



Tecnología Natural Language Aproximación

White Paper

Febrero 2007

ÍNDICE

1. El lenguaje natural como punto de conexión entre el usuario y la tecnología	3
2. Natural Language Goals : Claves para una nueva generación de tecnología	5
3. Natural Language Architecture : Desde la Tecnología hasta la Solución	7
4. Natural Language Dialogues : Interacciones naturales con un sistema informático	9
5. Natural Language Links : Diálogos y enlaces	10
6. Natural Language Contact Center : Un completo sistema de estadísticas que se introduce en la Lógica del Negocio	12

1. El lenguaje natural como punto de conexión entre el usuario y la tecnología

La historia de la tecnología en general y de la informática en particular se caracteriza por una tendencia progresiva hacia la simplificación en la utilización de los sistemas y componentes informáticos: desde la propia evolución de los sistemas operativos y los lenguajes de programación hasta los interfaces de usuario, pasando por el uso de terminales telefónicos, dispositivos móviles, electrodomésticos, automóviles, etc.

En todos estos ámbitos se han diseñado e implantado múltiples estrategias para facilitar la relación entre los sistemas y sus usuarios. Durante mucho tiempo, la usabilidad de estos entornos ha sido prácticamente ignorada.

Ahora bien, el propio lenguaje natural, el mismo lenguaje que utilizamos habitualmente, constituye el modo más cómodo, más ágil y más general como método de comunicación.

Así pues, la utilización del lenguaje natural ha constituido durante mucho tiempo un objetivo central como interfaz de comunicación. La necesidad de contar con una tecnología capaz de manipular directamente el lenguaje natural es especialmente relevante en ámbitos como Internet o los centros de contactación.

Pero además de un objetivo general de la tecnología, el lenguaje natural se ha convertido en uno de los mayores retos científicos en la historia de la informática. Esta complejidad se debe al hecho de que en la manipulación del lenguaje natural confluyen múltiples tecnologías, entre las que destacan el reconocimiento de voz, la comprensión de lenguaje natural, la gestión de diálogo, la representación y manipulación de bases de conocimiento y la síntesis de habla.

Durante los últimos años, y al amparo de grandes proyectos de investigación y desarrollo promovidos tanto desde el sector público como privado, se han obtenido ciertos avances en cada una de estas áreas por separado. Así por ejemplo, la síntesis de habla ha alcanzado un nivel muy alto. Los sistemas de reconocimiento de voz han pasado desde una primera generación de reconocimiento de palabras aisladas y para vocabularios muy reducidos, hasta la generación actual con sistemas capaces de reconocer habla conectada, usando grandes vocabularios (del orden de miles a decenas de miles de palabras) y con usuarios no entrenados. Esto ha permitido la utilización progresiva del reconocimiento de voz en el ámbito de la telefonía, permitiendo la aparición de los portales de voz. Otras tecnologías relacionadas con el lenguaje natural, tales como la extracción de información despertaron más interés que resultados han conseguido cumplir hasta el momento.

Auspiciadas por el éxito de la síntesis de habla y el reconocimiento de voz, han surgido estrategias que pretenden vender capacidades de comprensión del lenguaje natural y gestión del diálogo. Varios lenguajes y estándares han surgido a este nivel, entre los que cabe mencionar VoiceXML, CCXML, CallXML o SALT. No obstante, es necesario insistir en que no se trata en ningún caso ni de sistemas de comprensión de lenguaje natural ni de gestión de diálogo. Estos entornos permiten obtener una representación totalmente simplificada y dirigida por palabras clave (“key-word spotting”) de la frase reconocida, y permiten llevar a cabo transiciones entre posiciones de una red de estados finitos, simulando “diálogos dirigidos” o “diálogos basados en menús”. Por tanto, estos sistemas se deben caracterizar como extensiones funcionales al propio reconocimiento de voz, y en ningún caso como soluciones de comprensión del lenguaje natural o gestión del diálogo.

El enfoque de Natural Language ha ido dirigido hacia una tecnología que aborde un tratamiento completo y coherente de la comprensión del lenguaje natural y la gestión del diálogo. Siguiendo un símil psicológico y neurológico, muchos de los sistemas actuales reducen la capacidad lingüística a una respuesta refleja y automática dirigida únicamente por las palabras clave reconocidas. La tecnología Natural Language coloca en una posición central la capacidad intelectual de comprensión del lenguaje natural así como el control, intervención y colaboración en el diálogo.

2. Natural Language Goals: Claves para una nueva generación de tecnología

¿Por qué es tan complejo el lenguaje natural a la hora de abordar su manipulación mediante sistemas informáticos? Independientemente de la extrema complejidad neurológica que se encuentra en la base de nuestra facultad de comprensión del lenguaje, no deja de ser sorprendente la aparente facilidad con la que un niño de pocos años lo comienza a usar.

Esta aparente simplicidad del lenguaje no es tal, y realmente si lo analizamos con detalle veremos que el dominio del lenguaje realmente exige bastantes años de práctica y estudio, mediante los cuales podemos conseguir cierta riqueza léxica, conocer y usar correctamente las reglas de ortografía, puntuación, etc.

Por otro lado, el lenguaje natural presenta una serie de características que lo hacen especialmente complejo a la hora de ser manipulado automáticamente. Entre estas características se encuentran la ambigüedad, la flexibilidad, la conexión entre la semántica (el significado) y la pragmática (el contexto en el que usamos las expresiones), etc.

Evidentemente, abordar el tratamiento automático del lenguaje natural requiere por tanto abordar todos estos aspectos del mismo, y ese ha sido el reto y el objetivo de Natural Language. Las principales claves que cimientan esta nueva generación de tecnología son:

- + Una arquitectura modular y flexible, que organiza la funcionalidad del lenguaje en una serie de capas funcionalmente interdependientes. El modelo se ha inspirado en estructuras neuro-bio-psicológicas.
- + Teniendo en cuenta que el lenguaje es una facultad fundamentalmente simbólica, mientras que los ordenadores están orientados a procesos numéricos, ha sido necesario dotar al sistema de una capa de bajo nivel encargada de la comunicación entre estos dos mundos, así como de la aceleración de los procesos involucrados en la propia manipulación simbólica.
- + El objetivo final no es disparar una serie de eventos asociados a la aparición de una serie de palabras claves (el gran error de la generación de tecnología basada en *keyword spotting*). En lugar de ello, Natural Language se centra en la comprensión de las expresiones de lenguaje natural, llevando a cabo procesos de análisis léxico, gramatical y semántico profundo.
- + En aplicaciones informáticas con soporte de lenguaje natural, éste lenguaje es útil en la medida en que mediante el mismo podemos llevar a cabo operaciones (preguntar por

un horario, cancelar una cita, hacer una transferencia, comprar una entrada de un cine, ...). Por tanto, las capacidades de comprensión del lenguaje natural quedan englobadas en un motor superior, el módulo Logos de Natural Language, que aplica estrategias de Inteligencia Artificial para organizar la información recibida, y planificar la generación de la respuesta, también en lenguaje natural.

- + Como parte de las interacciones que realizamos en lenguaje natural el sistema requerirá acceder a sistemas de información para obtener información complementaria (qué horarios hay disponibles, ejecutar la cancelación de una cita, ...). Natural Language permite la conexión con sistemas de información externos usando distintas tecnologías (desde acceso a Sistemas de Bases de Datos hasta el uso de Servicios Web).
- + Por otro lado, el tiempo necesario para entender la frase del usuario, procesarla, y generar la respuesta correspondiente es un factor importante. La tecnología Natural Language ha conseguido llegar a la barrera del tiempo real en este campo. A título de ejemplo, en un simple portátil actual, la tecnología Natural Language nos permite responder a más de 1000 usuarios simultáneos con tiempos de respuesta medios inferiores a 1 milisegundo.
- + El núcleo de la tecnología Natural Language permite aplicarse tanto a aplicaciones basadas en lenguaje escrito (por ejemplo, vía interfaces Web) como basadas en lenguaje hablado (asumiendo las funciones de un centro de contactación), permitiendo la reutilización de los componentes (léxicos, gramáticas, modelos de diálogo, sistemas de generación, etc.) cuando una organización necesita ambos modelos de acceso.
- + El sistema de reconocimiento de voz de Natural Language no está concebido como un módulo funcionalmente separado de los motores de comprensión y gestión del diálogo. Al igual que ocurre en el cerebro, el reconocedor en tanto que oído está integrado a bajo nivel con el resto de componentes, y la comprensión final es una tarea en la que intervienen todos los módulos disponibles.
- + Versatilidad en la integración: La tecnología permite integrarse a distintos niveles con las plataformas que la organización ya dispone. En concreto, se soportan los estándares más implantados actualmente: desde APIs de funciones en librerías (para integraciones a bajo nivel) hasta el diseño del sistema de lenguaje natural y diálogo como un conjunto de servicios web (para integraciones a muy alto nivel basadas en XML), pasando por la integración usando TCP/IP.

3. Natural Language Architecture: Desde la Tecnología hasta la Solución

Aunque el núcleo de la tecnología Natural Language sigue los principios de integración y máxima conectividad, desde un punto de vista computacional se diferencia los siguientes módulos, cuyas principales funciones son:

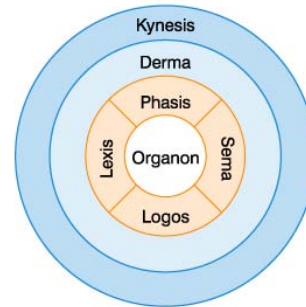


Figura 1. Arquitectura modular de la tecnología Natural Language

- + **Organon:** Este módulo constituye un soporte de bajo nivel para todo el sistema. Se encarga de enlazar las funciones de alto nivel que ofrece la tecnología, con el sistema informático mismo, realizando tareas como optimización del uso de la memoria, generación de un modelo simbólico de computación, etc.
- + **Lexis:** Este módulo concentra la funcionalidad relacionada con el nivel léxico del sistema, y aborda tareas que van desde la manipulación de diccionarios o léxicos genéricos y especializados, aplicación de modelos de ortografía y puntuación, análisis de la ambigüedad léxica, conexión con el modelo fonético para el reconocimiento de voz, etc.
- + **Phasis:** Las palabras y signos se organizan en expresiones que obedecen a reglas gramaticales. Ésta es la principal función de este módulo, que además ha de tener en cuenta fenómenos de ambigüedad gramatical, o incluso la manipulación de las probabilidades acústicas del reconocedor de voz para priorizar las interpretaciones posibles de una expresión.
- + **Sema:** El análisis de las expresiones del interlocutor deben generar una serie de modelos semánticos que representen el significado de dichas expresiones. La complejidad de esta tarea exige disponer de potentes motores para la especificación de tipos de datos muy flexibles, y contar con algoritmos muy potentes para manipular en tiempos inferiores al microsegundo dichas estructuras.
- + **Logos:** El significado o semántica de una expresión se ha de insertar en el contexto del diálogo que venimos manteniendo con el usuario, puede así mismo requerir el acceso a información externa (consultas a bases de datos o servicios web) y requerirá organizar la información que ha de transmitirse al usuario (solicitar más información al usuario para poder continuar, aclarar alguna cuestión, facilitarle los datos que estaba solicitando, etc.). Estas son las tareas que requieren un mayor grado de inteligencia en todo el sistema y de ellas se encarga el módulo Logos.

- + **Derma:** Este paquete está concebido como el interfaz para el desarrollador de nuevas aplicaciones. Como tal, por tanto, soporta los modelos de diseño necesarios para configurar todos los demás paquetes de la arquitectura Natural Language.

- + **Kinesys:** Hacer que todos los módulos anteriores actúen de forma coordinada en entornos masivos de explotación es un reto complejo. En primer lugar, se han de resolver los problemas de concurrencia masiva: en un único servidor se pueden mantener hasta 15.000 diálogos simultáneos, y evidentemente cada usuario ve su conversación como única. En segundo lugar, los sistemas basados en lenguaje natural han de integrarse con el resto de los sistemas de la organización que los necesita. Kinesys se encarga de ofrecer las distintas interfaces de integración (librería de funciones de bajo nivel, sockets TCP/IP, Servicios Web sobre XML). En tercer lugar, se han de tener en cuenta los requisitos de seguridad, robustez y fiabilidad. En este caso, Kinesys ofrece una arquitectura específica que permite la recuperación de todas las conversaciones de un servidor en otro de la red sin ningún tipo de pérdida en ninguna de las conversaciones.

4. Natural Language Dialogues: Interacciones naturales con un sistema informático

Una persona realiza una llamada telefónica al centro de atención al cliente de una compañía aérea y tras recibir el protocolario saludo, indica literalmente:

“Hola, buenas tardes. Necesitaría un billete para Barcelona, para volar mañana, saliendo desde Sevilla. Eso es para consultar los horarios y el precio Sevilla Barcelona.”

También podría haber accedido al portal de la compañía y haber escrito literalmente la misma petición. Pero podría haber solicitado exactamente la misma información de cientos de formas diferentes.



Figura 2. Entorno de diálogo

Este es uno de los múltiples fenómenos que hacen complejo el lenguaje natural. A continuación resumimos algunos de estos fenómenos.

Fenómenos relacionados con la comprensión	Organización y estructuración de la información	Diálogo inteligente y colaborativo
<ul style="list-style-type: none"> (a) Flexibilidad, naturalidad y fluidez (b) Repetición de información (c) Construcciones formales o idiomáticas (d) Elipsis, anáfora, ... (e) Corrección ortográfica (f) Modelos gramaticales profundos 	<ul style="list-style-type: none"> (a) Orden libre de los argumentos (b) Expresiones incompletas (c) Respuestas diferentes a las solicitadas por el sistema (d) Expresiones que generan múltiples comandos o tareas (e) Correcciones (tanto dentro de una frase como en frases posteriores del diálogo) (f) Dependencias semánticas no locales 	<ul style="list-style-type: none"> (a) Acomodación inicial de las expresiones incompletas a tareas por defecto (b) Reacomodación y seguimiento natural de tareas (c) Confirmaciones: implícitas y/o explícitas (d) Subdiálogos (e) Establecimiento de prioridades en comandos múltiples (f) Generación de respuestas colaborativas.

5. Natural Language Links: Diálogos y enlaces



Figura 3. Entorno de diálogo y enlaces

Estamos usando una interfaz Web para comprar unas entradas para una película. A la vez que vamos haciendo la consulta teniendo en cuenta las posibles películas que hay en cartelera y cuyo horario nos puede venir bien, el sistema nos podría mostrar los enlaces a las fichas con toda la información de cada una de las películas que vamos sugiriendo.

O bien queremos adquirir una serie de artículos a través de un catálogo de productos e iniciamos una conversación preguntando por el precio y colores de determinado artículo. A la vez que vamos haciendo las consultas el sistema nos podría ir mostrando directamente el detalle exacto del catálogo electrónico.

Un tercer ejemplo: nos conectamos al servicio de atención a clientes de una compañía aérea y simplemente indicamos "*Málaga Roma el 28*".

Desafortunadamente ese día no vuela esta compañía en esa ruta, por lo que a continuación indicamos "*y el 1 de marzo*". Además de darnos los detalles de los vuelos para ese día, el sistema nos muestra un enlace al aeropuerto Roma-Leonardo Da Vinci.

Todas estas funciones las puede realizar fácilmente la tecnología Natural Language. De hecho, distinguimos dos tipos de enlaces:

- Enlaces de movimiento directo: El sistema nos mueve directamente la página principal a un enlace. Por ejemplo, podría llevarnos al formulario de alta como nuevo cliente si detectase que no estamos registrado, o podría mostrarnos los detalles del producto que queremos comprar.
- Enlaces de movimiento sugerido: En este caso, nos sugiere una serie de links, junto con el eslogan y descripción correspondientes, y el usuario puede acceder directamente pulsando sobre el enlace.

Desde el punto de vista de la generación de los enlaces, el sistema distingue 3 tipos de enlaces:

1. **Enlaces forzados:** Esta categoría incluye los enlaces corporativos o patrocinados que queremos que el sistema sugiera de forma regular y sin atender al contexto de la conversación.
2. **Enlaces habituales:** Utilizando las estadísticas de acceso a enlaces, el sistema permite priorizar la generación de los enlaces que el resto de usuarios ha encontrado más útiles.
3. **Enlaces contextuales:** En este caso, el sistema utiliza la información del propio diálogo para seleccionar los enlaces. En el ejemplo anterior el enlace al aeropuerto se ha obtenido en función del lugar al que queremos viajar. También se puede utilizar el criterio correspondiente a la fecha del viaje, o incluso generar enlaces en función de la propia lógica del diálogo. Si un usuario pide hacer un vuelo a una ciudad en la que no opera la compañía, además de indicárselo mediante una respuesta, se le puede mostrar una página con la relación exhaustiva de destinos posibles.

Para todos estos modelos, el motor de selección de enlaces se puede configurar mediante una interfaz Web en la que se especifican los criterios de selección (número total de enlaces, porcentajes por tipos, datos concretos de los mismos enlaces, etc.)

6. Natural Language Contact Center: Un completo sistema de estadísticas que se introduce en la Lógica del Negocio

Realizar una conversación mediante esta tecnología (tanto mediante lenguaje hablado por llamada telefónica como lenguaje escrito en interfaces Web) no es una tarea que termine con la conversación en sí.

Para el usuario sí termina en ese momento, obteniendo la información que necesitaba, realizando la compra del billete correspondiente, o consultando su saldo. Pero el sistema almacena toda la información que se ha intercambiado con el usuario (salvaguardando siempre que sea necesario la identidad del interlocutor). Y esta información puede ser crucial para analizar la lógica del propio negocio.

Por ejemplo, podríamos preguntarle al sistema por el número de peticiones que recibimos para solicitar una cita en una franja horaria determinada y sin embargo tenemos que darle la cita en otra franja distinta a la solicitada o en el peor de los casos, perdemos ese cliente o usuario.

El modelo de contactación de Natural Language utiliza los siguientes criterios:

En primer lugar, un sistema queda parametrizado por los siguientes criterios:

- A) Servicios: Una aplicación puede atender varios servicios dentro de una organización, por ejemplo, "Atención a clientes", "Atención a empleados", "Quejas y reclamaciones", ...
- B) Idiomas: La organización puede ofrecer los mismos servicios en distintos idiomas.
- C) Modo de relación: Discrimina el tipo de acceso (vía telefónica, vía web, con o sin el uso de enlaces, ...)
- D) Finales transaccionales: Final correspondiente al diálogo entero. Por ejemplo, podemos detectar si el interlocutor cierra bruscamente la conversación (HangUp o interrupción), abandona por no uso (Time out debido a razones de pérdida de conexión, etc.) o finaliza correctamente la conversación.
- E) Finales operativos: Relación de funciones que se realizan durante el diálogo (consulta el saldo, realiza una transferencia y mira los movimientos del último mes, por ejemplo).
- F) Estrategias de diálogo: Posiciones estratégicas dentro del diálogo por las que pasa la conversación. Por ejemplo, durante la identificación mediante el DNI de una persona detectamos que no está registrado dicho valor en el sistema de información, etc.

En segundo lugar, el sistema registra los datos más significativos respecto a la conversación en sí (identificación de los puntos de acceso, tanto si es vía web como telefónica), la propia

conversación que se realiza, vinculando a cada frase los finales operativos y estrategias de diálogo correspondientes, y las horas de inicio y final.

Y en tercer lugar, se registran los enlaces pulsados de entre los sugeridos si estamos trabajando en modo diálogo más enlaces.

Toda esta información es registrada en un sistema de bases de datos. Desde una interfaz web es posible navegar por toda la información generada, utilizando los siguientes criterios:

- I. Diálogos: Que a su vez permite tres modos:
 - (a) Modo gráfico: Distintos tipos de gráficos (de barras, de tarta, de líneas, etc.) organizados por meses, días e incluso horas.

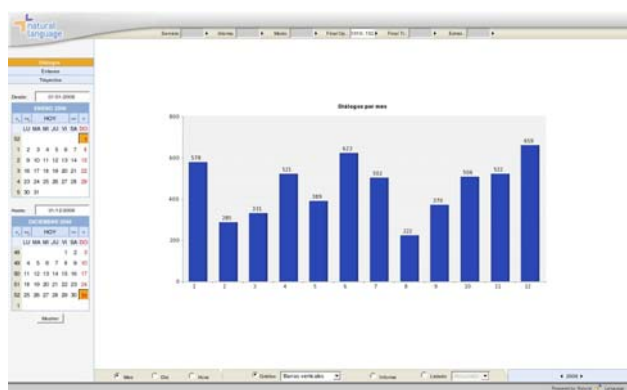


Figura 4. Gráfico de estadísticas de diálogos

- (b) Modo listado: Con acceso a cada uno de los diálogos, bien en modo resumido o completo.

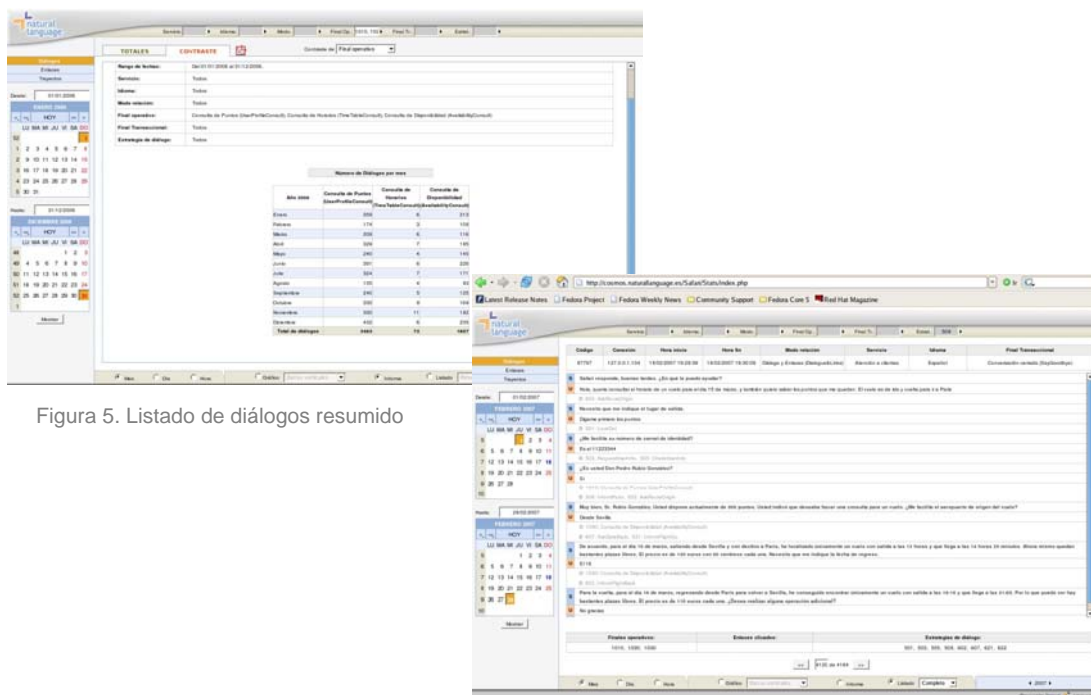


Figura 5. Listado de diálogos resumido

Figura 6. Listado de diálogos completo

- (c) Modo informe: Generando ficheros pdf con los informes tanto totales como de contraste de criterios (servicios, idiomas, modos de relación, finales transaccionales, finales operativos y estrategias de diálogo), y agrupados por fechas a nivel de años, meses o días.

- II. Enlaces: Sistema de estadísticas con la información de los enlaces más pulsados, agrupados así mismo por fechas.

- III. Lógica de negocio: Estadísticas que se introducen en la propias peculiaridades del sistema de información. Por ejemplo, para una compañía aérea, se le podría preguntar al sistema cuántas peticiones de billetes realizadas entre un periodo determinado (por ejemplo en el mes de enero de 2006) han pedido viajar en otro periodo distinto (por ejemplo, entre el 15 y el 30 de enero de 2007), en una serie de trayectos determinados (por ejemplo, en la ruta Madrid Londres), obteniendo del total de peticiones cuántas han podido atenderse y cuántas no por falta de disponibilidad.

Tecnología Natural Language

Aproximación

White Paper

Febrero 2007

Grupo Infinity

HEADQUARTERS
Parque de Negocios MAS BLAU
Edificio Muntadas
C/ Solsonés 2, esc. B
08820 - El Prat de Llobregat
Barcelona

Edificio Alfredo Mahou
Complejo AZCA
Pz. M. Gómez Moreno 2
Planta 15 B
28020 - Madrid

Edificio Corona Center
C/ Perú 49, 1ª Planta
41930 - Bormujos
Sevilla

Edificio Torre Arcos 1
Paseo de Tamarindos 400
Col. Bosque de las Lomas
05120 - Distrito Federal
México

+34 93 556 52 00