

1

Introducción a los sistemas de diálogo

Joaquim Llisterri

Departament de Filologia Espanyola
Universitat Autònoma de Barcelona

- 1.1 Los sistemas de diálogo
- 1.2 Tareas y módulos de un sistema de diálogo
- 1.3 Los conocimientos de un sistema de diálogo
- 1.4 Las aplicaciones de los sistemas de diálogo
- 1.5 Los sistemas de diálogo como campo de trabajo interdisciplinar

1.1 LOS SISTEMAS DE DIÁLOGO

Los sistemas de diálogo o sistemas conversacionales (SLS, *Spoken Language Systems*) son una tecnología concebida para facilitar la interacción natural mediante el habla entre una persona y un ordenador (Giachin 1997; Giachin, McGlashan 1997; Minker, Bennacef 2001; Tapias 2002). Constituyen como una interfaz o un intermediario entre un usuario y un sistema informático, que tiene la ventaja de no requerir el uso de una pantalla, un teclado o un ratón y de recurrir, en cambio, al medio de comunicación propio de los seres humanos. Por ello, suelen emplearse para servicios que se ofrecen a través del teléfono, aunque también se usan en todos aquellos casos en los que la lengua oral permite llevar a cabo una determinada tarea de forma más flexible y más eficaz que la escrita.

Los trabajos de quienes desarrollan sistemas de diálogo se enmarcan en las tecnologías del habla, ámbito interdisciplinar que tiene como finalidad proporcionar los conceptos, las técnicas y las herramientas que permiten a un ordenador reconocer e interpretar el habla humana, producirla con un elevado grado de naturalidad, identificar o verificar la identidad de un locutor o, incluso, traducir una conversación entre dos personas que emplean distintas lenguas (Boite *et al.* 2000; Escudier, Schwartz 2000; Holmes, Holmes 2001; Huang *et al.* 2001; Jurafsky, Martin 2000; Mariani 2002; O'Shaughnessy 1987). Sin embargo, como veremos más adelante, también es necesario incorporar conocimientos de otra disciplina, el procesamiento del lenguaje natural (PLN o NLP, *Natural Language Processing*), dedicada al tratamiento computacional de la lengua escrita; entre sus aplicaciones se encuentran, por ejemplo, la traducción automática de textos, la recuperación y extracción de información o las herramientas de ayuda a la escritura como los correctores ortográficos, gramaticales y de estilo. El procesamiento del lenguaje natural y las tecnologías del habla constituyen las llamadas tecnologías lingüísticas (TL o LT, *Language Technologies*) (Llisterri 2003; Llisterri, Martí 2002; Martí 2001).

Las capacidades ideales de un sistema de diálogo podemos encontrarlas ya en *Hal*, el ordenador imaginado por Arthur C. Clarke en 1950 y llevado al cine por Stanley

Kubrick en 1968¹ (Stork 1997): reconoce sin dificultades el habla espontánea, comprende todo tipo de enunciados sin restricciones de contenido, proporciona respuestas con sentido, gramaticalmente bien formadas y pragmáticamente adecuadas, tiene una voz completamente natural y permite, además, una interacción multimodal a través del habla y de la imagen. Aunque han pasado unas cuantas décadas y la investigación ha avanzado notablemente, los sistemas actuales todavía no se acercan a las múltiples posibilidades comunicativas de *Hal*, por los motivos que resumimos a continuación y que se detallan en el resto de capítulos de este libro.

En primer lugar, los sistemas de diálogo que podemos utilizar hoy en día están sujetos a las limitaciones del reconocimiento del habla (véase el capítulo 4). En este sentido, no disponemos aún de reconocedores que puedan procesar con un cien por cien de fiabilidad la variación propia del habla completamente espontánea, la multiplicidad de realizaciones fonéticas de locutores de diversa procedencia geográfica y social y las interferencias producidas por el entorno cuando el sistema tiene que utilizarse, por ejemplo, dentro de un vehículo en marcha o hablando por un teléfono móvil en plena calle.

Aun suponiendo que pudiéramos contar con los mejores reconocedores, para alcanzar las posibilidades de *Hal*, un sistema de diálogo debería comprender cualquier enunciado, con independencia del tema o de su complejidad sintáctica, semántica y pragmática. La comprensión del lenguaje, es decir, la creación automática de una representación del contenido de un enunciado a partir de los resultados del reconocedor de habla (véase el capítulo 5), está limitada, por el momento, a lo que se denominan dominios restringidos, ámbitos muy concretos a los que se aplica un determinado sistema de diálogo, como pueda ser la información de horarios de trenes, la reserva de billetes de avión o la información ciudadana.

En tercer lugar, las respuestas de *Hal* pueden confundirse con las de un humano, puesto que no sólo son gramaticalmente adecuadas, sino que abarcan cualquier tema y presentan las mismas características en cuanto a su riqueza y su flexibilidad que las del lenguaje natural. Para llegar a estos resultados debería disponerse de un sofisticado sistema de generación del lenguaje (véase el capítulo 7) capaz de emular los mecanismos mediante los que las personas conseguimos producir enunciados que respondan acertadamente a cualquier situación, algo que todavía no está al alcance de la tecnología actual.

Por otra parte, las conversaciones que *Hal* mantiene con sus interlocutores se caracterizan por su coherencia y su naturalidad, así como por el hecho de que, contrariamente a algunos de los sistemas actuales de diálogo, no es preciso que el interlocutor humano guíe al sistema confirmando que la interpretación de sus preguntas es la correcta. Esto implica un sistema de gestión del diálogo (véase el capítulo 6) también algo alejado de las posibilidades de hoy en día; pese a que se ha superado ya la etapa en la que las intervenciones del usuario de un sistema de diálogo debían ceñirse a unas pocas palabras y era casi imprescindible introducir a menudo preguntas como «¿Ha solicitado usted información sobre horarios de trenes entre Barcelona y Madrid?», parece claro que la interacción que en estos momentos puede realizarse con un ordenador no es exactamente igual que la que se refleja en los diálogos de *2001: A space*

1. Pueden escucharse algunas muestras de la voz de *Hal* en <http://www.palantir.net/2001/sounds.html>

odyssey o la que mantienen habitualmente dos personas en una conversación (véase el capítulo 2)

Finalmente, los métodos de conversión de texto en habla (véase el capítulo 8), pese a haber experimentado mejoras importantísimas en los últimos años, aún no permiten crear una voz sintetizada que pueda confundirse con la de una persona ni son capaces, por el momento, de ofrecer una realización prosódica de los enunciados idéntica a la del habla natural. Este es uno de los principales motivos por los que un usuario se percató inmediatamente de que, al llamar a un determinado servicio, quien le responde no es un interlocutor humano, sino un sistema de diálogo.

Es justo reconocer, sin embargo, que pese a no haber alcanzado el nivel imaginado por Clarke y Kubrick, los sistemas de diálogo se emplean en cada vez más aplicaciones reales, tanto en el entorno de la administración pública como en el de la empresa privada. Su gran ventaja es que permiten ofrecer un servicio telefónico continuado, durante 24 horas al día a lo largo de 365 días al año y que, al mismo tiempo, mejoran las prestaciones que proporcionan al cliente al limitar el trabajo de los operadores humanos a la solución de problemas más complejos de los que puede resolver con eficacia un sistema automático. En el apartado 1.4 reseñaremos brevemente algunas de las aplicaciones actuales de los sistemas de diálogo, complementando las que pueden encontrarse en los restantes capítulos del libro.

1.2 TAREAS Y MÓDULOS DE UN SISTEMA DE DIÁLOGO

Para llevar a cabo las funciones descritas en el apartado anterior, un sistema de diálogo debe constar de una serie de módulos, cada uno especializado en un aspecto concreto de la interacción entre la persona y la máquina. El objetivo fundamental de este libro es ofrecer una detallada información del funcionamiento de cada uno de ellos y, por lo tanto, en este capítulo introductorio, nos limitaremos a una sucinta revisión de las funciones y tareas de cada módulo que se integra en un sistema de diálogo, indicando al lector en qué capítulos puede profundizar en el estudio de un aspecto en particular.

Tomaremos como ejemplo un hipotético sistema de diálogo que facilitara automáticamente información sobre horarios de trenes.² Supongamos que se trata de una aplicación que, tras una bienvenida y unas breves instrucciones, realiza la pregunta «¿Qué horarios desea saber?». El usuario puede responder algo como «Eh... pues... quería saber a qué hora sale la... el último tren... eh... para Barcelona desde Madrid ... el sábado», enunciado que, en realidad, constituye una onda sonora en la que, contrariamente a lo que sucede en la lengua escrita, ni los sonidos ni las palabras aparecen claramente separados. En este sentido, el sistema de diálogo tendrá como primera tarea analizar estas variaciones en el tiempo de la amplitud y la frecuencia para intentar extraer de ellas el contenido de la información solicitada.

Este proceso inicial se realizará en el módulo de reconocimiento del habla (véase el capítulo 4), cuya tarea es convertir la onda sonora en una representación escrita que

2. Un sistema de este tipo se ha desarrollado, por ejemplo, en el marco del proyecto Basurde (Sistema de diálogo oral en dominios restringidos), en el que participaron varios de los autores de este libro (Bonafonte *et al.* 2000).

puedan utilizar los demás módulos del sistema. Para ello se necesita haber desarrollado previamente un conjunto de modelos fonéticos que recojan la variabilidad de cada segmento mínimo del habla —es decir, de cada alófono— y un modelo de lenguaje que establezca la probabilidad de combinación de una palabra con otra. Para obtener estos modelos se recurre al uso de corpus hablados, representativos tanto de los hablantes como del dominio de aplicación del sistema de diálogo (véase el capítulo 3). La tecnología utilizada en este módulo es, pues, el reconocimiento automático del habla (RAH o ASR, *Automatic Speech Recognition*), empleada también en otro tipo de aplicaciones como el dictado automático de textos, y considerada una parte esencial de las tecnologías del habla a las que aludíamos al principio.

Sin embargo, una representación escrita de lo dicho por el usuario no es suficiente para responder a su pregunta, pues además del reconocimiento es necesario que el sistema de diálogo alcance un cierto grado de comprensión. Debe señalarse aquí que cuando en el contexto de las tecnologías lingüísticas se habla de comprensión, se hace referencia a la generación automática de una representación semántica abstracta de un enunciado; así, en nuestro ejemplo, el sistema debería extraer la información de que Madrid es la ciudad de origen, Barcelona la de destino, la fecha es la que corresponda al primer sábado que aparezca en el calendario, la hora la correspondiente al último tren y de que se está realizando una petición de información. Esta tarea la realiza el módulo de comprensión o de interpretación semántica (véase el capítulo 5), recurriendo a las técnicas propias del campo conocido como comprensión del lenguaje natural (NLU, *Natural Language Understanding*), que forma parte del procesamiento del lenguaje, o de la comprensión del habla (SLU, *Spoken Language Understanding*), uno de los ámbitos de las tecnologías del habla.

Una vez que el sistema de diálogo ha averiguado qué tipo de información desea recibir el usuario, entran en juego los mecanismos que le permiten acceder a las bases de datos o bases de conocimiento que previamente se hayan incorporado. Con la información que de allí se extraiga, es necesario construir una respuesta, que deberá corresponder a una frase gramaticalmente bien formada del tipo «El último tren de Madrid a Barcelona el sábado 28 de junio sale a las 11 horas y 53 minutos». Esto supone una nueva tecnología: la generación del lenguaje (NLG, *Natural Language Generation*), que también forma parte del procesamiento del lenguaje natural y que constituye el núcleo del módulo de generación de respuestas (véase el capítulo 7).

En una última etapa, puesto que la consulta de información se realiza por teléfono, la frase creada por el módulo de generación de respuesta tendrá que convertirse en su equivalente sonoro para que el usuario del sistema de diálogo pueda escucharla. De esta tarea se ocupa la conversión de texto en habla (CTH o TTS, *Text-to-Speech Synthesis*), una rama de las tecnologías del habla que tiene como objetivo transformar textos escritos en su realización oral o, dicho de un modo más simple, efectuar la lectura en voz alta de un texto escrito. El módulo de conversión de texto en habla (véase el capítulo 8) es, por tanto, el último paso de un sistema de diálogo y el que se encarga, finalmente, de hacer llegar la información a la persona que espera una respuesta.

Todas estas tareas están, en cierto modo, coordinadas o controladas por un módulo específico, denominado con frecuencia módulo de gestión del diálogo (véase el capítulo 6), entre cuyas atribuciones se cuenta el seguimiento de la interacción entre la persona y el sistema informático en función de lo que el usuario desee realizar y de

la capacidad del sistema. Podemos imaginar que, una vez recibida la información sobre la hora de salida del último tren de Madrid a Barcelona el sábado, el usuario preguntará «¿Y el domingo?»; en este caso, el módulo de gestión del diálogo debería recordar la intervención anterior y deducir que los datos solicitados son los mismos, excepto un cambio en la fecha. Ante una pregunta como «¿No hay uno más temprano?», este módulo tendría que interpretar que el referente de «uno» es «tren de Madrid a Barcelona el sábado», y si el usuario respondiera «¿Y desde Alcalá?», la historia del diálogo permitiría saber que la ciudad de origen es Alcalá pero que la de destino sigue siendo Barcelona. En caso de dudas, si surgiera un problema en el módulo de reconocimiento del habla o en el de interpretación semántica, el gestor del diálogo sería el responsable de que el sistema solicitara una confirmación al usuario del tipo «¿Quiere usted saber los horarios de trenes de Alcalá a Barcelona?»

Un sistema de diálogo consiste, por lo tanto, con una estructura modular en la que cada módulo se ocupa de unas determinadas tareas en interacción con todos los demás. A grandes rasgos, el proceso descrito hasta ahora podría esquematizarse como se muestra en la figura 1.1.

Minker y Bennacef (2001: 155) proponen la arquitectura que se representa en la figura 1.2 y que consta, tal como hemos visto, de un módulo de reconocimiento de habla que se ocupa de transformar la señal sonora en una representación que pueda procesar el módulo de análisis semántico; una vez realizado este análisis, interviene el módulo de gestión del diálogo, conectado a una interfaz que permite la consulta a una base de datos y que envía la información recogida al módulo de generación de respuestas; este último, con la intervención de un conversor de texto en habla, facilitaría al usuario la información deseada. Como elemento central, estos autores incluyen el contexto del diálogo, que contiene información sobre el enunciado que se está procesando, la historia del diálogo y el estado actual de la interacción.

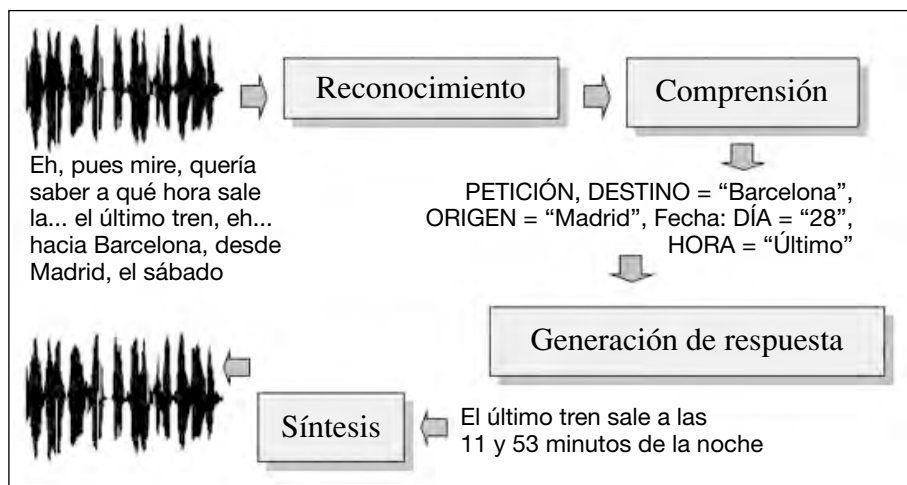


Figura 1.1

Representación esquemática del proceso de consulta a un sistema de diálogo.

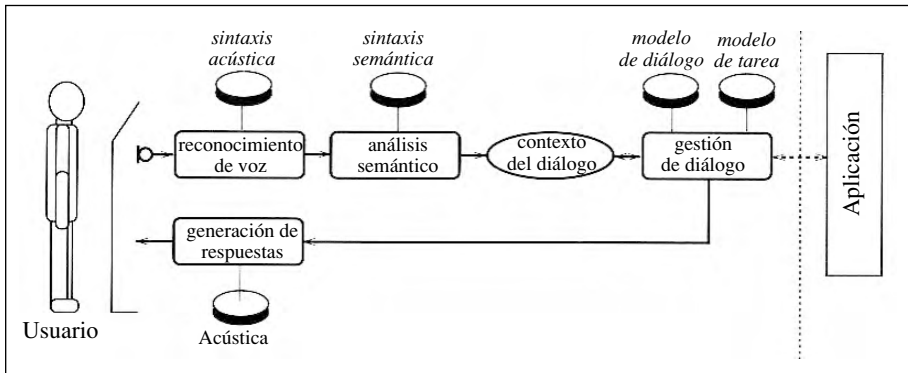


Figura 1.2

Arquitectura de un sistema de diálogo (Minker, Bennacef 2001: 155).

1.3 LOS CONOCIMIENTOS DE UN SISTEMA DE DIÁLOGO

Para llevar a cabo las tareas anteriormente descritas, es necesario contar con las tecnologías a las que nos hemos referido —reconocimiento automático del habla, comprensión del habla y del lenguaje natural, generación del lenguaje natural, conversión de texto en habla, gestión del diálogo—, pero también es imprescindible que el sistema de diálogo disponga de un conjunto de conocimientos sobre la lengua, sobre las tareas que deben realizarse, sobre el desarrollo de la interacción y sobre los potenciales usuarios. Siguiendo a Minker y Bennacef (2001: 9-10), resumiremos a continuación estos conocimientos. Antes, sin embargo, debe señalarse que su incorporación a un sistema de diálogo se lleva a cabo a través de modelos. La idea de un modelo responde al hecho de que estamos intentando reducir a una escala aceptable para un sistema informático la complejidad de la comunicación entre personas. Puesto que, por el momento, no es posible incorporar todos los conocimientos que un hablante activa al mantener una conversación con otro (véase el capítulo 2), debemos limitarnos a un modelo, una reproducción controlada de la realidad, que incorpore los elementos que estamos en condiciones de formalizar y de definir de modo que puedan ser objeto de un tratamiento computacional.

El primer tipo de conocimiento que se incorpora a un sistema de diálogo es, obviamente, el lingüístico. Para tal fin se diseñan modelos de lenguaje, empleados en el reconocimiento y en el análisis semántico. Los modelos de lenguaje usados para el reconocimiento del habla (véase el capítulo 4) consisten en información sobre la probabilidad de aparición conjunta de palabras, y se obtienen de corpus de diálogo como los descritos en el capítulo 2. En cambio, los modelos de lenguaje que se utilizan para el análisis semántico en el módulo de comprensión incorporan información léxica, sintáctica y semántica sobre la lengua en forma de gramáticas y lexicones computacionales (véase el capítulo 5).

El modelo de la tarea (véase el capítulo 6) recoge todos los conocimientos que se relacionan con la aplicación para la que se haya diseñado el sistema de diálogo; en este

sentido, es específico de cada caso en concreto y se define también, por lo general, a partir de corpus en los que o bien dos personas interactúan en una tarea real –por ejemplo un empleado de una agencia de viajes y un cliente que desea hacer una reserva–, o bien una persona mantiene un diálogo con un sistema simulado mediante el procedimiento conocido como el Mago de Oz (véanse los capítulos 3 y 6). Este conocimiento de la tarea es importante para que el sistema de diálogo pueda interpretar acertadamente un enunciado, puesto que no sólo entra en juego la información lingüística, sino también la correspondiente a los resultados que desea obtener el usuario. A grandes rasgos, suelen distinguirse las tareas relacionadas con la información –como la que ejemplificábamos en el apartado anterior– de las que suponen una transacción –la reserva de un billete, por ejemplo– y de las que implican una negociación –encontrar el billete más barato para un determinado trayecto.

En tercer lugar, parece claro que el sistema debe conocer también cómo se desarrolla la interacción en el contexto de una determinada aplicación: qué peticiones ha realizado un usuario, qué información puede solicitar a partir de la que ya se le ha proporcionado, qué posibilidades existen de que, en un momento dado, el diálogo se desvíe hacia otros temas –por ejemplo, de horarios a precios–, etc. Estas cuestiones son las que se intenta incorporar al modelo del diálogo (véase el capítulo 6), que se complementa en cada caso con la historia del diálogo, es decir, con los datos recopilados desde que un usuario ha iniciado su llamada. Así, por ejemplo, se puede discriminar si un enunciado como «Quiero saber los horarios de trenes de Madrid a Barcelona» es una petición nueva o la confirmación de que, efectivamente, ésta es la información deseada.

Finalmente, el modelo del usuario recoge todos aquellos conocimientos sobre el usuario que puedan ser relevantes para el funcionamiento del sistema de diálogo, tanto los que se refieren a sus rasgos lingüísticos como los que conciernen a su modo de interactuar con el sistema informático, determinado éste último por factores como la edad, el estado físico, el nivel socio-cultural o la familiaridad con el procedimiento. Este modelo puede variar a lo largo del diálogo, de modo que el sistema se adapte en la medida de lo posible al estilo del interlocutor y, en contextos como la enseñanza asistida por ordenador, a su nivel de conocimientos.

En los sistemas multimodales (véase el capítulo 6), que incorporan la posibilidad de usar otros canales de comunicación además del habla, es importante disponer también de un modelo del entorno o modelo del sistema, de modo que en cada momento se cuente con suficientes elementos para decidir el medio por el que se lleva a cabo la interacción o para integrar los datos procedentes de diferentes canales. En un sistema que, por ejemplo, combine el habla con una pantalla táctil, el usuario empleará uno u otro medio en función de la comodidad y la eficacia que le proporcione cada uno, y esperará también la respuesta por el canal que le ofrezca una respuesta más clara.

En la figura 1.2 puede observarse cómo cada uno de los módulos de un sistema de diálogo está asociado a uno o más modelos que reflejan los conocimientos necesarios para garantizar la interacción en el usuario y una aplicación informática. Así, los módulos de reconocimiento y de análisis emplean la información procedente de los modelos de lenguaje, y el de gestión del diálogo recurre al modelo del diálogo y al modelo de la tarea.

1.4 LAS APLICACIONES DE LOS SISTEMAS DE DIÁLOGO

Esencialmente, los sistemas de diálogo actuales se utilizan para obtener y gestionar información o para llevar a cabo transacciones, en ambos casos a través del teléfono. Aun así, como señalábamos al introducir al principio la noción de interfaz, también pueden catalogarse como sistemas de diálogo las aplicaciones que permiten interactuar mediante el habla con un ordenador al alcance del usuario, por ejemplo para manipular objetos en un sistema de diseño asistido por ordenador (Pineda *et al.* 2002), para gestionar y relacionar información compleja y disponer de una ayuda en la toma de decisiones (Allen *et al.* 2001) o para emular algunas de las funciones de un tutor humano en programas de enseñanza asistida por ordenador (Seneff, Wang, Zhan 2004). Otra aplicación sumamente relevante en la actualidad es la traducción automática del habla, de modo que, combinando un sistema de diálogo con las tecnologías propias de la traducción automática, se pueda llegar a conseguir que dos personas mantengan una conversación expresándose cada uno en su lengua (Wahlster 2000; Waibel 2000).

Los sistemas conversacionales que facilitan información se utilizan en diversos ámbitos: pueden proporcionar horarios y precios de transportes, información meteorológica, información ciudadana sobre horarios de servicios, sobre trámites en la administración o sobre la oferta cultural y de ocio, facilitar la consulta de notas, de resultados de concursos o de convocatorias, proporcionar el saldo de una cuenta bancaria o el gasto ocasionado hasta la fecha por un servicio telefónico, etc. Por lo general, se trata de aplicaciones en las que las informaciones que interesan al usuario se encuentran en una base de datos a la que accede el sistema de diálogo.

Un ejemplo de sistema de información meteorológica es aTTemp³, desarrollado conjuntamente por la Universidad Politécnica de Cataluña y por la Universidad Autónoma de Barcelona y promovido por el Departamento de Medio Ambiente y el Departamento de Universidades, Investigación y Sociedad de la Generalitat de Cataluña (Padrell, Hernando 2002). Actualmente se encuentra en funcionamiento y facilita, mediante una llamada telefónica, información sobre temperatura, precipitaciones, dirección y velocidad del viento y altura de las olas en cualquiera de los municipios de Cataluña; también permite al usuario programar avisos, en forma de llamada a un teléfono fijo o de mensaje a un móvil, cuando se den unas determinadas condiciones meteorológicas en una localidad seleccionada.

Algunos de estos servicios que facilitan información se han incorporado a los portales de voz, que cumplen la función de un clásico portal en Internet, substituyendo la consulta a través de un ordenador por el acceso telefónico. Un portal de voz puede incluir servicios personales como una libreta de direcciones, una agenda o un buzón de voz, servicios de información –noticias, cotizaciones bursátiles, información meteorológica, cartelera de espectáculos, guía de la ciudad, programación de televisión–, servicios de ocio y entretenimiento como horóscopos, juegos, participación interactiva en programas televisivos, servicios de asistencia al viajero –información de tráfico, turística o guías para llegar a un determinado lugar– y también los servicios de comercio electrónico como los que se describen más adelante. Diversas empresas españolas como Telefónica I+D (Baena, Serna 2002; Fernández, del Río, Serrano 2000), Ydilo,

3. Puede encontrarse más información sobre aTTemp en <http://gps-tsc.upc.es/veu/attempts/>

InfoSpeech o Porfinya⁴ se han especializado en el desarrollo de portales de voz y ofrecen su experiencia a empresas de muy diversos sectores.

Encontramos también sistemas de diálogo en los que, esencialmente, se gestiona información, como puede ser el caso de los servicios de atención al cliente que le orientan en la solución de un problema, de las centralitas automáticas que dirigen al usuario de la persona adecuada. Un sistema conversacional puede también ofrecer la posibilidad de gestionar datos pertenecientes al propio usuario, como sucede en los sistemas que permiten la consulta del correo electrónico a través del teléfono.

Mientras que esta última posibilidad la ofrecen ya operadoras como Telefónica (Villarrubia *et al.* 2003), otras empresas se han centrado en la automatización de los servicios de atención al cliente mediante sistemas de diálogo. Por ejemplo, el Grupo Voice Consulting⁵ dispone de sistemas basados en las tecnologías del habla para optimizar el rendimiento de centralitas telefónicas y de centros de atención al cliente.

Los sistemas de diálogo que permiten realizar transacciones son los que se emplean en todos los contextos relacionados con el comercio electrónico, como la venta de entradas, o de billetes de tren o de avión, y en los entornos relacionados con la banca electrónica. Un ejemplo de sistema conversacional a través del cual se pueden llevar a cabo transacciones lo encontramos, por ejemplo, en Cinentradas, un servicio desarrollado por Ydilo⁶ para la venta telefónica de entradas de cine que emplea un sistema de diálogo para atender a un mayor número de llamadas y reducir el tiempo de espera de los usuarios. En el campo de la banca electrónica podría citarse, entre otros, el servicio Línea BBVA ofrecido por el Banco Bilbao Vizcaya Argentaria y creado por la empresa Natural Vox,⁷ que permite solicitar transferencias o domiciliaciones, vender o comprar valores, contratar seguros o realizar aportaciones a planes de pensiones o a fondos de inversión, una vez, naturalmente, comprobada la identidad del cliente mediante una clave personal de acceso.

Con esta breve reseña no hemos agotado todas las posibilidades de aplicación que ofrecen los sistemas de diálogo. Nos hemos referido especialmente a sistemas a los que se accede mediante el teléfono, pero bien probablemente en el futuro asistiremos a un auge de los sistemas multimodales (véase el capítulo 6) integrados con la web. Los agentes conversacionales a los que se incorpore el reconocimiento y la conversión de texto en habla formarán parte de los sistemas de diálogo, convirtiéndose en asistentes virtuales interactivos que ayudarán en las tareas que necesite realizar un usuario. Una muestra de esta tecnología, aunque limitada a la lengua escrita, puede encontrarse, por ejemplo, en las páginas de la empresa de seguros DKV, en las que, para atender a los visitantes, se recurre a un asistente virtual desarrollado por Artificial Solutions con la tecnología proporcionada por Kiwilogic.⁸

4. Ydilo: <http://www.ydilo.com/>; InfoSpeech: <http://www.infospeech.com/>; Porfinya: <http://www.porfinya.com/>

5. Grupo Voice Consulting: <http://www.grupovoice.com/>

6. Esta aplicación y otras desarrolladas por Ydilo se describen en los documentos disponibles en <http://www.ydilo.com/esp/caseStudies.html> El servicio de Cinentradas se describe también en http://www.cervantes.es/seg_nivel/lect_ens/oesi/Estudios%20de%20caso/Cinentradas.PDF

7. El servicio se presenta en http://www.cervantes.es/seg_nivel/lect_ens/oesi/Estudios%20de%20caso/bbva.PDF. Sobre el sistema BpT (Banca por Teléfono) desarrollado por Natural Vox, así como sobre otros servicios vocales de esta empresa, puede encontrarse información en <http://www.naturalvox.com/>

8. DKV: <http://www.dkvseguros.com/>; Artificial Solutions: <http://www.artificial-solutions.com/>; Kiwilogic: <http://www.kiwilogic.de/>

1.5 LOS SISTEMAS DE DIÁLOGO COMO CAMPO DE TRABAJO INTERDISCIPLINAR

Pese a que tradicionalmente el desarrollo de los sistemas de diálogo ha estado vinculado al mundo de la ingeniería y de la informática, la necesidad de incorporar modelos del conocimiento lingüístico cada vez más sofisticados ha puesto de relieve el papel que pueden jugar los expertos en disciplinas lingüísticas si se quiere ir más allá de una aplicación restringida a unas pocas alternativas bien controladas de pregunta y respuesta (Llisterri 2003a, b; Llisterri *et al.* 2003).

Hemos visto, en primer lugar, que el módulo de reconocimiento debe enfrentarse a enunciados espontáneos, y es sabido que las características fonéticas del habla espontánea no siempre responden al ideal descrito en los libros de texto (véase el capítulo 4); por otra parte, el diseño de los modelos acústicos empleados en el reconocimiento requiere, como hemos indicado, la constitución y el tratamiento de corpus de lengua oral recogidos de modo que reflejen la variabilidad fonética a la que se enfrentará un sistema de diálogo en funcionamiento real (véase el capítulo 3). El módulo de conversión de texto en habla no podría realizar su tarea si no dispusiera de la suficiente información lingüística en cada uno de sus componentes (véase el capítulo 8) y, para ello, es esencial determinar, por ejemplo, la relación entre las grafías y sus correspondientes realizaciones sonoras y entre un enunciado escrito y su manifestación prosódica. En todos los casos que acabamos de mencionar interviene el conocimiento propio de un experto en fonética, cuya labor se centra en la descripción y el análisis de la materia sonora del habla, tanto en el aspecto segmental —es decir, los sonidos— como en el suprasegmental —los elementos prosódicos como el acento, la entonación y el ritmo.

Por su parte, el módulo de comprensión (véase el capítulo 5) necesita también acceder a modelos de lenguaje en los que se recoge información léxica, sintáctica y, especialmente, semántica. Este mismo tipo de conocimiento interviene también en el módulo de generación de respuestas (véase el capítulo 7), ya que ambos utilizan técnicas propias del procesamiento del lenguaje natural. Uno de los problemas que habitualmente se plantea en este campo es la falta de corpus anotados o etiquetados en el nivel morfológico, sintáctico y semántico que permitan modelar adecuadamente el conocimiento lingüístico que requiere el PLN (véanse los capítulos 5 y 6). Parece razonable pensar que el establecimiento de los criterios de anotación y la revisión de los resultados de las herramientas de etiquetado automático son tareas que únicamente pueden realizarse con una cierta precisión si se cuenta con la formación gramatical adecuada.

En tercer lugar, el módulo de gestión del diálogo y los modelos del diálogo y de la tarea (véase el capítulo 6) intentan, en cierto modo, reproducir el comportamiento del interlocutor humano que participaría en la interacción con un usuario, de modo que el intercambio sea lo más fluido y eficaz posible. La disciplina lingüística conocida como pragmática y, en especial, el análisis de la conversación, han realizado importantes contribuciones a la comprensión de cómo se lleva a cabo la comunicación verbal entre personas (véase el capítulo 2), por lo que parecería ilógico ignorar, en el diseño de un sistema de diálogo, las aportaciones procedentes de estos campos.

Finalmente, debe tenerse en cuenta, desde la perspectiva que nos ocupa en este apartado, que aunque las tareas que se realizan con un sistema de diálogo se lleven a cabo mediante el habla, es imprescindible lograr una comprensión o representación

semántica del enunciado, así como una respuesta bien construida en términos gramaticales. Por este motivo, como hemos visto, se recurre a las herramientas y los métodos propios del procesamiento del lenguaje natural –comprensión y generación–, que complementan a las tradicionalmente englobadas en las tecnologías del habla –reconocimiento y conversión de texto en habla.

Si las tecnologías lingüísticas se han desarrollado, en ocasiones, relativamente al margen de la lingüística, las tecnologías del lenguaje y las del habla han pasado también por una etapa caracterizada por una cierta falta de comunicación entre ambas. La necesidad de incorporar técnicas de comprensión y de generación a los sistemas de diálogo y de contar con información morfológica, sintáctica y semántica en la conversión de texto en habla y en el reconocimiento, por una parte, y la importancia que los métodos estocásticos, procedentes inicialmente del mundo del tratamiento del habla, han adquirido en el procesamiento del lenguaje natural, por otra, han contribuido indudablemente a un acercamiento.

En conclusión, nos encontramos ante un campo que aúna las tecnologías del habla con el procesamiento del lenguaje natural, y las tecnologías del lenguaje con las disciplinas lingüísticas. Este libro constituye una buena muestra de ello, pues han contribuido a su redacción cinco expertos procedentes del ámbito de las tecnologías y cinco que provienen del campo de la lingüística. Se imponen, pues, los equipos formados por especialistas con saberes muy diversos, capaces de analizar y modelar, en todas sus facetas, el proceso por el que las personas llegamos a comunicarnos.