UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

PROGRAMA DE POSGRADO
MAESTRÍA EN LINGÜÍSTICA HISPÁNICA
Programa de actividad académica

Denominación: Lexicografía basada en corpus (Praxis lexicográfica) Nombre de los profesores: Gerardo Sierra Martínez y Claudio Molina Salinas				
Clave:	Semestre: 2019-1	Campo de conocimiento: Lingüística hispánica		No. Créditos:
Carácter: Optativa	Horas		Horas por semana	Horas al semestre
Tipo: Teórica/práctica	Teoría: 2	Práctica: 2	4	64
Modalidad: Curso/seminario	Duración del curso: Semestral			

Actividad académica con seriación subsecuente: ninguna
Actividad académica con seriación antecedente: ninguna

Objetivo(s) del curso: Proporcionar al alumno los conocimientos teóricos y las herramientas necesarias para diseñar y construir un diccionario de especialidad a partir de una colección de textos de un área de especialidad. A lo largo del curso, se revisarán aspectos teóricos y aplicados de la lingüística implicados en la planificación, construcción y actualización de diccionarios, además de algunos otros aspectos teóricos y aplicados propios de la lingüística que resultan relevantes para la construcción de diccionarios.

Objetivo general: Formar especialistas en lexicografía de especialidad que estén familiarizados con el sustento metodológico y las experiencias prácticas necesarias para la elaboración de diccionarios, sus aplicaciones y la relación que establecen la teoría y la práctica con algunas otras subdisciplinas lingüísticas (lingüística de corpus, estadística lingüística, lingüística computacional, morfología léxica, semántica léxica, entre otras).

Objetivos específicos:

- Contar con las bases teóricas de lexicografía para el diseño y construcción de diccionarios de un área de especialidad.
- Conocer los diferentes tipos de diccionarios que existen, las diferentes posibilidades estructurales de estos y cómo se orienta su definición.
- Conocer y utilizar las herramientas disponibles para la creación de diccionarios a partir de corpus lingüísticos.
- Formalizar un primer análisis lexicográfico, elaborar un sistema conceptual y hacer una propuesta de normalización de un vocabulario específico.
- Resolver problemas lexicológicos y lexicográficos involucrados en la elaboración de un diccionario.
- Proponer un proyecto de obra lexicográfica (propuesta de planta o plantilla de diccionario).

Temario

Unidad	Tema	Horas	
		Teóricas	Prácticas
0	**La lexicografía. Competencias y límites** 0.1 Orígenes, desarrollo de la disciplina. 0.2 Teorías y evolución de la disciplina. Relación con otras áreas del conocimiento. 0.3 Tipos de unidades léxicas (palabra y término).	4	
1	**Delimitación del objeto de estudio: diseño de un diccionario de especialidad** 1.1 Tipología de diccionarios 1.2 Los diccionarios de especialidad, la variación lingüística 1.3 Delimitación de macroestructura y microestructura	4	
2	**Compilación del corpus** 2.1 Criterios de compilación de corpus	8	8

	2.2 Anotación y etiquetado 2.3 Estandarización de formatos 2.4 Gestores de corpus		
3	**Macroestructura: extracción terminológica** 3.1 Extractores terminológicos 3.2 Validación de términos 3.3 Relaciones entre términos. Las relaciones léxico-semánticas. 3.4 Bancos de datos	8	8
4	**Microestructura: obtención de definiciones** 4.1 Tipología de definiciones 4.2 Extracción de contextos definitorios 4.3 Generación de definiciones 4.4 Obtención de ejemplos	10	10
5	**Elaboración de un protodiccionario de especialidad** 5.1 Plataformas 5.2 Planteamiento y seguimiento de un proyecto lexicográfico	2	2
	Total de horas teóricas:	34	
	Total de horas prácticas:	30	
	Suma total de horas:	64	

Bibliografía básica:

Baroni, Marco, Adam Kilgarriff, Jan Pomikálek & Pavel Rychl´y. 2006. WebBootCat: a web tool for instant corpora. En EuraLex Conference.

Bergenholt, H.; Terp, E. 1995. *Manual of specialised lexicography: the preparation of specialised dictionaries*. Amterdam/Philadelphia: John Benjamins Publishing Company.

Dorantes M.A., Pimentel A., Sierra G., Bel-Enguix G., Molina C. (2017) "Extracción automática de definiciones analíticas y relaciones semánticas de hiponimia-hiperonimia con un sistema basado en patrones lingüísticos". *Linguamatica* 9 (2).

Gutierrez-Vasques, X. (2015). Bilingual lexicon extraction for a distant language pair using a small parallel corpus. NAACL-HLT 2015 Student Research Workshop (SRW).

Haensch, G.; Lothar, W. 1982. *La lexicografía. De la lingüística teórica a la lexicografía práctica*. Madrid: Gredos.

Kilgarriff, A., Husák, M., McAdam, K., Rundell, M., Rychly, P.: GDEX automatically finding good dictionary examples in a corpus. In: Proceedings of the XIII EURALEX International Congress (2008).

Kilgarriff, A., Marcowitz, F., Smith, S., & Thomas, J. (2015, May). Corpora and Language Learning with the Sketch Engine and SKELL. Revue française de linguistique appliquée, Vol. XX(1), 61–80.

Lara, L. F. 1997. *Teoría del diccionario Monolingüe*. México: COLMEX.

Lara, L. F. 2006. *Curso de lexicología*. México: COLMEX.

Lázaro, J. 2015. El ejemplo en terminología. Caracterización y extracción automática. PhD thesis, Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra, Barcelona.

Rey, A. 1992. *La terminologie: noms et notions*. Paris: P.U.F.

Riloff, Ellen & Jessica Shepherd. 1997. A corpus-based approach for building semantic lexicons. En 2nd Conference on Empirical Methods in Natural Language Processing (EMNLP'1997).

Rondeau, G. 1984. *Introduction à la terminologie*. Chicoutimi/Québec: Gaëtan Morin.

Sager, J. C. 1993. *Curso práctico sobre el procesamiento de la terminología*. Madrid: Pirámide y Fundación Germán Sánchez Ruipérez.

Sierra, G. 2009. Extracción de contextos definitorios en textos de especialidad a partir del reconocimiento de patrones lingüísticos. *Linguamatica* 2.

Sierra, G.; Lázaro, J.; Bel-Enguix, G. 2017. LEXIK. An Integrated System for Specialized Terminology. En *Advances in Computational Intelligence*.

Velardi, Paola, Roberto Navigli & Pierluigi D'Amadio. 2008. Mining the web to create specialized glossaries. IEEE Intelligent Systems 23(5).

Vivaldi, Jordi. 2009. Catálogo de herramientas informáticas relacionadas con la creación, gestión y explotación de corpus textuales. Tradumática 7.

Wilks, Yorick, Brian Slator y Louise Guthrie 1996. Electric Words. Dictionaries, Computers and Meaning, MIT Press: Cambridge.

Bibliografía complementaria:

Almeida, L. 2004. *Curso básico de terminología*. São Paulo: Editora da Universidade de São Paulo.

Alpizar, R. 1997. *¿Cómo hacer un diccionario científico técnico?* Buenos Aires: Memphis.

Espinosa-Anke, Luis, Roberto Carlini, Horacio Saggion & Francesco Ronzano. 2016. DEFEXT: a semi supervised definition extraction tool. En GLOBALEX Workshop, co-located with LREC'2016
Grishman, R. 1991. *Introducción a la lingüística computacional*. Madrid: Visor.
Hearst, Marti A. 1992. Automatic acquisition of hyponyms from large text corpora. En 14th Conference on Computational Linguistics.
Klavans, Judith L. & Smaranda Muresan. 2001. Evaluation of the DEFINDER system for fully automatic glossary construction. En American Medical Informatics Association Symposium (AMIA'2001).
Kouklakis, G., Mikros, G., Markopoulos, G., & Koutsis, I. (2007). Corpus Manager A Tool for Multilingual Corpus Analysis. In Proceedings from Corpus Linguistics Conference.
Landau, S. I. (2001), *Dictionaries: The art and craft of lexicography*, Cambridge: Cambridge University Press.
Lerat, P. 1997. *Las lenguas especializadas*. Madrid: Ariel lingüística.
McEnery, T. 1992. *Computational linguistics: a handbook; toolbox for natural language processing*. Wilmslow: Sigma Press.
Meyer, Ingrid. 2001. Extracting knowledge-rich contexts for terminography: A conceptual and methodological framework. En Recent advances in Computational Terminology.
Padró, Lluís & Evgeny Stanilovsky. 2012. FreeLing 3.0: Towards wider multilinguality. En Language Resources and Evaluation Conference (LREC'2012).
Picht, H.; Draskau, J. 1985. Terminology: An introduction. Guildford: University of Surrey.

Sugerencias didácticas:		**Métodos de evaluación:**	
Exposición oral	(X)	Proyecto final escrito	(X)
Exposición audiovisual	(X)	Trabajos y tareas fuera del aula	(X)
Ejercicios dentro de clase	(X)	Participación en clase	(X)
Ejercicios fuera del aula	(X)	Asistencia	(X)
Lecturas obligatorias	(X)		
Trabajo de Investigación	(X)		

Perfil profesiográfico: El profesor debe tener experiencia en lexicografía o terminología y estar familiarizado con las herramientas computacionales para la labor lexicográfica y el trabajo con corpus lingüísticos. Doctorado en lingüística o afines.