

2

MODELOS DE LENGUAJE Y TECNOLOGÍA DEL HABLA **(LANGUAGE MODELS AND SPEECH TECHNOLOGY)**

M^a Isabel Reyzábal Manso
Víctor Santiuste Bermejo
Universidad Complutense. Madrid

RESUMEN

En este artículo describimos los modelos de lenguaje que tratan de explicar cómo se procesa, qué elementos intervienen y en qué orden.

Realizaremos un rápido recorrido por las llamadas máquinas inteligentes y describiremos el modelo de lenguaje utilizado en tecnología del habla que actualmente proporciona mejores resultados.

Finalmente, tras mostrar que en el presente la investigación está orientada hacia el aspecto semántico del lenguaje, proponemos un modelo de comprensión lingüística oral.

ABSTRACT

In this paper we describe language models which aim to explain how language is processed, what elements take part in this process, and in what order.

We will take a quick look at the so-called intelligent machines and detail the language model used in speech technology that currently yields best results.

Finally, once we have shown that research is nowadays oriented to language's semantic side, we propose an oral language comprehension model.

1. INTRODUCCIÓN

Entre otras definiciones, la Enciclopedia Espasa se refiere al lenguaje como: “conjunto de sonidos articulados con que el hombre manifiesta lo que piensa o siente”. La consecuencia inmediata es la necesidad de un receptor de esos sonidos. Producción y comprensión van unidos. En este artículo nos ocuparemos de los modelos de lenguaje que tratan de explicar cómo se procesa, qué elementos intervienen y en qué orden. Finalmente propondremos un modelo de comprensión lingüística oral, además describiremos el modelo utilizado en tecnología del habla que actualmente proporciona mejores resultados.

Tratándose del lenguaje, encontramos dos posturas extremas: *genetistas y ambientalistas*, seguidores los primeros del racionalismo y los segundos del empirismo.

Los genetistas consideran que el lenguaje sigue un plan genéticamente predeterminado, inmutable, siguiendo etapas, cualitativamente igual en todos los individuos, teniendo el ambiente solamente una función desencadenante. Sin embargo, los ambientalistas dicen que el hombre únicamente está preparado para reaccionar frente a su medio.

La tesis innatista defiende que el sujeto incorpora a lo externo lo que la mente tiene, postura que consideramos correcta. Los psicolingüistas estudian la competencia comunicativa, en oposición a los conductistas que tratan de la conducta verbal.

Consideramos que el oyente recibe información, es un acto social, las consecuencias se enmarcan dentro de la Psicología Social, puesto que esa información se traduce en conocimientos, opiniones y actitudes.

La relación entre pensamiento y lenguaje ha sido estudiada por distintos autores, entre los más representativos podemos citar a Chomsky, Sapir, Piaget y Vigotsky. Considera el primero que lenguaje y pensamiento son independientes pero relacionados; por otro lado la hipótesis de Whorf-Sapir (determinismo lingüístico) dice que la cualidad del pensamiento es determinada por el lenguaje, Piaget (determinismo cognitivo) considera que el lenguaje se basa y es determinado por el pensamiento y, por último, Vigotsky considera el lenguaje como el instrumento básico del pensamiento, que a su vez es el más importante agente transformador de la realidad.

2. EL LENGUAJE

El lenguaje ha de ser estudiado desde distintas perspectivas. Como sistema lingüístico nuestras preguntas irán encaminadas a descubrir cómo funciona, conociendo la relación entre el signo (sonido en lenguaje hablado, grafismos en escritura) y el significado.

Si nos fijamos en su aspecto como conducta de comunicación, o comunicativa, debe ser estudiado en sus dos aspectos:

- Producción (un hablante y un oyente que puede comprender o no).
- Comprensión (modelo de comprensión: oral, escrita).

El lenguaje como sistema consta de distintos elementos entre los que se han de destacar los lingüísticos (signos) y los metalingüísticos (es el mismo elemento internalizado).

También es necesario hacer referencia para la interpretación lingüística a los elementos biológicos o neurales (Pinker, 1995). Defiende este autor que el lenguaje es un instinto puesto que la posesión del lenguaje es uno de los requisitos para considerarnos humanos.

Si queremos explicar cómo funciona el lenguaje debemos tener presentes las aportaciones de Wilhelm von Humboldt y Ferdinand de Saussure.

Wilhelm von Humboldt, precursor de las ideas de Chomsky, enuncia el principio: *El lenguaje hace uso infinito de medios finitos*, el beneficio en que empleamos un código para traducir combinaciones de ideas a combinaciones de palabras, según Pinker (1995), este código recibe el nombre de *Gramática Generativa* y no hay que confundirla con la gramática pedagógica o estilística

Saussure enunció el principio de la *arbitrariedad del signo*, es la relación convencional que existe entre sonido y significado. Es decir, no existe razón de necesidad para que a un significado le apliquemos un significado determinado.

El signo lingüístico está formado por un significante más un significado. Significantes son los elementos físicos perceptibles por los sentidos, en el caso de la lengua son los fonemas; significado es el equivalente a un concepto, la idea extraída de la realidad. Por ejemplo, al significante *perro* le aplicamos el significado *animal, mamífero, etc.*

Los principales elementos constitutivos de un lenguaje natural son los siguientes:

- *Léxico*, son las palabras que utilizamos para representar conceptos.
- *Reglas*, con ellas combinamos palabras, así expresamos relaciones entre conceptos. La consecuencia de que existan reglas es que el lenguaje se hace

ce extenso y además se convierte en un código autónomo respecto a las demás capacidades cognitivas.

La diferencia entre los sistemas combinatorios artificiales del tipo de sistema de encadenamiento de palabras y otro natural, como el que existe en el cerebro, es que *LAS ORACIONES NO SON CADENAS, SINO ÁRBOLES*.

En la gramática las palabras se agrupan en sintagmas del mismo modo que las hojas se unen para formar ramas, a cada sintagma se le da un nombre y los sintagmas más pequeños se pueden unir para formar otros mayores.

Las reglas en *lingüística* se representan mediante unos pocos símbolos:

- Regla que define el sintagma nominal:
SN \longrightarrow (det) N A *
- Regla que define la oración:
O \longrightarrow SN SV
- Regla que define el sintagma verbal:
SV \longrightarrow V SN

(Donde hay que leer: flecha significa: “consta de”; los paréntesis indican elementos opcionales, y, un asterisco: “tantos elementos como se quieran”).

Además necesitamos un diccionario mental que especifique a qué categoría pertenecen las palabras: Nombre (N), verbo (V), adjetivo (A), determinante (det).

Este conjunto de reglas constituye una gramática de estructura sintagmática (Chomsky, 1957, 1965). Sirve para definir cualquier oración adjuntando palabras a las ramas de un árbol invertido, el modular. Hay una superestructura invisible que mantiene unidas las palabras.

La gramática es autónoma respecto al sentido común del significado de las palabras, esto hace que podamos producir frases sin sentido y, gramaticalmente correctas.

Las ramas etiquetadas de la estructura sintagmática de una oración sirven como plan o esquema general para recordar la oración completa, esto hace que podamos manejar las dependencias incrustadas a larga distancia si.....entonces; o.....o.

O \longrightarrow o O o O
O \longrightarrow si O entonces O

Esta capacidad se denomina “RECURSIÓN”, permite incluir un ejemplar de un símbolo dentro de otro ejemplar del mismo símbolo y permite generar un núme-

ro infinito de estructuras. Hay un núcleo común que mantiene unido a cada “o” con su correspondiente “o” y a cada “si” con su correspondiente “entonces”.

La estructura sintagmática es una solución de ingeniería al problema de cómo seleccionar una trama de ideas que tenemos en la mente y codificarlas en forma de una estructura de palabras que sólo pueden salir por la boca de una en una. La estructura sintagmática es la materia de la que está hecha el lenguaje.

Los sistemas utilizados en tecnología del habla para imitar el lenguaje natural son recursivos.

El lenguaje posee unos principios. Cada categoría gramatical de palabras no se puede definir como clase de significados, sino como una clase de símbolos que se rigen según ciertas reglas formales (tal y como sucede con las piezas del ajedrez). Otro principio es que parece que todos los sintagmas de todas las lenguas del mundo tienen una misma anatomía:

- Un núcleo que da nombre al sintagma y determina su referencia.
- Unos participantes o argumentos que aparecen agrupados con el núcleo en un subsintagma (denominado N-barra o V-barra).
- Hay posibles modificadores que se sitúan fuera del N-barra o V-barra.
- Un especificador (o sujeto).

Presumimos que el lenguaje está en un módulo entre módulos. Es algo diferente y propio, *esto es la tesis de la separación*. Además consideramos que en el lenguaje se producen automatismos (capto lo que sé, si no, no proceso, esto es un principio general), igualmente afirmamos que el lenguaje es algo biológico y cultural. Esta tesis de la modularidad específica del lenguaje ha sido enunciada por J. Fodor (1970) y por S. Pinker. La teoría modular propone las siguientes pruebas de aval:

1º. El lenguaje es una capacidad distinta a otras:

- Hay inteligencia sin lenguaje (ejemplo en lesiones del área de Broca, que se produce afasia, no son sujetos retardados).
- Disfasia de desarrollo descrita por Gopnik (dificultad lingüística, pero buena capacidad intelectual).
- Lenneberg (tesis del punto crítico en la adquisición del lenguaje).

2º. Hay lenguaje, pero no inteligencia.

- Alzheimer, autismo, afasias, espina bífida, esquizofrenia.

3º. El lenguaje es un órgano de extrema perfección y complicación.

- El lenguaje consta de la gramática que se organiza en submódulos.
- Bajo el lenguaje hay un complejo circuito neural para lo que no importa, o es independiente el tamaño del cerebro y en el que hay diferencia en la localización.
- Hay complejidad genética en el lenguaje.

3. EL PROCESO DE COMPRENSIÓN

Llamamos comprensión a la facultad, capacidad o perspicacia para entender y penetrar en las cosas.

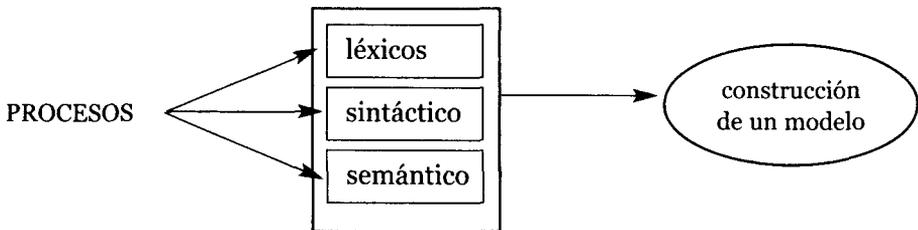
En el proceso de comprensión lingüística, hemos de tener en cuenta tres momentos:

percepción → análisis → uso

El análisis comprende la asignación de significado a las cadenas de entrada, además, el proceso de comprensión se produce como resultado de la interacción entre la información contenida en el texto (escrito o hablado) y los conocimientos previos del lector oyente.

Y el procesamiento del lenguaje se debería a la interacción con la información semántica y la referencia al contexto. Winogrand, (1972), Marcus, (1980) consideran que el análisis de una frase permite el acceso al conocimiento general del mundo y el acceso a un modelo mental.

El esquema para el análisis sería:



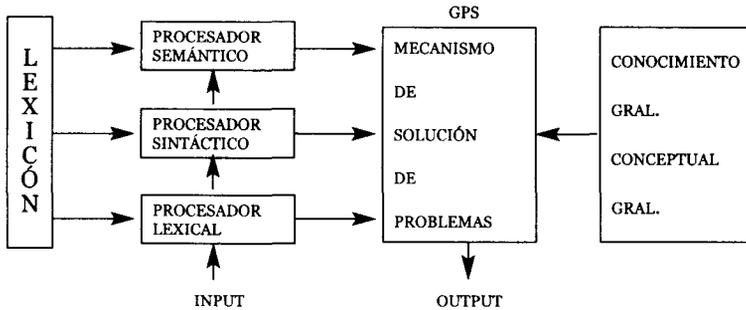
4. MODELOS

Consideramos que la base de la comprensión es la significación, la función primaria del lenguaje es la comunicación, el factor previo no es lingüístico, es comunicativo.

Hay distintos modelos de comprensión: el léxico (la palabra es la que condiciona), sintáctico (las normas de la sintaxis son las que hacen comprender) y el semántico (lo que el sujeto pone cuando comprende).

Pero, ¿cómo se procesa? Comenzamos por desmenuzar fonemas para posteriormente volver a integrarlos en un todo significativo, todo ello contando con la memoria imprescindible para el procesamiento. Los problemas son: ¿dónde poner cada elemento, cómo se produce el flujo de información entre los mismos? Hay dos principales modelos cognitivos que explican el procesamiento, es decir, todos los elementos que informan en la comprensión y cómo funciona.

- *El PRIMER modelo de Forster (1979) ha sido descrito por Santiuste (1982), se sitúa entre la teoría de la autonomía del subprocesador sintáctico y la teoría lineal interactiva:*

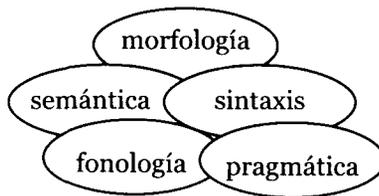


El lexicon es el procesador y el subprocesador es de abajo-arriba.

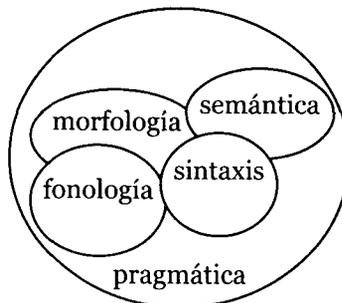
Forster, adscrito al grupo de investigación de influencia chomskiana, establece la división de subprocesadores (léxico, sintáctico y semántico) en una relación lineal, de manera que no hay procesamiento sintáctico si antes no lo ha habido léxico. Son, por lo tanto, autónomos. El procesamiento sería:

Léxico → reglas → significado

- Por el contrario se contempla en los trabajos de L. Tyler un *SEGUNDO* modelo que considera que el procesamiento es interactivo y en paralelo (formalista y funcionalista), estableciéndose una relación continua y paralela entre los componentes lingüísticos, los principales modelos son:
 1. Modelo *formalista*. Considera la pragmática como un componente más del lenguaje.



2. Modelo *funcionalista*. Considera la pragmática como organizador general total del lenguaje.

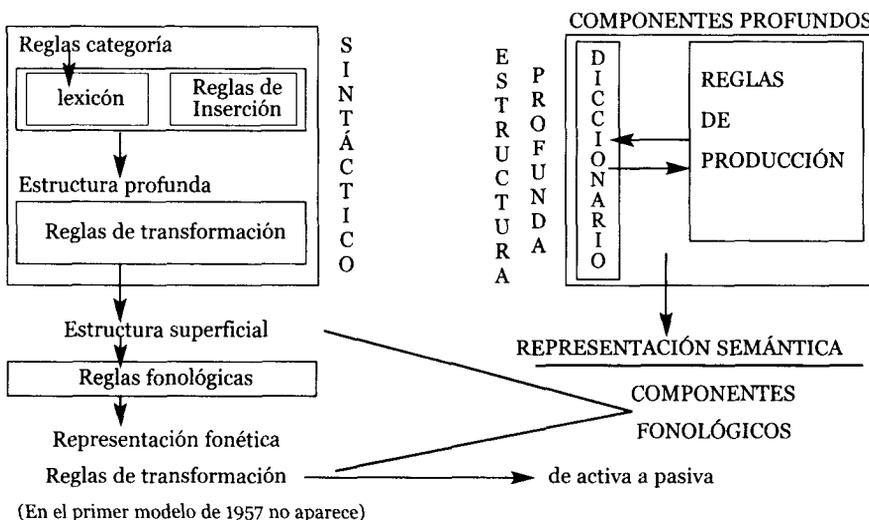


5. MODELOS EN EL DESARROLLO DE LA OBRA DE CHOMSKY

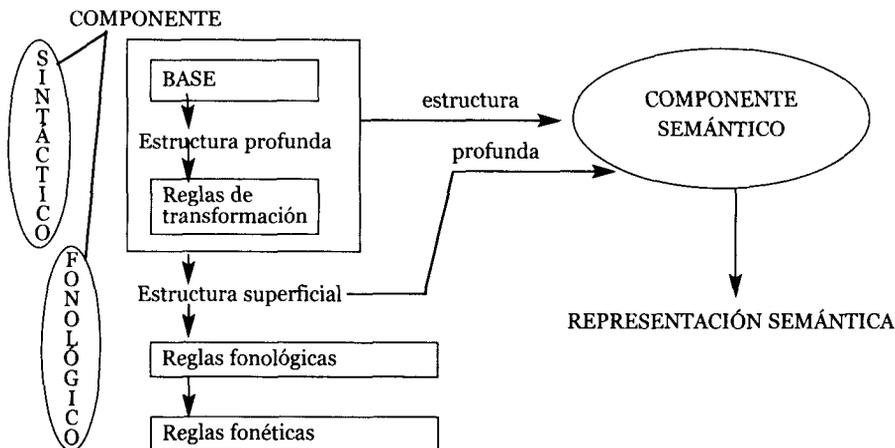
La gramática chomskiana pasa por tres momentos, en un cuarto defiende planteamientos minimalistas.

En 1957 la lingüística gira en torno a la sintaxis, no aparece la semántica, sí el léxico, la fonología y la sintaxis. En un segundo, “Aspectos de la teoría de la sintaxis”, incorpora la semántica, en el tercero, (1972), estudia semántica y gramática. Por último, en 1995, se sitúa en planteamientos minimalistas, esta información no es traducida al español hasta el 1999.

- **EL MODELO DE ASPECTS OF THE THEORY OF SYNTAX (1965)**
 Aparecen tres componentes: fonológico, sintáctico, semántico.



- **EL “MODELO STANDARD” (1972)**



Los componentes siguen siendo los mismos, modelo más simplificado, no hay tantos subcomponentes.

■ *LOS PLANTEAMIENTOS MINIMALISTAS (1995)*

Chomsky reflexiona y afirma que parece razonable pensar que el lenguaje puede aproximarse a ser un “sistema perfecto”; se hace dos preguntas:

1. ¿Cuáles son las condiciones generales que esperamos que satisfaga la facultad del lenguaje humano?
2. ¿Hasta qué extremo está determinada la facultad del lenguaje por estas condiciones sin que les subyazca ninguna otra estructura?

Si la respuesta a la segunda pregunta es positiva el lenguaje sería un “sistema perfecto”, satisface restricciones externas de la mejor forma posible y de un modo razonable.

Chomsky propone el PROGRAMA MINIMALISTA, éstos son sus puntos fundamentales:

- Hay un componente de la mente/cerebro humana dedicado al lenguaje (la facultad del lenguaje) que interactúa con otros sistemas.
- La facultad del lenguaje tiene, al menos, dos componentes:
 - ▶ Un sistema cognitivo que almacena información
 - ▶ Un sistema cognitivo de actuación que accede a esa información y la utiliza (pueden ser específicas del lenguaje).

COMPONENTES DEL LENGUAJE



El sistema cognitivo interactúa con los sistemas de actuación por medio de niveles de representación lingüística, hay dos sistemas “externos”: el sistema articulatorio-perceptual A-P y el sistema conceptual-intensional C-I. Cada uno con dos niveles de interfaz: la forma fonética FF (en A-P) y la forma lógica FL (en C-I). (El “doble interfaz” sonido con significado es la forma tradicional de describir el lenguaje, puede rastrearse al menos hasta Aristóteles).

Igualmente, distingue entre el lenguaje como *SISTEMA CONCEPTUAL* y el lenguaje del *SISTEMA DE COMPETENCIA PRAGMÁTICA*. (Pueden estar dañados selectivamente y disociados en el desarrollo, sus propiedades son diferentes).

6. OTROS MODELOS

Modelo de comprensión de textos de Kintsch y van Dijk considera que el texto está organizado en tres niveles: microestructura, macroestructura y superestructura; el primer nivel se encargaría de los significados locales y las relaciones lineales; el segundo tendría por cometido el significado global y la utilización de macroreglas, se dejaría para el tercer nivel el significado global y utilizaría macrorreglas. Consideran que la representación es textual y situacional.

Modelo "Garden-Path" (Frazier, 1978). El procesador del lenguaje es un sistema contenedor, posee varios componentes autónomos y cada uno realiza un nivel de análisis.

- Atendiendo a los modelos en los cuales la *INFORMACIÓN SINTÁCTICA* es prioritaria encontramos:

1. *Subsistemas independientes* (Cifton y Ferreiras, 1987).

Consta de un procesador sintáctico y un procesador temático, el primero se encargaría de las estructuras constituyentes (construye representaciones de las estructuras sintagmáticas), el segundo se encarga de las estructuras temáticas (selecciona la más plausible semántica y/o pragmáticamente).

2. *Modelo Serial de Guía Léxica* (Ford, Bresnan y Kaplan, 1982).

La información léxica determina el rango de estructuras potenciales de una frase y el orden en el cual se consideran las alternativas del análisis. La Gramática Funcional Léxica es su gramática de referencia.

- Entre los modelos guiados por el *CONTENIDO* encontramos:

1. *Modelo de procesamiento interactivo y paralelo: versión "fuerte"* (Fodor, Garret y Bever, 1968; Fodor, Bever y Garret, 1974). El analizador recupera la información léxica, ésta controla todas las hipótesis que se construyen en la estructura sintáctica de la frase. La información temática puede determinar las decisiones iniciales del análisis de las frases a la hora de resolver una ambigüedad.

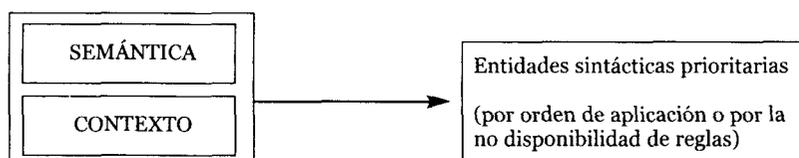
2. *Modelo de efectos contextuales: modelo interactivo "débil"* (Crain y Steedman, 1985). La ambigüedad local en el procesamiento del lenguaje se debe a la interacción entre la información semántica y la referencia al contexto. La Gramática Categorial o Combinatoria, es la que subyace a la teoría contextual.

Un esquema de las hipótesis interactivas sería:

“DÉBIL”



“FUERTE”



A la vista de los anteriores planteamientos nos centraremos en la convergencia entre autores para proseguir en la búsqueda de soluciones en aquellos aspectos en que discrepan.

Existe una opinión generalizada en considerar que la memoria resulta imprescindible para un buen procesamiento, entre los elementos implicados todos los autores coinciden en considerar la sintaxis, semántica y pragmática como componentes imprescindibles.

La colocación de cada elemento y cómo se produce el flujo de información es lo que debe aclararse, aunque los últimos trabajos conceden a la semántica mayor peso, como la DRT (teoría de la representación del discurso) elaborada en los años 80 y 90 por Hans Kamp y sus colaboradores y la SDRT (semántica dinámica), (Las-carides y Asher, 1999). O la semántica de los mundos posibles (Heim, 1992).

Se acepta que el flujo de información se produce de un modo interactivo, falta por determinar, si lineal o paralelo, aunque la segunda hipótesis parece la más plausible.

7. MÁQUINAS INTELIGENTES

Según definición de Norbert Wiener, cibernética es el estudio unificado del control y de la comunicación entre animales y máquinas. En los años 50 las tortugas de Grey Walter, que exhibían comportamientos sociales, o La Bestia de Hopkins que, guiada por un sónar y un ojo fotoeléctrico, era capaz de alimentarse mediante electricidad obtenida a través de un enchufe que era capaz de encontrar.

A la vez que los ordenadores analógicos y la cibernética, comienza el desarrollo de los ordenadores digitales de hoy basados en la separación de estructura y función (hardware y software, su modo de trabajar es la computación algorítmica),

comienza con la máquina de Turing en 1937. Se construye el ENIAC por el ejército americano para calcular trayectorias balísticas, bomba atómica etc. En los años 40 John von Neumann concibe una computadora basada en la lógica digital, opera ejecutando en serie, una tras otra, instrucciones que componen un algoritmo codificado en forma de programa el cual se encuentra almacenado en su memoria. El desarrollo de la electrónica es lo que ha permitido el gran auge de las actuales computadoras digitales.

En 1950 Turing y Claude Shannon diseñaron los primeros programas que permitían a un ordenador digital razonar y jugar al ajedrez. Newell, Simon y Shaw presentaron el Teórico Lógico, primer programa capaz de razonar sobre temas arbitrarios.

El término inteligencia artificial o IA lo acuña John McCarthy en los años 60, Marvin Minsky, Newel y Simon habían creado un programa que demostraba teoremas de geometría, tan sólo eran capaces de resolver los problemas para los que habían sido construidos.

Al desarrollarse la IA quedó eclipsada la cibernética y las redes neuronales, pero en 1969, Minsky y Papert mostraron las limitaciones de los perceptrones, modelo neuronal por excelencia de los años sesenta y la mayor parte de los recursos se destinaron a la IA.

Han transcurrido casi 50 años, tenemos ordenadores potentísimos y sin embargo no resultan más inteligentes. La IA culminó en los años 70 con la introducción de los **sistemas expertos**, son programas de computador en los que se codifica el conocimiento de expertos en una cierta materia en forma de reglas de decisión (diagnóstico de una enfermedad, diseño...).

En los años 80 se volvió a paradigmas de cómputos alternativos, redes neuronales, sistemas borrosos, algoritmos genéticos o a la computación evolutiva. Vuelven la ANS (Artificial Neural Systems) también conocidas como sistemas conexionistas.

La arquitectura von Neumann no resulta apropiada para sistemas de cálculo paralelo y distintos autores tratan de solventarlo; Hopfield, en el 82, introdujo nuevos puntos de vista combinando redes neuronales y modelos de vidrios de espín (spin-glass). Por último se encontró la forma de entrenar un perceptrón multicapa (Rumelhart, 1986), así se resolvían los problemas del perceptrón simple.

Los sistemas borrosos incidieron sobre la lógica digital (0 ó 1, verdadero o falso), el ser humano utiliza más criterios. Los sistemas borrosos son un tipo de lógica multivaluada como: SIENTONCES.....

Las redes neuronales artificiales emulan el hardware del cerebro y los sistemas borrosos se ocupan del software. En estos campos se utiliza el término *ABC* de la inteligencia (Marks, 1993) contemplándose sus tres facetas: *ARTIFICIAL*, *BIO-*

LÓGICA Y COMPUTACIONAL. Parece ser que las tareas que peor llevan a cabo las computadoras son las más fáciles para los organismos biológicos, los ordenadores pueden resolver problemas complejísima de aritmética, juegos de ajedrez, pero es muy difícil que posean capacidad perceptiva y movimiento. En la emulación de las facultades humanas se obtiene un alto nivel en cálculo, lógica y razonamiento, un bajo nivel en percepción, reconocimiento de patrones, asociaciones y aprendizaje.

Pretendemos emular el cerebro, para ello es preciso comparar la arquitectura de las computadoras con la que presenta el cerebro y tratar de emular sus capacidades.

La máquina de von Neumann es una máquina de procesamiento (hardware) actúa ejecutando en serie (una tras otra) una secuencia de instrucciones o programa (software) que almacena en su memoria. La máquina la componen cuatro unidades: unidad de entrada de información, unidad de salida, unidad de procesamiento (compuesta a su vez, por la unidad lógico-aritmética y la unidad de control) y memoria, es la máquina de cómputo en el sentido de Turing, hará la tarea para la que se programe y cambiando el programa de su memoria ejecutará tareas diferentes.

La mayoría de los ordenadores son máquinas manipuladoras de símbolos: la unidad de entrada suele ser el teclado o las unidades de disco, las de salida la pantalla o la impresora, la unidad de procesamiento o CPU (Central Processing Unit) es el microprocesador, la memoria puede estar distribuida entre el disco duro, cintas, CD-rom o la memoria central de semiconductor. El verdadero corazón del computador es el microprocesador, circuito electrónico que puede integrar millones de componentes electrónicos, además, entre sus características se encuentran las de ser baratos y versátiles.

El cerebro, cuyas capacidades queremos emular, no es una arquitectura von Neumann, no está formado por un microprocesador, ni constituido por unas cuantas CPUs, lo componen millones de procesadores elementales o neuronas interconectadas conformando redes de neuronas. La neurona es un pequeño procesador, sencillo, lento y poco fiable (diferente a los potentes microprocesadores). En nuestro cerebro se interconectan cien mil millones de neuronas operando en paralelo, cada neurona puede conectarse con otras 10.000 de promedio y puede procesar trabajando en paralelo de un modo impresionante.

Además la neurona no debe ser programada, aprende a partir de señales que recibe del entorno y opera con un esquema diferente al de los computadores poseen *autoorganización*, no existe un componente que gobierne el sistema como el CPU, las neuronas cuentan con la capacidad de autoorganizarse, aprendiendo del entorno y adaptándose a él emergiendo ricas propiedades de procesamiento.

El hombre trata de construir artefactos que sean capaces de imitar nuestras facultades y funciones, siendo la señal de voz y su "comprensión" (la asignación de

un significado) uno de los campos más estudiados, interesa poder comunicarse con las máquinas del modo más natural, y lo más natural es la voz.

Hay investigaciones que van más lejos en sus estudios sobre la interacción hombre-máquina, Rosaling W. Picard, del Mesia Labs Affecctif Computer, estudia esta interacción a través de sensores para que la máquina reciba señales y pueda interpretarlas.

8. TECNOLOGÍA DEL HABLA

La comunicación oral humana se basa principalmente en una gran cantidad de conocimiento previo, no sólo del lenguaje sino de cómo es el mundo, el principal problema en tecnología del habla es: *EL MODO DE TRANSMITIR ESTE CONOCIMIENTO A LA MÁQUINA*.

Ya hemos expuesto distintos modelos de lenguaje, en tecnología del habla el aspecto más relevante es el discurso, el procesamiento del lenguaje natural y el diálogo hombre-máquina, comenzaremos por el discurso.

Igualmente para que nuestros mensajes puedan ser interpretados usamos nuestros conocimientos lingüísticos sobre forma y estructura, pero cohesión y coherencia en el análisis del discurso desde una perspectiva dinámica. Cohesión es el uso de lazos y conexiones entre los diversos textos, Halliday y Hasan (1976) analizan bajo los nombres de referencia, sustitución, elipsis y relaciones léxicas.

Las formas correferenciales inducen al oyente a buscar interpretaciones fuera, (*la teoría semántica sobre referencia afirma que la relación de referencia se establece entre las expresiones de un texto y las entidades del mundo, y la de correferencia entre expresiones que aparecen entre distintas partes del texto*). Esta coherencia no es algo existente en el lenguaje, sino en los usuarios de esa lengua, *la persona da sentido a lo que escucha*.

Pragmática y análisis del discurso son disciplinas relacionadas estrechamente. Al analizar un texto o una conversación se encuentran implicados distintos fenómenos como actos de habla (centro de la teoría de Searle), presuposiciones, implicaturas y deixis; cada una ha dado origen a diferentes teorías, apareciendo distintas escuelas pragmáticas.

Se denomina "Etnometodología" al estudio de campo y análisis de la conversación. Es una disciplina que se interesa por el estudio de las interacciones y, especialmente, de las interacciones verbales cotidianas u ordinarias. A los seguidores o practicantes de esta corriente se les conoce como etnometodólogos o, simplemente, conversacionalistas.

Se distinguen dos corrientes: *Análisis del discurso y análisis de la conversa-*

ción, ambas estudian cómo se produce y cómo se comprende la coherencia y la organización secuencial en un discurso.

- *Análisis del discurso* (emplea metodología como los principios y conceptos teóricos propios de la lingüística).
- *Análisis de la conversación* (analiza grabaciones de conversaciones naturales utilizando métodos inductivos rigurosamente empíricos, evitando elaborar teorías prematuras, contrasta con la inmediata categorización de los datos típica del analista del discurso).

Levinson (1983) se sitúa en la perspectiva basada en el análisis de la conversación.

Desde este repaso, extraeríamos la conclusión que un *diseñador de aplicaciones en tecnología del habla, ha de tener en cuenta las características discursivas a la hora de desarrollar e implementar un sistema conversacional*. Ha de tener en cuenta las aportaciones que distintas áreas del análisis del discurso realizan al campo de la gestión del diálogo, de estas ramas destacar: *LOS ACTOS DE HABLA, LA PRAGMÁTICA Y EL ANÁLISIS DE LA CONVERSACIÓN*, toda la investigación actual en sistemas de gestión de diálogo gira en torno a conceptos e ideas tomadas de estos tres campos de interés.

Indudablemente, hay matices que distinguen un discurso de una conversación o de un diálogo, sin embargo en los Sistemas de Gestión de diálogo en Inteligencia Artificial, se considera igual a un sistema conversacional, oración, frase, enunciado y alude a la *emisión de un mensaje por parte de un hablante*.

Para que se considere que una oración es correcta desde el punto de vista semántico, en PLN (Procesamiento del Lenguaje Natural) se acepta que la oración ha de cumplir las condiciones de verdad. La descripción semántica de un lenguaje se convierte así en un mecanismo de estados finitos que nos permite saber cuáles son las condiciones de verdad para cada oración de ese lenguaje.

El cálculo de predicados de la lógica de primer orden ha jugado un papel fundamental en el tratamiento semántico en PLN, este tipo de lógica se puede definir como un pequeño modelo de cómo desarrollar el análisis semántico de un lenguaje, de forma que se aproveche la fuerza expresiva y, no ambigua de las matemáticas.

Esta lógica dice que las fórmulas bien formadas se especifican en la gramática, asociándose una regla de interpretación semántica para cada construcción sintáctica que permita la gramática. Dada una serie de oraciones se podrán describir sus condiciones de verdad con respecto al universo del discurso y las inferencias que se puedan extraer a partir de las condiciones. Sin embargo, *en lenguajes naturales el tratamiento semántico que proporciona la lógica de primer orden no es suficiente*.

Montague (1973) partió de la lógica de primer orden y desarrolló un modelo más potente: la lógica intensional, se trata de una representación intermedia entre la estructura sintáctica y la estructura del significado, esta noción de representación intermedia fue adoptada por Kamp (1981) en su “Teoría de Representación del Discurso” (DTR).

La “Teoría Situacional” (Barwise y Perry, 1983) avanza respecto a las corrientes montagueanas, formulando convenientemente la noción de “estado parcial del mundo”, es incorrecta la premisa común en las corrientes que siguieron a Montague indicando que la información es siempre total, para cada situación una proposición es verdadera o no.

La teoría de la representación del discurso (DTR) (Discourse Representation Theory) elaborada y desarrollada en los años 80 y 90 por Hans Kamp y sus colaboradores (Kamp, 1981; Kamp y Reyle, 1993) tiene como fin describir relaciones que se establecen entre la forma lingüística de un enunciado y su significado. Pretende establecer modelos de interpretación para fragmentos de discurso que van más allá de la oración. Con ello sería posible formular reglas de interpretación para ciertos términos no referenciales del discurso, como los pronombres y otros términos anafóricos.

Las Estructuras de Representación del Discurso (DRS) presentan las relaciones lógicas de predicación y cuantificación sobre variables individuales necesarias para la interpretación de los enunciados, las estructuras se relacionan entre sí dando lugar a un modelo de interpretación del discurso válido para fragmentos amplios del discurso.

La DRT es beneficiosa tanto para la interpretación del sintagma nominal (cuantificación, plurales, nombres genéricos, posesivos) como del sintagma verbal (tiempo, aspecto, modo y modalidad) es un modelo alternativo a la lógica de predicados. Igualmente permite tratar de forma adecuada a determinados fenómenos de enriquecimiento semántico del discurso, como presuposiciones e inferencias, que aporta contenido al significado estructural de las oraciones involucradas.

Kamp (1981) propuso una aproximación a la interpretación del discurso y de las anáforas pronominales que ponía de manifiesto las deficiencias del análisis montagueano. La teoría de Hans Kamp se presenta a sí misma como una teoría semántica para un fragmento del lenguaje natural, intenta describir las relaciones que se dan entre la forma lingüística de un enunciado y su significado.

Kamp pretende que su teoría semántica sea una representación del flujo de información que se produce en el acto lingüístico, entendido como una descripción parcial de la realidad. El significado de las oraciones que constituyen un discurso depende de la verdad del discurso, es decir, depende de un modelo de interpretación parcial que debe poder ser integrado en un modelo completo.

Otras teorías diferentes sobre la interpretación del discurso han tratado el

problema de las presuposiciones y de su papel en la interpretación, Teoría de los Estados mentales de Fauconnier (1984), el significado del discurso se representa mediante un conjunto estructurado de espacios mentales interconectados asimilables a las DRSs de Kamp o los Dominios de Discurso de Seuren (1985).

Algunas teorías usan la semántica de mundos posibles en la interpretación del discurso, Irene Heim (Heim, 1982 y 1992), trata los contextos del discurso como mundos posibles, o la Lógica Dinámica de Predicados (DPL) (Groenendijk y Stokhof, 1989), a la que se aplican diversas teorías semánticas (Dekker, 1992; Vermeulen, 1993).

Finalmente, nos encontramos con teorías semánticas dinámicas basadas en la Gramática de Montague, como la propuesta por J. Groenendijk y M. Stokhof (1987) o las teorías de Reinhard Muskens (1996), que combinan la Gramática de Montague y DRT.

Los enunciados tratados por los sistemas de gestión de diálogo forman parte de un diálogo. Las características de un diálogo *difieren de las de un discurso* (Ginzburg, 1998; Poesio, 1998; Rieser, 1999). Estas investigaciones han sido llevadas a cabo en campos como el análisis del discurso, la psicología cognitiva y la lingüística. Los diseñadores del sistema de gestión de diálogo tratan de encontrar pistas que les ayuden en su tarea de entender cómo se estructura un diálogo con el fin de poder modelarlo.

La SDRT (Asher, 1993) es una extensión de la DRT, incorpora información acerca de la estructura y función retórica del discurso y la extensión del SDRT (Lascarides y Asher, 1999) nace con el propósito de modelar la interpretación del diálogo. Lascarides y Asher concluyen su trabajo dejando un camino abierto a futuras investigaciones que incluyan la resolución de otros fenómenos propios del diálogo como son los actos del habla indirectos.

La DRT se ha utilizado como *modelo de representación semántica* durante los últimos años por los *módulos semánticos* que componen los sistemas conversacionales del tipo pregunta-respuesta, el sistema VERBMOBIL (Alexandersson y otros, 1998) (es un sistema de traducción automática, son modelos computacionales que presentan una gran robustez (traducido directamente del inglés "robustness", significa capacidad que tiene un sistema cualquiera para adaptarse a entornos adversos como: ruido, canal, locutor, tarea, etc.)

Resumiendo, las teorías dinámicas del significado, buscan establecer relaciones entre la interpretación de las expresiones lingüísticas y el contexto en el que aparecen. Entre las teorías que tratan de representar este proceso dinámico destacan:

- DRT teoría de la representación del discurso de Kamp (relaciones entre la forma lingüística de un enunciado y su significado).

- Semántica de situaciones de Barwise y Perry, (parte de la premisa: el significado de una oración simple declarativa es una relación entre preferencias y situaciones descritas).

El uso de voz en interfaces hombre-máquina (la voz el modo de acceder a los servicios que presta la máquina) ha permitido una revitalización, esto ha sido posible como parte del estudio de sistemas conversacionales con entrada y salida por voz. Actualmente, existen muchas aplicaciones que exigen el tratamiento de la estructura conversacional, aplicaciones cuya interfaz hombre-máquina y su interacción bidireccional, añade características del diálogo conocidas y manejadas por los sistemas de gestión de diálogo.

La entrada en el sistema se produce por medio de la expresión hablada espontánea reconocida por un módulo de reconocimiento de voz (RV) pasando a un módulo de procesamiento de lenguaje natural (PLN) que es el encargado de extraer significado de las muestras de lenguaje hablado espontáneo. El mayor reto es extraer el significado de la salida del reconocedor (el lenguaje espontáneo presenta características imprevisibles).

Una de las tareas más complejas consiste en extraer el significado de la cadena reconocida e impedir que el proceso de análisis se detenga (Young y Matessa, 1991).

Posteriormente un gestor de dialogo (GD) intervendrá en el proceso, es el encargado de analizar las características propias del nivel del discurso que debe de tenerse en cuenta a la hora de implementar sistemas naturales y sus subramas (pragmática, actos del habla, análisis de la conversación, turnos, anáfora, elipsis, silencios, solapamientos, reparaciones).

Las bases de datos y el administrador del sistema contribuirán a que se genere una respuesta. El procedimiento sería:

8.1. Arquitectura funcional

En la División de Tecnología del Habla de Telefónica I+D, el sistema de reconocimiento de voz se realiza mediante un sistema de MÓDULOS organizados, cada uno con una función: acondicionamiento de la señal de entrada, detector de actividad, extractor de características, reconocimiento de patrones y post-procesado. *No confundir con otros módulos que intervienen posteriormente*, nos referimos únicamente al RV (reconocedor de voz).

Así mismo, el reconocimiento se realiza a cuatro niveles: palabras aisladas, habla conectada, palabras aisladas basado en modelos de fonemas para vocabulario flexible y habla continua.

Denominamos “palabra” a la unidad básica en la que se apoya el reconocedor, pueden ser sílabas, demisílabas, fonemas, morfemas, palabras, conjuntos de palabras etc.

Basándose en un esquema de reconocimiento muy parecido al explicado aquí, numerosas comunidades de científicos han desarrollado y sacado a la luz sistemas de reconocimiento de habla continua. En Estados Unidos cabe resaltar los sistemas de IBM y el de los Laboratorios Bell de AT&T; los que se encuentran en las Universidades como en el sistema Summit del Massachusetts Institute of Technology, el sistema SPHINX de la Carnegie Mellon University, y el sistema DECIPHER del Stanford Research Institute; y los de otras compañías como el sistema Wall Street Journal de Dragon, y el sistema BYBLOS de BBN. En Europa podemos resaltar el sistema HTK, de la Universidad de Cambridge, el sistema SPICOS de Philips, el sistema del CSELT italiano, y el proyecto LIMSI francés.

8.2. Las ontologías

Las *ontologías* definen las clases de objetos que pueden ser manejados por el sistema, las relaciones entre ellos y algunos individuos especiales de esas clases. Las ontologías utilizan los llamados *SEMANTIC PARSING* (analizadores en los que es prioritario el aspecto semántico del lenguaje).

La Word Net es un sistema de referencia léxica on line en inglés, cuyo diseño está inspirado por las teorías psicolingüísticas vigentes sobre memoria léxica humana, los nombres verbos y adjetivos se organizan en conjuntos de sinónimos y cada uno de ellos representa los conceptos léxicos subyacentes. Existen diferentes relaciones que unen los conjuntos de sinónimos, se puede decir de Word Net que es un diccionario basado en principios psicolingüísticos.

Word Net divide el lexicón en 5 categorías: nombres, verbos, adjetivos, adverbios y palabras de función. De hecho, Word Net sólo contiene nombres, verbos, adverbios y adverbios.

Esta división se basa en las observaciones en el habla de pacientes afásicos (Garrett, 1982), que indican el probable almacenamiento por separado de las palabras, encontrándose agrupadas como parte de los componentes sintácticos del lenguaje.

La decisión de organizar los nombres como un sistema de herencia, refleja un juicio psicolingüístico sobre el lexicón mental, ¿qué tipos de evidencia aportan una base para tales decisiones?, el aislamiento de los nombres en un subsistema léxico separado recibe el apoyo de las observaciones clínicas de pacientes con afasia anómica; después de un ataque al hemisferio izquierdo, que afecta a la capacidad para comunicarse lingüísticamente, la mayoría de los pacientes quedan con un déficit de la habilidad de nombrar (Caramazza and Berndt, 1978).

En lo referente a adjetivos calificativos, Word Net contiene punteros entre adjetivos calificativos y el subsistema de los nombres que se refiere a sus atributos apropiados.

La Euro WordNet es una base de datos multilingüe con wordnets para varios lenguajes europeos: holandés, alemán, italiano, español, francés, checo y estonio. Las wordnets están estructuradas de la misma forma que la wordnet americana para inglés (Princeton WordNet, Miller y otros 1990), en agrupamientos, es decir grupos de sinónimos de palabras *con relaciones semánticas entre ellas*.

La mayoría de los sistemas adopta el denominado "parsing parcial" que consiste en la adopción de un enfoque determinista mediante el cual el parsing recorre la secuencia de entrada buscando los fragmentos que puedan emparejarse directamente con la tarea en el dominio, es decir solamente los fragmentos claves desde el punto de vista semántico. Son muchos los sistemas que construyen gramáticas para herramientas claves, entre otros E.V.A.R. (información horaria de trenes, Nöth y otros, 1999), Philips (Aust y otros, 1995), IRIS (López-Soto, 1999). A estos modelos se les conoce también como "*SEMANTIGRAMAS*" (Haas y otros, 1997).

El parsing parcial reduce el análisis sintáctico y semántico de una secuencia al procesar únicamente unidades específicas de un determinado plan criterio semántico en lugar de sintáctico.

Este "concepto semántico" varía según la tarea. Estos conceptos semánticos coinciden con los contenidos informativos que son vitales para la comprensión del mensaje en el dominio. El parser identifica y analiza dichos conceptos y les asigna una representación semántica que es con lo que opera el módulo de gestión de diálogo.

8.3. Un sistema de lenguaje natural, el Phoenix

Este sistema de comprensión del habla de la Carnegie Mellon University (School Of Computer Science, Pittsburg, Pennsylvania, USA), está orientado hacia la extracción de información relevante en una tarea, utiliza un PARSER (analizador) flexible basado en tramas.

Este sistema maneja fenómenos que son naturales en el habla espontánea por ejemplo: repeticiones y expresiones mal formadas gramaticalmente, así mismo, mantiene un archivo histórico de los rasgos principales del diálogo, puede resolver referencias elípticas, anáforas o de otro tipo indirectas, además, el contexto modela el sistema, maneja las correcciones y las preguntas que exceden de sus posibilidades, soluciona acoplamientos rígidos o laxo del reconocimiento del habla y del procesamiento del lenguaje natural. Todas estas características hacen que sea el utilizado en la actualidad por T I+D (Telefónica Investigación y Desarrollo), esta Empresa ha desarrollado y es propietaria de las adaptaciones para nuestra lengua, su base de datos alcanza 1.000.000 de palabras.

El Phoenix se ha utilizado para modelar el servicio de información de las líneas aéreas A.T.I.S. Nov. 1992. Cuando se le sometió al sistema de evaluación de idiomas hablados de A.R.P.A. respondió correctamente al 93'5% de inputs escritos y un 88'3 % de los hablados.

Una de sus características, es que un pre-procesador elimina las palabras que son idénticas a la palabra anterior de una frase, igualmente, la gramática es modificada con el fin de que los elementos puedan terminar con la palabra clave, utilizada para marcar las correcciones.

Como un atributo necesario del sistema de lenguaje hablado es la capacidad para responder a cualquier pregunta de forma razonable, si hay algo significativo sobre las palabras que no están analizadas gramaticalmente, pregunta al usuario sobre ellas y en cualquier otro caso genera una respuesta adecuada.

Con el fin de permitir a los usuarios elaborar soluciones incrementalmente, un sistema de lenguaje hablado debería permitir referencias implícitas y explícitas a preguntas y respuestas anteriores, el sistema debe mantener un registro del dialogo.

Es relativamente simple la resolución de la elipsis y la anáfora en este sistema. Las ranuras en las tramas, los intervalos de tiempo asignados a las tramas, son **semánticos**, así pues, sabemos que el tipo de objeto que se necesita para la resolución en el caso de la elipsis es añadir nuevos objetos, mientras que en el caso de la anáfora, simplemente tenemos que suponer que un objeto de tal tipo ya existe.

Al diseñar un sistema robusto, la interacción entre reconocimiento del habla y los componentes de comprensión del lenguaje representan una decisión estructural capital.

Según opinión de Ismael Cortázar, Jefe de División de Tecnología del Habla de T I+D, el analizador semántico que utilizan en su División es eficaz fundamentalmente por ser muy robusto, se basa en las RTN.

8.3.1. Las R.T.N. (*Recursive transition Networks*).

Es un método robusto para el análisis semántico, muy útil para el procesado de la salida de los Reconocedores de Habla Continua. El procedimiento para el uso de las R.T.N. se realiza mediante 3 fases:

- Descripción de las redes en el lenguaje RTN.
- Compilación de las descripciones generando un modelo binario.
- Generación de una Analizador Semántico, usando la librería ("parser") RTN, a la que se le pasa como parámetro (en tiempo de ejecución) qué modelo binario utilizar.

Las RTN no mantienen un contexto de diálogo: cada frase nueva se analiza sin considerar las frases anteriores, al construir las aplicaciones se usa otra herramienta conocida como GESTOR DE DIÁLOGO, que suele ser el “usuario” del parser RTN.

El análisis semántico mediante R.T.N. se basa en los conceptos:

- **FUNCIÓN:** Una función representa uno o varios conceptos semánticos, ante una función de entrada el “parser” elegirá la función que mejor se adapte.
- **RED:** Una red representa una entidad, parcial o total, dentro de un concepto semántico. Una función se compone de una o varias redes. Se dividen en tres tipos: REDES FUNCIÓN, REDES SUB-RED, REDES DATOS.
- **NO TERMINAL.** Entidad que define una agrupación de símbolos terminales. Se usa sólo para facilitar la definición de RTN, y no durante el “parsing”. Los no terminales han de definirse con posterioridad a todos sus usos.
- **TERMINAL.** Lo constituyen las palabras que pueden entrar en una frase. Tiene un código numérico y están recogidos en un diccionario.

Al proceso de definición viene aparejado la creación de los ficheros necesarios, cada uno define una función y cada función se resuelve mediante concatenación de alguna de las redes que la componen.

Una vez escritos los ficheros se pasa al proceso de generación del modelo binario, se realiza mediante el programa GANASEM, este programa tiene dos modos distintos de funcionamiento: como analizador léxico y como compilador.

Para realizar el análisis existe un conjunto de rutinas que facilitan la construcción de un analizador semántico, están contenidas en la librería LIBANASEM, la rutina principal es *doAnaSem*, que recibe como entrada un conjunto de códigos de palabras, tal y como los entrega el reconocedor de habla continua, y devuelve las redes y sub-redes, las que mejor han resuelto la frase.

El parser posee unos parámetros, por medio de unos valores iniciales comienza a utilizar la memoria, si no son suficiente, el parser seguirá funcionando, es decir, pedirá más memoria.

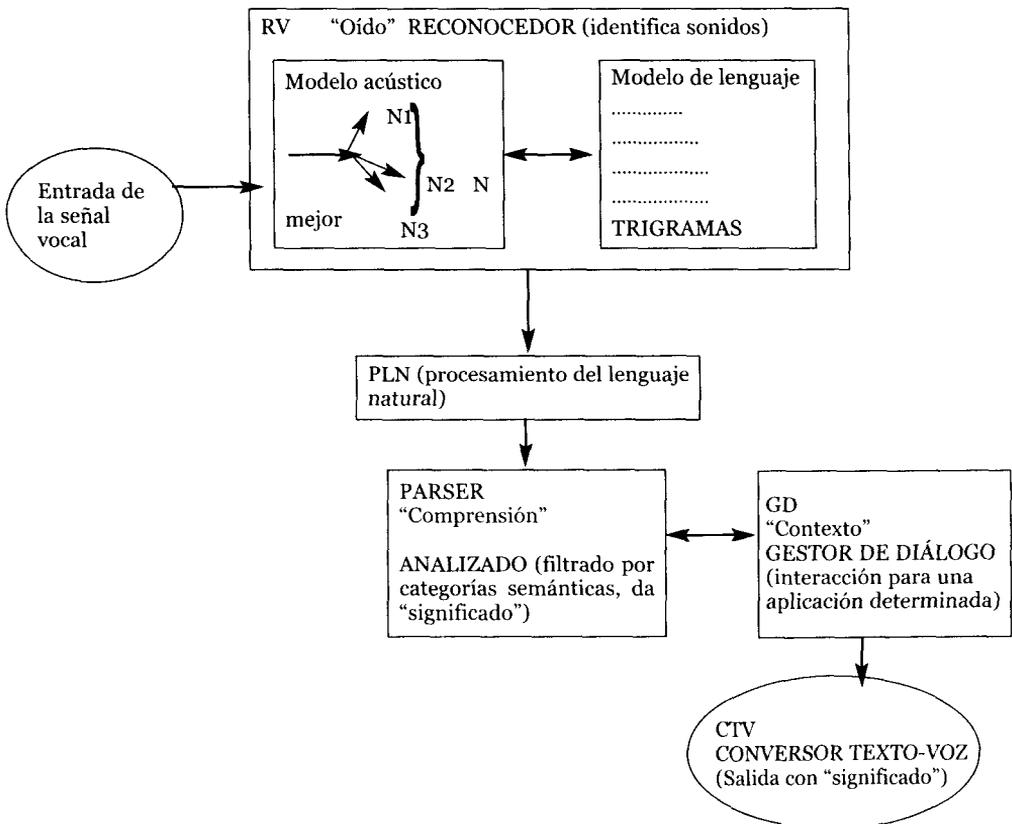
La librería del parser no hace ninguna suposición del entorno operativo, por ello utiliza un mecanismo de “callback” para pedir las funciones básicas. Posteriormente, el parser devuelve los resultados indicando la red y las sub-redes. Dado un conflicto, una frase puede ser interpretada de varias formas, los criterios para elegir la mejor son por orden de importancia para resolver el mayor número de palabras:

contiene una determinada red cuyo nombre se pasa como parámetro, o, contiene menos redes de nivel N (se comparan incrementalmente el nivel hasta que se resuelve el conflicto).

Para ayudar a definir las RTN, y analizar qué tal se adaptan a una aplicación se suministra la utilidad *anase*m, esta utilidad permite enviar frases (desde pantalla o fichero) y recibir los resultados del análisis (en pantalla o fichero). Hay que suministrarle uno o tres diccionarios, dado que puede simular lo que hace el reconocedor que a cada palabra reconocida le asigna cero, una o varias palabras a enviar como resultado de reconocimiento.

8.4. Modelo en tecnología del habla

Telefónica I+D está desarrollando aplicaciones dialogadas hombre-máquina utilizando el PHOENIX adaptado para el español, fundamentalmente porque es un sistema muy robusto, su analizador es semántico y el esquema del proceso en aplicaciones dialogadas lo reflejamos a continuación:



En primer lugar se realiza el esqueleto del parser (analizador), este es configurable, tengo que decir para qué lo voy a utilizar y se comienza a desarrollar la aplicación.

La señal vocal entra en el sistema (es preciso que la base de datos haya sido construida con un texto muy extenso) y es reconocida, para ello se utilizan los modelos acústicos y los modelos de lenguaje, estos dos modelos intercambian información, el primero envía al segundo la N-mejor (es decir, la mejor secuencia, son modelos basados en listas “n-best” que es comparada con los modelos de lenguaje estadísticos, siempre busca trigramas, es decir, secuencias de palabras que vayan precedidas de otras dos (desde el punto de vista matemático, un bigrama es una función con dos argumentos que se suelen representar como una tabla de doble entrada o matriz bidimensional, bigramas o trigramas son tablas que establecen la frecuencia de ocurrencia entre secuencias de dos unidades, en el primer caso, o tres, en el segundo).

Hasta este momento es la fase de reconocimiento (RV=reconocedor de voz), mientras el módulo PLN (procesamiento de lenguaje natural) y el resto de los componentes: sintaxis, gramáticas, los diccionarios, ficheros función proporcionan los datos necesarios para que continúe el proceso.

En un segundo momento es donde comienza lo que podríamos llamar, con muchas reservas, la fase de “comprensión”, se trata de “comprender y razonar”. La información pasa al analizador (PARSER) que es el encargado de dar un significado, filtra las secuencias de palabras con categorías semánticas (están etiquetadas, marcadas), busca palabras del reconocedor con significado para el usuario de la aplicación (recordamos que el parser ha sido desarrollado para una determinada aplicación) y las envía al gestor de diálogo (GD), este al recibirlas comprueba si cumple las reglas que están jerarquizadas y establecidas según la aplicación en el módulo de PLN. Si hay que completar alguna acción, hay una regla que detecta que faltan datos a fin de completar el árbol, decidiendo de lo que llega al parser.

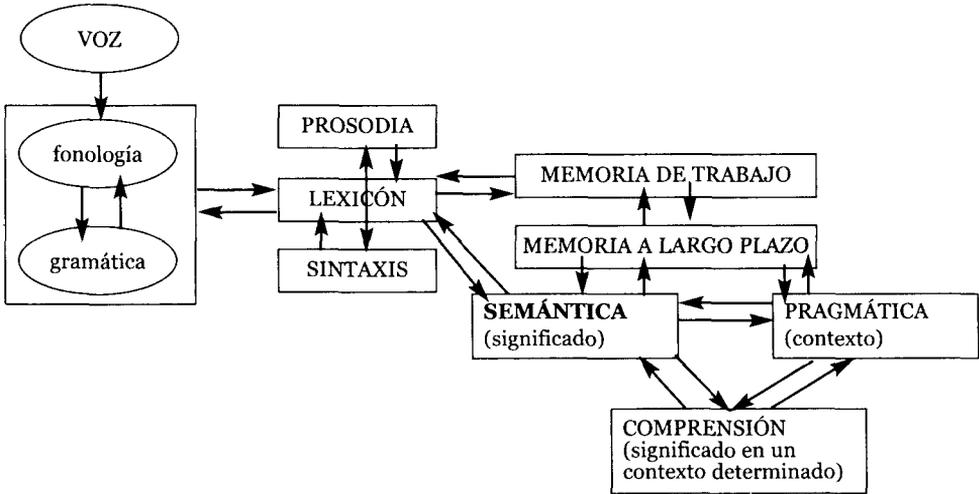
Por último, se produce la salida (síntesis) por medio del conversor texto-voz (CTV) y queda establecido el diálogo.

9. CONCLUSIONES

Ha quedado expuesto que las líneas actuales de investigación se centran en el aspecto semántico del lenguaje, si bien en tecnología del habla queda mucho que desarrollar, encontrándose en la actualidad en el “arco reflejo de la comprensión lingüística oral” (Reyzábal, 2001).

Considerando los modelos anteriormente descritos, nos inclinamos a adoptar un modelo de comprensión lingüística oral guiado por el contenido, formado por subsistemas, en éstos, el procesamiento se produce de un modo interactivo y bidireccionalmente en paralelo.

Proponemos la siguiente arquitectura para un modelo de comprensión lingüística oral (resaltamos el subsistema semántico porque consideramos: SI UN OYENTE NO LOGRA OBTENER UN SIGNIFICADO DE LO ESCUCHADO, LO ELIMINA. (Reyzábal 2001).



La voz es percibida, la fonología y la gramática son las encargadas de discriminar esos sonidos y aplicarles las reglas oportunas (trabajo computacional), pasan identificadas a nuestro lexicón (comienza el trabajo conceptual) éste interactuará con la sintaxis que le proporcionará el orden correcto de las palabras, la prosodia ayudará. La memoria de trabajo y la memoria a largo plazo siempre están activadas proporcionando información. Obtenemos un significado de lo escuchado integrándolo en un contexto y con los conocimientos previos. El siguiente paso es la comprensión. Todo está constantemente en interacción y los conocimientos son almacenados en la memoria a largo plazo, si interesa, o en la de trabajo.

REFERENCIAS BIBLIOGRÁFICAS

- Alexandersson, J., Buschbeck-Wolf B., Fujinami, T., Kipp, M. Maier, E. Reithinger, N. Schmitz, B. y Sigel, M. (1998). *Dialogue Acts in VERBMOVIL-2* Second Edition. Verbmobil-Report 226, DFKI GmbH Saarbrücken.
- Asher, N. (1993). *Reference to Abstract Objects in Discourse*. Kluwer Academic Publishers.
- Aust, H. M. Oeder, F. Seide, V. Stein Biss. (1995). The Philips automatic train time table information system. *Speech Communication*, 17: 249-262.
- Barwise, J. y Perry, J. (1983). *Situations and Attitudes*. Cambridge; Mass: The MIT Press.
- Chomsky, N. (1964). Degrees of grammaticalness, en Fodor, J., Katz, J. (Eds.) *The Structure of Language*. Englewood Cliffs: Prentice-Hall.
- Chomsky, N. y Miller, G. A. (1963). *Finitary models of language users*, en Luce, R. D. Bush, R. R. and Gallanter, E. (eds.), *Handbook of Matematical psychology*, Vol 2, New York, John Wiley.
- Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.
- Chomsky, N. (1965). *Aspects of the theory of Syntax*. Cambridge: MIT Press.
- Chomsky, N. (1995). *The Minimalist Program*, Cambridge, Massachusetts. (MIT Press).
- Clifton, C. Jr., & Ferreira, F. (1987). Modularity, en J. L. Gardfield. (Ed.) *Modularity in Sentence Comprehension: Knowledge Representation and Natural Language Understanding*. Cambridge, MA: MIT Press.
- Crain, S. & Stedman, N. (1985). On not being led up the garden path: The use of context by the psychological parser, en Dowty, D.; Karttunen, L. & Zwicky, A. (Eds.) *Natural language parsing*. Cambridge: Cambridge Univ. Press.
- Van Dijk, T. A. (1985). Introduction: discourse as a new cross-discipline, en van Dijk, T. A. (ed.) *Handbook of Discourse Analysis, Volumen I: Disciplines of Discourse*. New York: Academic Press, 1-10.
- Fauconnier, G. (1984). *Espaces Mentaux*. Paris: Minuit.
- Fodor, J. A. (1970). *The Modularity of Mind. An Essay on Faculty Psychology*. Cambridge, Mass: The M.I.T. Press.
- Ford, M., Bresnan, J. & Kaplan, R. (1982). A competence-based theory of syntactic closure, en Bresnan, J. (Ed.). *The mental representation of grammatical relations* Cambridge, MA: MIT Press, (pp.727-796).
- Forster, K. (1979). Levels of processing and the structure of the language processor, en Cooper, W. E. and Walker, E. C. T. (Eds.) *Sentence Processing: Psycholinguistic studie presented to Merrill Garrett*. London: Lawrence Erlbaum Associates Ltd.
- Forster, K. I. (1979). Levels of processing and the structure of the language processor, en COOPER, W. E. and WALKER, E. C. T. (Eds.) *Sentence processing*. N. York: Wiley.
- Frazier, L., Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition*, 6, 291-325.
- Frazier, L., Clifton, C. y Randall, J. (1983). Filling gaps: decision principles and structure in sentence comprehension. *Cognition*, 13, 187-222.
- Ginzburg, J. (1998). Clarifying Utterances, en Hulstijn, J. A Nijholt (eds.) *Proceedings of the Twente Workshop on the Formal Semantics and Pragmatics of Dialogues*. Enschede. Twente University, Faculteit Informatica, 11-31.
- Groenendijk, J. y Stockhof, M. (1989). Dynamic predicate logic: towards a compositional, non-representational semantics of discourse. *Institute for Language, Logic and Information*. Prepublication series LP-89-02, Universidad de Amsterdam.

- Haas, J., Nöth, E., Niemman, H. (1997). Semantigrams-Polygrams Detecting Meaning, en Proc. 2nd Sayel Workshop on Multi-Lingual Information Retrieval Dialogs, 65-70.
- Halliday, A.K. y Hasan, R. (1976). *Cohesion in English*. London: Longman.
- Heim, I. (1982). *The semantics of definite and indefinite noun phrases*. PhD thesis, Amherst: University of Massachusetts.
- Heim, I. (1992). Presupposition Projection and the Semantics of Attitude Verbs. *Journal of Semantics*, 9: 183-221.
- Kamp, H. y Reyle, U. (1993). *From Discourse to Logic*. Dordrecht: Kluwer.
- Kamp, H. (1981). A theory of truth and discourse representation, en Groenendijk, J., Jansen, T. y Stockhof, M. (eds.) (1984). *Truth, interpretation and information*. Dordrecht: Foris.
- Lascarides, A. y Asher, N. (1999). Cognitive States, Discourse Structure and the Content of Dialogue. *Preproceedings of the Amstelogue '99 Workshop on the Semantics and Pragmatics of Dialogue*, Amsterdam University.
- Levinson, S. C. (1981). Some pre-observations on the modelling of dialogue. *Discourse Processes*, 4 (1): 93-116.
- Levinson, S. C. (1983). *Pragmatics*. Cambridge: Cambridge University Press.
- López-Soto, T. (1999). *Estrategias de análisis gramatical y semántico para un sistema dirigido por voz*. Tesis Doctoral. Universidad de Sevilla.
- Marks II, R. J. (1993). Intelligence: Computational versus artificial. *IEEE Trans. On Neural Networks*, 4 (5), 737-739.
- Montague, R. (1973). *Formal Philosophy: Selected Papers of Richard Montague*, en Thomanson, R. (ed.) Yale University Press.
- Muskens, R. (1996). Combining Montague semantics and discourse representation. *Linguistics and Philosophy*, 19 (2), 143-186.
- Nöth, E., Batliner, A., Kiebling, A., Kompe, R. Niemann, H. (2000). Verbmobil: The Use of Prosody in the Linguistic Components of a Speech Understanding System. *IEEE Acoustics, Speech and Signal Processing*, 8 (5).
- Pinker, S. (1995). *El instinto del lenguaje. Cómo crea el lenguaje la mente*. Madrid: Alianza Editorial.
- Poesio, M. y Traum, D. (1998). Towards an axiomatisation of dialogue acts, en Hulstijn, J. y Nijholt, A. (eds.) *Proceedings of the Twente Workshop on the Formal Semantics and Pragmatics of Dialogues*. Enschede: Universiteit Twente, Faculteit Informatica, 207-222.
- Reyzábal, M. I. (2001): *Dificultades de comprensión lingüística oral, propuesta de mejora utilizando las nuevas tecnologías: Diálogo hombre-máquina mediante instrucciones expresadas en lenguaje natural*. Tesis Doctoral. Universidad Complutense de Madrid.
- Rieser, H. (1999). *The Structure of Task-oriented dialogue and the Introduction of New Objects*. *Preproceedings of the Amstelogue '99 Workshop on the Semantics and Pragmatics of Dialogue*, Amsterdam University.
- Rumelhart, D. E., McClelland, J. L. (1986) (des.). *Paralel Distributed Processing. Foundations*. Vol 1 MIT Press.
- Santituste, V. (1982). La autonomía del subprocesador sintáctico del lenguaje. *Revista de Psicología General y Aplicada*. 37, 3.
- Seuren, P. (1985). *Discourse Semantics*. Oxford: Blackwell.

PALABRAS CLAVE

Comprensión lingüística, modelos de lenguaje, tecnología del habla.

KEY WORDS

Linguistic comprehension, language models, speech technology.

PERFIL ACADÉMICO DE LOS AUTORES

Víctor Santiuste Bermejo. Catedrático de la Universidad Complutense de Madrid, Facultad de Educación, Departamento de Psicología Evolutiva y de la Educación. Líneas de investigación: Psicolingüística, Procesamiento lingüístico, afasias, dificultades de aprendizaje, comprensión y producción lingüística.

María Isabel Reyzábal Manso. Catedrática Universitaria Interina. Dpto. de Psicología Evolutiva y de la Educación. Facultad de Educación. Universidad Complutense de Madrid. Líneas de investigación: Tecnología del habla, psicolingüística, lenguaje hablado, estrategias de comprensión, dificultades de aprendizaje, modelos de comprensión lingüística, neuropsicología, aplicación de las nuevas tecnologías.

Dirección de los autores: Facultad de Educación - C.F.P.
Dpto. de Psicología Evolutiva y de la Educación
Rector Royo Villanova, s/n
28040 Madrid
E-mail: m.i.reyzabal@edu.ucm.es

Fecha recepción del artículo: 16. diciembre. 2004

Fecha aceptación del artículo: 25. abril. 2005