

Adquisición de un corpus de diálogo: DIHANA

Nieves Alcácer, María J. Castro, Isabel Galiano, Ramón Granell, Sergio Grau, David Griol

Departamento de Sistemas Informáticos y Computación
Universidad Politécnica de Valencia

{nalcacer,mcastro,mgaliano,rgranell,sgrau,dgriol}@dsic.upv.es

Resumen

En este trabajo se presenta el diseño y el proceso de adquisición de un corpus de diálogo de habla espontánea en castellano para el proyecto DIHANA. Para realizar esta adquisición se utilizó la técnica del Mago de Oz donde una persona simula el comportamiento de un gestor de diálogo, siendo automáticos los otros módulos del sistema de diálogo (reconocedor de voz, módulo de comprensión y sintetizador de voz).

En primer lugar, se realizó el diseño y planificación de la adquisición. Se definieron escenarios de diálogos y se desarrolló una plataforma de adquisición para uso del Mago de Oz. También se definió la estrategia a seguir por parte del Mago.

Un total de 900 diálogos fueron adquiridos por 225 usuarios. A todos ellos se les pasó una encuesta para evaluar el proceso de adquisición.

1. Introducción

El estudio y desarrollo de sistemas de diálogo automático es actualmente uno de los campos más destacados dentro de las tecnologías del lenguaje y del habla. Para el desarrollo de los distintos módulos que los componen [1], se hace imprescindible la existencia de ejemplos de diálogos de la tarea sobre la que se desea realizar el sistema de diálogo.

Uno de los objetivos tecnológicos del proyecto DIHANA [2] es la adquisición de un gran corpus oral de diálogo en lengua castellana de habla espontánea. La adquisición se plantea como una ampliación del corpus de diálogo adquirido para el proyecto BASURDE [3], siendo la tarea seleccionada en ambos casos la consulta telefónica de horarios y precios de trenes de grandes líneas. Otros proyectos de diálogo seleccionaron tareas similares [4, 5].

Los diálogos que se requieren para formar ese corpus tienen que cumplir la condición de que, aún siendo habla espontánea, deben estar sometidos a cierto control. Para ello se ha hecho uso de la técnica del Mago de Oz [6] y se ha planificado la adquisición minuciosamente mediante la definición de escenarios y una estrategia para el mago [7], además del desarrollo de una plataforma de adquisición [8].

La planificación de escenarios se explica con detalle en el siguiente apartado. Posteriormente, se expone como se ha desarrollado la adquisición y algunos detalles de la estrategia y de la plataforma. A continuación se muestran los resultados obtenidos de esta captación de datos y la evaluación que han realizado usuarios de la misma. Por último se comentará algunas conclusiones y cuales son las principales labores y líneas de investigación que se seguirán a partir del desarrollo de este trabajo.

El trabajo se ha desarrollado en el marco del proyecto DIHANA subvencionado por la CICYT número TIC2002-04103-C03-03.

<i>Objetivo:</i>	obtener horario; obtener precio
<i>Tipo de viaje:</i>	viaje de ida
<i>Situación:</i>	asistir a una boda
<i>Punto de salida:</i>	Valencia
<i>Punto de llegada:</i>	Barcelona
<i>Restricciones:</i>	salir el viernes; salir a partir de las 15:00; llegar el sábado; llegar antes de las 10:00; viajar en Talgo

Figura 1: Ejemplo de escenario.

2. Diseño de la adquisición

Como se ha comentado en la introducción, la planificación de la adquisición de los diálogos es un punto crucial para que las conversaciones adquiridas tengan la suficiente naturalidad y a su vez, se realicen en condiciones controladas. Estas condiciones de control vienen, en parte, reflejadas en el diseño de los escenarios de los diálogos [9]. Además de los diálogos, para ampliar el corpus acústico, se ha adquirido un conjunto de frases de la tarea más otro conjunto genérico de frases.

2.1. Escenarios

Los escenarios describen las condiciones y circunstancias concretas para que el informante sepa como actuar ante el sistema a la hora de dialogar con él. En la definición del escenario se incluye un *objetivo* (la información que debe obtener el informante), una *situación* que motiva el interés en la información, así como un esquema que condiciona la petición (*tipo de viaje, salida, llegada y restricciones*). Un ejemplo de escenario se muestra en la figura 1.

2.1.1. Tipos de escenarios

Los escenarios pueden ser clasificados en tres categorías o tipos (A, B, o C) dependiendo del objetivo que describen:

- El tipo A tiene por objetivo obtener el horario de trenes de viajes de ida.
- Los escenarios tipo B tienen por objetivo obtener el precio y, opcionalmente, el horario de trenes de viajes de ida.
- Los escenarios de tipo C son análogos a los de tipo B, pero se refieren a trayectos de ida y vuelta.

Las variables que definen la situación del escenario (hora, destino, origen, etc.) pueden ser instanciadas en la propia

descripción de la situación (variante *cerrada*), o deben ser instanciadas por el informante (variante *abierta*).

Para cada tipo de escenario se han definido 10 objetivos concretos y para cada uno de estos se han establecido 12 variantes (8 cerradas y 4 abiertas), en el caso de escenarios tipo A y tipo B; y 6 variantes (4 cerradas y 2 abiertas), en el caso de escenarios tipo C. De este modo se obtiene un total de 300 escenarios distintos.

2.1.2. Distribución de los escenarios

Las grabaciones de diálogos utilizando la técnica del Mago de Oz han sido realizadas por 225 informantes, cada uno de los cuales adquirió 4 escenarios. En la distribución de los escenarios entre los informantes se han respetado las siguientes condiciones:

- los escenarios adquiridos por un informante han de corresponder, como mínimo, a dos tipos diferentes;
- cada informante deberá realizar variantes abiertas y cerradas.

2.2. Frases

Además de realizar los escenarios, cada informante debe adquirir un subconjunto de frases acústicamente balanceadas y un subconjunto de frases de la tarea. Estas frases tienen interés, dentro de la tarea concreta, para la estimación de modelos acústicos y los modelos de lenguaje.

Por un lado, hay 200 frases acústicamente balanceadas. Cada informante ha pronunciado 8 de esas frases, de forma que, cada frase ha sido repetida 9 veces y el número total de frases leídas han sido 1800.

Además existen 1800 frases de la tarea distintas, de las cuales cada informante ha leído 8. Estas frases se han generado de forma sintética utilizando unos modelos sencillos de categorías.

3. Proceso de adquisición

3.1. Plataforma de adquisición y la estrategia del Mago

La solución típica para obtener una adquisición inicial consiste en emplear la técnica basada en un Mago de Oz en la que una persona sustituye al sistema de diálogo. La labor del Mago consiste en simular el sistema automático, ayudando al usuario a obtener respuesta a sus consultas, por lo que, en sucesivos turnos de diálogo, interacciona con el usuario siguiendo una estrategia dada. Para realizar este trabajo, el Mago puede estar asistido por un sistema de adquisición que facilite su labor.

La plataforma desarrollada en el marco del proyecto está formada por diferentes componentes que son utilizados en el proceso de adquisición: un gestor de comunicaciones (línea telefónica), un sistema de parametrización y reconocimiento automático del habla, un sistema de diálogo, un sistema de comprensión y un sistema de síntesis de voz. Todos estos componentes están automatizados, excepto el sistema de diálogo que se simulará mediante la técnica del Mago de Oz. La forma en que el usuario interacciona con el mago es la siguiente:

- El usuario realiza una consulta telefónica al sistema de manera que el Mago escucha lo que realmente dice el usuario. Las consultas son sobre alguno o algunos de los siguientes conceptos o frames: horarios, precios, duración o tiempo de recorrido, tipos de tren y servicios. Cada uno de estos frames constituye el concepto principal de la consulta que se está realizando.

- La señal vocal pasa al reconocedor automático del habla, el cual proporciona la secuencia de palabras que con mayor probabilidad ha dicho el usuario, junto con un valor numérico que da cuenta de la fiabilidad con que se ha reconocido dicha secuencia (figura 2, “Salida reconocedor”).
- A continuación el módulo de comprensión toma, de la secuencia de palabras del reconocedor, la información semántica relevante de la misma. Dicha información es utilizada para completar la información asociada al frame. (figura 2, “Salida módulo comprensión”).

Si esta información es insuficiente o no es muy precisa, el Mago solicita nueva información al usuario en un nuevo turno de diálogo, y así sucesivamente se repetirán los pasos anteriores, hasta que el Mago considere que la información disponible es suficiente para realizar una consulta a la base de datos.

Esta información vendrá recogida en una estructura a la que llamamos *pizarra* (véase figura 2). Inicialmente, la pizarra únicamente tendrá la información fecha actual y hora del sistema. Posteriormente, el Mago solicitará información al usuario en uno o más turnos, y estos datos irán actualizando la pizarra. Una vez el Mago tenga los datos mínimos completados del frame en cuestión, realizará la consulta a la base de datos, y los resultados de la misma se mostrarán en la plataforma (figura 2, “Resultado base de datos”).

En función de cómo se vaya actualizando la pizarra y el estado del diálogo, el Mago seguirá una estrategia determinada. Para ello el Mago interacciona con el usuario a través del módulo de generación de respuestas escritas, que serán comunicadas al usuario mediante un sintetizador (figura 2, “Generador de respuestas”).

La estrategia del Mago consistirá en ir supervisando la información generada durante el proceso de diálogo. Para ello podrá seguir alguna de las siguientes acciones:

- Completar información.
- Confirmación de datos.
- Comunicar que la información no es válida o no ha sido entendida.
- Consultar a la base de datos.

La elección de una u otra acción irá ligada a la confianza que se tenga en lo que ha dicho el usuario. Dicha confianza está automatizada en algunos casos. Por ejemplo, el reconocedor proporciona un valor ligado a la verosimilitud que se tiene en lo dicho por el usuario. En otros casos el Mago debe decidir acerca de lo dicho por el usuario. Cuando el Mago considera que ya puede realizar una consulta a la base de datos del sistema, se lo indica al usuario y la efectúa. La importancia de la estrategia radica en sistematizar el comportamiento y las respuestas del Mago (y más en nuestro caso cuando existe un mago por cada una de las 3 sedes).

En la figura 2 puede verse el aspecto de la plataforma de adquisición que ha sido desarrollada. Esta plataforma comprende las diferentes fuentes de conocimiento que han sido mencionados anteriormente.

3.2. Fases de la adquisición

La adquisición se ha realizado en diversas fases dependiendo del grupo al que pertenecían los informantes. Los primeros usuarios que adquirieron fueron investigadores del proyecto DIHANA, con el fin de realizar pruebas iniciales y así llevar a cabo una primera evaluación del sistema de adquisición, antes

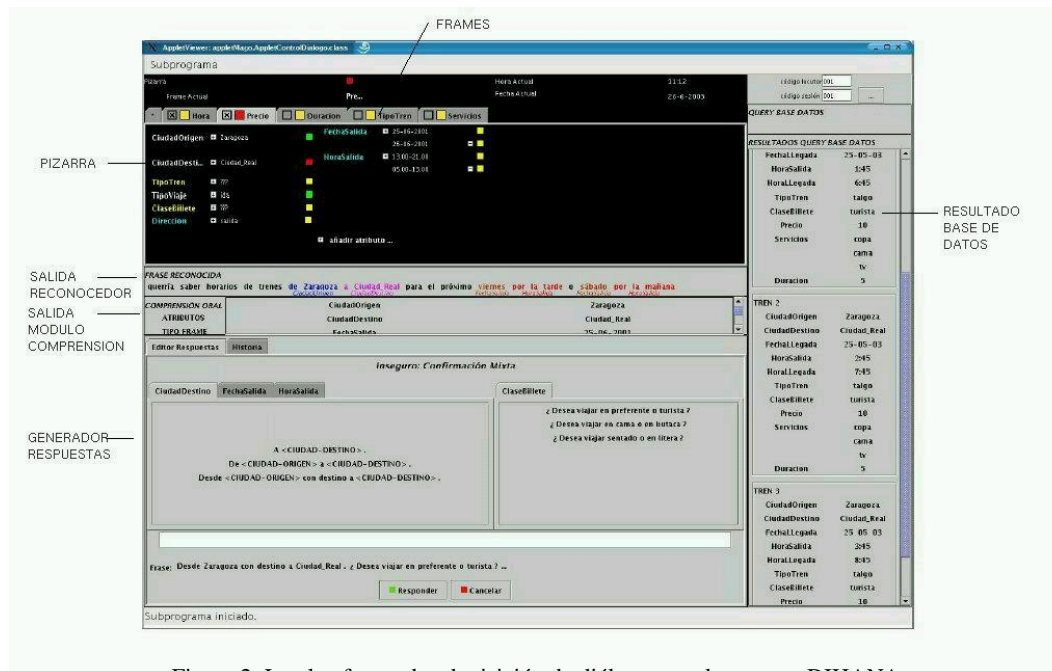


Figura 2: La plataforma de adquisición de diálogos en el proyecto DIHANA.

de que se pasara a adquirir con usuarios externos. Estos eran los únicos usuarios que conocían el funcionamiento interno del sistema de adquisición y, por lo tanto, conocían la existencia del Mago de Oz.

Posteriormente adquirieron usuarios externos al proyecto, formando este grupo compañeros de los investigadores del proyecto y alumnos. Todos ellos pensaban que estaban directamente adquiriendo a través de un sistema automático, sin la intervención de un Mago de Oz.

La experiencia se llevó a cabo paralelamente en las sedes de Bilbao, Valencia y Zaragoza.

4. Resultados

Las características generales del corpus de diálogo adquirido son las siguientes:

- 75 informantes por cada una de las 3 sedes: total 225 informantes.
- 4 diálogos (escenarios) por informante: total 900 diálogos.
- 16 frases por informante. 3600 frases leídas (1800 de la tarea y 1800 genéricas).

Por su parte, los informantes fueron estudiantes y personal de las universidades involucradas en el proyecto. La distribución por sexos fue 153 hombres y 72 mujeres y se intentó adquirir con personas que no tuvieran acentos locales excesivos.

A partir de los ficheros de audio de los diálogos se ha obtenido manualmente una transcripción ortográfica y un etiquetado de los fenómenos acústicos de los diálogos (ruidos y disfluencias).

El corpus posee 6278 turnos de usuario y 9129 turnos de Mago. Hay una media de 7 turnos de usuario y 10 turnos de Mago por diálogo. La media de palabras por turno de usuario es de 7.74 palabras y el vocabulario es de 823 palabras. Contabilizando los turnos de usuario y las frases adquiridas (3600 frases) hay 10.8 horas de grabación de voz de usuario.

Respecto a los ficheros en los que se almacenan los datos de los diálogos, hay que destacar que todos, excepto el de la señal acústica, están diseñados mediante el uso de etiquetas XML. La información almacenada es:

1. Las frases del diálogo del Mago.
2. Las frases del diálogo del usuario (transcripción ortográfica y salida del reconocedor del sistema de RAH).
3. Salida del módulo de comprensión.
4. Datos que se muestran en la pizarra.
5. Consultas a la base de datos y su resultado.
6. Audio.

5. Evaluación

Para realizar una primera evaluación sobre la experiencia, se diseñó una encuesta para medir el grado de satisfacción del usuario. La estructura de la encuesta está conformada por tres bloques diferenciados. El primero de ellos contiene seis preguntas para realizar una evaluación general del proceso de adquisición (véase la primera columna de la tabla 1). La tabla 1 muestra los resultados obtenidos para el primer bloque de preguntas generales.

Analizando la encuesta, puede concluirse que la evaluación general de la experiencia fue positiva. Hay que destacar la dualidad existente a la hora de valorar de forma correcta el ritmo de interacción del sistema y la relativa lentitud a la hora de proporcionar la respuesta al usuario.

Seguidamente en la encuesta, se evalúa la consecución del objetivo para cada uno de los cuatro escenarios que contiene la ficha de adquisición.

La figura 3 resume los resultados obtenidos en el bloque de consecución de objetivos de los escenarios. Las posibles respuestas eran *Muy fácil*, *Fácil*, *Así así*, *Difícil*, *No pude*.

En cuanto a la comparativa entre escenarios cerrados y abiertos, los resultados obtenidos fueron similares. Simplemente destacar un ligero incremento en la imposibilidad de obtener

el objetivo del escenario para el caso de los escenarios cerrados, que puede achacarse a su mayor complejidad.

Finalmente, la encuesta concluye solicitando a los informantes comentarios adicionales sobre la experiencia. En cuanto a las problemáticas más repetidas, cabe destacar:

- necesidad de mejora de la voz del sintetizador,
- reconocimiento de frases de mayor longitud,
- lentitud de respuesta,
- problemas de entendimiento y de generación de respuestas.

6. Conclusiones

En este trabajo se describe el proceso de adquisición completo de un corpus de diálogo en habla espontánea, constando este corpus de un total de 900 diálogos y 3600 frases, en los que han participado 225 usuarios en tres sedes diferentes.

El proceso de adquisición consistió en el diseño de los escenarios para los diálogos y en el desarrollo de una plataforma específica, utilizándose para simular el funcionamiento real de un sistema de diálogo la técnica del Mago de Oz. Los Magos (uno por sede) siguieron una estrategia previamente definida, de forma que el control del diálogo se enfocase en base a una iniciativa mixta, donde el control se reparte entre sistema y usuario. Además, se ha realizado una evaluación de la adquisición por parte de los usuarios, a través de encuestas.

A partir de este corpus se ha realizado una transcripción ortográfico-fonética con fenómenos acústicos, y actualmente, se están etiquetando los diálogos semiautomáticamente a nivel de comprensión.

7. Agradecimientos

Agradecemos, en primer lugar, la participación de los 225 usuarios que nos han permitido obtener el corpus de diálogo.

Por otra parte, agradecemos la colaboración de todos los componentes del proyecto DIHANA en el proceso de adquisición.

8. Referencias

- [1] Giachin, E. and McGlashan, S. : "Spoken Language Dialog Systems". Chapter 3 of *Corpus-Based Methods in Language and Speech Processing*. S.Young and G. Bloothoof (eds.), 67-117, Kluwer Academic Publishers. 1997
- [2] <http://www.dihana.upv.es>
- [3] Bonafonte, A., Aibar, P., Castell, E., Lleida, E., Mariño, J. B., Sanchis, E., Torres, M. I. 2000. Desarrollo de un sistema de diálogo oral en dominios restringidos. I Jornadas en Tecnología del Habla, Sevilla.
- [4] L. Lamel and S. Rosset and J. L. Gauvain and S. Bennacef and M. Garnier-Rizet and B. Prouts, L. Lamel and others, The LMSI Arise system, *Speech Communication*, volumen 31, número 4, páginas 339-354, agosto 2000.
- [5] R. Pieraccini and E. Levin and W. Eckert, AMICA: The AT&T Mixed Initiative Conversational Architecture, *Proceedings of Eurospeech'97*, páginas 1875-1878, Rhodes (Greece).
- [6] N. M. Fraser and G. N. Gilbert, *Simulating speech systems*, *Computer Speech & Language*, 1991, Volumen 5, páginas 81-99.
- [7] Galiano et al. Estrategia del Mago. Informe técnico DIHANA TIC2002-04103-C03.
- [8] M.I. Galiano, R. Granell, LL.F. Hurtado, A. Miguel, J.A. Sánchez, and E. Sanchis. La plataforma de adquisición de diálogos en el proyecto DIHANA. In *Proceedings of the SEPLN: XIX Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural*, pages 343-344. Septiembre 2003.
- [9] M.J.Castro et al. Report adquisición, Informe técnico DIHANA TIC2002-04103-C03.

	Bilbao	Valencia	Zaragoza	Total
¿Entendió al sistema cuando éste le hablaba?	(a) 65.15 % (b) 28.79 % (c) 6.06 % (d) 0 % (e) 0 %	(a) 72.86 % (b) 18.57 % (c) 8.57 % (d) 0 % (e) 0 %	(a) 78.67 % (b) 16.00 % (c) 5.33 % (d) 0 % (e) 0 %	(a) 72.51 % (b) 20.85 % (c) 6.64 % (d) 4 % (e) 0 %
¿El sistema comprendió lo que usted le decía?	(a) 1.52 % (b) 48.48 % (c) 48.48 % (d) 1.52 % (e) 0 %	(a) 10.00 % (b) 54.29 % (c) 25.71 % (d) 10.00 % (e) 0 %	(a) 2.67 % (b) 69.33 % (c) 26.67 % (d) 1.33 % (e) 0 %	(a) 4.74 % (b) 57.82 % (c) 33.18 % (d) 4.27 % (e) 0 %
¿Fue adecuado el ritmo de interacción?	(a) 9.09 % (b) 34.85 % (c) 28.79 % (d) 22.73 % (e) 4.55 %	(a) 14.29 % (b) 30 % (c) 34.29 % (d) 17.14 % (e) 4.29 %	(a) 4 % (b) 66.67 % (c) 22.67 % (d) 6.67 % (e) 0 %	(a) 9 % (b) 44.55 % (c) 28.44 % (d) 15.17 % (e) 2.84 %
¿Supo usted cómo actuar en cada momento del diálogo?	(a) 37.88 % (b) 36.36 % (c) 24.24 % (d) 1.52 % (e) 0 %	(a) 42.86 % (b) 27.14 % (c) 30.00 % (d) 0 % (e) 0 %	(a) 36.00 % (b) 41.33 % (c) 20.00 % (d) 2.67 % (e) 0 %	(a) 38.86 % (b) 35.07 % (c) 24.64 % (d) 1.42 % (e) 0 %
¿Con qué frecuencia el sistema fue lento en su respuesta?	(a) 10.61 % (b) 9.09 % (c) 25.76 % (d) 53.03 % (e) 1.52 %	(a) 7.14 % (b) 21.43 % (c) 21.43 % (d) 50.00 % (e) 0 %	(a) 1.33 % (b) 8 % (c) 24.00 % (d) 62.67 % (e) 4 %	(a) 6.16 % (b) 12.8 % (c) 23.7 % (d) 55.45 % (e) 1.90 %
¿El sistema se comportó del modo esperado durante la conversación?	(a) 10.61 % (b) 43.94 % (c) 40.91 % (d) 4.55 % (e) 0 %	(a) 21.43 % (b) 50.00 % (c) 27.14 % (d) 1.43 % (e) 0 %	(a) 12.00 % (b) 56.00 % (c) 28.00 % (d) 4 % (e) 0 %	(a) 14.69 % (b) 50.24 % (c) 31.75 % (d) 3.32 % (e) 0 %

Tabla 1: Se muestra el porcentaje de respuestas: (a)Siempre (b)Habitualmente (c)La mayor parte de las veces (d)A veces (e)Nunca, en cada una de las sedes en las que se realizó la adquisición (Bilbao, Valencia y Zaragoza) y el promedio de todo ello.

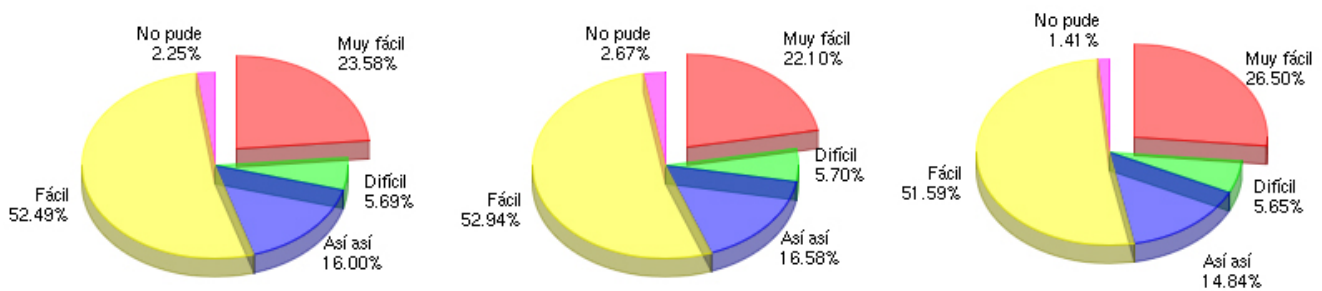


Figura 3: Resultados globales en la evaluación del objetivo del escenario. (1)Total (2)Escenarios cerrados (3)Escenarios abiertos