

**Universidad de Puerto Rico  
Recinto Universitario de Mayagüez  
Facultad de Artes y Ciencias  
Departamento de Estudios Hispánicos**

**LINGÜÍSTICA COMPUTACIONAL**

**I. Información General**

Curso:	Lingüística computacional
Código:	LING 5080
Núm. de créditos:	3
Horas contacto:	3
Prerrequisitos:	LING 4010 o ESPA 4201 o INGL 3225 o permiso del instructor

**II. Descripción**

Introducción al estudio de las propiedades computacionales del lenguaje humano y de los modelos de procesamiento de lenguas naturales. Análisis y evaluación de sistemas deterministas y no-deterministas de modelos computacionales de aprendizaje y procesamiento lingüístico: autómatas y transductores, sistemas de reglas formales, sistemas lógico-matemáticos y modelos probabilísticos. Representación de estructuras fonológicas, morfológicas, sintácticas y semánticas mediante analizadores basados en estos modelos computacionales. Revisión de herramientas en línea como corpus anotados, procesadores estructurales y redes semánticas. Discusión de aplicaciones de modelos computacionales en tecnologías de procesamiento lingüístico, tales como síntesis y reconocimiento de habla, comprensión de lenguas naturales, correctores ortográficos y gramaticales, traducción automática, motores de búsqueda, desambiguación léxica, extracción de información, respuesta automática a preguntas y agentes conversacionales.

**III. Objetivos**

Al terminar el semestre, los estudiantes podrán:

1. Producir expresiones regulares para buscar patrones en un corpus textual.
2. Construir algoritmos basados en autómatas para representar diferentes tipos de gramáticas de lenguajes formales.
3. Usar transductores de estado finito para análisis morfológico y fonológico.
4. Computar la probabilidad de una secuencia de palabras mediante el uso de modelos de n-gramas.
5. Etiquetar un corpus siguiendo diferentes modelos de procesamiento de lenguas naturales.
6. Identificar constituyentes, marcos de subcategorización y relaciones gramaticales de dependencia en estructuras sintácticas.
7. Usar analizadores sintácticos de diferentes algoritmos para representar estructuras sintácticas.
8. Utilizar métodos formales para la representación y composición de estructuras semánticas.
9. Representar computacionalmente los rasgos gramaticales y las relaciones semánticas entre elementos léxicos.
10. Identificar características de los algoritmos de segmentación y de resolución de referencialidad y coherencia en el procesamiento computacional del discurso.

**IV. Bosquejo del curso y distribución de tiempo**

1. Fundamentos conceptuales de procesamiento de lenguas naturales, modelos y algoritmos	1.5
2. Expresiones regulares y autómatas de estado finito	3
3. Morfología computacional; transductores	3
4. n-gramas y modelos probabilísticos	3
5. Etiquetaje de categorías y rasgos gramaticales	3
6. Fonología computacional	3
7. Gramáticas formales	3
8. Procesadores estructurales regulados	3
9. Procesadores estructurales estadísticos	3
10. Estructura de rasgos gramaticales y unificación	3
11. Complejidad y jerarquía de lenguajes formales	3
12. Significado y representaciones semánticas	3
13. Semántica computacional	3
14. Semántica léxica	1.5
15. Semántica léxica computacional	3
16. Discurso computacional	3
TOTAL	45 horas

**V. Recursos***Textos sugeridos:*

- Jurafsky, D. & Martin, J. (2009). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition*. 2da. edición.
- Moreno-Sandoval, A. (2001). *Lingüística computacional*.

*Lecturas suplementarias:*

- Además de los textos arriba indicados, se recomienda una selección temática de lecturas de la bibliografía.

*Audiovisuales:*

- Se recomienda la disponibilidad de computadoras personales y un proyector digital para el uso de plataformas electrónicas de manejo del curso y material suplementario en las páginas electrónicas incluidas en la bibliografía.

**VI. Estrategias de enseñanza**

El profesor utilizará los siguientes métodos:

- conferencias
- comentarios de lecturas especializadas
- análisis y discusión de datos y problemas

**VII. Estrategias de evaluación sugeridas**

• asignaciones	50%
• exámenes parciales	25%
• examen o proyecto final	25%

**VIII. Sistema de calificación**

90-100	A
80-89	B
70-79	C
60-69	D
0-59	F

**IX. Bibliografía**

- Allen, James (1995). *Natural Language Understanding*. 2da. edición. Redwood City, CA: Benjamins/Cummings.
- Barton, G. Edward; Berwick, Robert C. & Ristad, Eric Sven (1987). *Computational Complexity and Natural Language*. Cambridge, Mass.: The MIT Press.
- Berwick, Robert C. (1985). *The Acquisition of Syntactic Knowledge*. Cambridge, Mass.: The MIT Press.
- Berwick, Robert C.; Abney, Steven & Tenny, Carol (eds.) (1991). *Principle-Based Parsing: Computation and Psycholinguistics*. Dordrecht: Kluwer.
- Berwick, Robert C. & Weinberg, Amy S, (1984). *The Grammatical Basis of Linguistic Performance*. Cambridge, Mass.: The MIT Press.
- Bird, Steven; Klein, Ewan & Loper, Edward (2009). *Natural Language Processing with Python*. Sebastopol, CA: O'Reilly Media.
- Booij, Geert (2012). *The Grammar of Words*. 3ra. edición. Oxford: Oxford University Press.
- Brent, Michael (1997). *Computational Approaches to Language Acquisition*. Cambridge, Mass.: The MIT Press.
- Chomsky, Noam (1955/1975). *The Logical Structure of Linguistic Theory*. Berlin: Springer-Verlag.
- Chomsky, Noam (1957). *Syntactic Structures*. La Haya: Mouton.
- Chomsky, Noam (1965). *Aspects of the Theory of Syntax*. Cambridge, Mass.: The MIT Press.
- Clark, Alexander; Fox, Chris & Lappin, Shalom, eds. (2013). *The Handbook of Computational Linguistics and Natural Language Processing*. Oxford: Blackwell.
- Cook, Vivian & Newson, Mark (2007). *Chomsky's Universal Grammar*. 3ra. edición. Oxford: Blackwell.
- Davenport, Mike & Hannahs, S.J. (2011). *Understanding Phonetics and Phonology*. 3ra. edición. Londres: Hodder.
- Dickinson, Markus; Brew, Chris & Meurers, Detmar (2013). *Language and Computers*. Oxford: Wiley-Blackwell.
- Gorrell, Paul (1995). *Syntax and parsing*. Cambridge: Cambridge University Press.
- Grisham, Ralph (1986). *Computational Linguistics: An Introduction*. Cambridge: Cambridge University Press.

- Hausser, Roland (2001). *Foundations of Computational Linguistics: Human-Computer Communication in Natural Language*. Berlin: Springer-Verlag.
- Heim, Irene & Kratzer, Angelika (1998). *Semantics in Generative Grammar*. Oxford: Blackwell.
- Huang, Chu-Ren; Calzolari, Nicoletta; Gangemi, Aldo; Lenci, Alessandro; Oltramari, Alessandro & Prévot, Laurent (2010). *Ontology and the Lexicon: A Natural Language Processing Perspective*. Cambridge: Cambridge University Press.
- Jurafsky, Daniel & Martin, James. (2009). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition*. 2da. edición. New Jersey: Prentice Hall.
- Kumar, Ela (2011). *Natural Language Processing*. Nueva Delhi: IK International.
- Levelt, Willem J.M. (2008). *An Introduction to the Theory of Formal Languages and Automata*. Amsterdam: John Benjamins.
- Linz, Peter (2012). *An Introduction to Formal Languages and Automata*. 5ta. edición. Sudbury, Massachusetts: Jones & Bartlett Learning.
- Manning, Christopher & Schütze, Hinrich (1999/2003). *Foundations of Statistical Natural Language Processing*. Cambridge, Mass.: The MIT Press.
- McCarthy, John (2002). *A Thematic Guide to Optimality Theory*. Cambridge: Cambridge University Press.
- Mihalcea, Rada & Radev, Dragomir (2011). *Graph-Based Natural Language Processing and Information Retrieval*. Cambridge: Cambridge University Press.
- Moreno-Sandoval, Antonio (2001). *Lingüística computacional*. Madrid: Síntesis.
- Partee, Barbara; ter Meulen, Alice & Wall, Robert E. (1993/2007). *Mathematical Methods in Linguistics*. Dordrecht: Kluwer Academic Publishers.
- Prince, Alan & Smolensky, Paul (1993/2004). *Optimality Theory: Constraints Interaction in Generative Grammar*. Oxford: Blackwell.
- Pustejovsky, James (1998). *The Generative Lexicon*. Cambridge, Mass.: The MIT Press.
- Pustejovsky, James & Stubbs, Amber (2013). *Natural Language Annotation for Machine Learning*. Sebastopol, CA: O'Reilly Media.
- Roark, Brian & Sproat, Richard (2007). *Computational Approaches to Morphology and Syntax*. Oxford: Oxford University Press.
- Ruslan, Mitkov (2005). *The Oxford Handbook of Computational Linguistics*. Oxford: Oxford University Press.
- Sells, Peter; Shieber, Stuart & Wasow, Thowas, eds. (1991). *Foundational Issues in Natural Language Processing*. Cambridge, Mass.: The MIT Press.
- Smith, George (1991). *Computers and Human Language*. Oxford: Oxford University Press.
- Sproat, Richard (1992). *Morphology and Computation*. Cambridge, Mass.: The MIT Press.
- Tesar, Bruce & Smolensky, Paul (2000). *Learnability in Optimality Theory*. Cambridge, Mass.: The MIT Press.
- Tomita, Masaru (1991). *Current Issues in Parsing Technology*. Boston: Kluwer.
- Tordera-Yllescas, Juan C. (2011). *Lingüística computacional: tecnologies del habla*. Valencia: Publicacions de la Universitat de València.
- Tordera-Yllescas, Juan C. (2012). *Lingüística computacional: anàlisis, generació y traducción automática*. Valencia: Publicacions de la Universitat de València.

Tordera-Yllescas, Juan C. (2012). *El abecé de la lingüística computacional*. Madrid: Arco Libros.

van Eijck, Jan & Unger, Christina (2010). *Computational Semantics with Functional Programming*. Cambridge: Cambridge University Press.

## RECURSOS EN INTERNET

***Child Language Data Exchange System*** <http://chilides.psy.cmu.edu/>  
Corpus multi-medios de producción lingüística infantil.

***FrameNet*** <https://framenet.icsi.berkeley.edu/fndrupal/>  
Base de datos léxica legible por máquinas y humanos.

***International Phonetic Association*** <http://www.langsci.ucl.ac.uk/ipa/>  
Plataforma de recursos sobre el Alfabeto Fonético Internacional.

***The Lambda Calculator*** <http://dylanbumford.com/LambdaCalculator/>  
Programa interactivo para practicar cálculo lambda de tipos.

***Linguistic Data Consortium*** <https://www ldc.upenn.edu/>  
Repositorio y punto de distribución de recursos lingüísticos.

***Natural Language Toolkit*** <http://www.nltk.org/>  
Plataforma para construir programas en Python para trabajar con datos en lenguas naturales.

***OT-Help*** <http://people.umass.edu/othelp/>  
Herramientas para el estudio de tipología lingüística en versiones en paralelo y en serie de la Teoría de la Optimalidad y Gramática Armónica.

***Penn TreeBank Project*** <http://www.cis.upenn.edu/~treebank/>  
Banco de árboles sintácticos a partir de un corpus natural anotado.

***Praat: Doing Phonetics by Computer*** <http://www.fon.hum.uva.nl/praat/>  
Página electrónica del programa de análisis fonético-acústico de Boersma y Weenink.

***Rutgers Optimality Archive*** <http://roa.rutgers.edu/>  
Portal de la colección de artículos sobre la Teoría de la Optimalidad.

***Speech and Language Processing*** <http://www.prenhall.com/jurafsky-martin/>  
Portal del texto de Jurafsky & Martin con recursos para el instructor.

***The Stanford Parser*** <http://nlp.stanford.edu/software/lex-parser.shtml>  
Analizador sintáctico estadístico desarrollado por el Stanford NLP Group.

***WordNet*** <http://wordnet.princeton.edu/>  
Base de datos léxica del inglés.

**Ley 51 (Ley de Servicios Integrales de Personas con Impedimentos)**

Los estudiantes que reciban servicios de Rehabilitación Vocacional deben comunicarse con su profesor al inicio del semestre para que puedan recibir acomodo razonable y, de ser necesario, equipo de asistencia conforme con las estipulaciones del Recinto Universitario de Mayagüez.