Modelos específicos de comprensión en un sistema de Diálogo *

Fernando García, Emilio Sanchís, Lluís Hurtado, Encarna Segarra
Universidad Politécnica de Valencia
Camino de Vera s/n
46022 Valencia
{fgarcia,esanchis,lhurtado,esegarra}@dsic.upv.es

Resumen: Se presenta una aproximación para la estimación del componente de comprensión de un sistema de diálogo en la que este componente se hace depender del propio diálogo. En particular, se propone llevar a cabo una modelización específica para cada estado del proceso de diálogo. Este trabajo se desarrolla dentro del sistema de diálogo BASURDE, que contesta a través de la línea telefónica a consultas sobre horarios y precios de trenes de largo recorrido en castellano. Se presentan también algunos resultados experimentales.
Palabras clave: comprensión del habla, sistemas de diálogo, modelos estocásticos.

Abstract: We present an approach to the estimation of a dialogue-dependent understanding component of a dialogue system. Modelization which is specific to the dialogue state is proposed to improve the behaviour of the understanding process. This work is developed in the framework of the BASURDE Spanish dialogue system, which answers queries about train timetables by telephone in Spanish. Some experimental results are presented.
Keywords: Language Understanding, Dialogue Systems, Stochastic Models.

1. Introducción

Algunas de las características de los sistemas de diálogo hablado que se han desarrollado en los últimos tiempos (Lamel et al., 2000)(Glass y Weinstein, 2001)(CSDTK, 1999) son: acceso telefónico, dominio semántico restringido e iniciativa mixta en el control del diálogo. El diseño de sistemas de iniciativa mixta resulta mucho más complejo que el diseño de sistemas de diálogo guiados únicamente por el controlador de diálogo, pero presenta la ventaja de dar lugar a diálogos mucho más naturales. De momento, abordar el desarrollo de sistemas de la iniciativa mixta es posible siempre que se restrinja el dominio semántico de la aplicación. Por otra parte, el trabajo con dominios restringidos comporta que los corpora disponibles para este tipo de aplicaciones, etiquetados manualmente con alguna información sintáctica o semántica, sean de dimensiones muy reducidas.

Un sistema de diálogo hablado de estas características para la interrogación a una base de datos con información sobre un tema específico comprende varios componentes: el módulo reconocedor del habla, el módulo de comprensión del lenguaje, el controlador del diálogo, la interfaz con la base de datos, el módulo de generación de respuesta y el sintetizador de habla.

En este trabajo se presenta una aproximación a la construcción del componente de comprensión del habla, estudiando la viabilidad de aplicar modelos específicos en función del estado del diálogo (Xu y Rudnicky, 2000)(Hacioglu y Ward, 2001), y su aplicación al sistema de diálogo BASURDE (Bonafonte et al., 2000). La tarea en este sistema consiste en la consulta en castellano a través del teléfono sobre horarios y precios de trenes de la red española de ferrocarriles.

En el sistema BASURDE el reconocedor de voz proporciona la entrada del componente de comprensión, siendo la salida de éste, proporcionada al controlador del diálogo. Para representar el significado de las pronunciaciones del usuario se emplea el formalismo de los frames. Un frame determina, para cada uno de los turnos de usuario de los que consta el diálogo, el tipo de comunicación, así como, los datos que éste aporta.

Existen dos aproximaciones clásicas al problema de la comprensión del habla: la primera se basa en el uso de reglas para la detec-

* Este trabajo ha sido parcialmente financiado por los proyectos CICYT TIC 2000-0664-C02-01 y CICYT TIC 2002-04103-C03-03

ción de marcadores y palabras clave capaces de determinar el tipo de frame y hacerle corresponder sus atributos a partir de la frase de entrada; el segundo se basa en el uso de modelos que se aprenden automáticamente a partir de muestras. Algunos de ellos se basan en modelos estocásticos (Modelos ocultos de Markov y Gramáticas Regulares Estocásticas) (Bonneau-Maynard y Lefèvre, 2001)(Segarra et al., 2002)(Sanchis et al., 2002).

Una de las ventajas de emplear técnicas de aprendizaje automático es que permiten la adaptación a nuevas tareas y situaciones o incluso su aplicación a otras lenguas, de una forma sencilla; desarrollar nuevos modelos consiste en reestimar los modelos. Pero esto es posible sólo si se dispone de una suficiente cantidad de datos de entrenamiento.

En nuestro sistema de diálogo, la representación de la estructura de un diálogo está implementada a través de una red estocástica de actos de diálogo. El uso de esta estructura permite obtener una previsión del siguiente acto de diálogo esperado del usuario. Asumiendo que la respuesta de diferentes usuarios al mismo acto de diálogo del sistema es similar en algún sentido, esta información se va a emplear en el proceso de comprensión. En particular, se van a emplear diferentes modelos de comprensión en función del último acto de diálogo emitido por sistema, con el objetivo de obtener unos modelos más específicos, más dirigidos a la previsión del acto de diálogo correspondiente a la siguiente intervención del usuario.

2. La tarea BASURDE

La tarea BASURDE (Bonafonte et al., 2000) consiste en consultas telefónicas sobre los trenes españoles. El tipo de consultas (restricciones semánticas) son: consultas sobre horarios, precios y servicios de los trenes de largo recorrido. A partir del análisis de un conjunto de diálogos persona-persona se delimitó la tarea definiendo cuatro tipos de escenarios (hora de salida/llegada para ida, viaje de ida y vuelta, precios y servicios además de uno libre) con el fin de proceder a una adquisición de diálogos bajo condiciones controladas empleando la técnica del Mago de Oz. Un total de 215 diálogos fueron adquiridos de esta forma, conteniendo un total de 1440 turnos de usuario (14.902 palabras, siendo 14,4 el número medio de turnos por diálogo).

3. Etiquetas de actos de diálogo

La estructura del diálogo puede ser representada mediante actos de diálogo, para ello debemos definir un conjunto de etiquetas (Martinez et al., 2002). Por una parte, el número de etiquetas debe ser lo suficientemente grande como para modelizar las diferentes intenciones de los turnos, así como para cubrir todas la situaciones. Pero, por otra parte, si este número es demasiado alto, los modelos estimados para cada acto de diálogo pueden resultar mal estimados debido a la escasez de muestras. En BASURDE se ha definido un etiquetado a tres niveles, donde en un primer nivel se especifica el comportamiento del diálogo, y las etiquetas son genéricas para cualquier tarea de diálogo. En un segundo nivel las etiquetas hacen referencia a la semántica de las frases de entrada, y son específicas de la tarea. El tercer nivel hace referencia a las instancias particulares de los datos que se encuentran en las frases de entrada.

Se ha apendido una red estocástica de actos de diálogos a partir de las muestras de entrenamiento empleando estas etiquetas (Sanchis, Galiano, y Garcia, 2001). Esta red representa la estructura del diálogo, y puede ser empleada para que el controlador de diálogo genere el siguiente acto de diálogo del sistema y guiar con ello la estrategia del controlador. La predicción de la red estocástica también puede ayudar en la fase de reconocimiento y de comprensión generando hipótesis de cual va a ser el siguiente acto de diálogo del usuario.

En este trabajo nos centraremos en el primer nivel de etiquetado de diálogo, que corresponde al comportamiento del diálogo independiente de la tarea. Este primer nivel presenta las siguientes etiquetas de diálogo: *Apertura, Cierre, Indefinida, Espera, Nueva_consulta, No_entendido, Afirmación, Confirmación, Pregunta* y *Respuesta*. Nos limitamos a utilizar sólo este nivel en la definición de modelos específicos de comprensión ya que pensamos que las etiquetas que define representan una buena clasificación de los actos del usuario y además el número de muestras de cada clase sigue haciendo abordable su estimación de a partir de datos. En la Figura 1 se muestra un ejemplo de anotación de primer nivel.

S1:	Bienvenido al sistema automático de información de trenes regionales y de largo recorrido, qué desea? (S:*Apertura*)
U1:	Puede decirme a qué hora salen los trenes de Valencia a Barcelona? (U:*Pregunta*)
S2:	De Valencia a Barcelona,(S:*Confirmation*) qué dia quiere salir? (S:*Confirmación* S:*Pregunta*)
U2:	El próximo jueves (U:*Respuesta*)

Figura 1: Ejemplo de etiquetado de diálogo a un nivel

4. Categorización

Como el número de instancias de algunos atributos puede llegar a ser muy pequeño, pueden darse problemas de cobertura y de falta de muestras a la hora de aprender modelos estocásticos a partir de un corpus. Por ello, se trató de reducir el número de palabras en nuestro léxico con el empleo de lemas y categorías semánticas. Dada la compleja conjugación de los verbos españoles, se sustituyeron las diferentes formas verbales por su correspondiente infinitivo. Además se sustituyó cada instancia de nombre de ciudad por su categoría *nombre_ciudad*, las instancias de día de la semana por *diasem*; y los mismo con los números, meses, etc. . . . Se definieron siete categorías en el léxico, reduciendo el tamaño de 637 a 370 palabras diferentes.

5. Modelización de la comprensión

La modelización de la comprensión se divide en dos fases (Figura 2). La primera fase consiste en la traducción de la frase de entrada en términos de un lenguaje semántico intermedio: a cada frase de entrada le corresponde una cadena de unidades semánticas definidas sobre un cierto vocabulario semántico. Este lenguaje semántico se define de forma que las cadenas de unidades semánticas tienen una correspondencia secuencial con la frase de entrada, ello permite el uso de técnicas de traducción secuencial. En la segunda fase una serie de reglas traduce esta representación intermedia en una representación basada en frames. Como el lenguaje semántico intermedio esta cercano a la representación de frames, esta fase sólo requiere de unas pocas reglas para construir el frame. Un ejemplo de las acciones llevadas a cabo en esta segunda fase son las conversiones de las fechas relativas y horas en valores absolutos, p.e. "*próximo lunes*" por "*mm/dd/yy*" o "*por la mañana*" por "*intervalo de hora (5 a 12)*". Para la primera fase se hace uso de modelos estocásticos aprendidos automáticamente a partir de muestras.

Como ya hemos comentado, la cadena en el lenguaje semántico intermedio definido para la tarea es secuencial con la frase de entrada. Esto nos permite segmentar esta frase de entrada en un número de intervalos igual al número de unidades semánticas que hay en su correspondiente cadena semántica. Esto es, sea W el vocabulario de la tarea (conjunto de palabras) y V el alfabeto de unidades semánticas definido.

Cada frase de entrada en W^* tiene asociado un par (u,v), donde v es una secuencia de unidades semánticas y u es una secuencia de segmentos de palabras de la frase. A continuación podemos ver un ejemplo:

Par de entrada $(u,v)=(u_1u_2u_3u_4, v_1v_2v_3v_4)$
donde:

u_1: quisiera	v_1: consulta
u_2: horarios de trenes	v_2: (hora_salida)
u_3: a	v_3: marcador_destino
u_4: Alicante	v_4: ciudad_destino

La cadena semántica v para el modelo de lenguaje semántico de entrenamiento es:

consulta (hora_salida) marcador_destino ciudad_destino

Cuando se tiene un conjunto de entrenamiento de este tipo, el problema del aprendizaje de la traducción secuencial puede resolverse a través de autómatas de estados finitos. A continuación exponemos la aproximación seguida en este trabajo para llevar a cabo la modelización de esta fase de la comprensión (Segarra et al., 2002).

6. El modelos de dos niveles

Este modelo consiste en el aprendizaje de dos tipos de modelos a partir de un conjunto de pares de entrenamiento (u,v): un modelo para el lenguaje semántico $L_s \subseteq V^*$, y

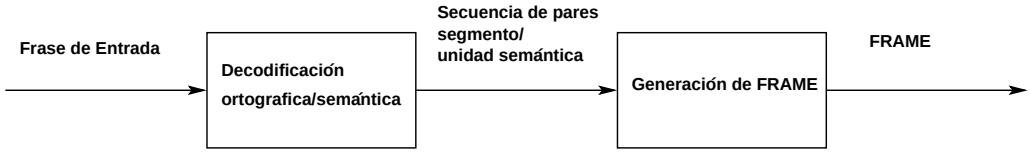

Figura 2: Esquema del proceso de comprensión

un conjunto de modelos, uno por categoría semántica $v_{\mathrm{i}} \in V$. El modelo regular A_{s} (un autómata de estados finitos) para el lenguaje semántico L_{s} se estima a partir de las cadenas semánticas $v \in V^*$ de la muestra de entrenamiento. El modelo regular $A_{\mathrm{v_i}}$ (un autómata de estados finitos) es estimado para cada categoría semántica $v_{\mathrm{i}} \in V$ a partir del conjunto de segmentos u_{i} obtenido de la muestra de entrenamiento asociado a cada una de estas unidades semánticas v_{i}. Estas estimaciones se llevan a cabo a través de técnicas automáticas.

Un modelo final A_{t} se obtiene a través de la aplicación de una sustitución regular σ del lenguaje semántico L_{s}. Sea $\sigma : V^* \rightarrow \mathbf{P}(W^*)$ una sustitución regular tal que $\forall v_{\mathrm{i}} \in V \quad \sigma(v_{\mathrm{i}}) = L(A_{\mathrm{v_i}})$. El modelo regular A_{t} es tal que $L(A_{\mathrm{t}}) = \sigma(L(A_{\mathrm{s}})) = \sigma(L_{\mathrm{s}})$. Esta sustitución σ convierte cada símbolo terminal $v_{\mathrm{i}} \in V$ del modelo regular A_{s} en su modelo regular $A_{\mathrm{v_i}}$ correspondiente. La creación de este modelo se muestra en la Figura 3.

Una de las ventajas de esta aproximación, es que podemos escoger la técnica de aprendizaje más adecuada para la estimación de cada modelo (el modelo semántico y el modelo de unidad semántica). La única restricción es que la representación de estos modelos debe ser dada en forma de un autómata de estados finitos.

Finalmente, el modelo obtenido A_{t} se emplea para analizar la frase de test $w = w_1$-$w_2 \ldots w_{|\mathrm{w}|}$. Este análisis se basa en el algoritmo de Viterbi. En la Figura 4 se muestra un ejemplo de traducción llevado a cabo por esta aproximación.

En este trabajo se ha inferido un modelo clásico de bigramas para el modelo semántico y también para los modelos de cada unidad semántica. Sin embargo, se podrían utilizar otras técnicas para la estimación de ambos modelos, como son el ECGI o el MGGI (Segarra et al., 2002).

El objetivo de la aproximación de 2 niveles es la búsqueda de la segmentación óptima l de las palabras de la frase de entrada $w = w_1, \ldots, w_{|w|}, \quad l = l_1 l_2 .... l_n = |w|, \quad w \in W$. Cada uno de estos segmentos l_i tiene asociada una unidad semántica (concepto) v_i, con lo que una secuencia de conceptos $v = v_1, \ldots, v_{|v|}, \quad v_i \in V$ está asociada a w y representa el significado de la frase.

Dada la secuencia de palabras w, el proceso estocástico consiste en encontrar la secuencia de conceptos v que maximiza la probabilidad *a posteriori*:

$$\widehat{V} = \underset{V}{\operatorname{argmax}} \Pr(V|W)$$

De acuerdo con la formula de Bayes la ecuación puede reescribirse de la siguiente forma:

$$\widehat{V} = \underset{V}{\operatorname{argmax}} \Pr(W|V) \Pr(V)$$

El termino $\Pr(W|V)$ es estimado como la probabilidad de cada segmento l dentro de su categoría:

$$\Pr(W|V) = \text{máx}_{\forall l_1, l_2, \ldots l_{n-1}} \{\Pr(w_1, ..., w_{l_1}|v_1) \cdot \Pr(w_{l_1+1}, ..., w_{l_2}|v_2) \cdot .. \cdot \Pr(w_{l_{n-1}+1}, .., w_n|v_n)\}$$

donde la probabilidad de cada segmento es estimada a través de la probabilidad los bigramas de las palabras dado el concepto asociado a la palabra k:

$$\Pr(w_i, ..., w_j|v_s) = \prod_{k=i}^{j} \Pr(w_k|w_{k-1}, v_s)$$

El termino $\Pr(V)$ es estimado en términos de la probabilidad de los bigramas de conceptos.

$$\Pr(V) = \Pr(v_1) \prod_{i=2}^{n} \Pr(v_i|v_{i-1})$$

Para realizar la decodificación semántica se construye un modelo integrado (autómata) A_t (Figura 3) empleando el modelo semántico A_s y el modelo de palabras A_{v_i} para cada

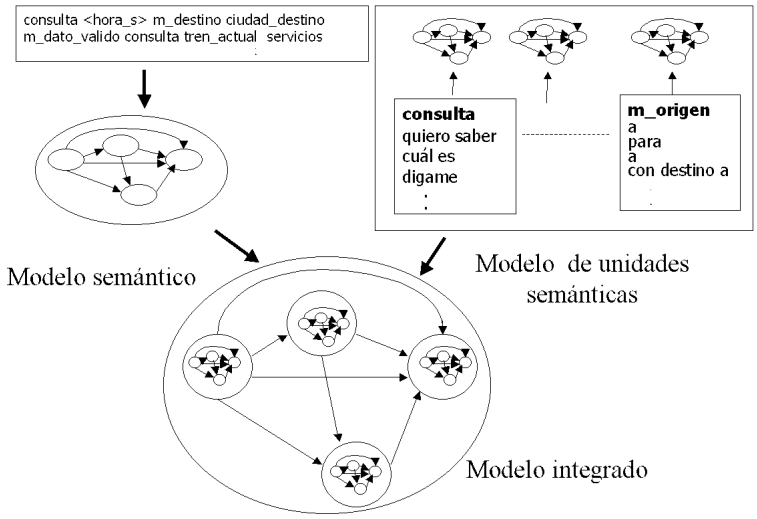

Figura 3: Creación de modelo integrado A_t a partir de los modelos semánticos A_s y los modelos de unidades semánticas $A_{v_i} \forall v \in V$.

Frase de entrada (9 palabras):
me podría decir los horarios de trenes para Barcelona
Frase de salida (9 unidades semánticas):
`consulta consulta consulta (hora_salida) (hora_salida) (hora_salida) (hora_salida) marcador_destino ciudad_destino`
Transducción:
`consulta (hora_salida) marcador_destino ciudad_destino`
Segmentación:
me podría decir: `consulta` los horarios de trenes: `(hora_salida)` para: `marcador_destino` Barcelona: `ciudad_destino`

Figura 4: Ejemplo de traducción.

concepto a partir de las pronunciaciones anotadas semánticamente del corpus de entrenamiento.

Se aplica el algoritmo de Viterbi para obtener la secuencia de unidades semánticas, y la segmentación asociada, correspondiente a la frase asociada.

7. *Modelización específica de la comprensión*

Para conseguir modelos específicos se han clasificado las muestras en función del estado del diálogo. Se dividieron las muestras de entrenamiento del usuario en 10 subconjuntos. Cada subconjunto es asociado con el primer nivel de etiquetas de diálogo, y contiene el turno de usuario que sigue a esta etiqueta. Por ejemplo, el conjunto *Apertura* contiene todos los turnos de usuario que se han pronunciado después de haber generado el sistema el acto de dialogo *Apertura*. Como ya se ha comentado en la introducción, pensamos que esta clasificación de las muestras permite una mejor modelización de los turnos de usuario.

Posteriormente después de analizados los subconjuntos obtenidos se comprobó que sólo cuatro de los subconjuntos, los que corresponden a las etiquetas *Apertura, Confirmación, Nueva_Consulta y Pregunta*, contienen un número de muestras suficiente para llevar a cabo una estimación aceptable del modelo estocástico.

La modelización especializada sólo se ha

aplicado en nivel superior de la técnica de 2 niveles (modelo semántico A_s). Justificamos esta elección en el hecho de que este nivel representa la semántica de la entrada, mientras que el nivel inferior representa sólo la instanciación de esta semántica en términos de secuencias de palabras. Gracias a esta elección, para inferir los modelos de cada unidad semántica (A_{v_i}) hemos aprovechado todas las muestras de dicha unidad en el corpus de entrenamiento.

En el proceso de decodificación, el controlador de diálogo selecciona el modelo específico adecuado (Figura 5), es decir, el que viene predeterminado por el último acto de diálogo generado por el sistema.

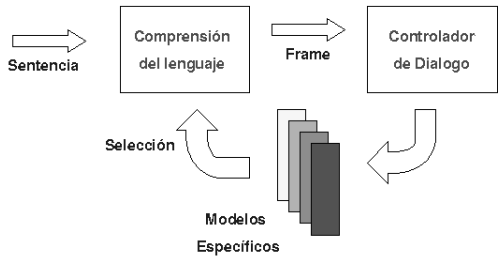

Figura 5: Selección de los modelos específicos A_t.

8. *Experimentación y Conclusiones*

Para estudiar la conveniencia de los modelos específicos de diálogo, se realizó una experimentación sobre BASURDE, comparando los resultados obtenidos con aquellos que fueron obtenidos empleando un modelo de lenguaje general. A partir de esta comparación se presentan una serie de conclusiones.

El corpus consiste en 215 diálogos y 1440 turnos de usuario. Para cada subconjunto de muestras correspondiente a cada una de las cuatro etiquetas de diálogo escogidas *Apertura, Confirmacion, Nueva_consulta* y *Pregunta* se crea un conjunto de entrenamiento del 75 % y uno de test del 25 %. El primer conjunto se emplea en el aprendizaje del modelo específico de esa etiqueta y el segundo se empleará como test, tanto para el modelo específico como para el modelo general. El modelo general se aprende a partir de todo el conjunto de muestras que no es de test. Nótese que el conjunto de aprendizaje para el modelo general es mucho mayor que el de los modelos específicos. En particular hay etiquetas para las cuales el conjunto de aprendizaje para el modelo general, supera el 90 % de las muestras.

Sobre los conjuntos de test definidos anteriormente se han hecho tres experimentos de comprensión: uno con la transcripción manual de las pronunciaciones del usuario (Texto), y otros dos con la salida de dos reconocedores diferentes. El primero (Rec. 1) con un Word Accuracy del 81,1 % y el segundo (Rec. 2) con un Word Accuracy del 84,8 %.

Para estos experimentos se han definido dos medidas de comprensión a nivel de frame:

- porcentaje de frames correctos (%fc); es decir el porcentaje de frames completos (tipo de frame y atributos) que son exactamente igual que el de referencia.
- porcentaje de atributos correctos del frame (nombre del frame y sus atributos) (%acf).

Los resultados (%fc y %acf) obtenidos empleando modelos específicos (esp.) y general (gen.) aparecen en la Tabla 1

Apertura						
	Texto		Rec. 1		Rec. 2	
	esp.	gen.	esp.	gen.	esp.	gen.
%fc	76,7	71,2	39,5	36,1	32,7	30,3
%acf	91,6	89,4	72,2	67,8	73,9	72,5

Confirmacion						
	Texto		Rec. 1		Rec. 2	
	esp.	gen.	esp.	gen.	esp.	gen.
%fc	91,4	87,1	67,1	64,3	73,3	69,8
%acf	95,1	94,2	75,2	73,9	81,9	82,3

Nueva_consulta						
	Texto		Rec. 1		Rec. 2	
	esp.	gen.	esp.	gen.	esp.	gen.
%fc	76,9	78,3	50,7	50,6	55,3	56,2
%acf	83,5	84,8	64,9	65,7	70,8	72,1

Pregunta						
	Texto		Rec. 1		Rec. 2	
	esp.	gen.	esp.	gen.	esp.	gen.
%fc	77,1	88,6	59,1	60,9	62,9	61,9
%acf	87,1	92,8	70,1	72,2	77,6	78,8

Cuadro 1: Resultados de los experimentos sobre modelos específicos y generales para las etiquetas *Apertura, Confirmación, Nueva_consulta y Pregunta.*

Aunque el porcentaje de turnos completamente comprendidos (%fc) no es muy alto, en algunos conjuntos, el porcentaje de atributos

y tipo de frame identificados (%acf) sí que es suficientemente bueno, incluso cuando las frases son la salida del reconocedor. Esto permite que en sucesivos turnos de diálogo, el sistema pueda completar informaciones sobre atributos y corregir errores.

En cuanto al comportamiento de los modelos específicos, se puede observar que es mejor en los conjuntos *Apertura* y *Confirmación*, que corresponden a tipos de intervenciones con estructuras más similares. En los otros casos no se consiguen mejoras, pero hay que tener en cuenta que el conjunto de muestras de aprendizaje es muy pequeño para los modelos específicos. Es de esperar que con una futura ampliación del corpus, se obtenga una mejora de resultados para estas clases y para otras clases que no han sido consideradas en estos experimentos. Además el uso de técnicas de interpolación puede servir para sacar mayor partido de ambos tipos de modelos (específicos y general).

Bibliografía

Bonafonte, A., P. Aibar, N. Castell, E. Lleida, J.B. Mariño, E. Sanchis, y M.I. Torres. 2000. Desarrollo de un sistema de diálogo oral en dominios restringidos. En *I Jornadas en Tecnología del Habla, Sevilla (Spain)*.

Bonneau-Maynard, H. y F. Lefèvre. 2001. Investigating stochastic speech understanding. En *IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*.

CSDTK. 1999. CMU Communicator Spoken Dialog ToolKit (CSDTK). http://www.-speech.cs.cmu.edu/communicator/.

Glass, J. y E. Weinstein. 2001. Speech builder: facilitating spoken dialogue system development. En *EUROSPEECH'01*, volumen 1, páginas 1335–1338.

Hacioglu, K y W. Ward. 2001. Dialog-Context Dependent Language Modeling Combining N-grams and Stochastic Context-Free Grammars. En *ICASSP*.

Lamel, L., S. Rosset, J.L. Gauvain, S. Bennacef, M. Garnier-Rizet, y B. Prouts. 2000. The LIMSI Arise system. *Speech Communication*, 31:339–353.

Martinez, C., E. Sanchis, F. García, y P. Aibar. 2002. A labeling proposal to annotate dialogues. En *Third International Conference on Language Resources and Evaluation (LREC))*, páginas 1577–1582, Las Palmas, Spain, 21-30 Mayo.

Sanchis, E., I. Galiano, y F. Garcia. 2001. A hybrid approach to the development of dialogue systems directed by semantics. En Jan Van, editor, *2nd SIGdial Wookshop on Discourse and Dialogue*, páginas 149–152, Aalborg, Denmark, September.

Sanchis, E., F. García, I. Galiano, y E. Segarra. 2002. Applying Dialogue Constraints to the Understanding Process in a Dialogue System. En Petr Sojka Ivan Kopeček, y Karel Pala, editores, *Fifth International Conference on Text, Speech and Dialogue—TSD 2002*, Lecture Notes in Artificial Intelligence LNCS/LNAI 2448, páginas 389–395, Brno, Czech Republic, September. Springer-Verlag.

Segarra, E., E. Sanchis, F. García, y L.F. Hurtado. 2002. Extracting semantic information through automatic learning techniques. *International Journal of Pattern Recognition and Artificial Intelligence*, 16(3):301–307.

Xu, W. y A. Rudnicky. 2000. Language Modeling for Dialog System. En *ICSLP*, Beijing, China.