

# La disciplina del Lenguaje Natural en el título de Ingeniero en Informática

Lidia Moreno / Antonio Molina  
Departamento de Sistemas Informáticos y Computación  
Universidad Politécnica de Valencia  
{lmoreno, amolina}@dsic.upv.es

## Resumen

*En este trabajo se presenta un estudio sobre la disciplina del Lenguaje Natural: relación de esta disciplina con otras, análisis de las recomendaciones realizadas por diversos organismos e instituciones nacionales e internacionales, programa teórico, programa práctico y sistema de evaluación de la asignatura de Lenguaje Natural que se imparte en el título de Ingeniero en Informática de la Universidad Politécnica de Valencia.*

## 1. INTRODUCCIÓN.

El lenguaje es uno de los aspectos fundamentales del comportamiento humano. En forma escrita sirve para transmitir el conocimiento de una generación a la siguiente durante largo tiempo. En forma hablada sirve como vehículo de comunicación principal en el comportamiento cotidiano con los demás.

El Procesamiento del Lenguaje Natural (PLN) es una parte esencial de la Inteligencia Artificial que investiga y formula mecanismos computacionalmente efectivos que faciliten la interrelación hombre/máquina y permitan una comunicación mucho más fluida y menos rígida que los lenguajes formales.

Todo sistema de PLN intenta simular un comportamiento lingüístico humano; para ello debe tomar conciencia tanto de las estructuras propias del lenguaje, como del conocimiento general acerca del universo de discurso. De esta forma, una persona que participa en un diálogo sabe cómo combinar las palabras para formar una oración, conoce los significados de las mismas, sabe cómo éstos afectan al significado global de la oración y posee un conocimiento del mundo en general que le permite participar en la conversación.

En este trabajo se realizará un estudio sobre la disciplina del Lenguaje Natural. Comenzaremos referenciando la posible relación de esta disciplina

con otras, como la Lingüística, Teoría Formal de Lenguajes, la Teoría de la Compilación, la Lógica o la Inteligencia Artificial. Además, se analizarán las recomendaciones realizadas por diversos organismos e instituciones, tanto en el ámbito nacional como internacional en lo concerniente al Lenguaje Natural. Tras este análisis inicial, se presenta brevemente el programa teórico y el práctico de la asignatura de Lenguaje Natural que se imparte en el título de Ingeniero en Informática de la Universidad Politécnica de Valencia. Para finalizar, se propone un sistema de evaluación de la asignatura.

## 2. RELACIÓN CON OTRAS MATERIAS.

Cuando hablamos de Lenguaje Natural hacemos referencia al lenguaje humano. Así pues, el Lenguaje Natural se opone a Lenguaje Formal o Artificial. Por lo tanto, pretender englobarlo de manera exclusiva en la I.A. es erróneo.

Por otra parte la Lingüística estudia y formaliza el lenguaje tanto en general como de las distintas lenguas que hablamos los humanos. Durante años el desarrollo de la Lingüística y el del Procesamiento del Lenguaje Natural se produce sin una notable transferencia de resultados en ningún sentido. Los lingüistas desarrollaban sus teorías sin pensar en una posible realización computacional de las mismas. Y los informáticos trataban las aplicaciones informáticas del Lenguaje Natural como un problema más de Ingeniería del Software. Esto no condujo a progresos en ambos campos, así que se unieron para dar sus frutos conjuntamente: en cualquier avance en la teoría lingüística se asocia su desarrollo computacional, y en cualquier sistema de Procesamiento del Lenguaje Natural subyace un modelo lingüístico.

El Procesamiento del Lenguaje Natural ha hecho uso de las técnicas desarrolladas en la Teoría Formal de Lenguajes o de la Teoría de la Compi-

lación<sup>1</sup> para procesar los Lenguajes Formales, como los Lenguajes de Programación. Pero la problemática que presenta el Lenguaje Natural es notablemente distinta a la de los Lenguajes de Programación<sup>2</sup>, por lo que hace que éstas técnicas sean insuficientes para la primera disciplina. La aplicación de técnicas provenientes de la I.A. hace que mejore la calidad y eficiencia de los sistemas actuales de Procesamiento del Lenguaje Natural.

Con relación a los mecanismos de representación tenemos que mencionar a la Lógica. Los formalismos lógicos constituyen el marco idóneo tanto para la representación como para el procesamiento de la información lingüística. Estos formalismos ofrecen una gran potencia expresiva, por su carácter declarativo, así como mecanismos eficientes si tenemos en cuenta la madurez alcanzada en el tratamiento de la unificación.

No es posible finalizar este análisis sin hacer referencia a la psicolingüística, ciencia que estudia los mecanismos humanos de comprensión y generación del lenguaje. Como muy bien se apunta en [Cortes93] haciendo referencia a la comprensión o generación del Lenguaje Natural, ¿deben las máquinas intentar reproducir las formas humanas de tratamiento, de acuerdo a lo que la psicolingüística proponga, o deben explorar sus propios métodos de procesamiento, más adecuados a sus peculiaridades?, dejando que el lector extraiga sus propias conclusiones.

### 3. RECOMENDACIONES DE ORGANISMOS E INSTITUCIONES.

Previo a la elaboración de un programa sobre una disciplina convendría conocer cuales son las recomendaciones de diversos organismos internacionales, con respecto a la formación en la materia. Nosotros hemos estudiado las recomendaciones sobre la formación en el Lenguaje Natural en los estudios de Informática realizadas por los siguien-

tes organismos e instituciones internacionales y nacionales:

- Propuesta Conjunta ACM / IEEE.

Se ha revisado el Curriculum'68 [ACM68], el Curriculum'78 [ACM79] con distintas propuestas que revisan este último [Koffman84, 85], así como ampliaciones a un segundo ciclo (Master) [ACM81], así como la propuesta elaborada por la IEEE [IEEE83] que dio lugar al documento [ACM88] y a los informes "Computing as a Discipline" [ACM89] y "Computing Curricula 1991" [ACM91]

- Recomendaciones de la Unesco.

Se ha estudiado el informe "A modular curriculum in Computer Science" [UNESCO84], publicado por el Comité Técnico de Educación TC3 de la *International Federation for Information Processing* (IFIP).

- Propuesta de la Universidad Carnegie Mellon.

Propuesta curricular que un equipo del Departamento de Computación de la Universidad de Carnegie Mellon (Pennsylvania, USA) presentó para el título en Informática, recogida en el documento [Shaw85].

- Diversas Universidades Extranjeras.

La Association for Computational Linguistics, en el año 1992, publica una relación extensa de cursos sobre lingüística computacional ofertados por distintas universidades del ámbito internacional [ACL92]. La oferta es enorme y presenta una gran variedad respecto a los objetivos (formación básica teórica y/o práctica, formación avanzada en el Lenguaje Natural o formación adyacente como fundamentos de los lenguajes formales, estadística, matemáticas y lógica para lingüistas,...), así como respecto al colectivo al que va dirigido (informáticos, lingüistas e intérpretes, estudiantes de tercer ciclo o de ciclos inferiores, en grupos de 10 a 75 alumnos), las referencias bibliográficas (generalmente coinciden en citar el texto de Allen, edición anterior a [Allen95]) y el lenguaje de programación (Lisp y/o Prolog). De entre esta oferta se han seleccionado para su análisis cursos correspondientes a tres universidades distintas: Departamento de Informática de la **Universidad Carnegie Mellon** (USA), el departamento de Lingüística Computacional de la **Universidad de Saarbrücken** (Alemania) y

<sup>1</sup> Incluso en algunos planes de estudio antiguos se impartía un módulo de Lenguaje Natural dentro de la asignatura de Compiladores.

<sup>2</sup> La comprensión y la generación del Lenguaje Natural requiere el tratamiento de gran cantidad y variedad de conocimientos, así como la utilización de técnicas inteligentes que mejoren la eficiencia.

el departamento de Informática de la **Universidad de Toronto** (Canadá).

- Diversas Universidades del Estado Español.

Se realiza un análisis del papel que se asigna al Lenguaje Natural en cada uno de los planes de estudios del título de Ingeniero en Informática de tres de las siguientes universidades públicas del Estado Español: Universidad Politécnica de Cataluña, Universidad del País Vasco [BOE95], y Universidad Politécnica de Madrid [IIUPM96].

Tras el estudio realizado sobre los documentos mencionados se pueden extraer interesantes conclusiones atendiendo a distintos parámetros:

- Tipo de materia y ciclo:

Generalmente aparece como una materia opcional, con un conjunto de temas básicos para introducir al alumno en la materia (segundo ciclo), y también se ofrece al alumno la posibilidad de profundizar en el tema, en función de sus intereses académicos y profesionales (tercer ciclo).

- *Tiempo estimado:*

Se considera adecuado un total de 6 créditos, lo que equivale a unas 60 horas.

- *Contenidos:*

Hay un consenso en todas las recomendaciones respecto a los conocimientos en términos generales, ya que es evidente que los contenidos dependerán de los créditos asignados a la asignatura y de la evolución científica de la misma. Actualmente aparecen dos temas innovadores, respecto a las recomendaciones de los años 80. Éstos son las gramáticas lógicas, los analizadores probabilísticos y los diccionarios.

Un resumen de las recomendaciones de los distintos organismos, apuntando la valoración asociada a los parámetros anteriores puede verse en la tabla 1.

Organismo	Materia <sup>3</sup> Ciclo Créditos	Contenidos
ACM/ IEEE	B + A - -	Sintaxis, Semántica, T.A. <sup>4</sup>
UNESCO	- - 20% IA	Análisis, Gramáticas transformacionales, ATN, Generación, T.A.
Carnegie Mellon	B + A - 15	Sintaxis, ATN, Gramáticas basadas en Unificación, Charts, desplazamiento-reducción, Análisis conexionista, Análisis Robusto, Semántica, Resolución referencias, Diálogo y discurso, Comprensión de textos, Adquisición del lenguaje, T.A. y Generación.
Saarbrücken	B - 4.5	Procesamiento habla, Morfología, Sintaxis, Chart, Gramáticas basadas en Unificación, Semántica y Pragmática.
Toronto	B - 4.5	Diccionario y Morfología, Sintaxis, ATN, Chart, Gramáticas basadas en Unificación y lógicas, Semántica, Pragmática, discurso cooperativo, Adquisición del lenguaje, T.A.
U.P.Cataluña	O + OP 2° 1+6	Léxico, Sintaxis, Semántica, ATN, Gramáticas basadas en Unificación, Lógicas y de rasgos, Charts, Análisis probabilístico, Pragmática y Generación.
U.P.Vasco	OP 2° 6	Léxico, Sintaxis, Semántica, Contexto, Generación.
U.P.Madrid	OP 2° 6	Sintaxis, Semántica, Contexto, Generación, T.A., Diálogo.

Tabla 1: Resumen de recomendaciones.

#### 4. PROGRAMA DEL CURSO.

En el plan de estudios del título de Ingeniero en Informática de la Universidad politécnica de Valencia se encuentra la asignatura *Lenguaje Natu-*

<sup>3</sup> B= Básica, A= Avanzada, T= Troncal, O= Obligatoria, OP= Opativa.

<sup>4</sup> Traducción Automática

ral dentro de la intensificación de Inteligencia Artificial formando parte de la especialidad de Software.

Ésta es una asignatura optativa de 4.5 créditos que se oferta al alumno en el segundo ciclo.

Atendiendo a la información analizada y expuesta en el punto anterior se ha elaborado una propuesta de programa que se ha puesto en marcha por primera vez en el presente curso 97/98.

A continuación, se presenta una propuesta de objetivos generales a alcanzar. A partir de los mismos se desarrolla una distribución temporal de la asignatura diferenciando entre teoría y prácticas. Y finalmente, se propone el sistema de evaluación de la asignatura.

#### 4.1. OBJETIVOS GENERALES.

Los objetivos generales a conseguir en la asignatura *Lenguaje Natural* se pueden resumir en los siguientes:

- Una sólida formación en los aspectos teóricos básicos que permitan una fácil asimilación de nuevas técnicas y métodos.
- Centrar el Lenguaje Natural en el ámbito de la Informática.
- Conocer los problemas del Lenguaje Natural y las técnicas básicas utilizadas en la resolución de los mismos.
- Conocer las aplicaciones del Lenguaje Natural
- Leer y comprender la literatura asociada al Lenguaje Natural .

Con objeto de alcanzar los objetivos propuestos se han distribuido los 4.5 créditos correspondientes a la asignatura en 3 créditos teóricos y 1.5 créditos prácticos. Lo que supone la siguiente distribución temporal:

- **Teoría** : 30 horas lectivas equivalentes a 2 horas semanales,
- **Práctica**: 15 horas lectivas equivalentes a 2 horas en semanas alternas,

impartidas en el semestre noveno o en el décimo. Es decir, durante un semestre correspondiente a quinto curso.

#### 4.2. PROGRAMA DE TEORÍA.

Con objeto de alcanzar los objetivos definidos en el espacio temporal previsto en el apartado anterior se propone el siguiente programa, que posteriormente se describe en la tabla 2.

Tema	Denominación	Horas
1	Preliminares del LN	7
2	Análisis Léxico	4
3	Análisis Sintáctico	7
4	Interpretación Semántica	6
5	Interpretación Contextual	4
6	Generación del lenguaje	2

Tabla 2: Programa de teoría.

##### Tema 1. Preliminares del Lenguaje Natural

###### Objetivos:

- Estudiar los conceptos básicos del L.N. y las fases de análisis en el proceso de comprensión de éste.
- Entender los problemas fundamentales de ambigüedad del lenguaje y estrategias de resolución de éstos.
- Conocer algunas aplicaciones del procesamiento del L.N.

###### Contenidos:

1. Introducción al L.N.: orígenes y tendencias actuales.
2. Fases de análisis y ambigüedad en el lenguaje.

##### Tema 2. Análisis Léxico.

###### Objetivos:

- Conocer la problemática de la construcción de Diccionarios.
- Determinar la representación de las informaciones léxicas.
- Aprender técnicas básicas de adquisición léxica.
- Optimizar el acceso a diccionarios

###### Contenidos:

1. Información léxica.
2. Representación de un lexicón.

3. Morfología.
4. Restricciones semánticas.

**Tema 3. Análisis Sintáctico.**

*Objetivos:*

- Conocer los componentes básicos de la sintaxis del lenguaje.
- Entender la necesidad de un formalismo para representar las estructuras del lenguaje con independencia de las estrategias de análisis.
- Desarrollar distintos algoritmos de análisis.
- Conocer las tendencias actuales respecto a formalismos gramaticales.

*Contenidos:*

1. Modelos de representación.
2. Algoritmos de análisis.

**Tema 4. Interpretación Semántica.**

*Objetivos:*

- Describir los elementos necesarios en la interpretación semántica.
- Conocer diversos lenguajes formales que permitan formular Formas Lógicas.
- Aprender los mecanismos de composición de la Forma Lógica.

*Contenidos:*

1. Especificación de la fase de interpretación semántica.
2. Lenguajes formales para la representación del significado.
3. Definición de un Lenguaje de Formas Lógicas basado en el cálculo de predicados.
4. Proceso de interpretación semántica: Composicionalidad.

**Tema 5. Interpretación Contextual.**

*Objetivos:*

- Estudiar los conceptos básicos de la interpretación contextual.
- Entender la necesidad de delimitar el marco de discurso.

- Conocer la problemática de la elipsis y la anáfora.

*Contenidos:*

1. Representación del conocimiento: contexto local y conocimiento del mundo.
2. Contexto de discurso local y referencia: anáfora y elipsis.

**Tema 6. Introducción a la generación del Lenguaje.**

*Objetivos:*

- Determinar posibles aplicaciones de la generación de L.N.
- Conocer las características y tareas de un generador de L.N.
- Especificar una arquitectura de sistemas de generación.
- Estudiar una interfaz de L.N.

*Contenidos:*

1. Objetivos, tareas y áreas de aplicación de un generador de lenguaje natural.
2. Fases del proceso de generación: profunda y superficial.
3. Tipos de generadores de lenguaje natural.
4. Conexión entre análisis y generación.

**Bibliografía Básica:**

[Allen95], [Gazdar89], [Sproat92], [Shieber86], [Zock88], [Covington94].

**4.3. PROGRAMA DE PRÁCTICAS.**

El objetivo del programa de prácticas consiste en proporcionar las herramientas necesarias para la construcción de un sistema de procesamiento del lenguaje natural. Este sistema se restringirá a un subconjunto de la gramática del castellano que, a partir de la oración escrita proporcione la interpretación semántica, en base a un lenguaje de forma lógica particular.

En particular, se desea potenciar la capacidad de abstracción del estudiante, para que sea capaz de solucionar un problema común desde distintas perspectivas en función del modelo que se utilice. Aprenderá así a captar los aspectos esenciales de

un problema, independientemente de la solución seleccionada que estará en función del entorno particular de implementación.

Los módulos básicos de un sistema o interfaz de lenguaje natural se representan en el esquema de la *Figura 1*. La entrada al sistema es una oración escrita en lenguaje natural, la cual es segmentada por el *tokenizador*<sup>5</sup> en unidades léxicas o *tokens*. La secuencia de unidades léxicas es procesada por el *módulo de análisis sintáctico-semántico*, el cual verifica si la oración es correcta gramaticalmente y, además, si su significado es consistente respecto de un dominio específico. Para aquellas oraciones válidas el módulo de análisis proporciona el árbol sintáctico y la forma lógica de la oración. La información léxica, morfológica y semántica necesaria para llevar a cabo el análisis se almacena en el *diccionario* o léxico. La *ontología* define la semántica del dominio.

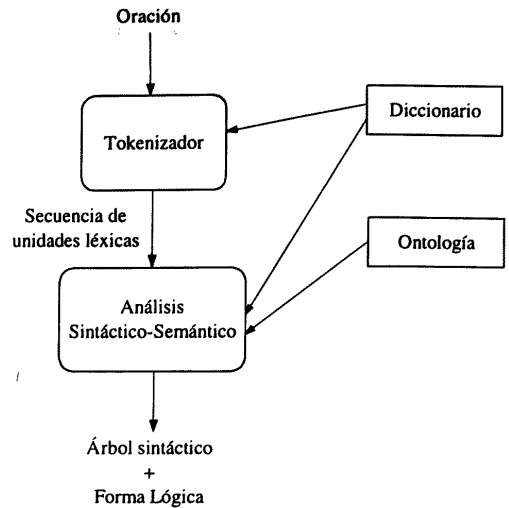
El tokenizador, la ontología de rasgos semánticos y los módulos de verificación semántica se proporcionarán al alumno.

Para la implementación puede utilizar el Sicstus Prolog 3.0 para windows 95 instalado en el laboratorio, o cualquier otro intérprete de Prolog, siempre que programe en 'Prolog estándar'.

La llamada al **tokenizador** es: *tokeniza(Term,L)* donde *Term* es el oración y *L* se instancia a la lista de unidades léxicas. Como ejemplo, la llamada *tokeniza('¿Qué ríos desembocan en el Mediterráneo?', L)* se resuelve con *L = [punc(191), may(qué), ríos, desembocan, en, el, may(mediterráneo), punc(63)]*

<sup>5</sup> El alumno puede utilizar el *tokenizador* que se encuentra en el directorio (página web) de la asignatura. Este módulo, implementado en Prolog, segmenta una oración en una lista de unidades léxicas: palabras, símbolos de puntuación y números positivos. La oración de entrada será un término Prolog, que previamente habrá sido leído desde el teclado o desde un archivo. Los símbolos de puntuación se traducen en términos de la forma **punc**(*n*), donde *n* es el código ascii del símbolo, las palabras que comienzan por mayúscula en **may**(*palabra*) y los números en **num**(*número*).

Si aparece algún símbolo no válido el tokenizador indica que el análisis léxico ha sido fallido.



*Figura 1: Esquema básico de un sistema de procesamiento de lenguaje natural.*

Cada entrada léxica del **diccionario** se corresponde con un predicado Prolog. Las distintas categorías léxicas que se utilizan corresponden con las especificadas en el estándar Parole, utilizado en diversos proyectos europeos de investigación relacionados con sistemas de procesamiento del lenguaje y etiquetado léxico de textos. Básicamente las entradas del diccionario recogen:

1. categoría léxica de cada palabra.
2. información morfológica asociada a cada palabra, necesaria para llevar a cabo el análisis sintáctico.
3. información semántica, para verificar si una oración es válida semánticamente.
4. significado asociado a cada palabra, mediante una forma lógica.

Para construir el analizador sintáctico son necesarios los tres primeros elementos. La forma lógica no necesita introducirla inicialmente, ya que es necesaria en el analizador semántico.

Además puede considerar algunas secuencias de palabras como unidades léxicas, lo que simplifica la fase de análisis. Por ejemplo, *comunidad autónoma*, *metro cúbico por segundo*, *Sistema ibérico*, etc. pueden verse como una única palabra.

Como ya se ha mencionado, se propone la construcción de un **analizador sintáctico** que reconozca un subconjunto gramatical del castella-

no, especificado a través de una Gramática de Cláusulas Definidas (DCG), según se explica en el tema 3 del programa de teoría. Para definir la gramática el alumno debe estudiar el conjunto de frases (*corpus*) proporcionado por el profesor. Este corpus es el que determina el dominio de la aplicación: en este caso se trata de una aplicación geográfica que maneja información sobre ríos, sistemas montañosos, mares y comunidades autónomas del territorio español. La gramática definida debe ser completamente independiente del dominio, aunque para verificar la validez semántica se debe hacer uso de la ontología de tipos específica. Un cambio del dominio de la aplicación no debe suponer un cambio de la gramática, sino de la ontología semántica y del diccionario.

El analizador debe aceptar aquellas oraciones que sean sintáctica y semánticamente correctas, en cuyo caso proporcionará el árbol de análisis, y rechazar las incorrectas.

Se debe verificar la concordancia de género y número.

Se incorporará una **ontología de tipos** para validar las oraciones semánticamente. Para la verificación de la compatibilidad entre tipos semánticos el alumno utilizará el módulo que le será proporcionado. El predicado *verificar(Rasgos, Restricciones)* verifica si la lista de *Rasgos* de una entidad candidata es compatible con la lista de *Restricciones* exigidas por otra entidad.

Por ejemplo, en la oración 'El río Turia pasa por Valencia' se debe verificar que, tanto el tipo semántico asociado al sujeto ('El río Turia'), como al objeto ('por Valencia') sean compatibles con las restricciones impuestas por el verbo 'pasar'. Según la ontología proporcionada el nombre 'río' tendrá asociado el tipo [río], 'Valencia' el tipo [autonomía] y el verbo 'pasa' impondrá al sujeto la restricción [río] y al objeto la restricción [atravesado]. Para verificar la consistencia semántica de la oración se realizarán dos llamadas al módulo de verificación: *verificar([río], [río])* y *verificar([autonomía], [atravesado])*.

Para aquellas oraciones correctas el sistema debe proporcionar el significado de la oración, en base al lenguaje de la forma lógica especificado en el tema 4 del programa de teoría: **Interpretación Semántica**.

La composición de la forma lógica asociada a una oración se realizará durante el proceso de aná-

lisis sintáctico, mediante unificación, según se explica en el tema 4.

Deberá incorporar a las reglas gramaticales implementadas los argumentos necesarios para recoger y componer la significado de cada estructura sintáctica. Además deberá completar las entradas léxicas con la forma lógica asociada a cada palabra. Por ejemplo, para la frase de entrada: '¿el mar Cantábrico baña las costas gallegas?', obtendríamos la forma lógica siguiente: 'sí\_no(existe(Y,mar1(Y)&Y=cantábrico,todo(X,costa2(X,galicia),bañar(Y,X)))'

#### 4.4. SISTEMA DE EVALUACIÓN.

Las tareas obligatorias a realizar por el alumno serán: la implementación en **Prolog** del módulo de *análisis sintáctico-semántico* y la construcción del *diccionario*, conforme se ha explicado en el apartado anterior, en base a los contenidos teóricos de la asignatura.

Se proponen una serie de ampliaciones entre las cuales el alumno podrá escoger una o varias para obtener una calificación superior:

- 1 Implementación de alguna **estrategia de análisis ascendente** explicadas en el tema 3, haciendo un estudio de eficiencia comparativo respecto a la estrategia descendente.
- 2 Construcción del diccionario mediante la estructura de datos **trie**, según se explica en el tema 2.
- 3 **Ampliación de la cobertura gramatical**, según los corpus proporcionados por el profesor, que permita que el analizador reconozca:
  - 3.1 interrogativas de objeto
  - 3.2 interrogativas de complemento preposicional
  - 3.3 oraciones de relativo
  - 3.4 tratamiento de la coordinación
- 4 Integración de herramientas en una **página web**, que permita:
  - Selección desde una lista de un analizador determinado,
  - Selección del dominio de aplicación: diccionario y ontología semántica
  - Introducción de la oración,

Ejecución del analizador seleccionado para la oración de entrada,

Visualización del árbol sintáctico y la interpretación semántica, de la oración introducida.

## 5. BIBLIOGRAFÍA.

- [Allen95] J. Allen. Natural Language Understanding. 2nd.ed. Benjamin Cummings series in Computer Science. 1995
- [ACL92] Survey of Computational Linguistics Courses. Association for Computational Linguistics, 1992.
- [ACM68] ACM Curriculum Committee on Computer Science. "Curriculum 68. Recommendations for academic programs in Computer Science", Communications of the ACM, vol.11, no.3, pag. 151-197, mar. 1968.
- [ACM79] ACM Curriculum Committee on Computer Science. "Curriculum 78. Recommendations for the undergraduate program in Computer Science", Communications of the ACM, vol.22, no.3, pag. 147-166, mar. 1979.
- [ACM81] Curriculum Committee on Computer Science. "Recommendations for master's level programs in Computer Science", Communications of the ACM, vol. 24, no. 3, pag. 115-123, mar. 1981.
- [ACM88] Task Force on the core of Computer Science. "Computing as a discipline", ACM Press, Baltimore, 1988.
- [ACM89] Task Force on the core of Computer Science. "Computing as a discipline", Communications of the ACM, vol. 32, no. 1, pag. 9-23, jan. 1989.
- [ACM91] Computing Curricula 1991. Report of the ACM/IEEE-CS Joint Curriculum Task Force on the core of Computer Science. ACM Press/IEEE Computer Society Press, 1991.
- [BOE95] Planes de estudios de la Universidad del País Vasco conducentes a las titulaciones de Ingeniero en Informática. Boletín Oficial del Estado. 13 de enero de 1993.
- [Briscoe91] T. Briscoe. Lexical Issues in Natural Language Processing. En Natural Language and Speech. E. Klein, F. Veltman Eds. Springer-Verlag, 1991
- [Cortes93] U. Cortés et al. Inteligencia Artificial. Edicions UPC, 1993.
- [Covington94] M.A. Covington. Natural Language Processing for Prolog Programmers. Prentice Hall. 1994
- [Gazdar89] G. Gazdar, C. Mellish. Natural Language Processing in PROLOG. Addison-Wesley. 1989
- [IEEE83] Computer Society Educational Activities Board. Model program in Computer Science and Engineering, IEEE Computer Society Press, 1983.
- [IIUPM96] Plan de Estudios de Ingeniero en Informática. Universidad Politécnica de Madrid. Junio 1996.
- [Iglesias93] C.A. Iglesias, Introducción a la generación del lenguaje natural. Informática y Automática, vol. 26-2, 1993
- [Koffman84] E.B. Koffman, P.L. Miller, C.E. Wardle. "Recommended curriculum for CS1, 1984", Communications of the ACM, vol.27, no.10.
- [Koffman85] E.B. Koffman, D. Stemple, C.E. Wardle. "Recommended curriculum for CS2, 1984", Com. of the ACM, vol.28, no.8.
- [Shaw85] M. Shaw (ed.). The Carnegie Mellon curriculum for undergraduate Computer Science, Springer-Verlag, 1985.
- [Shieber86] S. Shieber. An Introduction to Unification-Based Approaches to Grammar. CSLI Lecture Notes n. 4, 1986
- [Sproat92] R. Sproat. Morphology and Computation. The MIT Press, 1992.
- [UNESCO84] Unesco-IFIP. A modular curriculum in Computer Science, 1984.
- [Winograd83] T. Winograd. Language as a Cognitive Process, Syntax. Addison-Wesley, 1983.
- [Zock88] M. Zock, G. Sabah. Eds. Advances in Natural Language Generation. Vol I. Printer Publishers, London. 1988