



## PROGRAMA DE ACTIVIDAD FORMATIVA

### 1. Título de la Actividad

*Procesamiento de corpus para lexicografía y terminología*

### 2. Modalidad o tipo de actividad

*- Curso*

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

Lugar de dictado: Facultad de Filosofía y Letras. UNCuyo

Fecha: 30 y 31 de agosto y 1° de setiembre 2023

Días (de la semana): miércoles, jueves y viernes

Horario: 15 a 17 h

### 4. Duración en horas reales dictadas

Modalidad	Carga teórica	Carga práctica	Total	Porcentaje
Presencial	3	3	6	100%
No presencial	-	-	-	-
Total	3	3	6	100%

**Tutorías (horario y lugar):** Despacho (a designar) de la Facultad de Filosofía y Letras de la U. N. Cuyo, de 17 a 18 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*

### **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 (<http://www.tecling.com>).

**Nicolás Acosta** es estudiante de Letras de la Universidad Nacional de Cuyo. Colabora con el grupo Tecling.com en investigación en los campos de la gramática, la semántica y la lingüística computacional. Se especializa en el desarrollo de software orientado al análisis de textos, basado en el lenguaje de programación Perl. También colabora en investigaciones desarrolladas en el marco de los proyecto corpus EspaDA-UNCuyo, dirigido por la Dra. Viviana Bosio (<https://espada.uncu.edu.ar>).

**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. También colabora con el grupo de investigación Corpus EspaDA-UNCuyo

## **6. Destinatarios**

**6.1. Nivel:** abierto

**6.2. Procedencia:** abierta

**6.3. Cupo mínimo y cupo máximo de asistentes:** mínimo 3 y máximo 30

## **7. Fundamentación**

El curso de Procesamiento de corpus para lexicografía y terminología tiene un enfoque eminentemente práctico y orientado al tratamiento masivo de datos textuales. Se revisarán técnicas para el análisis del vocabulario separando el léxico general 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 de corpus bilingüe (paralelos y comparables). Se revisarán desde herramientas web amigables hasta el empleo de

instrucciones en la terminal de comandos. Al final el curso, los participantes tendrán una base metodológica como para emprender de manera autónoma un producto lexicográfico o terminológico basado en datos de corpus.

## **8. Objetivos**

Proporcionar a los estudiantes herramientas teórico-metodológicas para la creación de bases de datos léxicas o terminológicas basadas en corpus.

## **9. Contenidos**

1. Fundamentos teóricos y antecedentes históricos del cruce disciplinar entre la lexicografía, la terminología y la lingüística computacional.
2. La creación y gestión recursos léxicos y terminológicos
3. Aplicación de técnicas de lingüística computacional y lingüística de corpus.
4. Herramientas, métodos y técnicas para el procesamiento de datos textuales: análisis de vocabulario, extracción de terminología, alineación y explotación de corpus paralelos, expresiones regulares, lenguajes de *scripting* y estadística de corpus.

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

El curso se dicta 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 45 minutos para abordar los conceptos que se trabajarán. Esta exposición continuará con 15 minutos para consultas y diálogo. 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 curso, cada estudiante podrá comprender cómo distintas técnicas de la lingüística computacional y el procesamiento de lenguaje natural pueden servir para la producción de recursos terminológicos y lexicográficos, automatizando en diversas tareas. Estas herramientas se presentan en orden progresivo de complejidad, comenzando por las que tienen interfaz gráfica, para pasar luego a otras más avanzadas, tales como la línea de comandos, las expresiones regulares, los lenguajes de *scripting* y software para visualización de datos y estadística descriptiva.

## **Programa de actividades:**

**Miércoles:** Presentación general de la materia y del programa del curso. Creación y gestión de bases de datos léxicas y terminológicas.

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

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

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

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

## **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.

Sager, Juan Carlos. (1990). *A practical course in terminology processing*. Amsterdam/Philadelphia: John Benjamins.

### **Bibliografía complementaria**

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

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.

Foo, Jody. (2012). *Computational Terminology: Exploring Bilingual and Monolingual Term Extraction*. Licenciante Thesis - Linköping University, Department of Computer and Information Science, NLPLAB - Natural Language Processing Laboratory. Disponible en línea: <https://www.diva-portal.org/smash/get/diva2:508245/FULLTEXT01.pdf>

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

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.

Humbley, John (2022). The reception of Wüster's general theory of terminology. En Pamela Faber y Marie-Claude L'Homme. *Theoretical Perspectives on Terminology*.

- Explaining terms, concepts and specialized knowledge, John Benjamins Publishing Company, pp. 15-36.
- 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.
- Martínez de Sousa, J. (2009). *Manual básico de lexicografía*. Gijón: Trea.
- Pavel, Silvia & Nolet, Diane. (2002). *Manual de Terminología*. Translation Bureau. Canadá: Public Works and Government Services. Disponible en línea: [http://kit.consellodacultura.gal/web/uploads/adxuntos/archivo/593fa5d6bf451-manual\\_terminologia\\_canada.pdf](http://kit.consellodacultura.gal/web/uploads/adxuntos/archivo/593fa5d6bf451-manual_terminologia_canada.pdf)
- 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