

Universidad de Puerto Rico  
Recinto de Río Piedras  
Facultad de Humanidades  
Departamento de Programa Graduado de Lingüística

## PRONTUARIO

TÍTULO DEL CURSO	:	Lingüística Computacional
CODIFICACIÓN	:	LING 6340
CANTIDAD DE HORAS/CRÉDITO	:	45 horas / Tres créditos
PRERREQUISITOS, CORREQUISITOS Y OTROS REQUIMIENTOS	:	Ninguno
PROFESORA	:	Aida Vergne Vargas
MÉTODO DE CONTACTO	:	aida.vergnevargas@upr.edu
HORAS DE OFICINA	:	Por acuerdo
Durante el segundo semestre del año académico 2020-2021, este curso se ofrecerá en línea, por medio de reuniones sincrónicas.		
<b>DESCRIPCIÓN DEL CURSO</b>		
Equipo electrónico, lenguaje de máquina e investigaciones lingüísticas. Posibilidades y tipos de estudios en lingüística computacional. Ejercicios prácticos. <b>Curso PANORAMICO E INTRODUCTORIO presencial, en línea e híbrido. 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 nodeterministas 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</b>		

**automática, motores de búsqueda, desambiguación léxica, extracción de información, respuesta automática a preguntas y agentes conversacionales.**

**OBJETIVOS DE APRENDIZAJE**

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.

**LIBRO DE TEXTO PRINCIPAL**

No tiene.

**BOSQUEJO DE CONTENIDO Y DISTRIBUCIÓN DEL TIEMPO**

<i>Tema</i>	Distribución del tiempo		
	Presencial	Híbrida	En línea
I. 1. Fundamentos conceptuales de procesamiento de lenguas naturales, modelos y algoritmos	3 horas	3 horas (en línea)	3 horas
II. Expresiones regulares y autómatas de estado finito	3 horas	3 horas (presenciales)	3 horas
III. Morfología computacional; transductores	3 horas	3 horas (en línea)	3 horas
IV. n-gramas y modelos probabilísticos	3 horas	3 horas (presenciales)	3 horas
V. Etiquetaje de categorías y rasgos gramaticales	3 horas	3 horas (en línea)	3 horas
VI. Fonología computacional	3 horas	3 horas	3 horas

		(presenciales)	
VII. Gramáticas formales	3 horas	3 horas (en línea)	3 horas
VIII. Procesadores estructurales regulados	1.5 horas	1.5 horas (presenciales)	1.5 horas
IX. Procesadores estructurales estadísticos	1.5 horas	1.5 horas (en línea)	1.5 horas
X. Estructura de rasgos gramaticales y unificación	3 horas	3 horas (presenciales)	3 horas
XI. Complejidad y jerarquía de lenguajes formales	3 horas	3 horas (en línea)	3 horas
XII. Significado y representaciones semánticas	3 horas	3 horas (presenciales)	3 horas
XIII. Semántica computacional	3 horas	3 horas (en línea)	3 horas
XIV. Semántica léxica	3 horas	3 horas (presenciales)	3 horas
XV. Discurso computacional	3 horas	3 horas (en línea)	3 horas
XVI. NLTK	3 horas	3 horas (presenciales)	3 horas
<b>Total de horas contacto</b>	<b>45 horas</b>	<b>45 horas</b> (22.5 presenciales = 50% y 22.5 horas en línea = 50%)	<b>45 horas</b>

### ESTRATEGIAS INSTRUCCIONALES

Presencial	Híbrido	En línea
<ul style="list-style-type: none"> <li>• Conferencias del profesor</li> <li>• Lectura de artículos académicos</li> <li>• Discusión de lecturas</li> <li>• Videos instruccionales</li> <li>• Trabajos en grupo</li> <li>• Tareas individuales</li> <li>• Recolección y análisis de datos lingüísticos</li> <li>• Presentaciones orales</li> </ul>	<ul style="list-style-type: none"> <li>• Módulos instruccionales en línea</li> <li>• Lectura de artículos académicos en línea</li> <li>• Discusión de lecturas</li> <li>• Videos instruccionales</li> <li>• Trabajos en grupo</li> <li>• Tareas individuales</li> <li>• Recolección y análisis de datos lingüísticos</li> <li>• Presentaciones orales</li> </ul>	<ul style="list-style-type: none"> <li>• Módulos instruccionales interactivos</li> <li>• Conferencias asincrónicas del profesor</li> <li>• Lectura de artículos académicos en línea</li> <li>• Discusión de lecturas</li> <li>• Videos instruccionales</li> <li>• Trabajos en grupo</li> <li>• Tareas individuales</li> </ul>

	<ul style="list-style-type: none"> <li>• Videoconferencias asincrónicas y/o sincrónicas</li> </ul>	<ul style="list-style-type: none"> <li>• Recolección y/o análisis de datos lingüísticos</li> <li>• Presentaciones orales <ul style="list-style-type: none"> <li>• Videoconferencias sincrónicas</li> </ul> </li> </ul>
--	----------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

### RECURSOS MÍNIMOS DISPONIBLES O REQUERIDOS

Recurso	Presencial	Híbrido	En línea
Cuenta en la plataforma institucional de gestión de aprendizaje (Ej. Moodle)	Institución	Institución	Institución
Cuenta de correo electrónico institucional	Institución	Institución	Institución
Computadora con acceso a internet de alta velocidad o dispositivo móvil con servicio de datos	Estudiante	Estudiante	Estudiante
Programados o aplicaciones: procesador de palabras, hojas de cálculo, editor de presentaciones	Estudiante	Estudiante	Estudiante
Bocinas integradas o externas	No aplica	Estudiante	Estudiante
Cámara web o móvil con cámara y micrófono	No aplica	Estudiante	Estudiante

### TÉCNICAS DE EVALUACIÓN

Presencial	Híbrida	En línea
Asistencia y Participación...10%	Asistencia .....5%	Asistencia .....5%
Reseñas.....15%	Participación en foros de discusión.....5%	Participación en foros de discusión.....5%
Ejercicios.....15%	Reseñas.....15%	Reseñas.....15%
Presentaciones orales.....15%	Ejercicios.....15%	Ejercicios.....15%
Exámenes.....20%	Presentaciones orales.....15%	Presentaciones orales.....15%
Trabajo de investigación....25%	proyectos.....20%	Proyectos.....20%
<b>Total.....100%</b>	Trabajo de investigación.....25%	Trabajo de investigación....25%
	<b>Total.....100%</b>	<b>Total.....100%</b>

### ACOMODO RAZONABLE

Según la Ley de Servicios Educativos Integrales para Personas con Impedimentos, todo estudiante que requiera acomodo razonable deberá notificarlo al profesor el primer día de clase. Los estudiantes que reciban servicios de Rehabilitación Vocacional deben comunicarse con el (la) profesor(a) al inicio del semestre para planificar el acomodo razonable y el equipo de asistencia necesario conforme a las recomendaciones de la Oficina de Servicios a Estudiantes con Impedimentos (OSEI) del Decanato de Estudiantes. También aquellos estudiantes con necesidades especiales de algún tipo de asistencia o acomodo deben comunicarse con el (la) profesor(a). Si un alumno tiene una discapacidad documentada (ya sea física, psicológica, de aprendizaje o de otro tipo, que afecte su desempeño académico) y le gustaría solicitar disposiciones académicas especiales, éste debe comunicarse con la Oficina de Servicios a Estudiantes con Impedimentos (OSEI) del Decanato de Estudiantes, a fin de fijar una cita para dar inicio a los servicios pertinentes.

### **INTEGRIDAD ACADÉMICA**

La Universidad de Puerto Rico promueve los más altos estándares de integridad académica y científica. El Artículo 6.2 del Reglamento General de Estudiantes de la UPR (Certificación Núm. 13, 2009-2010, de la Junta de Síndicos) establece que “la deshonestidad académica incluye, pero no se limita a: acciones fraudulentas, la obtención de notas o grados académicos valiéndose de falsas o fraudulentas simulaciones, copiar total o parcialmente la labor académica de otra persona, plagiar total o parcialmente el trabajo de otra persona, copiar total o parcialmente las respuestas de otra persona a las preguntas de un examen, haciendo o consiguiendo que otro tome en su nombre cualquier prueba o examen oral o escrito, así como la ayuda o facilitación para que otra persona incurra en la referida conducta”. Cualquiera de estas acciones estará sujeta a sanciones disciplinarias en conformidad con el procedimiento disciplinario establecido en el Reglamento General de Estudiantes de la UPR vigente.

**Para velar por la integridad y seguridad de los datos de los usuarios, todo curso híbrido y a distancia deberá ofrecerse mediante la plataforma institucional de gestión de aprendizaje, la cual utiliza protocolos seguros de conexión y autenticación. El sistema autentica la identidad del usuario utilizando el nombre de usuario y contraseña asignados en su cuenta institucional. El usuario es responsable de mantener segura, proteger, y no compartir su contraseña con otras personas.**

### **NORMATIVA SOBRE HOSTIGAMIENTO SEXUAL**

“La Universidad de Puerto Rico prohíbe el discrimen por razón de sexo y género en todas sus modalidades, incluyendo el hostigamiento sexual. Según la Política Institucional contra el Hostigamiento Sexual en la Universidad de Puerto Rico, Certificación Núm. 130, 2014-2015 de la Junta de Gobierno, si un estudiante está siendo o fue afectado por conductas relacionadas a hostigamiento sexual, puede acudir ante la Oficina de Procuraduría Estudiantil, el Decanato de Estudiantes o la Coordinadora de Cumplimiento con Título IX para orientación y/o presentar una queja”.

### **SISTEMA DE CALIFICACIÓN**

A, B, C, D, F

## 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). *Chomskys 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: tecnologías del habla. Valencia: Publicacions de la Universitat de València.

Tordera-Yllescas, Juan C. (2012). Lingüística computacional: análisis, generación y traducción automática. Valencia: Publicacions de la Universitat de València.

### **Referencias electrónicas:**

Child Language Data Exchange System <http://childes.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.