

SAPLEN: UN SISTEMA DE DIÁLOGO EN LENGUAJE NATURAL PARA UNA APLICACIÓN COMERCIAL

Ramón López-Cózar Delgado
Antonio J. Rubio Ayuso

Dpto. de Electrónica y Tecnología de Computadores
Universidad de Granada

Resumen

SAPLEN es un sistema de procesamiento del lenguaje natural en fase de desarrollo, que intenta simular el diálogo llevado a cabo por el dependiente de un restaurante de comida rápida a la hora de atender las peticiones o consultas formuladas por los clientes. En el presente artículo se describen las bases que se están siguiendo para desarrollar el sistema, los objetivos que éste pretende conseguir y sus características principales, así como los formalismos de representación y de adquisición del conocimiento empleados. Finalmente, se realiza un análisis modular del sistema, y se lleva a cabo una descripción del estado actual de desarrollo del mismo.

Palabras clave

Procesamiento del lenguaje natural, diálogo, semántica, pragmática, adquisición de conocimiento, representación del conocimiento, frames, sistemas expertos.

1. Introducción.

El sistema SAPLEN (Sistema Automático de Pedidos en Lenguaje Natural) puede considerarse un Sistema Experto basado en reglas y guiado por objetivos [Rich E. y Knight k. 94], que intentamos sea capaz de comportarse de forma *similar* a como lo hace el dependiente de un restaurante de comida rápida a la hora de atender las peticiones de comida o consultas de los clientes.

Para el desarrollo del presente sistema está prestando su colaboración un restaurante de comida rápida de la ciudad de Granada, en el cual se han realizado grabaciones de conversaciones orales entre los dependientes del mismo y los clientes. Actualmente se dispone de un corpus de 117 diálogos transcritos a texto y etiquetados, que se usan para *enseñar* al sistema el comportamiento que se desea adopte.

En cada diálogo, cada frase consta de una serie de etiquetas que identifica el origen de la misma (*dependiente* o *cliente*) y sus características

intencionales (illocutivas) [Seto et al. 94] (pedir comida, pedir bebida, consultar el precio de algún producto, etc.).

El sistema no está diseñado para realizar el cobro correspondiente a la lista de pedidos realizados; su misión consiste en determinar cuales son los productos deseados por los clientes, y facilitarles el ticket correspondiente, en el que figura el coste total a pagar.

2. Objetivos generales.

El objetivo general del sistema es lograr un comportamiento similar al modelo del mundo real, para ello debe ser capaz de proporcionar la información solicitada y formular las preguntas de forma semejante a como lo hacen los dependientes del restaurante. Para lograr este objetivo, el sistema debe ser capaz de procesar las conversaciones espontáneas, debiendo ser tolerante a los fenómenos que suelen ocurrir en la comunicación espontánea entre las personas (repetición innecesaria de palabras, uso continuo de información contextual, alteración del orden gramatical habitual, introducción de incorrecciones gramaticales, anáforas, rectificaciones, pausas, etc.).

Se tiene previsto dotar al sistema de una determinada capacidad de aprendizaje que le permita *aprender* a partir de su interacción con el usuario la existencia de productos de los cuales no tiene conocimiento previo. Así por ejemplo, si el sistema desconoce la existencia de la bebida denominada *San Francisco* y algún cliente realiza un pedido de la misma, el sistema iniciará un diálogo encaminado a determinar si el producto en cuestión es una comida, bebida, o complemento (*ketchup* p.e.), y a añadir (en su caso) dicho producto a su base de conocimiento de productos susceptibles de ser pedidos en el restaurante¹, de forma que la próxima vez que un cliente realice un pedido del producto mencionado, el sistema informe de la no disponibilidad del mismo en el restaurante.

El sistema debe ser capaz de proporcionar sugerencias a los clientes respecto a los pedidos que éstos realizan, y se debe comportar de forma *natural* en aquellas situaciones en las que no comprende lo dicho por el usuario. Podemos considerar al sistema como un agente o entidad autónoma con capacidad para llevar a cabo tareas específicas, capaz de comunicarse con el exterior (personas u otros agentes autónomos [González 96]) en lenguaje natural (LN).

Otra característica que se está teniendo en cuenta durante el desarrollo del sistema es la capacidad para procesar frases provenientes de conversacio-

¹ En dicha base de conocimiento figuran, por una parte, aquellos productos que están disponibles en el restaurante para el que hipotéticamente *trabajaría* el sistema, y por otra, aquéllos otros que aún no estando disponibles en el mismo, suelen solicitarse en otros restaurantes.

nes orales, por lo tanto, la interpretación semántica de las frases del usuario se realiza sin tener en cuenta los signos de puntuación (puntos, comas, etc.). En algunas ocasiones esta restricción puede implicar tratar determinadas ambigüedades (al igual que sucede en el mundo real), en tal caso, el sistema formulará las preguntas oportunas al usuario para resolverlas.

3. Bases para el desarrollo del sistema.

Las ideas básicas que se están teniendo en cuenta para desarrollar el sistema son las siguientes:

i) **Existencia de información innecesaria en las frases del usuario:** En muchos casos no todas las palabras de las frases son necesarias para obtener el mensaje (contenido semántico) de las mismas, de forma que se puede prescindir de ellas y obtener el mensaje a partir de las palabras realmente significativas (palabras clave). Precisamente, el mecanismo utilizado por el sistema para la obtención del contenido semántico de las frases del usuario se basa en un examen de las palabras clave, y en un análisis de las relaciones entre las mismas [Seto et al. 94].

ii) **Economía en el uso del lenguaje:** A partir del corpus de diálogos obtenidos puede observarse que los clientes suelen usar el menor número posible de palabras para realizar sus pedidos, y en muchos casos, suelen comunicarse con el dependiente a modo de *telegramas*.

iii) **Ambigüedad reducida:** Los clientes suelen usar frases breves, simples y concisas, siendo la ambigüedad que se encuentra en los diálogos bastante reducida.

iv) **Uso de un vocabulario de palabras muy reducido:** Debido a lo restringido de la aplicación para la cual está diseñado, el sistema sólo necesita manejar un número de palabras clave muy reducido.

v) **Existencia de un patrón común que suele utilizarse para realizar los pedidos/consultas/modificaciones:** Los pedidos/consultas/modificaciones suelen amoldarse al siguiente patrón general:

<acción> [<cantidad>**] [**<producto>**] [**<contenido>**] [**<sabor>**]
[**tamaño**] [**<complementos_producto>**]**

donde

<acción> puede ser "*pedir*", "*consultar*" o "*modificar*".

<cantidad> puede ser "*uno*", "*dos*", "*tres*", "*algunos*", "*varios*", etc.

<producto> puede ser "*bocadillo*", "*hamburguesa*", "*cerveza*", ...

<contenido> puede ser "*atún*", "*tortilla*", "*jamón*", ...

<sabor> puede ser "*naranja*", "*limón*", "*piña*", ...

<tamaño> puede ser "*pequeño*", "*mediano*", "*grande*", ...

<complementos_producto> puede ser "*sin hielo*", "*con tomate*", ...

[] significa opcional.

Utilizando dicho patrón común puede obtenerse el contenido semántico de la gran mayoría de las frases de los clientes, no obstante, si el sistema

duda en cuanto a la obtención del mismo, formula las preguntas convenientes al usuario hasta obtenerlo.

4. Características.

Podemos mencionar las siguientes características del sistema:

i) **Mantenimiento de un diálogo natural con el usuario:** En la medida de lo posible, y teniendo en cuenta lo restringido del dominio de la aplicación, el usuario debe encontrarse cómodo con el sistema y no *creer* que está dialogando con una máquina (test de Turing). Algunos ejemplos de estos sistemas los podemos encontrar en [Hatazaki et al. 94], [Zue et al. 94], [Seto et al. 94] y [Yamada et al. 94]. El sistema se está desarrollando de forma que intente siempre llevar la iniciativa durante la conversación con el usuario, sin embargo, éste es libre para decir lo que desee en cualquier momento o para tomar la iniciativa en la conversación.

ii) **Flexibilidad en la obtención de la información necesaria para el sistema:** El lenguaje es fundamentalmente una actividad colectiva. Cuando dos personas hablan entre sí, cada una realiza acciones individuales (pronunciación de frases, identificación de sonidos, formación de interpretaciones, etc.) pero muchas veces esas acciones son realmente realizadas por el par formado por ambas [Clark 94], [Hirst et al. 94]. En este sentido, en el sistema SAPLEN la información necesaria para completar los pedidos del usuario puede obtenerse a partir de una única acción de éste:

CLIENTE: "*quiero un bocadillo de atún con tomate*"

o a partir de sucesivas acciones en el diálogo, mediante la colaboración del par cliente-sistema:

CLIENTE: "*quiero un bocadillo*"

SAPLEN: "*¿de qué lo quieres?*"

CLIENTE: "*de atún con tomate*"

El sistema puede obtener la información que necesita para completar los pedidos de forma discontinua, es decir, si p.e. el sistema pregunta al usuario el *tamaño* deseado para una *cerveza*, y el éste responde realizando p.e. el pedido de un *bocadillo* para el cual no indica su *contenido* (pedido *incompleto*), el sistema cambiará de objetivo y se interesará por este último pedido (*focus shifting*, [Ferrari 91]), de forma que una vez haya completado el pedido correspondiente al *bocadillo*, volverá a interesarse por el pedido correspondiente a la *cerveza*.

iii) **Consulta masiva a las Bases de Conocimiento:** Durante su funcionamiento, el sistema consulta frecuentemente la información almacenada en las Bases de Conocimiento de las que dispone:

- Base de Conocimiento de la Interface de entrada (B.C.I.E.).
 - Diccionario Semántico (D.S.).
 - Base de Conocimiento de Reglas Sintácticas (B.C.R.Si.).
 - Base de Conocimiento de Reglas Semánticas (B.C.R.Se.).

- Base de Conocimiento de la Interface de salida (**B.C.I.S.**).
- Base de Conocimiento de Reglas para la Formación de Palabras (**B.C.R.F.P.**).
- Base de Conocimiento de Reglas para la Formación de Frases (**B.C.R.F.F.**).
- Base de Conocimiento de Productos del Restaurante (**B.C.P.R.**).

iv) **Utilización de diversos tipos de conocimiento:** El sistema usa el conocimiento **lingüístico** almacenado en forma de reglas en la B.C.R.Si. y en la B.C.R.Se. para obtener la interpretación semántica correspondiente a la entrada del usuario, y la información almacenada en forma de reglas en la B.C.I.S. para poder formar correctamente las palabras y frases. Por otra parte, utiliza el conocimiento **estratégico** almacenado en forma de reglas en el Módulo de Iniciativa (M.It.) del Módulo Central del Sistema (M.C.S.) y en el Generador de Consultas (G.Co.) del Generador de Lenguaje Natural (G.L.N) (ver Apartado 6. *Análisis Modular*). Finalmente, emplea el conocimiento **del dominio** almacenado en el Módulo Contextual (M.C.), en la B.C.R.Se., y en la B.C.P.R.

v) **Interface gráfica:** El sistema consta de una interface gráfica (a modo de ticket de compra) que permite al usuario rastrear en todo momento la información comprendida por el sistema, así como conocer la relación de productos pedidos.

vi) **Limitaciones del sistema:** De momento, el sistema es capaz de procesar las frases sencillas y simples correspondientes a las conversaciones comunes de los clientes del restaurante (la gran mayoría según el corpus de diálogos). Sólo es útil para la aplicación para la que está diseñado, sin embargo, las técnicas empleadas son generales y pueden utilizarse en otras aplicaciones.

vii) **Interacción con el usuario:** El sistema se está desarrollando de forma que acepte las frases correspondientes a los pedidos/consultas/modificaciones de los clientes a partir del teclado, no obstante, se intenta que sea capaz de comprender las frases espontáneas (habladas) de éstos. Así por ejemplo, el sistema no está diseñado para detectar los signos de puntuación en las entradas del usuario, de esta forma, una vez desarrollado por completo el sistema, podría estudiarse la posibilidad de utilizarlo como parte integrante de un sistema de pedidos en restaurante mediante voz.

5. Representación del conocimiento.

El formalismo de representación del conocimiento utilizado por el sistema es el denominado **FRAMES** (*marcos*), introducido por Minsky en 1975. Cada frame representa una clase de elementos, y se compone de una serie de *slots* (ranuras) con valores asociados y posibles restricciones sobre los mismos, de forma que cuando los slots necesarios de un frame se han rellenado, se ha instanciado un elemento de la clase [Rich E. y Knight K.

94], [Allen 95]. Cada slot de un frame puede estar vacío, relleno con un valor simple, o bien, contener a su vez otro frame.

El sistema PEGASUS, por ejemplo [Zue et al. 94], utiliza frames para representar el significado de las consultas (habladas) de los usuarios acerca de viajes en avión (p.e. horarios existentes para viajar de una ciudad a otra).

En el sistema SAPLEN las frases del usuario se han dividido en tres tipos ilocutivos (*illocutionary act types*) [Nuccetelli 89], [Nagata et al. 94]): *pedir* (p.e. "un bocadillo de atún"), *consultar* (p.e. "¿de qué tenéis los bocadillos?") y *modificar* ("... bueno, pues entonces, en lugar de uno de atún me pones dos").

Una iteración (acción) del usuario puede generar uno o varios frames. El sistema utiliza frames del tipo:

F_COMIDA_x

<acción_x>:

<cantidad_x>:

<comida_x>:

<contenido_x>:

<tamaño_x>:

<temperatura_x>:

<complementos_x>:

para cada pedido/consulta/modificación correspondiente a alguna comida, y frames del tipo:

F_BEBIDA_x

<acción_x>:

<cantidad_x>:

<bebida_x>:

<tamaño_x>:

<sabor_x>:

<alcohólica_x>:

<carbonatada_x>:

<temperatura_x>:

<complementos_x>:

para cada pedido/consulta/modificación correspondiente a alguna bebida, donde *x* hace referencia a la posición que ocupa el frame en la memoria del sistema. Como hemos comentado anteriormente, en SAPLEN una iteración del usuario puede generar uno o varios frames. Cada uno de ellos se almacena en la memoria del sistema en una posición *i*, $i \geq 0$, de forma que el primer frame que se forme se almacena en la posición 0, el segundo en la posición 1, el tercero en la posición 2, y así sucesivamente.

En este trabajo, hemos denominado slot *principal* a aquél que no puede estar vacío en el frame al que pertenece. en caso de que sí pueda estarlo, se denomina *secundario*. Diremos que un frame está *completo* si están rellenos todos sus slots principales, en caso contrario, diremos que está *incompleto*. Un pedido de comida o bebida está *completo* si todos los frames que éste genera están completos, en caso contrario, diremos que el pedido está *incompleto*.

Un frame correspondiente al pedido/consulta/modificación de alguna comida tiene como slots principales: <cantidad_x>, <comida_x>, <contenido_x>, y <tamaño_x>, los demás slots del frame se consideran secundarios. Análogamente, un frame correspondiente al pedido/consulta/modificación de alguna bebida tiene como slots principales: <cantidad_x>, <bebida_x>, <tamaño_x> y <sabor_x>, los demás slots del frame se consideran secundarios.

En este artículo llamaremos frame *i* a aquél que ocupa la posición *i* en la memoria, y frame *j* a aquél que ocupa la posición *j*. Diremos que un slot del frame *i* y el mismo slot del frame *j* son *compatibles*, cuando el slot del frame *j* tiene el mismo valor que el slot del frame *i*, o bien, está vacío. Diremos que el frame *i* *subsume* al frame *j* ($i < j$) si todos los slots de ambos frames son compatibles. Cuando un frame *i* subsume a otro frame *j*, se produce un *proceso de unificación* de ambos frames, lo que provoca el traspaso de la información almacenada en el frame *i* al frame *j*, y la anulación del frame *i*. Una vez producida la unificación, el frame *j*, generado después que el frame *i* (ya que $i < j$) consta de la suma de la información de ambos frames.

El mecanismo utilizado por SAPLEN para representar y almacenar la información proveniente de las sucesivas iteraciones del usuario se basa en un proceso continuo de unificación de frames como el que se acaba de describir: el sistema estudia la posibilidad de unificar cada nuevo frame con todos los frames anteriores a él que no estén anulados aún. Dicha tarea se realiza por un módulo localizado en el Sistema de Memoria (S.M.) denominado Unificador de Frames (U.F.).

6. Análisis modular.

El sistema se compone, en un primer nivel, de los siguientes módulos:

- Interface de Entrada (I.E.).
- Módulo Central del Sistema (M.C.S.).
- Sistema de Memoria (S.M.).
- Base de Conocimiento de Productos del Restaurante (B.C.P.R.).
- Base de Datos de Identidad del Sistema (B.D.I.S.).
- Diccionario Semántico (D.S.).
- Interface de Salida (I.S.).

6.1 Interface de Entrada (I.E.).

La I.E. recibe la interacción del usuario en LN a partir del teclado, y proporciona una interpretación semántica de la misma al M.C.S. La I.E. consta de los siguientes módulos:

- Módulo Contextual (M.C.).
- Procesador de Lenguaje Natural (P.L.N.).
 - Detector de Palabras Clave (D.P.C.).
 - Agrupador de Palabras Clave (A.P.C.).
 - Analizador Semántico (A.S.).
- Base de Conocimiento de la Interface de Entrada (B.C.I.E.).
 - Base de Conocimiento de Reglas Sintácticas (B.C.R.Si).
 - Base de Conocimiento de Reglas Semánticas (B.C.R.Se.).

El M.C. tiene como misión proporcionar al P.L.N. la información referente al contexto dentro del cual se solicitan los pedidos, se formulan las preguntas, o se realizan las modificaciones de pedidos.

El D.P.C. determina la secuencia de palabras clave correspondiente a la entrada del usuario mediante la consulta a la B.C.P.R. y al D.S. [Castellón et al. 89], en el cual se encuentran las palabras clave que se utilizan para realizar el reconocimiento de las palabras relevantes de las frases del usuario. Dichas palabras clave se encuentran agrupadas en clases, y para algunas de ellas, figura además una descripción en cuanto a género y número.

La secuencia de palabras clave obtenida por el D.P.C. se transmite al A.P.C., el cual realiza las agrupaciones de las mismas a partir de la consulta a la B.C.R.Si. En esta base de conocimiento se localiza un conjunto de reglas que intentan realizar correctamente las agrupaciones de las palabras clave, en caso de no lograrlo, el sistema solicitará la ayuda del usuario, así por ejemplo, si éste introduce: "*bocadillo uno cerveza*", existe una regla que intenta determinar si el referente de *uno* es *bocadillo* o *cerveza*, haciendo uso del cálculo de una medida de distancia entre las palabras involucradas, del género, y del número de las mismas. En este caso, *uno* se asignará a *bocadillo* por existir coincidencia en cuanto a género y número), pero si un usuario algo "despistado" introduce p.e.: "*bocadillos una batidos*" el sistema no puede localizar el referente de *una*, así que responde: "*no te he comprendido, ¿quieres un bocadillo o un batido?*".

Finalmente, el A.S. proporciona la interpretación semántica de la entrada del usuario al M.C.S., para lo cual utiliza dichas agrupaciones de palabras clave, y la información almacenada tanto en el D.S. como en la B.C.R.Se. En esta base de conocimiento se encuentra almacenado un conjunto de reglas que permiten al sistema detectar incorrecciones semánticas existentes en las frases del usuario.

La gramática utilizada por SAPLEN es una gramática semántica [Rich E. y Knight K. 94]. En una de gramática de este tipo, existe generalmente una acción semántica asociada a cada una de las reglas, de forma que el

resultado del análisis y de la aplicación de todas las acciones semánticas asociadas es el significado de la frase. Este fuerte emparejamiento entre las acciones semánticas y las reglas de la gramática funciona porque éstas están diseñadas teniendo en cuenta conceptos semánticos clave.

De entre las principales ventajas de este enfoque podemos citar que, por una parte, pueden evitarse muchas ambigüedades que surgirían durante un análisis estrictamente sintáctico, ya que algunas de las interpretaciones no tendrían sentido semánticamente, y por otra, pueden ignorarse aspectos sintácticos que no afectan a la semántica (p.e. algunas discordancias de género - "*quiero una bocadillo*" p.e. -).

Las principales desventajas que presenta son dos: en primer lugar, el número de reglas necesarias puede ser muy grande, ya que se pueden perder muchas generalizaciones sintácticas, y como consecuencia, el proceso de análisis puede ser muy costoso.

A partir de los experimentos realizados con gramáticas semánticas en varios dominios, se ha llegado a la conclusión de que pueden ser muy útiles para producir interfaces en lenguaje natural para un dominio restringido. Como solución global al problema de comprensión del lenguaje natural no son aconsejables, debido a su inoperancia a la hora de capturar generalizaciones lingüísticas importantes.

6.2 Módulo Central del Sistema (M.C.S.).

El M.C.S. recibe de la I.E. la interpretación semántica correspondiente a la iteración del usuario y proporciona a la I.S. el nuevo estado en que se encuentra el sistema. Además, proporciona al S.M. la nueva información comprendida por éste, y eventualmente, actualiza la B.C.P.R. a partir del aprendizaje de nuevos productos mediante la conversación con los clientes. Para poder realizar estas funciones, el M.C.S. recibe además las siguientes informaciones:

- De la B.C.P.R. recibe la información correspondiente a los productos del restaurante.
- De la B.D.I.S. recibe la información correspondiente al sistema en sí; y
- Del S.M. obtiene la historia del diálogo con el usuario.

Dentro del M.C.S. podemos distinguir los siguientes módulos:

- Módulo de Saludos del Usuario (M.Sa.).
- Módulo de Iniciativa (M.It.).
- Módulo de Incomprensión (M.Ip.).
- Módulo de Aprendizaje (M.Ap.).
- Módulo de Pedidos (M.Pe.).
- Módulo de Consultas del Usuario (M.Co.).
- Módulo de Modificaciones del Usuario (M.Mo.).
- Módulo de Identificación del Sistema (M.Id.).

Todos estos módulos reciben la interpretación semántica proporcionada por la I.E., y la historia del diálogo proporcionado por el S.M. Cada uno de ellos se activará (o no) en función de la interpretación semántica que recibe (p.e. el M.Sa. sólo se activa si detecta la existencia de un saludo del usuario). Cuando un módulo se activa envía al S.M. la fracción de la interpretación semántica que lo activó; de esta forma se posibilita aislar los fragmentos de la interpretación semántica que no han sido comprendidos por el sistema. Esta es la idea del mecanismo de **aislamiento de incomprensiones** utilizado por SAPLEN, según el cual, p.e. la entrada del usuario: "*bocadillos una batidos Y UNA FANTA DE NARANJA*" se divide en dos partes: las palabras en minúscula representan la primera de ellas, en la que se ha producido una incomprensión; las palabras en mayúscula representan la segunda parte, que se corresponde con un fragmento comprendido por el sistema. El M.Ip. del M.C.S. detectará dicha incomprensión y enviará una orden al G.L.N. para que el sistema informe de ella al usuario.

El M.It. recibe la información proveniente de la B.C.P.R. para determinar las preguntas que se deben formular al usuario para completar los pedidos incompletos.

El M.Ap. recibe la información proveniente de la B.C.P.R. para determinar las preguntas que se deben formular en el caso de determinar la existencia de nuevos productos (información que añadirá a la mencionada base de conocimiento).

El M.Co. también accede a la B.C.P.R. para determinar el estado que transmitirá a la I.S., p.e., si el usuario formula una pregunta acerca de una bebida no disponible en el restaurante, generará el estado correspondiente a "*INFORMAR BEBIDA NO DISPONIBLE*".

6.3 Sistema de Memoria (S.M.).

El S.M. recibe del M.C.S. la nueva información comprendida por el sistema y proporciona a éste la historia del diálogo. Consta de los siguientes módulos:

- Memoria del Sistema.
- Creador de Frames (C.F.).
- Unificador de Frames (U.F.).
- Modificador de Slots (Mo.Sl.).

El C.F. se encarga de recibir del M.C.S. la nueva información comprendida por el sistema y de crear en la memoria del sistema el número necesario de frames para almacenarla.

El U.F. examina la Memoria del Sistema tras cada iteración del usuario, para realizar, en caso de ser necesario, el proceso de unificación comentado anteriormente.

El Mo.SI. recibe del M.C.S. la nueva información comprendida por el sistema y se encarga de modificar los slots correspondientes de los frames afectados por las modificaciones realizadas por el usuario.

6.4 Base de Conocimiento de Productos del Restaurante (B.C.P.R.).

Como se ha comentado anteriormente, en esta base de conocimiento se encuentra la relación de productos susceptibles de ser pedidos en el restaurante. Inicialmente, sólo consta de los productos existentes en el mismo y de algunos susceptibles de ser pedidos en cualquier restaurante aunque no estén disponibles, pero se permite la adición de nueva información. Las consultas a esta base de conocimiento pueden ser de dos tipos: por una parte, se produce una consulta **explícita** cada vez que el usuario realiza una consulta acerca de algún producto, y por otra parte, se realiza una consulta **implícita** (por iniciativa del sistema) para comprobar si el restaurante dispone de los productos solicitados (p.e. si el usuario realiza el pedido: "*una cerveza grande*" y el restaurante no dispone supuestamente de cervezas de tal tamaño, el sistema determinará la no disponibilidad a partir de la inexistencia en esta base de conocimiento del par $\langle \textit{cerveza, grande} \rangle$).

En esta base de conocimiento se almacena además cuales son los *complementos* habituales para cada comida y bebida. También se almacena el género (masculino o femenino) de los productos, de ahí que el G.L.N acceda a ella a la hora de determinar el artículo que debe acompañar a cada uno de aquéllos. Todos los productos disponibles en el restaurante se encuentran almacenados mediante instancias de las clases **COMIDAS** y **BEBIDAS** definidas como sigue:

COMIDAS

género:
comida:
relación_comida:
contenido:
relación_contenido:
temperatura:
complementos:

BEBIDAS

género:
bebida:
tamaño:
sabor:
temperatura:
complementos:

6.5 Base de Datos de Identidad del Sistema (B.D.I.S.)

En esta base de datos, aún por desarrollar, se encontrará almacenada la información sobre el sistema en sí, que permitirá informar a quienes se pongan en contacto con él (personas u otros sistemas de procesamiento de lenguaje natural) acerca de su identidad, finalidad, posibilidades, etc.

6.6 Diccionario Semántico (D.S.)

Como se ha comentado anteriormente, en el D.S. se encuentran las palabras clave que se utilizan para realizar el reconocimiento de las palabras relevantes de las frases del usuario. Dichas palabras clave se encuentran agrupadas en clases, y para algunas de ellas, figura además una descripción en cuanto a género y número.

6.7 Interface de Salida (I.S.)

A partir del nuevo estado generado por el M.C.S., la I.S. envía a la pantalla la nueva iteración en LN del sistema. Consta de los siguientes módulos:

- Generador de Lenguaje Natural (G.L.N.).
 - Generador de Informes (G.If.).
 - Generador de Consultas (G.Co).
 - Generador de Saludos (G.Sa).
 - Generador del Ticket de Pedidos (G.T.P.).
- Base de Conocimiento de la Interface de Salida (B.C.I.S.).
- Base de Conocimiento de Reglas para la Formación de Palabras (B.C.R.F.P.).
- Base de Conocimiento de Reglas para la Formación de Frases (B.C.R.F.F.).
- Controlador de Referencias (C.R.).

El G.L.N. se encarga de generar la próxima salida en LN del sistema a partir del nuevo estado en que éste se encuentra (proporcionado por el M.C.S.) y de la información proporcionada por el Controlador de Referencias. La demás información que necesita se encuentra almacenada en la B.C.I.S. y en la B.C.P.R. La generación en LN se realiza en dos fases: en la primera el sistema determina qué se debe decir (**generación profunda**) y en la segunda determina cómo se debe decirlo (**generación superficial**). La generación superficial utiliza como entrada el resultado de la generación profunda, al que aplica una serie de reglas para frases de aspecto *natural* y sintácticamente correctas [Iglesias 93].

El G.If. se encarga de formar los informes que el sistema proporciona a modo de respuestas a las consultas del usuario, y de proporcionarle a éste cualquier otra información respecto a los pedidos que realiza (p.e. la no

disponibilidad de algún producto, los diferentes tamaños disponibles para una determinada bebida, etc.).

La misión del G.Co. consiste en formar las preguntas que el sistema debe formular al usuario, encaminadas p.e. a determinar el tamaño y/o sabor, en su caso, de las bebidas pedidas por aquél.

El G.Sa. se encarga de generar los saludos (de encuentro y de despedida) que el sistema proporciona al usuario, en función de la hora en que se produce la conversación.

La misión del G.T.P. consiste en mostrar en la esquina superior derecha de la pantalla la lista abreviada (a modo de ticket de compra) correspondiente a los pedidos realizados por el cliente durante la conversación.

Tanto el G.If. como el G.Co. necesitan acceder a la B.C.R.F.P. y a la B.C.R.F.F. En la B.C.R.F.P. se encuentran las reglas necesarias para realizar la flexión adecuada de las palabras en cuanto a género y número, y en la B.C.R.F.F. se encuentran las reglas necesarias para que el sistema sea capaz de formar correctamente la frases. Ambos Generadores deben acceder además a la B.C.P.R. para conocer el género de los productos acerca de los cuales van a informar o realizar alguna consulta, pudiendo así determinar correctamente el artículo (*el, la, lo, los, las*) que debe preceder a cada uno de ellos. El G.Sa. no necesita consultar dichas bases de conocimiento, ya que los saludos que genera están predefinidos, seleccionándose uno, de entre los posibles, en función de la hora en que se realiza la conversación. El G.T.P. sólo necesita consultar la B.C.R.F.P., ya que no tiene que formar frases.

El G.Co. recibe además la información proporcionada por el C.R., el cual tiene como finalidad determinar la forma en que se formulan las preguntas al usuario, basándose para ello en el cálculo de una medida de **distancia temporal** al referente correspondiente. De esta forma, si p.e. un cliente realiza el pedido: "*un bocadillo*" (sin indicar de *qué* lo quiere), se crea un frame incompleto que será detectado por el M.It., el cual generará el estado oportuno para que el G.Co. formule la pregunta correspondiente a "**FALTA CONTENIDO**". Como por ahora la referencia a *bocadillo* es muy reciente en el tiempo, el G.Co. se limita a ordenar al G.If. que enumere las diversas alternativas entre las cuales el cliente debe seleccionar el *contenido* deseado, pero si aquél no realiza la elección, y en su lugar hace otro pedido/consulta/modificación, el sistema ejecutará un **cambio de foco** (*focus shifting*) para atender el nuevo pedido/consulta/modificación. Una vez atendido éste, el sistema volverá a intentar completar el pedido que aún tiene pendiente (correspondiente al *bocadillo*), pero ahora la distancia temporal ha aumentado, así que el G.Co. formula la pregunta correspondiente a "**FALTA CONTENIDO**" haciendo uso del referente (*bocadillo*), y a continuación, llama de nuevo al G.If. como en el caso anterior.

7. Estado actual de desarrollo del sistema.

El sistema se está desarrollando en lenguaje C, y por el momento consta de 30.000 líneas de código aproximadamente.

La I.E. está bastante desarrollada, pero aún es necesario incluir más reglas sintácticas y semánticas en la B.C.I.E.

Por el momento, el sistema sólo reconoce pedidos y consultas, ya que aún no se ha creado ni el Módulo de Modificaciones (M.Mo.) del M.C.S., ni el Modificador de Slots (Mo.Sl.) del Sistema de Memoria (S.M).

En la B.C.P.R. sólo están almacenados algunos productos que permiten comprobar el funcionamiento del sistema durante su desarrollo, así que es necesario completarla con los productos disponibles en el restaurante real. Gracias a que en esta base de conocimiento se almacena la información correspondiente a las relaciones existentes entre los diversos productos, la adición/eliminación/modificación de productos no requiere modificaciones en el resto del sistema.

La B.D.I.S. no se creado aún.

El D.S. consta, de momento, de algunas palabras imprescindibles para poder comprender los pedidos y consultas de los clientes, pero es conveniente ampliarlo, p.e. con las palabras que describen objetos que suelen existir en el restaurante (*mesa, silla, cocina, etc.*).

El desarrollo del S.M. se encuentra bastante avanzado en lo que se refiere a creación y unificación de frames, sin embargo, debido a que no se ha iniciado aún la etapa de *modificación*, no se ha creado el Mo.Sl.

El M.C.S. se encuentra en un estado de desarrollo intermedio: los Módulos de Saludos del usuario, Iniciativa, Incomprensión, Pedidos y Consultas del Usuario se encuentran bastante avanzados, no obstante, todavía no se han iniciado los Módulos de Aprendizaje (M.Ap.), Modificaciones del Usuario (M.Mo.) e Identificación del sistema (M.Id.).

En general, el sistema consta ya de una sensación aceptable de *naturalidad* en los informes y preguntas que genera, no obstante, es necesario realizar diversas mejoras.

8. Aplicación práctica.

En nuestra opinión, la integración del sistema en un sistema de pedidos mediante voz constituye su aplicación más interesante. Para ello, sería necesario disponer de una interface voz-texto que proporcione la secuencia de palabras pronunciadas por el usuario, y por otra, de una interface que realice el proceso contrario, es decir, la transformación de las secuencias de palabras a voz. En teoría, el sistema se podría instalar en una máquina colocada en el restaurante, ante la cual, los clientes realizasen sus pedidos. También se podría instalar, teóricamente, en un sistema de pedidos mediante voz vía teléfono, lo que permitiría que los clientes formularan sus pedidos en lenguaje natural remotamente.

9. Conclusiones.

En este trabajo hemos presentado las bases que seguimos para desarrollar el sistema, las características principales del mismo, y los objetivos generales que intentamos conseguir.

A partir del corpus de diálogos etiquetados, hemos obtenido el conjunto de palabras clave necesario para realizar la comprensión de las frases, el comportamiento que se desea adopte el sistema, y la estrategia que éste debe seguir.

Como mecanismo de representación del conocimiento hemos elegido los frames, ya que, por una parte, en la bibliografía puede encontrarse un gran número de artículos, libros, y trabajos de investigación relacionados con esta metodología, y por otra, se adaptan muy bien a la naturaleza del problema que intentamos resolver.

Hemos definido qué entendemos por slots principales y secundarios, slots compatibles, frames completos e incompletos, y finalmente, hemos descrito los mecanismos de subsumción y unificación seguidos por nuestro sistema.

Hemos presentado un análisis modular con el que hemos intentado mostrar la estructura general del sistema, así como la comunicación que se establece entre sus diversos componentes.

Hemos expuesto un mecanismo de aislamiento de incomprensiones, que permite separar, en determinadas ocasiones, las partes de la frase que no han sido comprendidas de aquéllas otras que sí lo han sido.

Hemos definido el mecanismo de tratamiento de referencias usado por nuestro sistema (*distancia temporal al referente*), y hemos descrito la implementación que hemos realizado de la teoría de cambio de foco (*focus shifting*).

Finalmente, hemos mencionado dos posibles aplicaciones prácticas del sistema descrito.

10. Trabajo futuro.

En primer lugar, hemos de iniciar el desarrollo de los Módulos de Modificación, Aprendizaje e Identificación del sistema.

Para lograr una mayor sensación de naturalidad por parte del sistema, se requiere dotar de mayor potencia al G.L.N., de forma que tenga mayores posibilidades de expresión.

Una vez desarrollado el sistema, debemos comprobar el funcionamiento del mismo a partir del corpus de diálogos, y a continuación en el mundo real, es decir, en el restaurante.

Finalmente, sería conveniente mejorar la interface gráfica de la que dispone el sistema, de forma que permita la representación de algunos gestos de la cara del hipotético dependiente del restaurante (que representen p.e. la incomprensión de la frase del usuario), así como la incorporación de gráficos

correspondientes a los productos que el cliente está pidiendo en cada momento. De esta forma, lograríamos que el sistema resultara más amigable para los clientes.

11. Bibliografía.

[Allen 95] J. Allen. "Natural language understanding". The Benjamin/Cummings Publishing Company Inc., 1995

[Castellón et al. 94] Irene Castellón, Alicia Manzanera, Antonia Martí. "La interpretación semántica en un sistema de diálogo". Boletín SEPLN nº 8, diciembre 1989.

[Clark 94] H. H. Clark. "Managing problems in speaking". Speech Communication 15 (1994) 377-378

[Díaz J. y Rodríguez P.] Julia Díaz García y Pilar Rodríguez Martín. "Prógenes: La interfaz en lenguaje natural". Boletín SEPLN nº 11, diciembre 1991.

[Ferrari 91] Ferrari, Giacomo. "Towards a realistic dialogue model". Boletín SEPLN nº 11, diciembre 1991.

[González 96] José Carlos González. "Interoperabilidad del software: Una perspectiva desde la Inteligencia Artificial". Actas II Jornadas de Informática, XVIII Escuela de Verano de Informática AEIA, 1996.

[Hatazaki et al. 94] K. Hatazaki, F. Ehsani, J. Noguchi, T. Watanabe. "Speech dialogue system based on simultaneous understanding". Speech Communication 15 (1994) 377-378

[Hirst et al. 94] G. Hirst, S. McRoy, P. Heeman, P. Edmonds, D. Horton. "Reparing conversational misunderstandings and non-understandings". Speech Communication, 15 (1994) 377-378

[Iglesias 93] Iglesias, Carlos A. "Introducción a la generación de lenguaje natural" Revista Informática y Automática, vol. 26-2/1993.

[Nagata et al. 94] M. Nagata, T. Morimoto. "First steps Towards statistical modelling of dialogue to predict the speech act type of the next utterance". Speech Communication 15 (1994) 377-378.

[Nuccetelli 89] Susana Nuccetelli. "Reconocimiento de fuerzas ilocutivas en la interacción Hombre-Máquina". Boletín SEPLN nº 8, diciembre 1989.

[Rich E. y Knight K. 94] "Inteligencia Artificial, Segunda edición". Mc-Graw Hill, 1994.

[Seto et al. 94] S. Seto, H. Kanazawa, H. Shichi, Y. Takebayashi "Spontaneous Speech dialogue system TOSBURG II and its evaluation". Speech Communication 15 (1994) 377-378

[Yamada et al. 94] M. Yamada, F. Itoh, K. Sakai, Y. Komori, Y. Ohora, M. Fujita. "A spoken dialogue system with active/non-active word control for CD-ROM information retrieval". Speech Communication 15 (1994) 377-378

[Zue et al. 94] V. Zue, S. Seneff, J. Polifroni, M. Phillips, C. Pao, D. Goodine, D. Goddeau, J. Glass. "PEGASUS: A spoken dialogue Interface for on-line air travel planning". *Speech Communication* 15 (1994) 377-378