

UNIVERSIDAD DE LOS ANDES  
MÉRIDA – VENEZUELA  
2005

## Introducción a la lingüística computacional Programa de postgrado en Lingüística Programa de postgrado en Modelado y Simulación de Sistemas

Profesores

Dra. Melva J. Márquez R.

Dr. Jacinto A. Dávila Q.

Dr. Luis Raúl Isea

### Introducción

El presente curso tiene como objetivo principal presentar a los estudiantes de postgrado en las áreas de la lingüística y la informática, una panorámica teórica y práctica de la lingüística computacional (LC) y de las tecnologías de la información empleadas en el procesamiento del lenguaje natural (PLN), especialmente de textos escritos.

Para ello, nos proponemos lograr los siguientes objetivos específicos:

1. Dar a conocer la evolución que ha tenido la lingüística computacional desde sus inicios
2. Introducir conceptos fundamentales que permitan su familiarización: gramáticas formales, formalismos gramaticales, parsing, corpus, etiquetaje, etc.
3. Comprender los fundamentos, posibilidades y limitaciones de los modelos computacionales del lenguaje que dominan actualmente: el simbólico y el probabilístico.
4. Introducir a los estudiantes en el PLN a través de dos lenguajes de programación de alto nivel: Prolog, de orientación lógica y Python, de orientación a objetos.

### Temario

1. Presentación del curso. Delimitación del campo de la lingüística computacional.
2. Breve historia de la LC. El procesamiento del corpus lingüístico: Papel del corpus es la investigación lingüística y su importancia para las industrias de la lengua; características; diseño; representatividad; codificación; etiquetarios.
3. Recuperación y extracción de información.
4. Técnicas de procesamiento del lenguaje. La representación del conocimiento lingüístico: aproximación simbólica; aproximación estadística
5. Algoritmos y pseudocódigos. Formalismos gramaticales.
6. Gramáticas formales. Tipos. Método de parsing.
7. Estructura de un PLN simbólico: Introducción al PROLOG
8. PROLOG /Ejercicios
9. Modelos probabilísticos. Introducción al lenguaje Python (entre la orientación a objetos y el script)
10. Lenguaje Python /Ejercicios

### Actividades

Los temas se darán de la siguiente manera: En primer lugar, se expondrá el tema y luego se debatirán aquellos aspectos que resulten de interés para los participantes. Los ejercicios se realizarán tanto dentro como fuera de clase.

## Evaluación

La nota final del curso será obtenida a partir de la realización de dos ejercicios finales que se planteen en las clases prácticas (40% para cada ejercicio final) y de un trabajo escrito de reflexión (20%) sobre uno de los temas planteados durante las clases.

## Duración del curso

El curso se dictará 3 horas/semana durante 10 semanas los días miércoles p.m., a excepción de los temas 7, 8, 9 y 10, cuyas horas serán dictadas de mutuo acuerdo con el profesor.

## Bibliografía recomendada

### Tema 1

Abaitua, J. (1999): *La lingüística computacional*. Página web:  
<http://www.serv-inf.deusto.es/abaitua/konzeptu/cl2.htm>

Gómez Guinovart, X. (2000): "Perspectivas de la lingüística computacional", *Novática: Revista de la Asociación de Técnicos de Informática*, Número especial del 25 aniversario (*Horizonte 2025*), 145, 85-87. Publicación electrónica en: <http://www.ati.es/novatica/2000/145/javqom-145.pdf>

Jurafsky, Daniel y Martin, James H. (2000) *Speech and Language Processing - An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. (S&LP), New Jersey: Prentice Hall, Upper Saddle River. (URL) "Chapter 1. Introduction", Jurafsky y Martin (2000) (URL)

### Tema 2

Abadal, E. El control de la terminología en la recuperació de la informació. En Cabré, M. Teresa, Lluís Codina y Rosa Estopa (Eds.) (2000). *Terminología i documentació*. Barcelona (España): Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra.

Biber, Douglas, Susan Conrad y Randi Reppen (1998). *Corpus linguistics: investigating language structure and use*. Cambridge: Cambridge University Press.

Buchanan, B. G. (2002-2003): "Brief History of Artificial Intelligence", A chronology of significant events in the History of AI initially prepared for use in Introduction to AI classes. Publicación electrónica en: <http://www.aaai.org/Pathfinder/bbhist.html>

Martí Antonín, M. A. y Llisterri, J. (2001): "La ingeniería lingüística en la sociedad de la información", *Digithum, Revista digital d'humanitats*, 3, UOC. Publicación electrónica en: <http://www.uoc.edu/humfil/articles/esp/llisterri-marti/llisterri-marti.html>

McEnery, Tony y Andrew Wilson (2001) *Corpus Linguistics, an introduction*. 2<sup>nd</sup> ed.- Edinburgh: Edinburgh University Press. (Edinburgh Textbooks in Empirical Linguistics).

Moreno Sandoval, A. (1998): *Lingüística computacional. Introducción a los modelos simbólicos, estadísticos y biológicos*, Madrid: Síntesis.

### Tema 3

Clarke, I., Sandberg, O., Wiley, W., y Hong, T. Freenet: A Distributed Anonymous Information Storage and Retrieval System.

Gonzalo, J. (2003) *Recuperación y extracción de información*. En M. A. Martí Antonín (coord.), *Tecnologías del lenguaje*, Barcelona: Edicions de la UOC, 157-192.

Krzysztof G., y Kulkarni, A. (2001) *FIRE: An Information Retrieval Interface for Intelligent Environments*. Verona. *Proceedings of International Workshop on Information Presentation and Natural Multimodal Dialogue*.

### Tema 4

Badia Cardús, T. (2003): "Técnicas de procesamiento del lenguaje", en M. A. Martí Antonín (coord.), *Tecnologías del lenguaje*, Barcelona: Edicions de la UOC, 193-248.

Llisterra, J. (2003): "[Lingüística y tecnologías del lenguaje](http://liceu.uab.es/~joaquim/publicacions/TecnoLing_Lynx02.pdf)", *Lynx. Panorámica de Estudios Lingüísticos*, Departament de Teoria dels Llenguatges, Universitat de València, 2, 9-71. Publicación electrónica: [http://liceu.uab.es/~joaquim/publicacions/TecnoLing\\_Lynx02.pdf](http://liceu.uab.es/~joaquim/publicacions/TecnoLing_Lynx02.pdf)

Martí Antonín, M. A. (coord.) (2003): *Tecnologías del lenguaje*, Barcelona: Editorial UOC.

Manning, D. D. y Schütze, H. (1999): *Foundations of Statistical Natural Language Processing*, Cambridge, Mass./London, England: The MIT Press. Capítulo 1 en disponible en PDF: <http://www.aclweb.org/archive/fsnlp-ch1.pdf> y otros materiales disponibles de forma electrónica: <http://nlp.stanford.edu/fsnlp/>

Moreno Sandoval, A. (1998): *Lingüística computacional. Introducción a los modelos simbólicos, estadísticos y biológicos*, Madrid: Síntesis.

### Tema 5

Material suministrado por el profesor. Ejercicios en clase.

### Tema 6

Pollard C., y Sag, I. (1993) *Head-Driven Phrase Structure Grammar*. Stanford, CA: CSLI Lecture Note Series, Center for the Study of Language and Information

Shieber, S. M. (1989 [1986]): *Introducción a los formalismos gramaticales de unificación*, Barcelona: Teide [An introduction to unification-based approaches to grammar, Stanford, Ca.: CSLI]

Sells, P. (1985) *Lectures on Contemporary Syntactic Theories*. Stanford, CA: CSLI Lecture Note Series, Center for the Study of Language and Information

### Tema 7

Convington, M. (1994) *Natural Language Processing for Prolog Programmers*. Prentice Hall: Englewood Cliffs.

Gazdar, G. y Mellish, Ch. (1989): Natural Language Processing in PROLOG. An Introduction to Computational Linguistics, Wokingham, England: Addison-Wesley.

## Tema 8

Ejercicios

## Tema 9

Sitio principal de Python ([URL](#))

Tutoriales en Español ([URL](#)) ([URL Otro!](#)) y otros recursos ([URL Inglés](#)) ([URL Español](#))

Tutorial para quienes no tenga experiencia programando ([URL](#))

Otros libros y recursos sobre Python ([URL](#))

## Tema 10

Ejercicios