistemas de diálogo que incluye una visión panorámica de su evolución histórica, así como una descripción de la arquitectura básica de un sistema de diálogo convencional. A continuación se presenta un sistema de diálogo en el que hemos incorporado la capacidad de procesamiento multilingüe, comentándose la finalidad y funcionamiento de sus componentes principales. Posteriormente se describe la aproximación que hemos seguido para incorporar este tipo de procesamiento, tratando diversos aspectos relacionados con reconocimiento del habla, comprensión del lenguaje, control del diálogo, generación de frases y síntesis de voz. Finalmente se presentan las conclusiones y se indican algunas líneas de trabajo futuro.

Palabras clave: Sistema de diálogo, información multilingüe, reconocimiento del habla, procesamiento de lenguaje natural, comprensión del habla, gestión del diálogo, generación de lenguaje natural, síntesis de voz.

Abstract

This paper presents our current work on spoken dialogue systems concerned with multilingual processing, in order to allow that users of a dialogue system can interact with it using different languages, for example Spanish, English, French, etc. Initially the paper introduces the performance of this kind of system and describes the evolution of this technology. Next it presents the dialogue system in which we have included the multilingual processing, commenting the objectives and performance of its main components. The paper describes later the approach followed to enable the multilingual processing in the system and explains how it has been setup regarding speech recognition, language understanding, dialogue management, sentence generation and speech synthesis. Finally, the paper presents the conclusions and indicates some possibilities for future work.

Keywords: Spoken dialogue systems, multilingual information, speech recognition, natural language processing, speech understanding, dialogue management, natural language generation, speech synthesis

1. Introducción

Los sistemas de diálogo son programas de computador diseñados para interactuar oralmente con los usuarios con objeto de proporcionar determinados servicios, como por ejemplo, información y reserva de viajes en avión (Seneff y Polifroni, 2000), información meteorológica (Zue et al. 2000; Padrell y Hernando, 2002; Wang et al. 2000; Nakano et al. 2001), compra de productos (López-Cózar et al. 2002a), información y reserva de viajes en tren (Billi et al. 1997), etc. Dado que el habla constituye la forma más natural de comunicación entre las personas, la finalidad de estos sistemas es aumentar la rapidez, efectividad y facilidad a la hora de realizar este tipo de tareas de forma automática. Los sistemas iniciales estaban muy limitados en cuanto al tipo de frases que los usuarios podían usar y al tipo de tareas a realizar; no obstante, durante las tres últimas décadas su complejidad ha aumentado notablemente, permitiendo una interacción más natural y cercana a la humana. A pesar de los avances realizados, la funcionalidad de estos sistemas está limitada a dominios de aplicación restringidos, los cuales permiten usar gran cantidad de conocimiento acerca de las posibles palabras y tipos de frases que probablemente usarán los usuarios para comunicarse con los sistemas.

La implementación de un sistema de diálogo constituye un problema complejo que se suele descomponer en varios subproblemas, los cuales se corresponden con módulos que realizan funciones concretas. Dichos módulos se suelen agrupar en dos bloques conocidos como *interfaz de entrada* e *interfaz de salida* (Figura 1), interconectados mediante un módulo de gestión del diálogo cuya finalidad es controlar el funcionamiento de ambas interfaces y gestionar la interacción usuario-sistema.

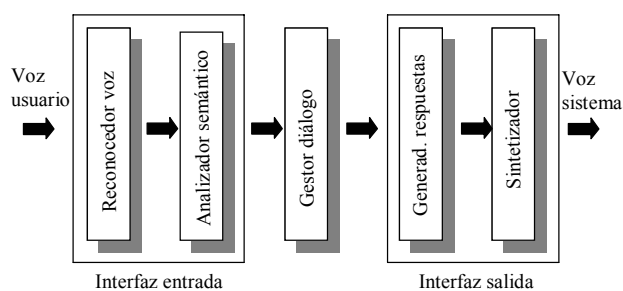


Figura 1. Estructura genérica de un sistema de diálogo unimodal

Cuando el sistema es de tipo *unimodal* la interacción se realiza únicamente mediante voz, en cuyo caso, la interfaz de entrada suele estar constituida por dos módulos: un *reconocedor del habla* (encargado de

recibir la voz del usuario y transformarla en una secuencia de palabras en formato de texto) y un *analizador semántico* (cuya finalidad es procesar las palabras proporcionadas por el reconocedor del habla para obtener su significado). Dichos procesos requieren hacer frente a un gran número de problemas para intentar evitar que el reconocedor del habla proporcione palabras no pronunciadas por el usuario y el analizador semántico obtenga un significado erróneo de las mismas (López-Cózar et al. 2002b). Cuando el sistema de diálogo es de tipo *multimodal* en la interfaz de entrada se utilizan *canales* de información adicionales a la voz, relacionados por ejemplo con los gestos o movimiento de los labios del usuario (Fisher et al. 2000; Gibbon et al. 2000; Darrell et al. 2002).

La interfaz de salida está constituida por módulos encargados de proporcionar la información solicitada por el usuario y otros tipos de mensajes necesarios para el control del diálogo. Para ello, se usan unas modalidades de comunicación u otras dependiendo de que el sistema de diálogo sea unimodal o multimodal. Cuando el sistema es unimodal, esta interfaz suele estar constituida únicamente por dos módulos: un *generador de respuestas* (cuya finalidad es construir frases en formato de texto a partir de la representación interna de la información procesada por el sistema) y un *síntetizador de voz* (cuya finalidad es generar la salida oral del sistema a partir de dichas frases). La generación de la salida oral se suele realizar mediante un proceso de transformación texto-a-voz de las frases en formato de texto, aunque son posibles otras técnicas, como por ejemplo concatenación de frases pregrabadas. Análogamente al caso de la interfaz de entrada, cuando el sistema es multimodal se utilizan canales de salida adicionales a la voz, como por ejemplo, gráficos o imágenes de un agente animado con aspecto humano que realiza gestos y mueve sus labios sincronizadamente con la salida oral (House et al. 2001; Wahlster et al. 2001).

Los sistemas de diálogo multilingües se caracterizan por permitir la interacción de usuarios que hablan idiomas diferentes, es decir, un mismo sistema multilingüe puede ser usado por un usuario que hable castellano, inglés o francés, por ejemplo (Zue et al. 1997). Este tipo de sistemas de diálogo resulta de gran interés en la actualidad, como prueba la existencia de diversos proyectos de investigación dedicados al desarrollo de esta tecnología. Entre ellos, cabe citar el proyecto E-Matter¹ (*E-Mail Access through the Telephone using speech TEchnology Resources*), en el que participa la empresa española Telefónica I+D junto con diversas universidades de este país. El objetivo de este

¹ <http://www.ub.es/gilcub/e-matter>

proyecto es desarrollar un sistema de diálogo multilingüe que facilite el acceso telefónico al correo electrónico, permitiendo que el usuario pueda acceder a su buzón de correo, escuchar los mensajes entrantes, seleccionar un determinado mensaje y escucharlo, enviar un mensaje de respuesta, etc., basándose todo ello únicamente en la interacción oral. La funcionalidad deseada para este sistema es la siguiente: Una vez que el usuario selecciona un mensaje, el sistema detecta automáticamente el idioma en que está escrito, realiza una corrección de posibles errores gramaticales y a continuación lo lee utilizando un conversor texto-a-voz correspondiente a dicho idioma. Se pretende que el sistema pueda leer mensajes escritos en cualquiera de las lenguas habladas en España (castellano, catalán, gallego o vasco) y en dos idiomas de la Unión Europea (inglés y francés), permitiendo la fácil adición de otros idiomas.

Otro proyecto relacionado con el procesamiento multilingüe del habla, desde un punto de vista más genérico, es el denominado C-ORAL-ROM² (*Integrated Reference Corpora for Spoken Romance Languages*), en desarrollo durante los años 2001-2004. Se trata de un proyecto internacional en el que participa la Universidad Autónoma de Madrid (España), el Instituto Cervantes (España) y otras instituciones de Portugal, Francia e Italia. La iniciativa tiene dos motivaciones principales; por una parte, subrayar la importancia estratégica de los recursos semánticos para el desarrollo de la sociedad de la información y las tecnologías del lenguaje humano, y por otra, potenciar los recursos semánticos disponibles en lenguas románicas, dado que son muy escasos en comparación con los existentes en el idioma inglés y las lenguas germánicas. De ahí que una de las principales tareas del proyecto sea desarrollar un conjunto de herramientas y recursos semánticos (tanto orales como en formato de texto) para cuatro idiomas romances: castellano, francés, portugués e italiano.

Tras esta introducción, el resto del artículo está organizado de la siguiente forma. En la sección 2 se presenta el sistema de diálogo experimental; se hace referencia al corpus de diálogos usado para desarrollar dicho sistema y se describe la estructura modular del mismo, explicando someramente el funcionamiento de sus componentes principales. En la sección 3 se presenta la nueva versión del sistema en la que estamos trabajando actualmente para incorporar el procesamiento multilingüe; se comentan los cambios realizados en los diversos

módulos del sistema a fin de incorporar este tipo de procesamiento y se muestran dos ejemplos de la interacción permitida actualmente mediante los idiomas castellano e inglés. Finalmente, la sección 4 presenta las conclusiones e indica algunas líneas de trabajo futuro.

2. Sistema de diálogo VIAJERO

El proyecto TEL1999-0619 de la Comisión Interministerial de Ciencia y Tecnología (España) llevado a cabo durante los años 1999-2001 ha permitido desarrollar un sistema de diálogo unimodal, llamado VIAJERO, cuya finalidad es atender telefónicamente a clientes de empresas de transporte de viajeros en autobuses de línea. El sistema ha sido diseñado para procesar reservas de billetes realizadas por los clientes y proporcionar información acerca de precios de viajes, horarios de salida y llegada de autobuses, etc., usando el idioma castellano durante la interacción (López-Cózar et al. 2000). Gracias a la participación de una empresa colaboradora interesada en la implantación del sistema, se ha grabado un corpus de conversaciones telefónicas entre clientes y telefonistas de la empresa que contiene 1.957 diálogos. La Figura 2 muestra un ejemplo de estos diálogos (*T* = telefonista, *C* = cliente).

T: Empresa "X", buenos días. Le atiende José Francisco
C: Hola, buenos días. Mire, yo quería información de los autobuses de ... bueno, los horarios para mañana por la noche a Madrid
T: Pues por la noche ... a las 21:00 horas, 23:00 y 1:30 de la madrugada
C: Vale ... Problema no hay de billetes, ¿no?
T: Hombre, eso no se lo puedo yo asegurar. Si viene usted diez minutos antes puede que se encuentre usted sin billete
C: Si, claro. Vale, pues ya iré ... ya me acercaré a comprarlo
T: Muy bien
C: Vale, muchas gracias
T: Vale, hasta luego. Adiós
C: Hasta luego

Figura 2. Diálogo de ejemplo telefonista-cliente

Usando dicho corpus se han seleccionado aleatoriamente 500 diálogos a fin de realizar un estudio de los mismos que permita conocer el vocabulario, las estructuras sintácticas y semánticas que los usuarios suelen usar para hablar con los telefonistas, y el tipo de frases que el sistema debe generar para proporcionar información e interactuar con los usuarios (Graff y Bird 2000; Andreas et al. 2000). Los 500 diálogos seleccionados contienen 3.677 frases de clientes y un total de 17.839 palabras (1.004 palabras se corresponden con nombres de ciudades y aproximadamente un 2,5% de las mismas son disfluencias, como "eh", por ejemplo). Las frases de los usuarios han sido clasificadas en los 9 tipos mostrados en la Tabla 1. Las frases de los tipos 1 y 2 han sido usadas conjuntamente para crear una

*bigramática*³ (Huang et al. 2001; Jelinek 1999; Nasr et al. 1999) que se usa al principio del diálogo a fin de permitir el reconocimiento de consultas y reservas. A partir de los restantes tipos de frases se han creado 7 bigramáticas (una por cada tipo) que se usan cuando el sistema solicita un determinado dato al usuario; por ejemplo, cuando solicita el número de teléfono del usuario se usa la bigramática creada mediante las frases del tipo 7. Además, en cada uno de los nueve tipos de frases se han incluido 150 frases adicionales (creadas manualmente) para permitir que el usuario pueda solicitar ayuda, solicitar la repetición del anterior mensaje generado por el sistema o indicar que éste no le ha comprendido correctamente. Adicionalmente a las 8 bigramáticas indicadas anteriormente, se ha creado una bigramática más cuya finalidad es permitir que el usuario pueda identificar qué datos han sido reconocidos incorrectamente por el sistema, para lo cual se han usado otras 50 frases creadas manualmente.

	Tipo frase	Subtipo
1	Consulta	Existencia autobús
		Existencia plazas
		Reserva
		Precio billete
		Horario salida
		Horario llegada
		Duración viaje
2	Reserva plazas	-
3	Ciudad	-
4	Día de semana	-
5	Día de mes	-
6	Hora	-
7	Número de teléfono	-
8	Confirmación	-
9	Número plaza reservadas	-

Tabla 1. Tipos de frases de clientes del corpus de frases en castellano

La Figura 3 muestra la estructura modular del sistema VIAJERO. La voz del usuario constituye la entrada de un reconocedor de habla continua independiente del locutor, basado en unidades fonéticas independientes del contexto modeladas mediante modelos ocultos de Markov semicontinuos (Rabiner y Juang 1986, 1993). El reconocedor del habla procesa la señal de voz del usuario y proporciona como resultado una secuencia de palabras que constituye la entrada del analizador semántico. Cada palabra w_i de la secuencia va acompañada por un número real $c_i \in (0,1)$ que

³ Modelado de lenguaje usado para estimar $P(w_i | w_{1..i})$, es decir, la probabilidad de aparición de cualquier palabra w en la posición i de cualquier frase, teniendo en cuenta que se conoce la palabra que ocupa la posición $i-1$ en la frase.

proporciona información acerca de la confianza del reconocedor en el reconocimiento correcto de dicha palabra (López-Cózar et al. 1999). Si c_i no supera un determinado umbral de confianza, el sistema considera que la palabra w_i ha sido reconocida de forma incorrecta, en cuyo caso la palabra se descarta. Si c_i supera el umbral de confianza, el sistema considera que la palabra w_i ha sido reconocida correctamente. Este mecanismo basado en valores de confianza no es infalible, dado que es posible que el valor de confianza de una palabra reconocida incorrectamente supere el umbral de confianza. Por este motivo el sistema solicita adicionalmente confirmaciones explícitas al usuario, a fin de verificar la corrección de las palabras reconocidas. El sistema consta de un diccionario que contiene las 1.341 palabras distintas que pueden ser reconocidas.

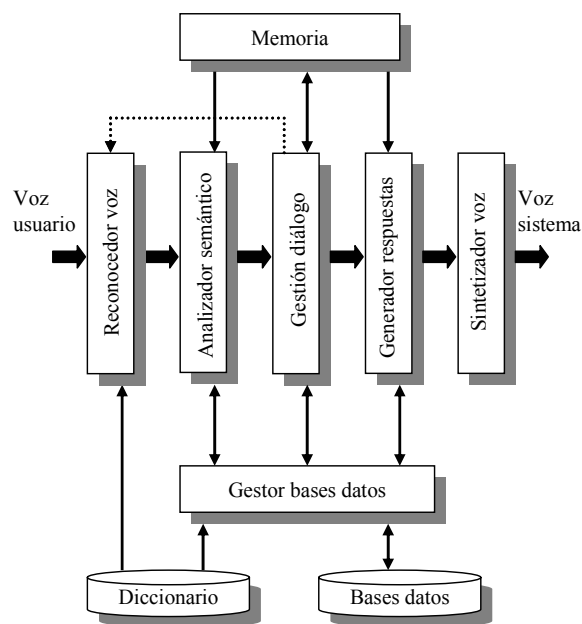


Figura 3. Estructura modular del sistema VIAJERO

El analizador semántico procesa las palabras existentes en la entrada proporcionada por el reconocedor del habla y aplica un conjunto de reglas de análisis para obtener el significado de las mismas (Boros et al. 1999; Schadle et al. 1999). El resultado del análisis se almacena mediante unas estructuras de representación del conocimiento denominadas *frames*⁴ (Allen, 1995), que modelan abstractamente entidades del mundo real junto con sus relaciones. Cada entidad consta de un conjunto de atributos (usualmente denominados *slots*) que se rellenan con

⁴ El almacenamiento de una representación semántica puede ocupar uno o varios frames.

valores determinados para representar una instancia concreta de dicha entidad.

El módulo de gestión del diálogo recibe los frames creados por el analizador semántico y emplea una estrategia de interacción de tipo mixto (Rosset et al. 1999; Relano et al. 1999), implementada mediante reglas “if ... then ... else ...”, para decidir la próxima acción que debe realizar el sistema, como por ejemplo, consultar las bases de datos para proporcionar la información solicitada por el usuario, confirmar alguna palabra, etc. Tras realizar la acción correspondiente, el módulo de gestión del diálogo decide el tipo de respuesta que debe generar el sistema, proporcionando al generador de respuestas la información obtenida de las bases de datos, en su caso. Para optimizar el funcionamiento del sistema, el módulo de gestión del diálogo indica al reconocedor del habla la bigramática que debe utilizar para procesar la siguiente frase del usuario. Al inicio del diálogo se usa una bigramática poco restrictiva que permite reconocer consultas y reservas de billetes, mientras que en determinados estados del diálogo se usan bigramáticas más restrictivas. Por ejemplo, si el valor de confianza correspondiente a la ciudad de destino pronunciada por el usuario no supera el umbral de confianza, el sistema genera el siguiente mensaje: “*Por favor, diga de nuevo el nombre de la ciudad a la que desea viajar*”, utilizándose para reconocer la siguiente frase del usuario una bigramática que sólo permite reconocer nombres de ciudades (y algunas otras palabras necesarias para la gestión del diálogo).

El generador de respuestas se encarga de construir en formato de texto los diversos tipos de mensajes que el sistema debe generar, para lo cual dispone de 34 patrones de frases distintos (Glass et al. 1994; Baptist y Seneff, 2000). Cada patrón consta de una serie de conceptos, expresiones y huecos, de forma que durante la generación de las respuestas se expanden los conceptos y las expresiones, y se rellenan los huecos con las palabras correspondientes. Además, el generador emplea reglas para flexionar correctamente las palabras (en cuanto a género y número) y usar pronombres en las frases teniendo en cuenta la información contextual.

El sintetizador de voz realiza la conversión a voz de los mensajes construidos en formato de texto. Dicho módulo está implementado mediante la plataforma de síntesis de voz multilingüe FESTIVAL⁵, desarrollada en la Universidad de Edimburgo (Reino Unido).

El módulo de memoria tiene como finalidad almacenar los frames creados por el analizador

semántico, usando un mecanismo de unificación de frames que permite procesar las informaciones parciales proporcionadas por el usuario. Por ejemplo, si en una determinada interacción el usuario indica que desea viajar a “*Madrid*” y en una interacción posterior indica que desea salir desde “*Granada*”, se generan dos frames para almacenar ambos datos⁶, los cuales se combinan mediante el proceso de unificación dando como resultado un único frame que contiene la suma de las informaciones contenidas en ambos frames, esto es, un frame cuyo contenido semántico indica que el usuario quiere viajar a “*Madrid*” saliendo desde “*Granada*”.

Las bases de datos almacenan las reservas realizadas por los usuarios así como la información que el sistema puede proporcionar, relacionada con los trayectos de autobuses (rutas, horarios de salida y llegada, duración de los viajes, precios, etc.).

3. Sistema VIAJERO II

Actualmente estamos trabajando en una nueva versión del sistema VIAJERO, llamada VIAJERO II, cuya finalidad es incorporar el procesamiento multilingüe para permitir la interacción de los usuarios mediante varios idiomas. Para implementar esta funcionalidad hemos considerado que a partir de una misma frase expresada en idiomas diferentes es posible obtener una misma representación semántica, denominada *interlingua*⁷ (Waibel et al. 1991; Roe et al. 1991; Hutchins y Somers, 1992; Wahlster, 1993, 1994; Takezawa et al. 1993). En dicha representación únicamente cambian las palabras específicas del idioma, manteniéndose inalterada la función pragmática de la frase, esto es, la intención del usuario. Ello permite crear una independencia entre el idioma utilizado por el usuario y el utilizado por el sistema, lo que podría permitir que el usuario se comunique con el sistema en un determinado idioma (castellano, por ejemplo), y el sistema se comunique con el usuario en un idioma distinto (como por ejemplo, inglés).

La aproximación al problema de la interacción multilingüe basada en la *interlingua* ha sido usada previamente con éxito en el desarrollo de diversos sistemas de diálogo, entre los que cabe destacar el sistema VOYAGER (Glass et al. 1995), diseñado para proporcionar información geográfica del área de Cambridge (Estados Unidos). Este sistema usa un analizador sintáctico denominado TINA (Seneff,

⁵ <http://www.cstr.ed.ac.uk/projects/festival>

⁶ El usuario también puede proporcionar ambos datos en una misma interacción, en cuyo caso sólo se genera un frame.

⁷ Este tipo de representación se usa habitualmente en los sistemas de traducción automática entre idiomas.

1992) para procesar frases pronunciadas en japonés, francés o italiano. Por ejemplo, para la frase en inglés "Where is the library near Central Square?" y su versión en italiano "Vicino a Central Square, dove che sta la biblioteca?", este analizador obtiene los árboles sintácticos mostrados en las Figuras 4 y 5, respectivamente.

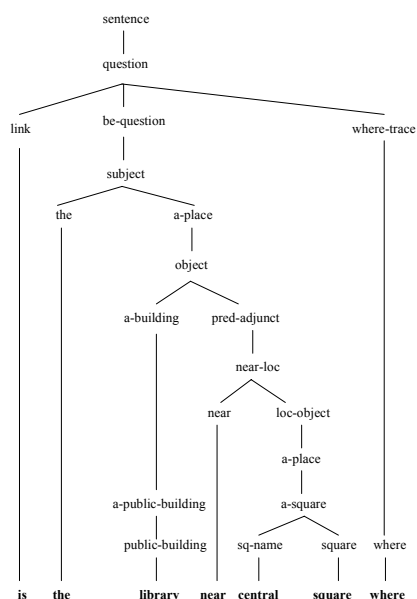


Figura 4. Árbol sintáctico de la frase "Where is the library near Central Square?"

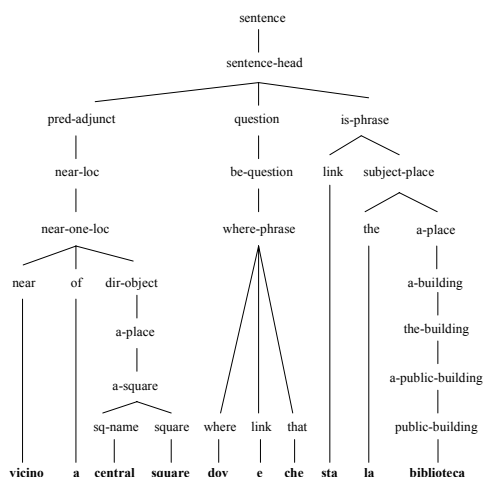


Figura 5. Árbol sintáctico de la frase "Vicino a Central Square, dove che sta la biblioteca?"

La ventaja de la aproximación basada en la interlingua reside en que a partir de ambos árboles de análisis se puede obtener la misma representación semántica de la frase (mostrada en la siguiente Figura), en la que únicamente cambian determinadas palabras específicas de cada idioma.

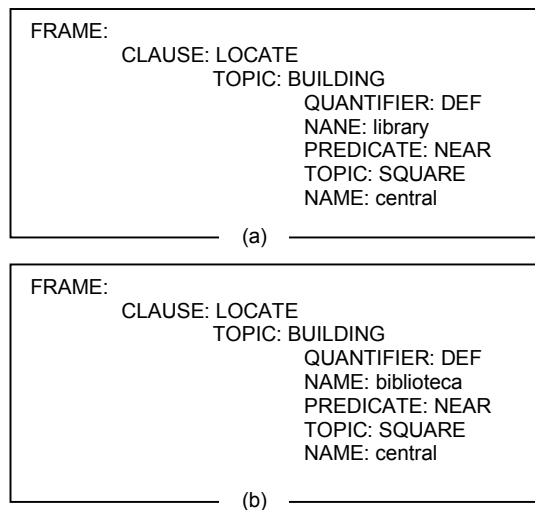


Figura 6. Representaciones semánticas obtenidas. (a) A partir de árbol de Fig. 4. (b) A partir de árbol de Fig. 5

A fin de facilitar el procesamiento de información multilingüe y la adición de nuevos idiomas en el sistema VIAJERO II, los diversos módulos del mismo se han creado independientes, en la medida de lo posible, del idioma utilizado en la interacción usuario-sistema. La información dependiente de cada idioma se almacena de forma externa utilizando tablas y reglas que son usadas por el analizador semántico y el generador de respuestas. Así, no es necesario modificar los módulos si se añade un nuevo idioma, sino que únicamente se han de añadir las correspondientes tablas y reglas.

3.1 Interfaz de entrada

En una primera etapa de desarrollo del sistema VIAJERO II hemos incorporado el inglés como idioma adicional al castellano. Ello ha requerido analizar un corpus de frases pertenecientes a dicho idioma, relacionadas con la tarea que debe realizar el sistema. El análisis ha permitido conocer los ítems léxicos y los modelos de lenguaje que debe usar el reconocedor del habla (Glass et al. 1999; Adda-Decker, 2001; Ma et al. 2002; Tian et al. 2002), así como las reglas necesarias para que el analizador semántico pueda extraer el significado de las frases en inglés.

3.1.1 Reconocimiento del habla

Los tipos de frases mostrados en la Tabla 1 han sido usados para compilar 8 bigramáticas de palabras (Huang et al. 2001). Para reconocer cada frase del usuario, el sistema usa un reconocedor de habla continua basado en HTK (*Hidden Markov Model Toolkit*) (Hain et al., 1999; Woodland et al., 1999), que usa modelos acústicos correspondientes al idioma de la frase, y selecciona una bigramática u otra en función del estado actual del diálogo.

Para incorporar el procesamiento del idioma inglés, se ha usado un corpus de frases construido mediante la traducción a inglés de un total de 400 frases de entre los 9 tipos de frases existentes en el corpus de frases en castellano (mostrados en la Tabla 1). Además, se ha traducido el 50% de frases creadas en castellano para permitir el control del diálogo por parte del usuario. Por consiguiente, el corpus de frases en inglés consta de $400 + 75 = 475$ frases, las cuales han sido usadas para compilar las nueve bigramáticas de palabras que usa el reconocedor del habla del sistema. Durante el procesamiento de cada frase en inglés, el reconocedor usa modelos acústicos entrenados para este idioma así como una de las nueve bigramáticas compiladas a partir del corpus de frases en inglés, la cual es seleccionada en función del estado actual del diálogo.

3.1.2 Comprensión de frases

Para incorporar el procesamiento multilingüe hemos implementado una metodología de análisis simple, pero efectiva, que permite extraer el significado de las frases pronunciadas por los usuarios. Para analizar cada frase, se usa un conjunto de reglas de análisis semántico de la forma:

$$RS \rightarrow C_1 C_2 C_3 \dots C_n$$

donde RS representa el contenido semántico de la regla y cada C_i representa un componente de la regla, que puede ser una clase de palabras (concretamente, palabras-clave⁸), o bien, otra regla semántica. Cada clase de palabras es un conjunto de palabras que tienen la misma función pragmática en las frases. El sistema usa un total de 34 clases de palabras, algunas de las cuales se muestran en la Figura 7.

Cuando una clase de palabras aparece entre corchetes en las reglas, dicha clase es opcional durante el análisis semántico; es decir, no es necesario que una palabra de la clase aparezca en la frase para que la regla se pueda aplicar en el análisis. El símbolo “[”

en las reglas representa *alternancia*, es decir, la regla se puede aplicar cuando en la frase aparece una palabra de alguna de las clases situadas a ambos lados del símbolo. Las palabras que no son consideradas palabras-clave (p. e. los artículos) se ignoran durante el análisis semántico; así, la regla FECHA mostrada en la Figura 8 puede ser aplicada para analizar la frase⁹: “el lunes”.

```

AYUDA = {"ayuda", "ayudes", ...}
BUS = {"bus", "autobus", "autobuses", ...}
CIUDAD = {"madrid", "sevilla", "barcelona", ...}
DESEO = {"quiero", "me gustaria", "necesito", ...}
DESTINO = {"ir a", "viajar a", ...}
DIA_S = {"lunes", "martes", "miercoles", ...}
DIA_M = {"uno", "primero", "dos", "tres", ...}
USABLE = {"hay", "teneis", "coger", ...}
ERROR = {"error", "equivocado", "equivocacion", ...}
HORA = {"una", "dos", "tres", ...}
INTERROG = {"que", "como", "cuanto", ...}
INTERV = {"proximo", "este", ...}
MES = {"enero", "febrero", "marzo", ...}
PLAZA = {"asiento", "plaza", ...}
PROCEDENCIA = {"desde", "de", ...}

```

Figura 7. Clases de palabras para el idioma castellano

```

DESEO_VIAJE ← [DESEO] DESTINO CIUDAD [FECHA]

DESEO_INFO ← EXISTE_BUS? |
              DURACION_VIAJE? |
              PRECIO? |
              HORAS_SALIDA? |
              HORAS_LLEGADA?

EXISTE_BUS? ← INTERROG BUS EXISTE [VIAJE]

DESEO_AYUDA ← [DESEO] AYUDA

FECHA ← [INTERV] DIA_S |
         [INTERV] DIA_M MES |
         DIA_S [INTERV]

CIUDAD_DESTINO ← CIUDAD P = ciudad_destino ? |
                  DESTINO CIUDAD

CIUDAD_ORIGEN ← CIUDAD P = ciudad_origen ? |
                  PROCEDENCIA CIUDAD

```

Figura 8. Reglas semánticas para el idioma castellano

Determinadas reglas pueden ser aplicadas (o no) en función del prompt generado por el sistema de diálogo, el cual permite resolver posibles ambigüedades que pueden existir a la hora de aplicar

⁸ Palabras necesarias para extraer el significado de las frases.

⁹ En este artículo se considera que una *frase* es el conjunto de todas las palabras generadas por el usuario en un turno del diálogo, con independencia de que sintácticamente constituyan una o varias frases.

las reglas. Por ejemplo, en las reglas CIUDAD_DESTINO y CIUDAD_ORIGEN mostradas en la Figura 8 se usa este tipo de información, denotada mediante un superíndice sobre el nombre de la clase de palabras CIUDAD.

Supongamos, por ejemplo, que en respuesta al prompt del sistema: “¿Desde dónde quiere salir?” (representado en las reglas como $P=ciudad_origen?$) el usuario pronuncia la frase “Madrid”. Si no se usara la información contextual, esta frase podría ser analizada por las dos reglas indicadas anteriormente, es decir, se produciría una ambigüedad en la aplicación de las reglas. Gracias a la información contextual proporcionada por el prompt se puede resolver esta ambigüedad, permitiéndose únicamente la aplicación de la regla CIUDAD_ORIGEN.

A modo de ejemplo, la siguiente Figura muestra el resultado del análisis semántico de la frase¹⁰ “quiero ir a madrid este viernes que autobuses hay”, obtenido mediante la aplicación de las reglas DESEO_VIAJE y EXISTE_BUS?.

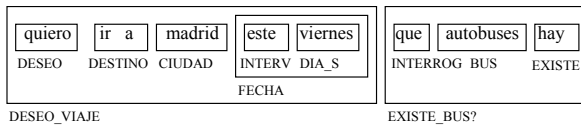


Figura 9. Representación semántica de una frase pronunciada en castellano

La siguiente Figura muestra cómo se almacena la representación semántica mostrada en la Figura 9, concretamente mediante dos frames (1.1 y 1.2).

FR	REGLA	CIUDAD DESTINO	FECHA
1.1	DESEO_VIAJE	madrid	este viernes

FR	REGLA
1.2	EXISTE_BUS?

Figura 10. Almacenamiento de la representación semántica mediante frames

Para incorporar el procesamiento del idioma inglés, se ha analizado el corpus de frases con objeto de clasificar las palabras-clave en clases de palabras. Algunas de las clases obtenidas se muestran en la siguiente Figura. Al igual que ocurre con las clases de palabras para el idiomas castellano, estas nuevas

clases de palabras están constituidas por palabras inglesas que tienen la misma función pragmática en el idioma inglés.

```

AVAIL = {"are", "is", "take", ...}
ARRIVAL = {"go to", "get to", "travel to", ...}
BUS = {"bus", "buses", ...}
CITY = {"madrid", "sevilla", "seville", "barcelona", ...}
DAY_W = {"monday", "tuesday", ...}
DAY_M = {"first", "second", "third", ...}
DESIRE = {"want", "would like", "need", ...}
DEPARTURE = {"from", "at", ...}
HOUR = {"one", "two", "three", ...}
HELP = {"help", "assistance", ...}
INTERV = {"this", "next week", ...}
MONTH = {"january", "february", "march", ...}
MISTAKE = {"mistake", "wrong", "error", ...}
OBJECT = {"bus", "buses", "seat", "seats", ...}
QUESTION = {"what", "how", "when", "which", ...}

```

Figura 11. Clases de palabras para el idioma inglés

El análisis de las frases del corpus ha permitido, asimismo, adaptar al idioma inglés las 24 reglas de análisis semántico construidas previamente para el idioma castellano. La parte izquierda de las reglas es exactamente igual en ambos casos; la única diferencia se encuentra en la parte derecha, en la que cambia el nombre de las clases de palabras y, en algunos casos, el orden de éstas para acomodarse al orden gramatical del idioma inglés. La siguiente Figura muestra la versión inglesa de las reglas mostradas en la Figura 8.

```

DESEO_VIAJE ← [DESIRE] ARRIVAL CITY [FECHA]

DESEO_INFO ← EXISTE_BUS? |
              DURACION_VIAJE? |
              PRECIO? |
              HORAS_SALIDA? |
              HORAS_LLEGADA?

EXISTE_BUS? ← INTERROG BUS AVAIL [TRAVEL]

DESEO_AYUDA ← [DESIRE] HELP

FECHA ← [INTERV] DAY_W |
        [INTERV] DAY_M |
        DAY_W [INTERV]

CIUDAD_ORIGEN ← CITY P = ciudad_origen ? |
                DEPARTURE CITY

CIUDAD_DESTINO ← CITY P = ciudad_destino ? |
                  ARRIVAL CITY

```

Figura 12. Reglas semánticas para el idioma inglés

A modo de ejemplo, la Figura 13 muestra el resultado del análisis de la frase: “i need to go to

¹⁰ La frase está escrita como ha sido proporcionada por el reconocedor de voz del sistema, esto es, con todas las palabras en minúscula y sin signos de puntuación.

madrid this friday what bus can I take". Como se puede observar, el resultado es similar al mostrado en la Figura 9; las palabras cambian pero el contenido semántico permanece inalterado pues se han aplicado las mismas reglas de análisis semántico: FECHA, DESEO_VIAJE y EXISTE_BUS?.

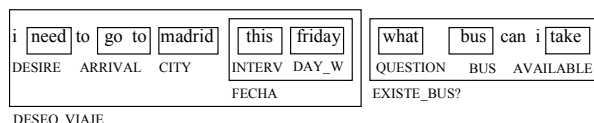


Figura 13. Análisis semántico de una frase pronunciada en inglés

3.2 Módulo de gestión del diálogo

Los frames obtenidos a partir del análisis de las frases pronunciadas en las interacciones de un usuario dado se almacenan en una estructura de tipo pila (*stack*) como la mostrada en la siguiente Figura.

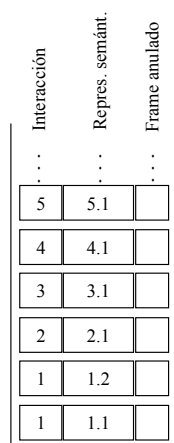


Figura 14. Pila para almacenamiento de frames

Esta estructura permite el procesamiento de las informaciones parciales que el usuario puede proporcionar a lo largo del diálogo. Tras insertar cada nuevo frame en la cima de la pila, se inicia un procedimiento de unificación que permite combinar frames, dando como resultado un frame que contiene la suma de las informaciones contenidas en los frames combinados. La gestión del diálogo se realiza en función del contenido de la pila: el frame situado en la cima determina la acción que debe realizar el sistema; por ejemplo, si se trata de un frame correspondiente a una regla de tipo DESEO_VIAJE pero el usuario no ha especificado la fecha del viaje, el sistema solicitará este dato al usuario en su siguiente turno del diálogo.

El funcionamiento de la pila y el mecanismo de unificación de frames es independiente del idioma

utilizado en la interacción, lo que permite que la gestión del diálogo sea exactamente igual en castellano e inglés.

La incorporación del procesamiento del idioma inglés no ha cambiado sustancialmente la gestión del diálogo. La única diferencia existente cuando hay más de un idioma disponible es la inclusión de un *prompt*¹¹ adicional al principio del diálogo, mediante el que el sistema solicita que el usuario especifique el idioma que desea utilizar durante la interacción (castellano o inglés, de momento). Mediante esta selección, el módulo de gestión del diálogo decide: el vocabulario, las bigramáticas y los modelos acústicos (para castellano o inglés) que debe utilizar el reconocedor del habla; el vocabulario, las clases de palabras y las reglas que debe usar el analizador semántico; los patrones que debe usar el generador de respuestas para generar las frases; y la voz (castellana o inglesa) que debe usar el sintetizador de voz para generar la salida oral del sistema.

3.3 Interfaz de salida

Para construir las respuestas en ambos idiomas, el sistema utiliza 34 patrones que describen los diversos tipos de frases que puede generar. A modo de ejemplo, la siguiente Figura muestra los patrones usados para proporcionar información en castellano acerca de la salida de autobuses. Una posible frase que se puede generar mediante estos patrones es la siguiente: *"El autobús con destino Madrid, sale desde el andén número tres, a las 18:30 horas."*. La generación de este tipo de mensajes se realiza expandiendo el patrón *salida_bus*, la cual se realiza expandiendo, a su vez, los patrones *destino* y *salida*. La expansión del patrón *destino* genera el fragmento *"El autobús con destino Madrid,"*, instanciando el concepto ciudad con la palabra *"Madrid"*.

```
salida_bus → <destino> <salida> "."
destino → "El autobús con destino" <ciudad> ","
salida → "sale" <origen> <hora>
origen → "desde el andén número" <anden> ","
hora → "a las" <hora_salida> "horas"
```

Figura 15. Patrones para generación de frases en castellano

Análogamente, el patrón *salida* genera el fragmento *"sale"* seguido de la expansión de los patrones *origen* y *hora*, el primero de los cuales genera el segmento *"desde el andén número tres,"* instanciando el

¹¹ Turno del diálogo en el que el sistema solicita un determinado dato o información al usuario.

concepto anden con la palabras “tres”, y el segundo genera el fragmento “a las 18:30 horas” instanciando el concepto hora_salida con la palabra “18:30”.

La instanciación de los conceptos se realiza usando datos que han podido ser proporcionados por el usuario, o bien, han podido ser obtenidos mediante consultas a las bases de datos del sistema. Si algún concepto no puede ser instanciado por no haberse obtenido un dato necesario, el patrón correspondiente no se expande, evitándose así la generación de frases incorrectas. Así, si en el ejemplo anterior el concepto anden no hubiera sido instanciado, el patrón origen no se hubiera expandido, en cuyo caso se hubiera generado la frase: “El autobús con destino Madrid, sale a las 18:30 horas.”.

Para construir las respuestas en inglés se han creado nuevas versiones de los patrones usados para el idioma castellano, teniendo en cuenta el distinto orden gramatical. La Figura 16 muestra la versión inglesa de los patrones mostrados en la Figura anterior, los cuales permiten generar, por ejemplo, la siguiente frase “The bus for Madrid, leaves from gate number three, at 18:30.”.

Los patrones usados en estos ejemplos son muy simples, no obstante, pueden llegar a ser bastante complejos si han de especificar cómo expandir los conceptos teniendo en cuenta aspecto como número, género, caso, contracciones de palabras, etc. Por ejemplo, en francés las palabras “de le” deben contraerse formando la palabra “du”; en castellano las palabras “a el” deben contraerse formando la palabra “al”; y en inglés las palabras “a other” deben contraerse formando la palabra “another”. Además, en los patrones se debe tener en cuenta el uso de pronombres; por ejemplo, en inglés se debe decidir entre usar “a” o “an” para hacer referencia al concepto de unidad, el cual se expresa de forma distinta en segmentos como “A cake” y “An animal”.

```

bus_departure → <destination> <departure> “.”
destination → “The bus for” <city> “,”
departure → “leaves” <source> <time>
source → “from gate number” <gate> “,”
time → “at” <departure_time>

```

Figura 16. Patrones para generación de frases en inglés

La respuesta del sistema construida en formato de texto mediante los patrones se envía al sintetizador de voz para su reproducción de forma oral. La adición

del idioma inglés tampoco ha requerido cambios importantes respecto a la versión previa del sistema. La única diferencia existente ahora consiste en que el módulo de control debe decidir qué idioma debe utilizar el sintetizador en función del idioma seleccionado por el usuario al principio del diálogo, lo cual se traduce en ajustar un determinado parámetro del sintetizador con el valor adecuado.

3.4 Diálogos de ejemplo

En esta sección se muestran dos ejemplos de diálogos realizados en laboratorio, el primero usando el idioma castellano y el segundo el idioma inglés. En el primer ejemplo el usuario interactúa con el sistema únicamente para obtener información.

- (1) **SISTEMA:** Buenas tardes. Bienvenido a la empresa “X”. Tras escuchar la señal, diga la palabra, español, si desea hablar en español. Hello, welcome to “X”. Please say, English, after the tone if you want to speak in English.
<sonido de tono>
- (2) **USUARIO:** español
- (3) **SISTEMA:** ¿Necesita instrucciones de uso?
- (4) **USUARIO:** no
- (5) **SISTEMA:** de acuerdo, ¿Cómo puedo ayudarle?
- (6) **USUARIO:** necesito saber qué autobuses hay para ir a Madrid este jueves
- (7) **SISTEMA:** de acuerdo, destino Madrid, día jueves. ¿Desde qué ciudad quiere salir?
- (8) **USUARIO:** desde Granada
- (9) **SISTEMA:** voy a confirmar los datos de la consulta, destino Madrid, día jueves, salida desde Málaga. ¿Es todo correcto?
- (10) **USUARIO:** no
- (11) **SISTEMA:** de acuerdo. ¿Cuál es el dato incorrecto, la ciudad de destino, el día del viaje, o la ciudad de salida?
- (12) **USUARIO:** la ciudad de salida
- (13) **SISTEMA:** de acuerdo, ¿Desde qué ciudad quiere salir?
- (14) **USUARIO:** desde Granada
- (15) **SISTEMA:** de acuerdo, voy a confirmar otra vez los datos de la consulta, destino Madrid, día jueves, salida desde Granada. ¿Está todo correcto ahora?
- (16) **USUARIO:** sí
- (17) **SISTEMA:** de acuerdo, hay autobuses que salen de Granada y van a Madrid el jueves a las siguientes horas: 10, 12, 18 y 20. ¿Desea algo más?
- (18) **USUARIO:** sí, ¿Cuánto vale el viaje?
- (19) **SISTEMA:** el viaje de ida y vuelta cuesta 30 euros. ¿Algo más?
- (20) **USUARIO:** no
- (21) **SISTEMA:** de acuerdo. Muchas gracias por haber utilizado nuestro servicio automático de atención a clientes. Buenas tardes.

Figura 17. Ejemplo de interacción en castellano

En el diálogo se puede observar la efectividad de la estrategia de confirmación para corregir errores de reconocimiento del habla ocurridos durante la interacción; por ejemplo, el error de reconocimiento ocurrido en el turno 8 es detectado y corregido por el usuario gracias a la confirmación explícita (Lavelle et al. 1999) generada por el sistema en el turno 9. En el

turno 11 se observa que el sistema tiene conocimiento de los datos proporcionados por el usuario hasta el momento actual del diálogo. El sistema solicita confirmaciones explícitas en los turnos 9 y 15 a fin de asegurar que las consultas a las bases de datos se realizan usando los datos correctos. Asimismo, el sistema muestra al usuario que *recuerda* las acciones que ha realizado previamente; por ejemplo, en el turno 15 incluye los segmentos “otra vez” y “ahora” a fin de indicar que se trata de un segundo intento de confirmación de los mismos datos.

A continuación se muestra el ejemplo de interacción en inglés. En este caso, el usuario interactúa con el sistema para obtener información y realizar una reserva de plazas.

- (1) **SISTEMA:** Buenas tardes. Bienvenido a la empresa “X”. Tras escuchar la señal, diga la palabra, español, si desea hablar en español. Hello, welcome to “X”. Please say, English, after the tone if you want to speak in English.
<sonido de tono>
- (2) **USUARIO:** english
- (3) **SISTEMA:** Do you need use instructions?
- (4) **USUARIO:** yes
- (5) **SISTEMA:** This system provides information about bus departures, arrivals and prices, and allows making travel reservations. Help is available at any time by saying the word, help. If the system misunderstands please say the word, mistake, at any time. How can I help you?
- (6) **USUARIO:** help
- (7) **SISTEMA:** You can enter a query or make a travel reservation. How may I help you?
- (8) **USUARIO:** What buses are there for going to Madrid?
- (9) **SISTEMA:** ok, destination Madrid. Where are you leaving from?
- (10) **USUARIO:** Granada
- (11) **SISTEMA:** ok, from Granada. What day of the week you want to travel?
- (12) **USUARIO:** this Thursday
- (13) **SISTEMA:** ok, I need to confirm these data. From Granada to Madrid this Thursday. Is it all right?
- (14) **USUARIO:** yes
- (15) **SISTEMA:** There are buses from Granada to Madrid this Thursday at 10, 12, 18 and 20 hours. Do you need anything else?
- (16) **USUARIO:** yes, I would like to make two reservations for the bus leaving at 12
- (17) **SISTEMA:** ok, making two reservations for the bus leaving from Granada to Madrid this Thursday, May 5th, at 12. Please say your telephone number
- (18) **USUARIO:** 6 5 0 3 5 6 5 4 3
- (19) **SISTEMA:** did you say 6 5 0 3 5 6 5 4 3?
- (20) **USUARIO:** yes
- (21) **SISTEMA:** ok, you have two reservations, numbers 34 and 35, in the bus leaving from Granada to Madrid this Thursday at 12:00. The tickets must be paid at the bus station counter one hour before the bus departure. Visa and Master Card are accepted. Can I do anything else for you?
- (22) **USUARIO:** repeat please
- (23) **SISTEMA:** you have two reservations, numbers 34 and 35, in the bus leaving from Granada to Madrid this Thursday at 12. The tickets must be paid at the bus station counter one hour before the bus departure. Visa and Master Card are accepted. Anything else?
- (24) **USUARIO:** no, thanks

(25) **SISTEMA:** All right. Thank you very much for using our system. Goodbye.

Figura 18. Ejemplo de interacción en inglés

Como se puede observar en el diálogo, no se producen errores de reconocimiento del habla durante la interacción. Asimismo, se observa que el sistema tiene capacidad para inferir la fecha del viaje, no indicada por el usuario explícitamente, a fin de confirmar la reserva del viaje. El número de teléfono se solicita para asociar la reserva al usuario y para que la empresa pueda contactar con él en caso necesario. También se observa que el sistema utiliza expresiones distintas para no resultar excesivamente reiterativo, usando un repertorio de expresiones similares, en castellano e inglés, de entre las que selecciona una de forma aleatoria en cada momento, por ejemplo: “How can I help you?”, “How may I help you”, “Do you need anything else?”, “Anything else?”, “Can I do anything else for you?”, etc.

4. Conclusiones y trabajo futuro

En este artículo se ha realizado una introducción a la tecnología de los sistemas de diálogo que se ha centrado principalmente en el procesamiento multilingüe, el cual permite la interacción usuario-sistema mediante diversos idiomas. En primer lugar se ha presentado una visión panorámica de la evolución histórica de dicha tecnología, se ha descrito el funcionamiento y la arquitectura modular de un sistema de diálogo convencional, y se ha hecho referencia a dos proyectos de investigación actuales, uno relacionado con el desarrollo de un sistema de diálogo multilingüe, y el otro relacionado con el desarrollo de herramientas y recolección de material multilingüe. En segundo lugar se ha descrito brevemente un sistema de diálogo previamente desarrollado, distinguiendo la finalidad y funcionamiento de sus componentes principales. En tercer lugar se ha presentado la nueva versión de dicho sistema en la que estamos trabajando actualmente para incorporar la capacidad de procesamiento multilingüe. La sección muestra los cambios realizados en la versión previa del sistema para permitir la interacción en castellano e inglés, y finaliza mostrando dos diálogos de ejemplo realizados en ambos idiomas.

La metodología empleada para incorporar la capacidad de procesamiento multilingüe se basa, principalmente, en la obtención de la misma representación semántica de una frase dada, con independencia del idioma en que dicha frase haya sido pronunciada por el usuario. Como se puede observar en los diálogos mostrados en las Figuras 17 y 18, la metodología empleada es efectiva pues permite al usuario interactuar con el sistema en castellano y en inglés. No obstante, la

capacidad de interacción en inglés se ha desarrollado en base a un corpus de frases traducidas que no ha sido revisado por un hablante nativo de este idioma. Por este motivo, dichas frases quizás no se correspondan exactamente con la forma en que una persona nativa de habla inglesa interactúe con el sistema en una situación real. Por consiguiente, en trabajos próximos tenemos previsto usar un corpus de frases cuyos locutores sean personas nativas del idioma inglés, a fin de refinar las bigramáticas utilizadas durante el reconocimiento del habla y las reglas utilizadas durante el análisis semántico de las frases. Asimismo, tenemos previsto que una persona nativa de habla inglesa revise los mensajes en inglés generados por el sistema, a fin de refinar también los diversos patrones utilizados para generar frases en dicho idioma.

Una cuestión importante que no ha sido tratada en este trabajo es la relacionada con la evaluación del sistema desarrollado. En trabajos próximos tenemos previsto realizar evaluaciones de la interacción basada en los diversos idiomas disponibles, a fin de corregir posibles errores o deficiencias de los diversos módulos del sistema, como paso previo a su implantación en el mundo real.

Referencias

- Adda-Decker M. 2001. Towards multilingual interoperability in automatic speech recognition. *Speech Communication*, 35, pág. 5-20
- Allen J. 1995. *Natural language understanding*. The Benjamin/Cummings Publishing Company Inc.
- Andreas W., Harald L., Dafydd G. 2000. Enhancing Speech Corpus Resources with Multiple Lexical Tag Layers. *Second International Conference on Language Resources and Evaluation*, pág. 403-407
- Baptist I., Seneff S. 2000. GENESIS-II: A versatile system for language generation in conversational system applications. *Proc. de ICSLP*, pág. 271-274
- Billi R., Castagneri G., Danielli M. 1997. Field trial evaluations of two different information inquiry systems". *Speech Communication*, 23, 1-2, pág. 83-93
- Boros M., Heisterkamp P. 1999. Linguistic Phrase Spotting in a Single Application Spoken Dialogue System, *Proc. de Eurospeech '99*, 1983-1986
- Darrell T., Fisher J., Wilson K. 2002. Geometric and statistical approaches to audiovisual segmentation for untethered interaction. *Proc. de Workshop on Multimodal Dialogue and Document Processing*, Taipei, Taiwan
- Fisher J. W., Darrell T., Freeman W. T., Viola P. 2000. Learning statistical models for audio-visual fusion and segregation. *Advances in Neural Information Processing Systems*, 13
- Gibbon D., Mertins I., Moore R. K. 2000. *Handbook of multimodal and spoken dialogue systems*, Kluwer Academic Publishers
- Glass J., Polifroni J., Seneff S. 1994. Multilingual language generation across multiple domains. *Proc. de ICSLP '94*, pág. 983-986
- Glass J., Flammia G., Goodine D., Phillips M., Polifroni J., Sakai S., Seneff S., Zue V. 1995. Multilingual spoken-language understanding in the MIT Voyager system. *Speech Communication*, 17 (1-2), pág. 1-18
- Glass J., Hazen T., Hetherington I. L. 1999. Real-time telephone-based speech recognition in the Jupiter domain. *Proc. de ICASSP '99*, pág. 61-64
- Graff D., Bird S. 2000. Many Uses, Many Annotations for large speech corpora: Switchboard and TDT as case studies. *Second International Conference on Language Resources and Evaluation*, pág. 427-433
- House D., Beskow J., Granström B. 2001. Timing and interactions of visual cues for prominence in audiovisual speech perception. *Proc. de Eurospeech '01*, pág. 387-390
- Huang X., Acero A., Hon H. 2001. *Spoken Language Processing. A Guide to Theory, Algorithm and System Development*, Prentice Hall
- Hutchins W., Somers H. 1992. *An Introduction to Machine Translation*. Academic Press
- Jelinek F., 1999. *Statistical Methods for Speech Recognition*, MA, MIT Press
- Hain T., Woodland P.C., Niesler T.R., Whittaker E.W.D.; 1999; The 1998 HTK System for Transcription of Conversational Telephone Speech; *Proc. of International Conference on Acoustics, Speech and Signal Processing*

- Lavelle C. A., de Calmés M., Pérennou G. 1999. Confirmation strategies to improve correction rates in a telephonic inquiry dialogue system. Proc. de Eurospeech '99, pág. 1399-1402
- López-Cózar R., Rubio A. J., García P., Segura J. C. 1999. Uso de valores de confianza y expectativas en el sistema de diálogo SAPLEN. Procesamiento del lenguaje natural, nº 24, pág. 37-41
- López-Cózar R., Rubio A. J., García P., Díaz-Verdejo J. E., López-Soler J. M. 2000. Sistema telefónico de atención a viajeros. I Jornadas en Tecnologías del Habla, CD-ROM, ISBN 84-95118-58-0, Dep. Legal SE-1396-02
- López-Cózar R., De la Torre A., Segura J. C., Rubio A. J., V. Sánchez. 2002a. Testing dialogue systems by automatically generating conversations. Interacting with Computers, 14, pág. 521-546
- López-Cózar R., Rubio A. J., Díaz-Verdejo J. E., López-Soler J. M. 2002b. Validación de un sistema de diálogo mediante el uso de diferentes umbrales de poda en el proceso de reconocimiento automático de voz. Procesamiento del Lenguaje Natural, nº 29, pág. 205-211
- Ma B., Guan C., Li H., Lee C-H. 2002. Multilingual speech recognition with language identification. Proc. de ICSLP, pág. 505-508
- Nakano N., Minami Y., Seneff S., Hazen T. J., Cyphers D. S., Glass J., Polifroni J., Zue V. 2001. Mokusei: A telephone-based Japanese conversational system in the weather domain. Proc. de Eurospeech '01, pág. 1331-1334
- Nasr A., Esteve Y., Béchet F., Spriet T., de Mori R. 1999. A language model combining N-grams and stochastic finite state automata, Proc. de Eurospeech '99, pág. 2175-2178
- Padrell J., Hernando J. 2002. ACIMET: Access to meteorological information by telephone. Proc. de ICSLP, pág. 2713-2716
- Rabiner L. R., Juang B. H. 1986. An Introduction to Hidden Markov Models. IEEE ASSP Magazine, Enero
- Rabiner L. R., Juang B. H. 1993. Fundamentals of Speech Recognition. Prentice-Hall
- Relaño Gil J., Tapias D., Villar J. M., Gancedo M. C., Hernández L. A. 1999. Flexible mixed-initiative dialogue for telephone services. Proc. de Eurospeech '99, pág. 1179-1182
- Roe D., Pereira F., Sproat R., Riley M., Moreno P., Macarron A. 1991. Toward a spoken language translator for restricted-domain context-free languages. Proc. de Eurospeech '91, pág. 1063-1066
- Rosset S., Bennacef S., Lamel L. 1999. Design strategies for spoken language dialog systems. Proc. de Eurospeech '99, pág. 1535-1538
- Schadle I., Antoine J. Y., Memmi D. 1999. Connectionist language models for speech understanding: The problem of word order variation, Proc. de Eurospeech '99, pág. 2035-2038
- Seneff S. 1992. TINA: A natural language system for spoken language applications. Computational linguistics, 18(1), pág. 61-86
- Seneff S., Polifroni J. 2000. Dialogue management in the Mercury flight reservation system. Proc. de ANLP-NAACL, Seattle, Mayo
- Takezawa T., Yato F., Sagayama S., Tashiro T., Nagata M., Kumeratsu A. 1993. ATR's speech translation system: ASURA. Proc. de Eurospeech '93, pág. 1291-1294
- Tian J., Häkkinen J., Viikki O. 2002. Multilingual pronunciation modelling for improving multilingual speech recognition. Proc. de ICSLP '02, pág. 497-500
- Wahlster W. 1993. Verbmobil, translation of face-to-face dialogs. Proc. de Eurospeech '93, pág. 29-38
- Wahlster W. 1994. Overview of user modelling for Verbmobil. Keynote presentation to User Modeling biannual workshop
- Wahlster W., Reithinger W., Bocke A. 2001. SmartKom: multimodal communication with a lifelike character. Proc. de Eurospeech '01, pág. 1524-1546
- Waibel A., Jain A., McNair A., Saito H., Hauptmann A., Tebelkis J. 1991. JANUS: A speech-to-speech translation system using connectionist and symbolic processing strategies. Proc. de ICASSP '91, pág. 793-796
- Wang C., Cypherws S., Mou X., Polifroni J. Seneff S., Yi J., Zue V. 2000. MuXing: A telephone-access Mandarin conversational system in the weather domain. Proc. de ICSLP, pág. 715-718

Woodland P.C., Hain T., Moore G.L., Niesler T.R.,
Provey D., Tuerk A., Whittaker E.W.D.; 1999;
The 1998 HTK Broadcast News Transcription
System: Development and Results; DARPA
Broadcast News Workshop

Zue V., Seneff S., Glass J., Hetherington L., Hurley
E., Meng H. Pao C., Polifroni J., Schloming R.
Schmid P. 1997. From interface to content:
translingual access and delivery of on-line
information. Proc. de Eurospeech '97, pág. 2227-
2230

Zue V., Seneff S., Glass J., Polifroni J., Pao C.,
Hazen T., Hetherington L. 2000. Jupiter: A
telephone-based conversational interface for
weather information. IEEE Trans. on Speech and
Audio Proc., 8(1), pág. 85-96