

# **I.6 LA COMUNICACIÓN HOMBRE - MAQUINA UTILIZANDO LENGUAJE NATURAL RESTRINGIDO**

*Juan Martín González Vázquez\**

## **Resumen**

Se describe un modelo para el procesamiento del lenguaje natural, el cual tiene como módulos principales un decodificador/encodificador, un validador de completez, un estructurador de respuestas y un ejecutor.

## **INTRODUCCIÓN**

La tecnología actual para la fabricación de computadoras y las características de estas rebasan cualquier expectativa planteada en sus anales, su aplicabilidad ya es comprobada en la gran mayoría de nuestras actividades, ahora bien si es tan útil e importante su uso, como podemos obtener el mayor provecho de ella, sin necesidad de recurrir a un experto en su uso o a algún tipo de educación especializada para ser uno de ellos, ya que pese a los esfuerzos hechos a la fecha es necesario definir procedimientos complejos y traducirlos a un lenguaje que sea manejado por la marca y versión de computadora que queremos utilizar, o adaptarnos a los esquemas de productos ya hechos para aplicaciones mas o menos parecidas a las nuestras, y sin omitir la posibilidad de que no existan tales productos para nuestra aplicación específica.

---

\* Juan Martín González Vázquez elaboró este trabajo en abril de 1985, siendo subdirector de informática en la Secretaria de Programación y Presupuesto, en colaboración con el Instituto Nacional de Estadística, Geografía e Informática en la Dirección General de Política Informática en la Dirección de Desarrollo Informático, Proyecto Censo Agropecuario. Instituto Politécnico Nacional Unidad Profesional Interdisciplinaria De Ingeniería Ciencias Sociales y Administrativas División de Ingeniería Unidad de Investigación y Desarrollo del Departamento de Computación

Dado lo anterior caemos en la cuenta que una de las principales barreras para el uso de una computadora, es precisamente la comunicación con ella, de donde surge la necesidad de dispositivos que hagan esta comunicación sencilla y eficaz y en donde no sea necesario más que nuestro conocimiento en el área, la problemática que nos afecta y una computadora.

Aparentemente lo expuesto es una imagen de un usuario que su único problema es que necesita información, y no tiene los recursos de tiempo y/o conocimientos en el manejo de una computadora para obtenerla. Sin embargo no es el único tipo de usuario, ya que la vida profesional de algunas áreas en la calle o en el hogar existen actividades que son fácilmente sistematizables y susceptibles de ser automatizadas por lo que necesitaríamos mas que información, acciones y ejemplos claros. Los tenemos en armadoras de autos, robots manipuladores de sustancias peligrosas, controladores de tráfico, cajeros automáticos, cajas registradoras, conmutadores telefónicos, juegos infantiles, automóviles, etc., los cuales muchas veces requieren recibir sus procedimientos por medios altamente sofisticados y nuevamente nos encontramos con el problema de comunicación.

A lo que surge un sentimiento general de necesidad de un dispositivo de comunicación eficaz, y se discute sobre la conveniencia de que consista en una interfase en el idioma de la persona que lo use y que sea capaz de hacer que la computadora nos entienda de una manera sencilla y eficaz, aunque a su vez surge otra corriente que manifiesta que para manejar los problemas de su área no le es funcional manejar el idioma sino la jerga con todo y simbología de su área común, como podría ser el caso de un matemático o un químico, etc.; los propios informáticos cooperan en esta discusión pidiendo un lenguaje altamente taquigráfico para sus conceptos de manejo de la computadora e información.

Por lo que es necesario establecer las reales características del lenguaje natural, las cuales tienen el requisito primordial de ser generales y soportar los requerimientos de todos sus posibles

usuarios, ya que a todos les asiste la razón al declarar "natural" su pretendida forma de comunicación con la computadora por lo que se definen las siguientes características:

- 1) Debe ser *independiente* de un idioma determinado; lo que significa que la formación de enunciados tendrá las reglas, signos y criterios de consideración para la unidad léxica que determine la persona que use el sistema y este debe ser capaz de digerir los aspectos antes mencionados para el idioma en cuestión y aplicarlos a futuros contactos.
- 2) Ser capaz de *codificar mensajes propios del sistema* en el idioma que ha aprendido.
- 3) Debe ser capaz de determinar la *validez del mensaje* en función del contexto en el que se da, el idioma definido y los conocimientos proporcionados al sistema.
- 4) Debe ser capaz de *estructurar procedimientos propios y/o enseñados* además de poder inferir posibles procedimientos complementarios, para la total ejecución del mensaje y también supervisar que no viole ninguna restricción que por medio de la instrucción o la experiencia posea.
- 5) También será capaz de *determinar todos los recursos de la computadora* necesarios para procesar el mensaje, administrarlos y por último informar al emisor del mensaje, el resultado de tal ejecución en el idioma propio del usuario.

Podemos concluir que con lo antes expuesto el manejar "Lenguaje Natural" es algo más que el manejo de la traducción de un idioma a otro y que lleva involucrado desde las características más involucradas con la computadora hasta conceptos tan humanos como experiencia, enseñanza, aprendizaje, etc. y no obstante la presente lo declara como restringido y la razón es que los motivos de la comunicación no abarcan conceptos como sentimiento, instintos, corazonadas y cuestiones difíciles de cuantificar y de representar como un estado en el computador, tales como alegría, tristeza, temor, amor, etc. y se reduce el esquema a obedecer y manejar las propias habilidades en función de la satisfacción de un requerimiento. No obstante no se niega la

posibilidad de crear el concepto, representación y respuesta de los ejemplos anteriores.

## II MANEJADOR DE LENGUAJE NATURAL RESTRINGIDO

Ahora bien para la temática de lenguaje natural restringido planteamos el siguiente modelo al que nominalmente lo constituyen las siguientes partes:

- 1) *Decodificador / Encodificador*: Se encarga de traducir el mensaje fuente a una notación estándar del sistema que sea procesable por las otras unidades que lo constituyen, además de procesar los mensajes del sistema y trasladarlos al idioma del usuario.
- 2) *Validador de Completes*: Determina si la estructura decodificada tiene algún sentido basado en la que se ha aprendido y que la acción a la que apunta no esta censurada por alguna restricción aprendida o esta limitada por algún concepto que invalide al usuario como ejecutor de un proceso, así, al mismo tiempo que si es necesaria más información para darle el curso adecuado al mensaje la solicite.
- 3) *Estructurador de Respuestas*: Se encarga de subdividir la acción genérica en subacciones más específicas que estén en coordinación con lo que puede hacer la computadora, verificar los procedimientos necesarios para cumplirlos y en caso de que no existan declarados verificar si los puede inferir de los que si lo están. Así como declarar las variables necesarias para el paso de información entre procedimientos. Determinar nuevamente si uno de los procedimientos incorporados no choca con las restricciones conocidas y finalmente generar los requerimientos de información necesarios así como de recursos materiales para ejecutar el procedimiento.
- 4) *Ejecutor*: Se encarga de verificar que los requerimientos estén disponibles y en la cantidad deseada y que estas cantidades

tampoco infrinjan las restricciones registradas. Una vez que tenga todo dispuesto debe ejecutar cuando así sea necesario los módulos pertinentes y/o generar los programas, corridas que se requieran para lograr la respuesta pedida, monitorear la ejecución y estructurar un informe del resultado para pasárselo al Decodificador / Encodificador y este a su vez al usuario.

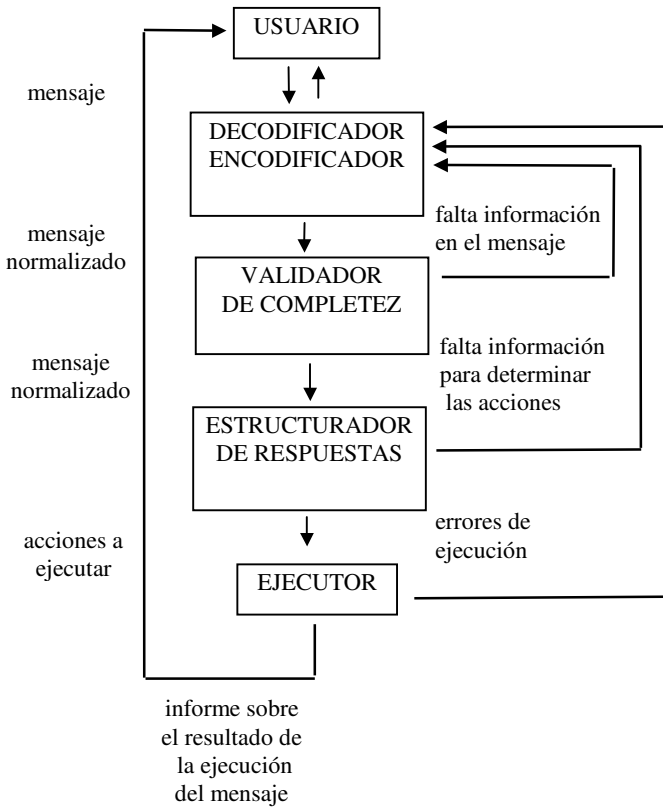


Figura 1. Manejador de Lenguaje Natural Restringido

## 1 DECODIFICADOR / ENCODIFICADOR

Lo Constituyen Las Sigüientes Partes:

1) *Analizador Léxico*: En primera instancia necesita que se le proporcione los caracteres que puede considerar como validos no obstante cuando aparezca un carácter desconocido pedirá autorización para ponerlos en su acervo de caracteres validos, otra cosa que también pide son los criterios para separar una unidad léxica de otra, las cuales se almacenan en un lugar expofeso para ello. Independientemente que tendría las especificaciones para procesar español.

2) *Categorizador y Validador*: En función a una gramática. Esta parte no obstante de tener lo necesario para procesar español orientado a proceso de datos es capaz de "Aprender" una nueva gramática en función a ejemplos; la cual también se almacena y enriquece y en sí esta parte tiene como salida las unidades que separa el analizador léxico validadas en su estructura y categorizadas genéricamente por lo que representan en la gramática.

3) *Encodificador*: Funciona en forma inversa al categorizador, es decir parte de una serie de unidades separadas y categorizadas, las cuales al seguir la gramática las estructura de acuerdo a esta en un mensaje en el idioma que representa la gramática.

4) *Formateador*: Lo único que hace es centrar mensajes y manejar la pantalla de manera de lograr cierta estética dentro del mensaje desplegado a parte de generar el medio ambiente para grabar a disco, cinta, o papel dichos mensajes.

## 2 VALIDADOR DE COMPLETES

Sus partes principales son:

1) *Resolvidor de indefiniciones e integrador de significado*: Es resolvidor de indefiniciones porque en su momento ciertas

unidades léxicas que no haya podido categorizar en el paso anterior, en esta parte las integra a la base de conocimiento por lo que pide al usuario información sobre las relaciones a establecer en la base de conocimientos con otras unidades así como características que la ubiquen en un contexto determinado y/o limitaciones que afecten de alguna forma su operación bajo alguna situación especial y en su momento posibles valores por ausencia. La función del integrador de significado consiste en integrar el mensaje original todas aquellas particularidades de las unidades que influyen o se vean influenciadas por la operación de la unidad en cuestión así como las reglas o relaciones que haya que considerar para su secuenciación.

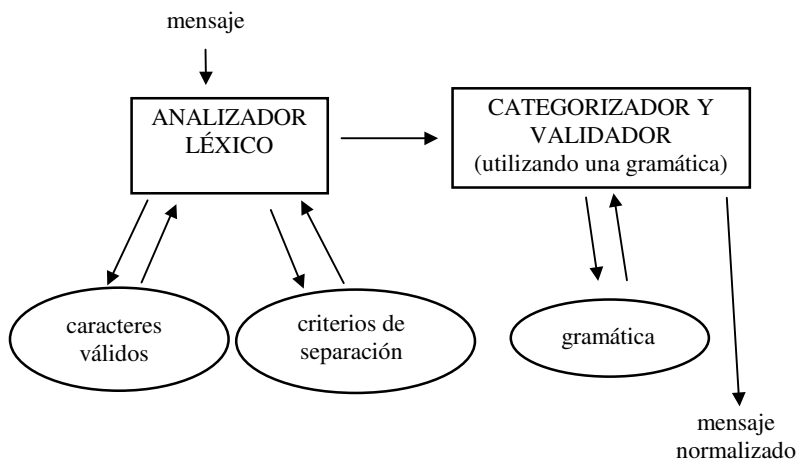


Figura 2. Decodificador /Encodificador

2) *Validación del significado contra el contexto*: Es la consideración global de todas las unidades ensambladas de acuerdo al significado e integradas como las nuevas características generales del mensaje que se contrastan contra los elementos de la base de restricciones y se determina la validez del mensaje en esta etapa en función de las violaciones a las restricciones, además de que se pide en caso de ambigüedad en las partes integradas de significado, información para hacer la selección adecuada.

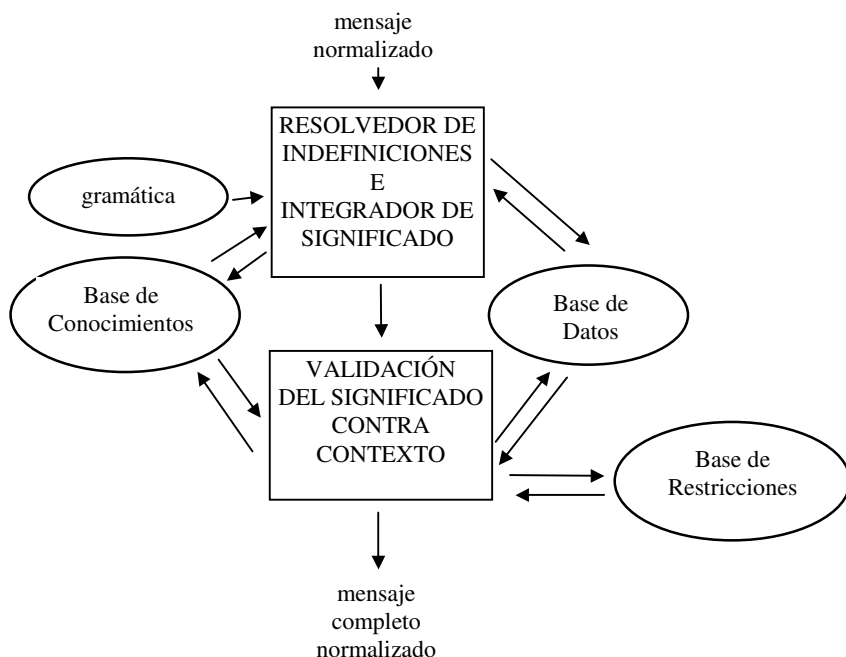


Figura 3. Validador de Completez

### 3 ESTRUCTURADOR DE RESPUESTAS

Tiene cuatro módulos que completan su función y que son:

1) *Determinador de los procedimientos a utilizar*: Esta parte integra los procedimientos que proporcionen las salidas deseadas para cada una de las unidades definidas por el significado y el contexto y en caso de que una unidad no tenga procedimientos declarados debe de tratar de integrar procedimientos que no obstante estén declarados para otras unidades, logren satisfacer las salidas necesarias o en su momento solicitar expresamente la definición de un proceso en particular.

2) *Discretizador de procedimientos*: Esta parte se encarga de declarar las variables y/o constantes en los procedimientos de tal forma que todos tengan una secuencia de entradas y salidas validas para la consecución de la salida total, pero en caso de que alguno de los requerimientos tipo variable o constante no pueda ser ubicado en el procedimiento anterior o posterior para completar una entrada o salida determinada se requerirá su definición al usuario a fin de que se integre a un determinado procedimiento como salida además de que se declare el lugar de donde se puede obtener.

3) *Determinador de la validez del procedimiento*: En esta etapa se verifica que los accesos, entradas y salidas en función a la operación y contenido no violen ninguna restricción declarada para un procedimiento determinado, o un usuario.

4) *Generador de Requerimientos de ejecución*: Una vez que se han ubicado las fuentes que dan información a los procedimientos y los lugares en donde esta se plasma se procede a hacer una lista con todos los requerimientos lógicos y físicos, entendiendo por esto los lugares y formatos necesarios para los datos así como los dispositivos asociados a los datos de entrada y salida y algunos otros insumos necesarios para recuperar la información y/o plasmarla tanto en calidad como en cantidad y si en algún momento cierto recurso es necesario, pero no esta definido se procede a definirlo, verificando que no choque su definición con alguna restricción declarada.

## 4 EJECUTOR

Esta componente del modelo tiene las siguientes partes:

1) *Generador de flujo logístico*: Se encarga de establecer los controles de documentación de versiones, actualización y requerimiento de equipo y administrar de alguna forma la concurrencia en el requerimiento de recursos y/o tiempo de máquina y tramitar la mejor operación de los procesos generados por los mensajes del usuario y regidos por las restricciones declaradas por la física de la máquina y/o restricciones a causa del tipo de usuario o procedimiento.

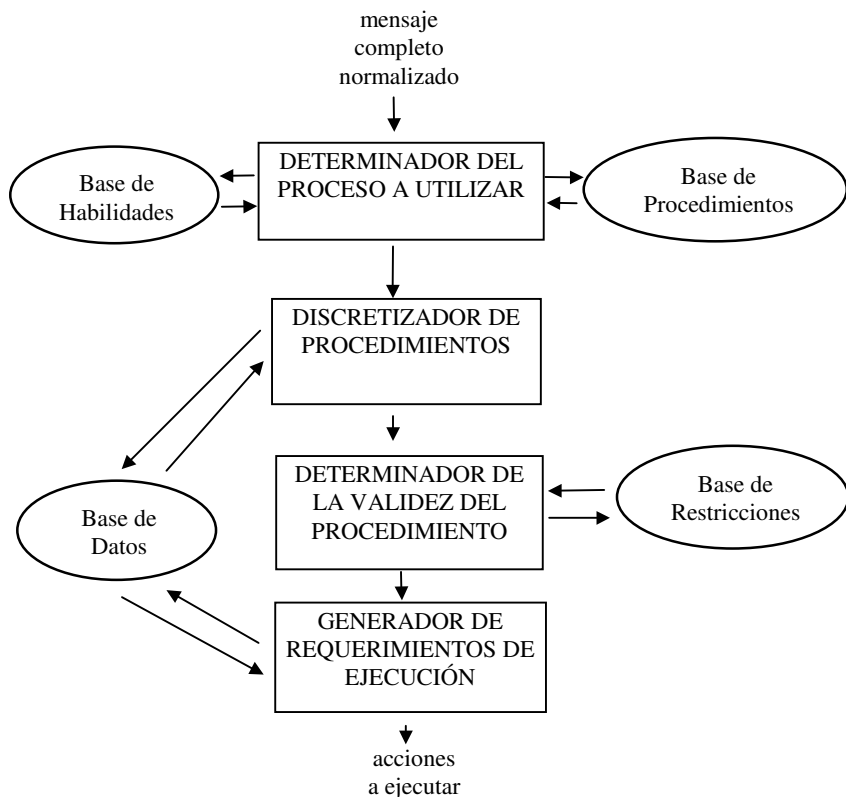


Figura 4. Estructurador de Respuestas

- 2) *Generador de programas, corridas de ejecución y/o llamado a módulos generales*: Una vez que se tiene lo que se va a hacer y con que se va a hacer, es necesario generar o transformar los procedimientos en algún lenguaje comprensible por la máquina, o en algún llamado a uno de los módulos generalizados del sistema, así como generar también los postulados de control que logren que la máquina tenga conciencia real de lo que necesitamos para la ejecución y finalmente poner en ejecución el proceso recolectando toda la información posible para dar un informe final del resultado de este, pero con la característica de auxiliarse de la base de conocimiento y la gramática para estructurar el

informe en forma de unidades léxicas separadas y categorizadas para entregarlas al encodificador /decodificador para que lo transmita al usuario.

3) *Estructurador del informe de resultado de la ejecución.*

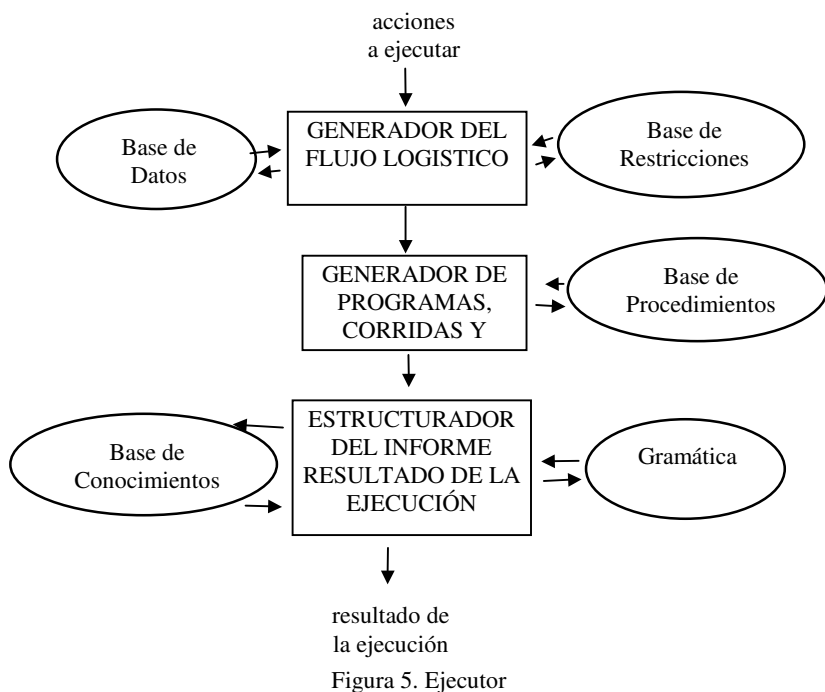


Figura 5. Ejecutor

Cabe aclarar que en todas las etapas del sistema en general se considera la posibilidad del aprendizaje y un módulo general de "olvido" que evita el crecimiento inmoderado de los datos acumulados con objeto transitorio, además de que los programas, y lenguaje de control generado se borran al término de la corrida.

## CONCLUSIONES

Se presentó un modelo para el procesamiento del Lenguaje Natural Restringido. Es innegable que pese a la complejidad descrita en el modelo, su ámbito es aún restringido aunque su

utilidad es grande no obstante el esquema es fácilmente adaptable cuando necesitamos que las salidas sean acciones ya que en lugar de información como salida serán señales para la operación de mecanismos subordinados a la computadora y se le añade que a cada paso de ejecución solicite información que se enruta como mensaje y cambia si es necesario el curso de acción del procedimiento en general.

## BIBLIOGRAFÍA

- [1] Galindo Soria, Fernando, "*Notas del Curso, Lenguajes Formales*", UPIICSA, 1981
- [2] Galindo Soria, Fernando, "*Notas del Curso, Compiladores*", UPIICSA, 1982
- [3] Galindo Soria, Fernando, "*Notas del Curso Sistemas Operativos I*", UPIICSA, 1983
- [4] Galindo Soria, Fernando, "*Notas del Curso Sistemas Operativos II*", UPIICSA, 1984
- [5] Galindo Soria, Fernando, "*Notas del Seminario de Sistemas de Bases de Datos*", SPP UPIICSA, 1984
- [6] Donovan, "*Sistemas Operativos*", ed. Diana, 1979
- [7] F. de Saussure, "*Curso de Lingüística General*", ed. Losada, 1945
- [8] Chomsky, Noam, "*Estructuras Sintácticas*", de. Siglo XXI, México, 1974
- [9] Chomsky, Noam, "*Aspectos de la Teoría de la Sintaxis*", ed. Siglo XXI, México 1974
- [10] Greene, Judith, "*Psicolingüística*", ed. Trillas, México, 1980
- [11] González Vázquez, Juan Martín, "*Diseño de un Lenguaje de Control Icónico*", UPIICSA, México, 1984.
- [12] González Vázquez, Juan Martín, "*Aspectos Elementales del Lenguaje Natural*", UPIICSA, México, 1982