



**UNIVERSIDAD DE BUENOS AIRES
FACULTAD DE FILOSOFÍA Y LETRAS**

DEPARTAMENTO: LETRAS

**SEMINARIO: INTRODUCCIÓN AL ANÁLISIS AUTOMÁTICO
DE LA GRAMÁTICA DEL ESPAÑOL**

MODALIDAD DE DICTADO: VIRTUAL (según Res. (D) N°
732/20 y normativa específica dispuesta a los efectos de organizar el
dictado a distancia)

PROFESOR/A: KOZA WALTER

BIMESTRE DE VERANO

AÑO: 2022

CÓDIGO N°:

UNIVERSIDAD DE BUENOS AIRES
FACULTAD DE FILOSOFÍA Y LETRAS
DEPARTAMENTO DE LETRAS
SEMINARIO: INTRODUCCIÓN AL ANÁLISIS AUTOMÁTICO DE LA GRAMÁTICA DEL ESPAÑOL
MODALIDAD DE DICTADO: VIRTUAL¹
CARGA HORARIA: 64 HORAS
BIMESTRE DE VERANO 2022
CÓDIGO N°:

PROFESOR: WALTER KOZA

a. Fundamentación y descripción

En esta asignatura, de carácter teórico-práctico, se abordan contenidos fundamentales de la morfología y la sintaxis del español, desde una perspectiva formal y dentro del marco del procesamiento del lenguaje natural basado en reglas. A tales efectos, se presentan, por un lado, diversas problemáticas referentes a los mecanismos de formación de palabras y la manera en que estas se combinan y establecen diversas funciones sintácticas. Por otro lado, se plantean métodos de formalización de dichas estructuras con vistas al diseño de algoritmos computacionales que permitan su identificación en textos de lenguaje natural y su generación automática. Se espera que, al finalizar el curso, los estudiantes adquieran conocimientos básicos para el desarrollo de recursos de análisis automáticos y, así, desarrollar investigaciones en el plano gramatical.

b. Objetivos:

Objetivo general:

- Analizar fenómenos morfológicos y sintácticos del español desde una perspectiva formal con vista al modelamiento computacional para su reconocimiento en textos de lenguaje natural y generación automática

Objetivos específicos:

- Conocer los procesos de formación de palabras, sintagmas y oraciones
- Conocer las metodologías del análisis automático basado en reglas
- Desarrollar algoritmos computacionales básicos para el análisis automático de aspectos específicos de la gramática del español

c. Contenidos:

Unidad 1: Representación formal del conocimiento lingüístico

- Introducción al procesamiento del lenguaje natural
- Lenguajes naturales y lenguajes formales

¹ Programa adecuado a las pautas de funcionamiento para la modalidad virtual establecidas en Res. (D) N° 732/20 y otra normativa específica dispuesta a los efectos de organizar la cursada en el contexto de la emergencia sanitaria que impide el desarrollo de clases presenciales en la Universidad.

- La jerarquía Chomsky-Schützenberger
- El análisis automático de textos basado en reglas: aplicaciones

Unidad 2: Análisis automático de la morfología del español

- Aspectos de la morfología flexiva y derivativa del español
- Gramáticas morfológicas informáticas para generación y reconocimiento automático
- Construcción de diccionarios electrónicos

Unidad 3: Análisis automático de la sintaxis del español

- El nivel sintagmático: sintagmas léxicos y funcionales
- La oración simple: formalización de la estructura argumental del predicado
- Gramáticas sintácticas informáticas para generación y reconocimiento automático

d. Bibliografía, filmografía y/o discografía obligatoria, complementaria y fuentes, si correspondiera:

Unidad 1: Representación formal del conocimiento lingüístico

Bibliografía obligatoria

- Bonino, R. & Rodrigo, A. (mimeo). NooJ: Tutorial en español.
- Chomsky, N. (1969). *Syntactic structures*. Mouton. La Haya.
- Moreno Sandoval, A. (1998). *Lingüística computacional*. Madrid: Editorial Síntesis. [selección de capítulos]
- Silberztein, M. (2016). *Formalizing natural languages: The NooJ approach*. Londres: ISTE. [selección de capítulos]
- Silberztein, M. (2003). *NooJ manual*. [En línea] Disponible en: <http://nooj4nlp.org/downloads.html>

Bibliografía complementaria

- Bosque, I. & Demonte, V. (1999). *Gramática Descriptiva de la Lengua Española, vol. 2 Las construcciones sintácticas fundamentales. Relaciones temporales, aspectuales y modales*. Espasa. Madrid.
- Bosque, I. & Gutiérrez-Rexach, J. (2011). *Fundamentos de sintaxis formal*. Madrid: AKAL Editores. [selección de capítulos]
- Chomsky, N. (1970). *Aspects of the theory of syntax*. Massachusetts: MIT Press.
- Di Tullio, A. (2007). *Manual de gramática del español*. Buenos Aires: Waldhuter Ediciones.
- Gallego, A. (2015). *Perspectivas de sintaxis formal*. Madrid: AKAL.
- Periñán, C. (2012). En defensa del procesamiento del lenguaje natural fundamentado en la lingüística teórica. *Onomazein*, 26, 13-48.
- RAE (2009). *Nueva gramática de la lengua española*. Madrid: Espasa.
- Solias Arís, T. (2015). *Métodos formales en lingüística*. Madrid: Editorial Síntesis [selección de capítulos]

Unidad 2: Análisis automático de la morfología del español

Bibliografía obligatoria

- Bonino, R. & Rodrigo, A. (mimeo). NooJ: Tutorial en español.
- Mella Llopart, M. (1983). Análisis morfológico automático del español. *Procesamiento del lenguaje natural*, 1, 18-23.
- Silberztein, M. (2016). *Formalizing natural languages: The NooJ approach*. Londres: ISTE. [selección de capítulos]
- Silberztein, M. (2003). *NooJ manual*. [En línea] Disponible en: <http://nooj4nlp.org/downloads.html>
- Varela, S. (2005). *Morfología léxica: la formación de palabras*. Madrid: Gredos. [selección de capítulos]

Bibliografía complementaria

- Bosque, I. & Demonte, V. (1999). *Gramática Descriptiva de la Lengua Española, vol. 2 Las construcciones sintácticas fundamentales. Relaciones temporales, aspectuales y modales*. Espasa. Madrid.
- Bosque, I. & Gutiérrez-Rexach, J. (2011). *Fundamentos de sintaxis formal*. Madrid: AKAL Editores.
- Di Tullio, A. (2007). *Manual de gramática del español*. Buenos Aires: Waldhuter Ediciones.
- Fábregas, A. (2013). *La morfología. El análisis de la palabra compleja*. Editorial Síntesis. Madrid. [selección de capítulos]
- RAE (2009). *Nueva gramática de la lengua española*. Madrid: Espasa.

Unidad 3: Análisis automático de la sintaxis del español

Bibliografía obligatoria

- Bonino, R. & Rodrigo, A. (mimeo). NooJ: Tutorial en español.
- Gross, G. (2014). *Manual de análisis lingüístico*. Barcelona: Editorial UOC. [selección de capítulos]
- Gross, M. (1984). Lexicon-Grammar and the syntactic analysis of French. En *International Conference on Computational Linguistics (COLING)*. California: Stanford University.
- Koza, W. & Mare, M. (en prensa). Computational modeling of a generative grammar for clitic order in River Plate Spanish. *RESLA*.
- Silberztein, M. (2016). *Formalizing natural language: The NooJ approach*. Londres: ISTE. [selección de capítulos]
- Silberztein, M. (2003). *NooJ manual*. [En línea] Disponible en: <http://nooj4nlp.org/downloads.html>

Bibliografía complementaria

- Bosque, I. & Demonte, V. (1999). *Gramática Descriptiva de la Lengua Española, vol. 2 Las construcciones sintácticas fundamentales. Relaciones temporales, aspectuales y modales*. Espasa. Madrid.
- Bosque, I. & Gutiérrez-Rexach, J. (2011). *Fundamentos de sintaxis formal*. Madrid: AKAL Editores.
- Di Tullio, A. (2007). *Manual de gramática del español*. Buenos Aires: Waldhuter Ediciones.
- Gross, M. (1994). The lexicon-grammar of a language. *Encyclopedia of language*. Oxford: Pergamon Press.
- La Serna, N. (2004). Un analizador sintáctico eficiente para gramáticas del español. *RISI*, 1(1), 19-26.
- RAE (2009). *Nueva gramática de la lengua española*. Madrid: Espasa.
- Rivera, R., Barrón, C. & Calvo, H. (2016). Análisis y comparación de diferentes métodos de reconstrucción de árboles semánticos. *Research in computing science*, 120, 81-88.

e. Organización del dictado de seminario

El seminario se dicta en modalidad virtual mientras duren las restricciones actuales. Su funcionamiento se adecua a lo establecido en la Res. (D) N° 732/20 y a la normativa específica dispuesta a los efectos de organizar el dictado a distancia.

El dictado de clases se realiza a través del campus virtual de la Facultad de Filosofía y Letras y de otros canales de comunicación virtual que se consideren pertinentes para favorecer el intercambio pedagógico con los/las estudiantes.

La carga horaria total es de 64 horas.

Modalidad de trabajo

Este curso se desarrolla a través de la discusión y práctica de los programas de análisis lingüístico en el aula, sobre la base de lecturas, exposiciones orales y ejercicios prácticos; para ello, es necesario que los estudiantes cuenten con computador y el programa de análisis computacional NooJ. Las sesiones se estructuran en tres partes. En primer lugar, se expone un tema, el cual es comentado y complementado por los otros participantes a la luz de la bibliografía del curso. En segundo lugar, se realiza una actividad práctica, individual o grupal, para ello, se les instalarán las herramientas de análisis automático presentadas en el curso. Por último, el profesor cierra la sesión con una síntesis de los temas tratados.

Descripción de las actividades y tareas

La asignatura contempla la elaboración de implementaciones computacionales básicas para el abordaje de problemas lingüísticos de índole gramatical. Los estudiantes deberán diseñar los recursos pertinentes (diccionarios, gramáticas, corpus, etcétera) y evaluar sus alcances y limitaciones. Para ello, es necesario que cuenten con computadora, en la primera sesión se darán las indicaciones para la instalación de la herramienta NooJ (Silberztein, 2016).

f. Organización de la evaluación

El sistema de regularidad y aprobación del seminario se rige por el Reglamento Académico (Res. (CD) N° 4428/17) e incorpora las modificaciones establecidas en la Res. D 732/20 para su adecuación a la modalidad virtual de manera excepcional:

Regularización del seminario:

Es condición para alcanzar la regularidad del seminario aprobar una evaluación con un mínimo de 4 (cuatro) durante la cursada. Para ello los/las docentes a cargo dispondrán de un dispositivo definido para tal fin.

Aprobación del seminario:

Los/as estudiantes que cumplan el requisito mencionado podrán presentar el trabajo final integrador que será calificado con otra nota. La calificación final resultará del promedio de la nota de cursada y del trabajo final integrador.

Si el trabajo final integrador fuera rechazado, los/as interesados/as tendrán la opción de presentarlo nuevamente antes de la finalización del plazo de vigencia de la regularidad. El/la estudiante que no presente su trabajo dentro del plazo fijado, no podrá ser considerado/a para la aprobación del seminario.

VIGENCIA DE LA REGULARIDAD: El plazo de presentación del trabajo final de los seminarios es de 4 (cuatro) años posteriores a su finalización.

RÉGIMEN TRANSITORIO DE ASISTENCIA, REGULARIDAD Y MODALIDADES DE EVALUACIÓN DE MATERIAS: El cumplimiento de los requisitos de regularidad en los casos de estudiantes que se encuentren cursando bajo el Régimen Transitorio de Asistencia, Regularidad y Modalidades de Evaluación de Materias (RTARMEM) aprobado por Res. (CD) N° 1117/10 quedará sujeto al análisis conjunto entre el Programa de Orientación de la SEUBE, los Departamentos docentes y los/las Profesores a cargo del seminario.

g. Recomendaciones

- Contar con computadora
- Poseer conocimientos básicos de gramática



Firma

Walter Koza
Aclaración

A handwritten signature in black ink, appearing to read 'Miguel Vedda', written in a cursive style.

Miguel Vedda
Director del Departamento de Letras