

17-18

MÁSTER UNIVERSITARIO EN CIENCIA
DEL LENGUAJE Y LINGÜÍSTICA
HISPÁNICA

GUÍA DE ESTUDIO PÚBLICA



LINGÜÍSTICA COMPUTACIONAL

CÓDIGO 24402743



Ámbito: GUJ - La autenticidad, validez e integridad de este documento puede ser verificada mediante el "Código Seguro de Verificación (CSV)" en la dirección <https://sede.uned.es/valida/>



7F3AD149536BC3E238F6ACC7179D183A

17-18

LINGÜÍSTICA COMPUTACIONAL
CÓDIGO 24402743

ÍNDICE

PRESENTACIÓN Y CONTEXTUALIZACIÓN
REQUISITOS Y/O RECOMENDACIONES PARA CURSAR ESTA ASIGNATURA
EQUIPO DOCENTE
HORARIO DE ATENCIÓN AL ESTUDIANTE
COMPETENCIAS QUE ADQUIERE EL ESTUDIANTE
RESULTADOS DE APRENDIZAJE
CONTENIDOS
METODOLOGÍA
SISTEMA DE EVALUACIÓN
BIBLIOGRAFÍA BÁSICA
BIBLIOGRAFÍA COMPLEMENTARIA
RECURSOS DE APOYO Y WEBGRAFÍA



Nombre de la asignatura	LINGÜÍSTICA COMPUTACIONAL
Código	24402743
Curso académico	2017/2018
Títulos en que se imparte	MÁSTER UNIVERSITARIO EN CIENCIA DEL LENGUAJE Y LINGÜÍSTICA HISPÁNICA
Tipo	CONTENIDOS
Nº ETCS	5
Horas	125.0
Periodo	SEMESTRE 2
Idiomas en que se imparte	CASTELLANO

PRESENTACIÓN Y CONTEXTUALIZACIÓN

Presentación del equipo docente

Fernando Sánchez León es doctor en Filosofía y Letras por la Universidad Autónoma de Madrid (1997), universidad en la que ha sido profesor de Lingüística General y de Lingüística Computacional. Actualmente, es miembro del Departamento de Tecnología de la Real Academia Española, institución a la que está vinculado, bien como colaborador, bien como personal de plantilla, desde 1996. Asimismo, ha impartido cursos de doctorado sobre Lingüística Computacional en el programa de doctorado de Lingüística Teórica y sus Aplicaciones en el I. U. Ortega y Gasset, dependiente de la Universidad Complutense de Madrid (cursos 1999-2000, 2001-2002, 2002-2003, 2004-2005 y 2005-2006). Ha sido Honorary Research Fellow en el Department of Linguistics and Modern English Language de la Universidad de Lancaster (Reino Unido) durante el período académico 1998-2002. Trabaja desde hace casi 30 años en Lingüística Computacional, disciplina en la que se ha formado gracias a su participación en proyectos europeos dedicados a la traducción automática (EUROTRA, 1987-1992), corrección gramatical (GramCheck, 1994-1995) y lingüística de corpus (CRATER, 1994-2005). En los últimos años, ha centrado su actividad en el desarrollo de herramientas y recursos de calidad para el procesamiento lingüístico automático en español, en sus dimensiones sincrónica y diacrónica (CREA y CORDE, 1995-2000), así como en el desarrollo de técnicas de extracción de términos a partir de corpus especializados (CCT, 2003-2006). Asimismo, trabaja en la relación entre lexicones computacionales y diccionarios electrónicos en el contexto de las obras lexicográficas de la Real Academia Española.

Este es un curso en el que se pretende abordar los problemas computacionales del procesamiento lingüístico, fundamentalmente desde un punto de vista teórico, aunque sin renunciar a que el alumno tenga una experiencia práctica en el uso de herramientas y recursos habituales en la Lingüística Computacional.



REQUISITOS Y/O RECOMENDACIONES PARA CURSAR ESTA ASIGNATURA

Se espera que el alumno tenga un conocimiento medio de los tradicionales niveles de descripción de la gramática, sus problemas y métodos. Asimismo, el alumno deberá tener capacidad lectora en inglés, pues buena parte del material didáctico está escrito en esta lengua. Finalmente, para la realización de ejercicios y prácticas, el alumno debe disponer de un ordenador.

EQUIPO DOCENTE

Nombre y Apellidos
Correo Electrónico
Teléfono
Facultad
Departamento

RICARDO MAIRAL USON
rmairal@flog.uned.es
8645/6819
FACULTAD DE FILOLOGÍA
FILOLOGÍAS EXTRANJERAS Y SUS LINGÜÍSTICA

COLABORADORES DOCENTES EXTERNOS

Nombre y Apellidos
Correo Electrónico

FERNANDO SÁNCHEZ LEÓN
fsanchez@invi.uned.es

HORARIO DE ATENCIÓN AL ESTUDIANTE

El alumno será asesorado a lo largo de todo el curso a través de tutorías, en las que se le ayudará a resolver cualquier duda relativa a la materia impartida. Con este fin, se establece el horario que se indica más adelante, durante el cual el alumno podrá consultar telefónicamente al profesor cualquier asunto que sea de su interés. Asimismo, el profesor contestará a la mayor brevedad posible cualquier consulta que le sea remitida por correo electrónico.

Horario de atención al alumno

Fernando Sánchez León

Días de la semana: Martes y Jueves

Horas: 16:30 a 18:30 horas

Dirección postal: Departamento de Tecnología

Centro de Estudios de la Real Academia Española

c/ Serrano, 187-189

28002 Madrid

Teléfono: 91 745 55 35

Correo electrónico: fsanchez@rae.es



COMPETENCIAS QUE ADQUIERE EL ESTUDIANTE

RESULTADOS DE APRENDIZAJE

El curso pretende que el estudiante conozca los hitos más relevantes de la Lingüística Computacional, los problemas que se plantea y las aplicaciones que se desarrollan en esta área de investigación. Pretende, asimismo, que se familiarice con la formalización de problemas lingüísticos y para ello profundice en el estudio (y en la experimentación) de los formalismos de representación del conocimiento lingüístico más utilizados en la actualidad, resolviendo problemas concretos en varios de los niveles de descripción gramaticales.

El programa del curso persigue los siguientes objetivos en relación con la adquisición de conocimientos:

1. Que el estudiante revise y comprenda los fundamentos, problemas y métodos de los modelos computacionales del lenguaje dominantes en la actualidad: el probabilístico y el simbólico, con especial hincapié en el segundo de ellos.
2. Que conozca someramente la historia de la Lingüística Computacional, así como las aplicaciones más reseñables del Procesamiento del Lenguaje Natural.
3. Que estudie y, por tanto, conozca en profundidad, los formalismos más usuales para la representación del conocimiento lingüístico en los niveles morfológico, sintáctico y semántico.

CONTENIDOS

METODOLOGÍA

El curso incluye dos tipos de materiales, unos preparados expresamente para el mismo y otros que introducirán al alumno en otras tantas áreas de investigación dentro de la disciplina o que complementan el primer tipo de materiales. El objetivo de ambos es la transmisión de conocimientos teóricos. Sin embargo, los materiales del primer tipo incluyen ejercicios prácticos y proponen problemas cuya resolución será responsabilidad del alumno. Se pretende con ello una participación activa del alumno en el proceso de aprendizaje y un fortalecimiento de los aspectos prácticos del mismo, que permitan un seguimiento del proceso de aprendizaje adecuado por parte del profesor. Asimismo, el trabajo final de implementación de una gramática reúne, de algún modo, una parte importante de los saberes que el curso pretende transmitir al alumno, de forma que su realización pasa por la asimilación y la puesta en práctica de buena parte de los conocimientos adquiridos en el curso, a la vez que permite al alumno desarrollar sus aspectos más creativos en la elección y resolución de un problema de sintaxis de las lenguas naturales.

El estudiante debe enfocar la asignatura como un proceso continuo y constante, al que debe dedicar, si es posible todos los días, un tiempo razonable. Este enfoque favorecerá la comprensión y la capacidad de reflexión sobre el material que se le entrega. De hecho,



algunos materiales son relativamente densos, tanto en la exposición, como en el tema mismo que abordan, lo que requiere una lectura pausada de los mismos y no un intento de asimilación veloz y voraz de los contenidos.

En general, el material no se ha desarrollado para el estudio memorístico sino más bien para la reflexión y el desarrollo de la capacidad de análisis y el razonamiento crítico. A estos objetivos contribuirá un tratamiento pausado y constante del material.

Se recomienda vivamente la lectura del material en el orden que se sugiere, pues la presentación de conceptos y problemas es incremental, por lo que otros itinerarios pueden llevar a la incomprensión y, por ende, al desaliento del estudiante.

Se recomienda seguir las pautas habituales de estudio, de forma que el trabajo con los textos se reduzca a las tareas de:

a) identificación de la estructura lógica del texto, con objeto de ir anticipando las cuestiones que el capítulo plantea;

b) lectura(s) del texto, buscando las ideas principales, los conceptos nuevos, su definición y su relación con los conceptos anteriores, del texto o de toda la asignatura;

c) resumen de las ideas principales del texto, realizando una reflexión, en los términos propios del alumno de las aportaciones del mismo;

d) asimilación de estas ideas, intentando fijar los conceptos fundamentales, relacionándolos con los conceptos aprendidos anteriormente y reproduciéndolos mentalmente;

e) realización de los ejercicios y resolución de los problemas planteados en todos aquellos capítulos que los incluyen. Unos y otros deberán realizarse después de haber leído y asimilado todos los conceptos que plantea el capítulo correspondiente;

f) relectura del texto: esta segunda lectura permitirá al alumno asimilar la terminología del dominio. Efectivamente, solo tras haber comprendido y asimilado los conceptos nuevos, se está en condiciones de asegurar la adecuada asimilación terminológica, que es, por otra parte, una de las características del discurso vinculado al conocimiento de una materia.

Una vez terminado el estudio de un tema y realizados los ejercicios y resueltos los problemas (si procede), el estudiante estará en condiciones de plantear sus dudas al profesor de la asignatura y de hacerle llegar su solución a los problemas o el resumen razonado del tema estudiado. Es decir, todos los temas conllevan trabajo que debe entregarse: para aquellos que no contienen ejercicios deberá realizarse el resumen razonado de ideas principales; en aquellos que contienen ejercicios, la resolución de estos será prueba de comprensión de los contenidos del tema. El profesor atenderá las dudas y comentará su solución a los problemas.

Cuando el estudiante inicie el estudio del material didáctico correspondiente al capítulo sobre sintaxis, decidirá, junto con el profesor, el tema de trabajo práctico de implementación que deberá realizar como parte de los requisitos de evaluación del curso. En todo momento, por medio de las tutorías, recibirá consejos y comentarios sobre su trabajo hasta la entrega del mismo al final del curso.



SISTEMA DE EVALUACIÓN

BIBLIOGRAFÍA BÁSICA

Los comentarios se introducen durante el curso y conforme a lo consignado en la Guía de la asignatura.

Jurafsky, D., and J. H. Martin (2000). *Speech and language Processing*. New Jersey: Prentice Hall.

Allen, J. F. (1994). *Natural Language Understanding*, 2nd edition, Reading, MA: Addison-Wesley.

Moreno Sandoval, A. (1998). *Lingüística Computacional*. Madrid: Síntesis.

En todas las lecturas se indica, con un número entre corchetes, el capítulo del temario al que se refieren. Asimismo, aquellas referencias que llevan además un asterisco se incluyen en la documentación del curso. La bibliografía, para facilitar la relación con el temario de la asignatura, se indica en el orden de lectura.

- Sánchez León, F. (2008) "Panorama general e histórico, definición(es), aproximación, modelos". [1, *]
- European Commission (1997?) "Language Engineering: Harnessing the power of language". [1, *]
- Oficina del Español en la Sociedad de la Información (OESI) "Tecnologías del habla" [Introducción en línea disponible en: http://oesi.cervantes.es/TLTODOS/tecnologias_del_habla.htm] El alumno deberá navegar por las páginas de este curso sobre tecnologías del habla. El curso original se ha eliminado de la red, por lo que el alumno deberá navegar por estos nuevos contenidos.[2]
- Llisterri, J. (2003) "Las tecnologías del habla: Entre la ingeniería y la lingüística", en *Actas del Congreso Internacional "La Ciencia ante el Público. Cultura humanística y desarrollo científico y tecnológico"*. Universidad de Salamanca, 28-31 de octubre 2002. Salamanca: Instituto Universitario de Estudios de la Ciencia y la Tecnología. Edición en CD-ROM. pp. 44-67. [2, *]
- King, M. (1988?) "A General Introduction to Natural Language Processing". [3, *]
- Jurafsky, D. and Martin, J. H. (2008) *Speech and Language Processing*, Prentice Hall, New Jersey, 2nd edition [3, *].
- Liddy, E. D. (1998) "Natural Language Processing", en *Encyclopedia of Library and Information Science*, Marcel Decker, Inc., 2nd edition. [3, *].
- Sánchez León, F. (2008) "Introducción a Prolog" [4, *].
- Sánchez León, F. (2008) "ELIZA. Una terapeuta en casa" [4, *].
- Grefenstette, G. and P. Tapanainen (1994) "What is a word, What is a sentence? Problems of tokenization", in *3rd International Conference on Computational Lexicography*, Budapest,



pp. 79-87. [5, *]

- Sánchez León, F. (2008) “Autómatas de estados finitos”. [6, *]
- Sánchez León, F. (2008) “Fonología/Morfología de estados finitos”. [6, *]
- Karttunen, L., J-P. Chanod, G. Grefenstette and A. Schiller (1996) “Regular Expressions for Language Engineering”. *Natural Language Engineering*, 2(4): 305-238. [6, *]
- Sánchez León, F. (2008) “Gramáticas y lenguajes formales”. [7, *]
- Sánchez León, F. (2008) “*Parsing*: estrategias”. [7, *]
- Sánchez León, F. (2008) “Introducción a los formalismos gramaticales basados en unificación”. [7, *]
- McConnel, S. (2002) “PC-PATR Reference Manual: A unification based parser”. [7, *]
- Agirre, E. and Ph. Edmonds (2006) “Introduction”, in *Word Sense Disambiguation: Algorithms and Applications*, Springer. [8, *]
- Sánchez León, F. (2008) “Anotación morfosintáctica de corpus”. [9, *]

BIBLIOGRAFÍA COMPLEMENTARIA

Se presenta únicamente la bibliografía complementaria relativa a los dos primeros temas, pues el resto se ha integrado en los capítulos correspondientes del material didáctico.

- Gómez Guinovart, J. (1998): “Fundamentos de Lingüística Computacional: bases teóricas, líneas de investigación y aplicaciones”. En J. Baró i Queralt y P. Cid Leal (eds.), *Anuari SOCADI de Documentació i Informació*, Barcelona: Societat Catalana de Documentació i Informació, 135-146. Interesante artículo introductorio en el que se presentan los grados de imbricación entre Lingüística e Informática y las distintas líneas de trabajo en función de esa imbricación. Con todo, téngase en cuenta que han transcurrido casi 20 años desde su publicación y que, por tanto, algunas limitaciones tecnológicas que plantea el artículo han sido superadas. Incluye una sección de fuentes de información relacionadas con esta disciplina.
- Jurafsky, D. y Martin, J. H. (2000): Capítulo 1 “Introduction”. Repaso general a los problemas, modelos y algoritmos para el procesamiento lingüístico, junto con un breve apunte histórico.
- Automatic Language Processing Advisory Committee (1966): *Language and Machines. Computers in Translation and Linguistics*. El alumno interesado en conocer las razones que se adujeron para el descalabro de la Traducción Automática a mediados de los 60, puede encontrar el informe ALPAC en la red <http://books.nap.edu/openbook.php?isbn=ARC000005>



- Moreno Sandoval, A. (1998): *Lingüística computacional*. Síntesis, Madrid. Se trata de una introducción básica muy asequible, por este motivo se ha incluido en la bibliografía recomendada. No obstante, debe manejarse con cautela, pues, aunque el planteamiento de la obra es adecuado, contiene algunos errores de fondo. El alumno deberá contrastar con el profesor de la asignatura aquellos aspectos contradictorios con las ideas del resto de la documentación del curso.

RECURSOS DE APOYO Y WEBGRAFÍA

Los programas y lenguajes de programación que se utilizarán para ejemplificar aspectos del curso se incluyen en el material didáctico que se pone a disposición del alumno. Debe hacerse notar que este no es un curso de programación para lingüistas y que, por tanto, será difícil, en la mayoría de los casos, usar algunos de los programas y lenguajes para algo más que probar en casa los ejemplos que se explican en la documentación del curso.

En la elección de programas y lenguajes se ha optado por aquellos que se encuentran en el dominio público (son gratuitos) y que tienen versiones bajo sistema operativo Windows (en cualquiera de sus versiones). Se ha supuesto, por tanto, que el alumno tiene alguna versión de este sistema operativo instalada en su ordenador. Todo el material, sin embargo, se ha desarrollado y probado bajo Linux, sistema operativo para el que, lógicamente, existen versiones de estos mismos programas.

Para instalar los programas incluidos en la carpeta que contiene el material didáctico, el alumno deberá seguir las instrucciones que se incluyen en cada caso. Todos los programas se encuentran en directorio software. Todos tienen un desinstalador, por lo que podrán eliminarse del disco duro, si se desea, cuando termine el curso.

A título meramente indicativo, además de un programa de descompresión de archivos, el mencionado directorio contiene un intérprete de Perl (ActivePerl), un intérprete de Prolog (SWI-Prolog) y el conjunto de programas para el desarrollo de procesadores morfológicos y sintácticos con gramáticas de unificación PC-PARSE. Asimismo, se incluyen programas ejemplo, desarrollados expresamente para el curso, con los que verificar determinados aspectos prácticos del curso.

Competencias específicas

El programa del curso pretende estas u otras competencias semejantes:

- Comprensión de los fundamentos, problemas y métodos de los modelos computacionales del lenguaje dominantes en la actualidad: el probabilístico y el simbólico, con especial hincapié en el segundo de ellos.
- Estudiar someramente la historia de la Lingüística Computacional, así como las aplicaciones más reseñables del Procesamiento del Lenguaje Natural.
- Identificación de los formalismos más usuales para la representación del conocimiento lingüístico en los niveles morfológico, sintáctico y semántico.
- Ser capaz de resolver problemas en un entorno de programación lógica, del que se presentarán los rudimentos en el curso.



- Aplicar los programas que forman parte del material didáctico del curso para probar con sus propios datos un conjunto de aplicaciones.
- Conseguir o adoptar una gramática propia sintáctica de algún fenómeno lingüístico de su interés. El material didáctico incluye algunos temas orientativos.
- Adopción de programas lógicos mínimos para que realicen una tarea distinta a aquella para la que originalmente fueron implementados.
- Desarrollo, principalmente gráfico, de autómatas finitos para la descripción de ciertos fenómenos lingüísticos.

Actividades formativas

ACTIVIDAD FORMATIVA	HORAS
PRESENCIALIDAD	
TRABAJO AUTÓNOMO DEL ESTUDIANTE	105 0
INTERACCION CON EL PROFESOR Y CON OTROS ESTUDIANTES	10 0
TRABAJO EN GRUPO CON OTROS ESTUDIANTES	10 0

IGUALDAD DE GÉNERO

En coherencia con el valor asumido de la igualdad de género, todas las denominaciones que en esta Guía hacen referencia a órganos de gobierno unipersonales, de representación, o miembros de la comunidad universitaria y se efectúan en género masculino, cuando no hayan sido sustituido por términos genéricos, se entenderán hechas indistintamente en género femenino o masculino, según el sexo del titular que los desempeñe.

