



## SECRETARÍA DE POSGRADO

### PROGRAMA

#### 1. Título del seminario: **Lingüística de Corpus**

**Equivalente a Seminario de Investigación I**

**2. Carrera:** Maestría en Lingüística Aplicada, orientación Enseñanza de Lenguas Extranjeras

**3. Modalidad (o tipo de actividad): semi presencial (sincrónico)**

**4. Fecha:** 1er cuatrimestre 2021 (abril, mayo y junio)

**5. Duración:** 20 horas reloj

#### 6. Disertantes

##### 6.1. Docente/s responsable/s (nombre y filiación institucional):

Magdalena Zinkgraf (Fadel, Uncoma)

María Angélica Verdú (Fadel, Uncoma)

##### 6.2. Curriculum vitae sintético de cada uno de los profesores participantes

**Mg. Magdalena Zinkgraf** es Profesora Adjunta en Lengua Inglesa III del Profesorado en Inglés y del Taller de Comunicación Escrita en Inglés de FADEL, UNComa. Posee un Master in ELT and Applied Linguistics King's College, London y el Diploma en Estudios Avanzados (DEA) de UNED, Madrid, (Proyecto de investigación: un estudio de corpus de colocaciones en ensayos de estudiantes avanzados de inglés). Es docente y miembro del Comité Académico de las carreras de Especialización y Maestría en Lingüística Aplicada, Orientación Enseñanza de Lenguas Extranjeras. Desde 1996 se ha formado como investigadora en la adquisición de lengua materna y extranjera. Desde 2014 ha dirigido dos proyectos sobre el estudio de la enseñanza y aprendizaje de secuencias formulaicas en diversos contextos y niveles de inglés como lengua extranjera. Ha dictado cursos de transferencia en relación con sus investigaciones y cursos de escritura académica a investigadores. Ha publicado artículos en revistas nacionales e internacionales y capítulos de libros con relación a la temática de investigación y al estudio de corpus. También ha participado de numerosos congresos y jornadas nacionales e internacionales.

**Mg. María Angélica** es egresada del Profesorado en Inglés de la Universidad Nacional del Comahue, donde se desempeña como Jefe de Trabajos Prácticos en Lengua Inglesa I y III, en la carrera de Profesor de Inglés. Es Máster en Lingüística aplicada a la enseñanza del inglés como lengua extranjera de la Universidad de Jaén, España y posee título de Experta universitaria en e-learning 2.0 y entornos virtuales para la enseñanza de idiomas por la Universidad Tecnológica Nacional, Bs. As. Argentina. Durante 2017-8 realizó una adscripción en la cátedra Inglés 4 aplicando conceptos de la Lingüística de corpus. Ha disertado en cursos y congresos nacionales e internacionales. Es investigadora en la Facultad de Lenguas de la Universidad Nacional del Comahue desde 2014 y sus áreas de investigación se centran en el estudio de la enseñanza y aprendizaje del lenguaje formulaico en contextos de enseñanza de inglés como lengua extranjera.

## 7. Fundamentación

La lingüística de corpus (LC) ha ganado protagonismo en el estudio del español en las últimas décadas pero ha sido desarrollada en mayor medida en la corriente británica y estadounidense de estudios de diversas variedades del inglés. En consecuencia, esta disciplina ha acumulado un bagaje de conocimiento que brinda a los estudiosos en el campo de la lingüística y de la lingüística aplicada herramientas y principios fundamentales para investigar diferentes aspectos del lenguaje en uso en un marco de solidez empírica. A partir de la formulación de preguntas o problemáticas referidas al uso del lenguaje, las herramientas que la LC nos acerca hacen posible explorar de manera científica estos interrogantes, empleando tanto un análisis cualitativo como cuantitativo de los datos compilados en los corpus. Las múltiples aplicaciones de la LC vuelven esta metodología de estudio esencial, especialmente en estos tiempos modernos en que la computación y la informática, así como el almacenamiento de textos digitales, permiten una búsqueda y un análisis complejo de grandes cantidades de textos mediados por la tecnología. La LC ofrece un marco sistemático y colaborativo de investigación en torno a intereses de estudio -y sus cambios a través del tiempo- y una serie de técnicas finitas al servicio del análisis lingüístico de cuestiones puntuales de uso, ampliando el horizonte de estudio en la exploración de corpus ya existentes y en la creación de nuevos corpus específicos. Dado que sólo el investigador en su carácter de humano tiene las facultades para tomar decisiones en cuanto a abordar cuestionamientos lingüísticos, extraer la información, interpretar los datos y discernir entre ellos, una formación en los principios de esta disciplina es de gran utilidad para los especialistas en el lenguaje.

## 8. Objetivos

Una vez finalizado el seminario, los participantes podrán:

- interpretar datos de un corpus usando técnicas como línea de concordancia, colocación y palabra clave,
- describir e incorporar los principios fundamentales que inspiran la LC a sus propios estudios,
- aplicar las técnicas de análisis de corpus al análisis cualitativo y cuantitativo de distintos tipos de datos usando *Antconc* u otro software de concordancia disponible (*Wordsmith*, etc),
- formular una pregunta de investigación orientada a analizar un fenómeno lingüístico empleando corpus ya existentes,
- discernir los lineamientos principales para el diseño de su propio corpus en relación con un interrogante de investigación en torno a un aspecto lingüístico a estudiar,
- diseñar un estudio empírico de corpus empleando la metodología que caracteriza a LC según una pregunta de investigación orientadora,
- dar cuenta de la metodología que caracteriza a LC y sus aplicaciones a diversas disciplinas
- hacer uso crítico y consciente de la LC en sus propios estudios y en la evaluación de otros

## 9. Contenidos

### **Línea temática 1: Fundamentos y nociones teóricas**

Breve historia de la Lingüística de Corpus (LC) y su relación con la investigación en la lingüística aplicadas. Concepciones de LC: basada en corpus (*corpus-based*) e impulsada por corpus (*corpus-*

*driven*). Utilidad, validez y limitaciones de la LC. Caracterización de corpus, tipos y descripción. Corpus disponibles y usos. La web como corpus.

### **Línea temática 2: Análisis e interpretación**

Técnicas básicas de análisis de corpus: *Types, tokens, type-token ratio, wordlists* (listas de palabras y frecuencias), líneas de concordancia (*KWIC*), colocaciones, palabras y secuencias clave. Interpretación de estos datos (análisis cualitativo y cuantitativo). Búsqueda en corpus existentes. Procedimientos y criterios para el diseño, recolección, tamaño y organización de los corpus. Extracción de datos para la construcción de un corpus. Limitaciones de los estudios de corpus.

### **Línea temática 3: Aplicaciones en la lingüística**

Interacciones entre el estudio de diversas disciplinas (Sociolingüística, Análisis del Discurso (crítico), Pragmática, Estilística, Lingüística Forense y otras) y LC. Enseñanza y aprendizaje de lenguas extranjeras y LC (DDL -*data-driven language learning*). Otras aplicaciones de LC: Diccionarios, gramáticas de referencia y libros de texto construidos a partir de corpus. LC vs. Lingüística computacional.

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

Este seminario introduce aspectos teórico-prácticos en torno a la Lingüística de Corpus (LC), partiendo de actividades prácticas y de análisis de la bibliografía propuesta, de las cuales se derivan conceptos teóricos fundamentales, en tanto que se exploran las aplicaciones de la LC en diversos ámbitos de la lingüística. Dado que se trata de un seminario de investigación, este curso brinda las herramientas para que los participantes analicen críticamente estudios realizados en el campo de LC y, como trabajo final, elaboren un posible proyecto de investigación de algún fenómeno lingüístico, empleando técnicas de la LC en un corpus propio o analizando corpus ya existentes. Los encuentros sincrónicos se destinan a debatir cuestiones exploradas en la bibliografía, compartir reflexiones y resolver problemáticas de las actividades propuestas. De acuerdo a sus intereses, seleccionan material bibliográfico específico para luego presentar sus conclusiones en encuentros sucesivos. En la plataforma PEDCO se ofrecen tareas complementarias para completar y entregar de manera virtual, tendientes a explorar conceptos y técnicas de las líneas temáticas y desarrollar los objetivos propuestos arriba.

## **11. Evaluación final**

*Para acreditar el seminario se necesita el 80% de asistencia/participación/etc.*

La evaluación de este taller estará determinada por la asistencia de los participantes, su participación y trabajo en clase y la entrega de los cuatro (4) trabajos prácticos obligatorios y un trabajo final con su respectiva re-escritura, de ser necesario.

### **1) Trabajos prácticos obligatorios**

TP1. Seleccionar un capítulo de un manual y presentar un resumen sobre la aplicación de LC a ese campo/ estudio/ aspecto

TP2. Análisis en grupo/pares de líneas de concordancia de un corpus en torno a algún aspecto lingüístico empleando *AntConc* o algún otro software a tal fin

TP3. Seleccionar un estudio empírico de corpus y presentar un resumen sobre los principios de LC aplicados según el interrogante planteado y otras cuestiones de diseño metodológico del estudio  
TP4. Presentación del anteproyecto del diseño de su propio estudio de corpus (Trabajo Final)

## 2) Trabajo Final:

### Opción A)

Exploración de un corpus preexistente para responder una pregunta o hipótesis de algún fenómeno lingüístico. Describir el corpus seleccionado de acuerdo a los criterios en la línea temática 1.

### Opción B)

Diseñar corpus propio para responder una pregunta o hipótesis de algún fenómeno lingüístico. Detallar el tipo de corpus a diseñar siguiendo la caracterización/tipología en la línea temática 2. Describir método de recolección de datos para conformar el corpus.

En ambos casos se entregará un plan de trabajo que incluya las siguientes secciones:

- Introducción (relevancia de realizar un estudio de corpus en el marco de la investigación que se propone).
- Marco teórico (breve estado de la cuestión: investigaciones previas relevantes al estudio propuesto que justifiquen un nuevo estudio).
- Pregunta o hipótesis sobre el análisis de algún aspecto lingüístico.
- Metodología (Dependiendo de OPCIÓN A) O B) esta sección detallará participantes, instrumentos de recolección de datos y tipo de análisis de datos empleando técnicas de la LC o analizando corpus ya existentes).

Los requisitos que el trabajo final deberá cumplir son los siguientes:

- Fecha de entrega final: 18 de junio
- Documento Word
- Extensión: 3 a 5 páginas A4
- Normas APA 2019-2020

Se establece **una pre-entrega (4 de junio)** con su devolución para acompañar el proceso de escritura del trabajo final, además de la presentación del bosquejo en el último encuentro presencial.

## 12. Bibliografía obligatoria y/o complementaria

- Adolphs, S., & Lin, P. M. S. (2011). Corpus linguistics. En J. Simpson (Ed.) *The Routledge Handbook of Applied Linguistics* (pp. 597-610). Taylor and Francis Balkema.
- Anthony, L. (2018). *AntConc* [software]. <http://www.laurenceanthony.net/>
- Baker, P., Gabrielatos, C., Khosravini, M., Krzyzanowski, M., McEnery, T. & Wodak, R. (2011). “¿Una sinergia metodológica útil? Combinar análisis crítico del discurso y lingüística de corpus para examinar los discursos de los refugiados y solicitantes de asilo en la prensa británica.” *Discurso & Sociedad*, vol. 5(2), pp. 376-416.
- Baker, P., Hardie, A. & McEnery, T. (2006). *A Glossary of Corpus Linguistics*. Edinburgh University Press
- Biber, D. & Reppen, R. (2015). *The Cambridge Handbook of English Corpus Linguistics*. Cambridge University Press.
- Cheng, W. (2012). *Exploring Corpus Linguistics: Language in Action*. Abingdon, Routledge.

- Davies, M. (2012b). “Expanding horizons in historical linguistics with the 400-million-word corpus of historical American English”. *Corpora* 7. 121–157.
- Ghadessy, M., Henry, A. & Roseberry, R.L. (Eds.) (2001). *Small Corpus Studies and ELT: Theory and Practice*. John Benjamins.
- Jiménez-Yáñez, R. M. (2014). “Metadiscursos y persuasión: estudio de editoriales de periódicos españoles sobre la muerte de Osama Bin Laden?”. *Discurso & Sociedad*, vol 8(4), pp. 589-622
- Lew, R. (2009). The Web as Corpus Versus Traditional Corpora: Their Relative Utility for Linguists and Language Learners. En Baker, P. (Ed.) (2009). *Contemporary Corpus Linguistics* (pp. 289-300). Continuum
- McEnery, T. (Director) & McGlashan, M. (Series Ed.). (2013). *Corpus: some key terms* (ESRC Centre for Corpus Approaches to Social Science (CASS) Briefings). Lancaster University
- McEnery, T. & Hardie, A. (2012) Accessing and Analysing corpus data. En *Corpus Linguistics: Method, Theory and Practice* (pp.25-55). Cambridge University Press
- McEnery, T. & Xiao, R. (2010). What corpora can offer in language teaching and learning. En E. Hinkel (Ed.), *Handbook of Research in Second Language Teaching and Learning* (Vol. 2, pp. 364-380). Routledge.
- Meunier, F. & S. Granger (Eds.). (2008) *Phraseology in Foreign Language Teaching and Learning*. John Benjamins Publishing Company
- Parodi, G. (2008). Lingüística de corpus: una introducción al ámbito. *RLA. Revista de Lingüística Teórica y Aplicada*, 46 (1), 93-119. <http://dx.doi.org/10.4067/S0718-48832008000100006>
- Pérez, J. & Zinkgraf, M. (2017). What Goes around Comes around: A Corpus Study of Formulaic Sequences in Advanced Students’ Written Production. En: M. Fernández Beschedt, A. E. Castro, P. Formiga & Z. Risso Patrón (Eds.). *Zonas de contacto: culturas, lenguas y educación: Actas del 4º Congreso Nacional “El conocimiento como espacio de Encuentro”*, Fadel– UNCo. (pp. 88-98) [https://drive.google.com/open?id=1Hzg4Hln9iJYvU\\_mjuKWQKbBTargzgSb4](https://drive.google.com/open?id=1Hzg4Hln9iJYvU_mjuKWQKbBTargzgSb4)
- Reppen, R. & Vlach-Simpson, R. (2020.) Corpus linguistics. En N. Schmitt & M. Rodgers (Eds.), *An Introduction to Applied Linguistics* (3<sup>ra</sup> edición. pp. 89 – 108). Routledge.
- Rojo, G. (2009). “Sobre la construcción de diccionarios basados en corpus”. *Tradumatica*, 7/2009.
- Scott, Mike (2004). *WordSmith Tools* [software]. Oxford University Press.
- Siyanova-Chanturia, A. & Spinam S. (2019) Multi-Word Expressions in Second Language Writing: A Large-Scale Longitudinal Learner Corpus Study. En *Language Learning* 70 (pp. 420–63)
- Zinkgraf, M. / Verdú, M. A. (2015) “Formulaic Sequences Involving ‘Fact’ in EAP Production: a Corpus Study”. En L. Anglada, N.L. Sapag, D. Banegas & M. A. Soto (Eds.). *EFL Classrooms in the New Millennium: Selected Papers from the 40th FAAPI Conference*. (pp. 58-69). <http://www.faapi.org.ar/downloads/FAAPI2015.pdf>
- Zinkgraf, M., Rodeghiero, A. C. & Pérez, J. (2015). “Más allá de la palabra: un estudio de corpus de secuencias formulaicas en ensayos de estudiantes universitarios de ILE”. En D. Riestra, S. M. Tapia & Goicoechea (Comps.). *Cuartas Jornadas Internacionales de Investigación y Prácticas en Didácticas de las Lenguas y Literaturas, Tomo II*. (pp. 752-778)
- Zinkgraf, M., Rodríguez, S., Castro, A. E. & Verdú, M. A. (2015). New perspectives on vocabulary: Formulaic sequences in the structuring of advanced learners’ written production. En A. Castro, P. Formiga & Z. Risso Patrón (Eds.), *Plurilingüismo, diversidad e interculturalidad* :



*un acercamiento desde la educación, traducción e investigación : Actas del III Congreso Nacional El conocimiento como espacio de encuentro.* Fadel– UNCo. EDUCO - Universidad Nacional del Comahue

- Zufferey, S. (2020). *Introduction to Corpus Linguistics*. Iste, London.

**13. Observaciones (si corresponde) -----**