



## SECRETARÍA DE POSGRADO

### PROGRAMA DE ACTIVIDAD FORMATIVA DE POSGRADO

#### 1. Título de la Actividad

Lexicografía Basada en Corpus

#### 2. Modalidad o tipo de actividad

- Seminario

#### 3. Lugar de dictado, fecha de realización y horarios

Lugar de dictado: Facultad de Filosofía y Letras. UNCuyo

Fecha: Del 31 de octubre al 04 de noviembre de 2022

Días (de la semana): lunes a viernes

Horario: 16 a 19 h

#### 4. Duración en horas reales dictadas

Modalidad	Carga teórica	Carga práctica	Total	Porcentaje
Presencial	7	8	15	100%
No presencial	-	-	-	-
Total	7	8	15	100%

**Tutorías (horario y lugar):** Despacho (a designar) de la Facultad de Filosofía y Letras de la U. N. Cuyo, de 14:30 a 15:30 h los días miércoles, jueves y viernes.

## 5. Disertantes

### 5.1. Docente responsable

-- *Rogelio Nazar, Doctor en Ciencias del Lenguaje y Lingüística Aplicada. Profesor del Departamento de Lingüística de la Pontificia Universidad Católica de Valparaíso, Chile. rogelio.nazar@pucv.cl*

### 5.2. Docente/s invitados

*Irene Renau, Doctora en Ciencias del Lenguaje y Lingüística Aplicada. Profesora del Departamento de Lingüística de la Pontificia Universidad Católica de Valparaíso, Chile. irene.renau@pucv.cl*

### 5.3. Curriculum vitae sintético (no más de 15 líneas) de cada uno de los profesores participantes

**Irene Renau** es profesora adjunta de la Pontificia Universidad Católica de Valparaíso y miembro correspondiente de la Academia Chilena de la Lengua. Su investigación y publicaciones se centran en el léxico del español, en especial desde la perspectiva del análisis de corpus y la creación de diccionarios y herramientas de aprendizaje del vocabulario. Es investigadora del grupo Tecling de tecnologías del lenguaje.

**Rogelio Nazar** es profesor adjunto de la Pontificia Universidad Católica de Valparaíso. Su área de especialidad es la lingüística computacional, e investiga en lexicología, terminología y análisis del discurso. Actualmente dirige el grupo de investigación Tecling (<http://www.tecling.com>), dedicado al procesamiento del lenguaje natural e integra el grupo de investigación Corpus EspaDA-UNCuyo (<https://espada.uncu.edu.ar>).

## 6. Destinatarios

**6.1. Nivel:** posgrado

**6.2. Procedencia:** abierta

**6.3. Cupo mínimo y cupo máximo de asistentes:** mínimo 5 y máximo 10

## 7. Fundamentación

Este seminario presenta un panorama reciente de los diferentes enfoques prácticos de la lexicografía basada en datos de corpus. Está centrado en la revisión de técnicas para el análisis del vocabulario, lo que incluye tanto el estudio del léxico general como el de la terminología especializada. Se plantea un recorrido por distintas técnicas de análisis computacional de textos dispuestas en un orden progresivo de complejidad, desde el análisis de concordancias en un corpus monolingüe hasta análisis estadísticos más complejos y en más de una lengua, incluyendo la utilización de lenguajes de scripting, que permiten el estudio de las relaciones paradigmáticas y sintagmáticas. Al final del



seminario, los participantes tendrán las herramientas teóricas y metodológicas necesarias para crear de manera autónoma un producto lexicográfico basado en datos de corpus.

## 8. Objetivos

Proporcionar a los estudiantes herramientas teórico-metodológicas para la creación de diccionarios electrónicos basados en corpus.

## 9. Contenidos

1. Fundamentos teóricos y antecedentes históricos del cruce disciplinar entre la lexicografía y la lingüística computacional.
2. La lexicografía computacional y la creación y gestión recursos léxicos: herramientas computacionales para el desarrollo de bases de datos lexicográficas
3. Panorama de las áreas de aplicación lingüística computacional.
4. Herramientas, métodos y técnicas para el procesamiento de datos textuales: expresiones regulares, lenguajes de *scripting* y software para análisis estadístico de datos.

## 10. Metodología de trabajo y/o actividades de los asistentes

El seminario se dictará en una sala de clases dotada de computadoras para que los estudiantes puedan seguir las actividades prácticas. Las sesiones estarán constituidas por una primera sesión expositiva de 60 minutos para abordar los conceptos que se trabajarán. Esta exposición continuará con 15 minutos para consultas y diálogo, 15 minutos de pausa, y, finalmente, el resto del tiempo de la sesión para el trabajo práctico individual o grupal de los participantes bajo la supervisión de los docentes.

## 11. Descripción de las actividades prácticas indicando lugar donde se desarrollan, modalidad de supervisión y modalidades de evaluación

A lo largo de las sesiones del seminario, cada estudiante podrá comprender cómo distintas técnicas de la lingüística computacional y el procesamiento de lenguaje natural pueden



servir para la producción de recursos lexicográficos, automatizando en diverso grado tareas que sería excesivamente costoso llevar a cabo de manera manual. Estas herramientas se

disponen en orden progresivo de complejidad, presentando primero las que tienen interfaz gráfica y amigable, para pasar luego a otras más avanzadas, tales como la línea de comandos, las expresiones regulares, los lenguajes de *scripting* y software para la visualización de datos y estadística descriptiva.

### **Programa de actividades:**

**Lunes:** Presentación general de la materia y del programa del curso.

**Martes:** Primeras experiencias de trabajo práctico con herramientas para la explotación de datos textuales.

**Miércoles:** El proyecto de creación de un diccionario digital. Primeras experiencias con software para la gestión lexicográfica.

**Jueves:** Ejercicios de procesamiento de datos textuales con expresiones regulares, lenguajes de *scripting* y software para análisis estadístico de datos.

**Viernes:** Integración y ejercitación práctica de lo aprendido a través de un proyecto lexicográfico específico. Ejercicios en grupo destinados a solucionar problemas de procesamiento de datos lexicográficos.

## **12. Evaluación final: modalidad y requisitos de aprobación**

-Porcentaje mínimo de asistencia para que el alumno acceda a la evaluación: 70%.

-La modalidad de aprobación es por medio de un trabajo escrito. La fecha de entrega será una semana posterior a la finalización del dictado de clases, es decir el viernes 11 de noviembre. La fecha de devolución de los resultados por parte de los docentes será de dos semanas después de la entrega, es decir el viernes 25 de noviembre de 2022. Está prevista una fecha de recuperatorio el 2 de diciembre, para los casos en que sea necesario.

## **13. Bibliografía**

### **Bibliografía fundamental**

Atkins, S. & Rundell, M. (2008). *The Oxford Guide of Practical Lexicography*. Oxford: Oxford University Press.

Manning, C.; Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. Cambridge: MIT Press.

Porto Dapena, J. Á. (2002) *Manual de técnica lexicográfica*. Madrid: Arco Libros.



Rojo, G. (2021). *Introducción a la lingüística de corpus en español*. Londres y Nueva York: Routledge, 2021.

### **Bibliografía complementaria**

Baayen, R. H. (2008). *Analyzing Linguistic Data. A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press.

Butler, Ch. (1985). *Statistics in Linguistics*. New York: Basil Blackwell.

De Schryver, G.-M. (2003). Lexicographers' dreams in the electronic dictionary age. *International Journal of Lexicography* 16(2): 143-199.

De Schryver, G.-M. (2012). Lexicography in the crystal ball: facts, trends and outlook (pp. 93-146). En R. Fjeld & J. M. Torjusen (eds.), *Proceedings of the 15th EURALEX International Congress*. Oslo: University of Oslo.

Escandell, V. (2007). *Apuntes de semántica léxica*. Madrid: UNED.

Fillmore, C. J. (1975). An alternative to checklist theories of meaning (pp. 123-131). *Proceedings of the First Annual Meetings of the Berkeley Linguistics Society*. Berkeley: Berkeley Linguistic Society.

Fontenelle, T. (2008). *Practical Lexicography. A Reader*. Oxford: Oxford University Press.

Gries, S. (2009). *Quantitative corpus linguistics with R: a practical introduction*. Londres & New York: Routledge, Taylor & Francis Group.

Hammond, M. (2003). *Programming for Linguists: Perl for Language Researchers*. Oxford: Wiley-Blackwell.

Hanks, P. (2004). Corpus Pattern Analysis (pp. 87-97). En G. Williams & S. Vessier (eds.), *Proceedings of the Eleventh EURALEX International Congress, Euralex 2004*, Lorient: Université de Bretagne-Sud.

Hanks, P. (2013). *Lexical Analysis: Norms and Exploitations*. Cambridge, MA: MIT Press.

Herdan, G. (1964). *Quantitative Linguistics*. Londres: Butterworths.

Landau, S. 2001. *Dictionaries. The Art and Craft of Lexicography*. Cambridge: Cambridge University Press.

Lara, L. F. (1997). *Teoría del diccionario monolingüe*. México, D. F.: El Colegio de México.



UNCUYO  
UNIVERSIDAD  
NACIONAL DE CUYO



FACULTAD DE  
FILOSOFÍA Y LETRAS

POSGRADO  
SECRETARÍA DE  
POSGRADO

Marcus, Solomon; Nicolau, Edmond; Stati, Sorin. (1978). *Introducción en la lingüística matemática*. Barcelona: Teide.

Martínez de Sousa, J. (2009). *Manual básico de lexicografía*. Gijón: Trea.

Porto Dapena, Á. (2014). *La definición lexicográfica*. Madrid: Arco Libros.

Sinclair, J. (2004). *Trust the text. Language, corpus and discourse*. Londres: Routledge.

Stubbs, M. (1996). *Text and Corpus Analysis*. Oxford: Blackwell.

Verzani, J. (2002). SimpleR: Using R for Introductory Statistics [en línea]. Disponible en:  
<https://cran.r-project.org/>

Zgusta, L. (1971). *Manual of Lexicography*. París: Mouton.

ROGELIO NAZAR

Firma y aclaración docente responsable