



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

DEPARTAMENTO: LETRAS

MATERIA: MODELOS FORMALES NO TRANSFORMACIONALES

RÉGIMEN DE PROMOCIÓN: EF

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

PROFESOR: MIGUEL VEDDA

CUATRIMESTRE: 2°

AÑO: 2021

CÓDIGO N°: 0581

UNIVERSIDAD DE BUENOS AIRES
FACULTAD DE FILOSOFÍA Y LETRAS
DEPARTAMENTO DE LETRAS
MATERIA: MODELOS FORMALES NO TRANSFORMACIONALES
MODALIDAD DE DICTADO: VIRTUAL¹
RÉGIMEN DE PROMOCIÓN: EF
CARGA HORARIA: 96 HORAS
CUATRIMESTRE y AÑO: 2º CUATRIMESTRE DE 2021
CÓDIGO N° 0581

PROFESOR: MIGUEL VEDDA (DIRECTOR DEL DEPARTAMENTO)

EQUIPO DOCENTE²:

JEFE DE TRABAJOS PRÁCTICOS REGULAR: BALBACHAN, FERNANDO
AYUDANTE DE 1º CATEGORÍA REGULAR: CARRANZA, FERNANDO

a. Fundamentación y descripción

En la tradición de la lingüística formal, se le reconoce a la obra de Chomsky *Estructuras Sintácticas* (1957) un rol fundacional. Una de las principales tesis del modelo del '57 se refiere a la complejidad de los lenguajes naturales, proponiendo a la sintaxis como un componente autónomo que genera oraciones mediante reglas de estructura de frase – también conocidas como Gramática Independientes de Contexto (*Context Free Grammar*, CFG) – y la aplicación *a posteriori* de reglas transformacionales.

La irrupción de la Gramática Generativo-Transformacional del '57 marcó un cambio de paradigma en la investigación lingüística. Sin embargo, a partir de la aproximación formal a las propiedades de los lenguajes naturales, han surgido diversas propuestas superadoras de la postura chomskyana inicial – el mismo Chomsky habría de dejar de lado el componente transformacional en sus modelos posteriores del '81 (Principios y Parámetros), y '95 (Programa Minimalista).

El presente programa ha sido pensado para alumnos tanto de la carrera de Letras de la orientación en Lingüística como de la carrera de Filosofía con orientación en Lógica, y para todos aquellos que tengan un fuerte interés en el estudio de la estructura sintáctica y del

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.

² Los/as docentes interinos/as están sujetos a la designación que apruebe el Consejo Directivo para el ciclo lectivo correspondiente.

lenguaje natural y, en especial, en los alcances de la utilización de los modelos formales para el abordaje del mismo.

b. Objetivos

El objetivo fundamental de este curso es centrarse en el relevamiento minucioso de las teorías gramaticales NO transformacionales que evolucionaron a partir de la primera gramática chomskyana y que representan una importante línea de investigación en Procesamiento de Lenguaje Natural o *Natural Language Processing* (NLP), poniendo especial énfasis en las propiedades formales, poder expresivo y limitaciones de cada modelo en el estado del arte. Asimismo, se busca que el alumno trabaje con los rudimentos de implementaciones de dichas gramáticas bajo la forma de *parsers* (*analizadores sintácticos*), demostrando aplicaciones prácticas en español de los modelos teóricos.

c. Contenidos

La materia será dictada en 5 unidades que delinean un preciso recorrido teórico a través de los Modelos Formales No Transformacionales, desde sus orígenes en la lingüística chomskyana hasta las propuestas más recientes que evidencian marcadas mejoras en el Procesamiento de Lenguaje Natural y que atestiguan un cambio de paradigma científico desde los modelos simbólicos a los enfoques estadísticos y, más recientemente, hasta los modelos basados redes neuronales. Adicionalmente, el curso incluye una reflexión integral acerca de las implicancias lingüísticas y epistemológicas de una teoría gramatical.

Unidad I: Bajo la égida de Chomsky

Se introducirán la obra fundacional de Chomsky y las nociones formales con las que se ha de trabajar más adelante. Puntualmente se expondrá la jerarquía chomskyana respecto de las gramáticas (desde las más simples modelos de autómatas a las Máquinas de Turing, pasando por las Gramáticas Independientes de Contexto CFG). A su vez se expondrán las limitaciones en el poder expresivo de cada una de las Gramáticas, recurriendo tanto a lenguajes formales como a lenguajes naturales: por ej. el problema de la recursividad en la gramáticas de autómatas y la concordancia básica sujeto-verbo y el orden variable de constituyentes en las CFGs.

- 1.- Gramáticas: reglas de estructura de frase y transformacionales
 - 1.1 Dependencia e independencia del contexto
 - 1.2 Transformaciones sintácticas
- 2.- La jerarquía de Chomsky respecto de las gramáticas formales
- 3.- Tipos de autómatas: autómata finito, autómata descendente, autómata lineal

- 4.- Lenguajes regulares y autómatas
- 5.- Gramática Independiente de Contexto (CFG). Ventajas y limitaciones

Unidad II: La superación del modelo chomskyano

Se desarrollarán modelos formales alternativos a la CFG que trabajan con predicados lógicos de primer orden y el operador lógico *Unificación*. En particular se presentará la Gramática de Cláusula Definida – *Definite Clause Grammar* (DCG). Se puntualizará su utilidad como la primera propuesta de superación de las limitaciones expresivas de las Gramáticas Independientes de Contexto, aunque también se marcarán los problemas detectados en este tipo de modelos: por ej. la concordancia sujeto-verbo y la subcategorización verbal en español.

- 1.- Gramática de Cláusula Definida (DCG)
 - 1.1 Lógica de predicados de primer orden
 - 1.2 El operador lógico Unificación
- 2.- Ventajas sobre CFG. Aplicaciones prácticas en lenguaje natural
- 3.- Poder expresivo y limitaciones

Unidad III: Modelos optimizados fuera del área de injerencia chomskyana

Se abordarán las gramáticas basadas en la estructura de *rasgos (features)*. El gran salto cualitativo que representan estas gramáticas respecto del paradigma chomskyano es que intentan resolver los problemas tradicionales de los lenguajes naturales -como constituyentes discontinuos u orden variable en español- a través de operaciones de unificación de matrices de *features*, propiedades de herencia y restricciones de linealidad. Específicamente se trabajará con la gramática denominada Gramática de Estructura de Frase gobernada por Núcleos *Head-driven Phrase Structure Grammar-HPSG*, desarrollada por Ivan Sag.

- 1.- El concepto de *rasgo (feature)*
- 2.- Matriz de *feature*
- 3.- Implementación: herencia, jerarquías de tipos y restricciones de linealidad
- 4.- Head-driven Phrase Structure Grammar-HPSG Aplicaciones en español

Unidad IV: El paradigma estadístico

El paradigma estadístico aplica diversas técnicas probabilísticas a grandes *corpora* de entrenamiento, con vistas a inducir categorías y fenómenos específicos del lenguaje natural a partir de la detección de patrones estadísticamente significativos en la *tabula rasa* que constituyen los *corpora*. Sin embargo, el paradigma estadístico es más que una mera

aplicación de técnicas y modelización matemática: estos enfoques aportan evidencia de plausibilidad psicolingüística a un renovado debate acerca de la naturaleza misma del lenguaje.

- 1.- Principios fundantes del paradigma estadístico
 - 1.1 Cadenas de Markov y la equiparación de probabilidad y gramaticalidad
 - 1.2 Información distribucional e información mutua
- 2.- El debate epistemológico en torno al Argumento de la Pobreza de los Estímulos (APS).
La investigación actual sobre *corpus*
 - 2.1 Anotación morfosintáctica de *corpus*
 - 2.2 Técnicas de clustering e inducción de fenómenos sintácticos
 - 2.3 Aprendizaje automático o *machine learning*

Unidad V: El paradigma conexionista de redes neuronales

Evolución histórica hacia los modelos actuales de redes neuronales: validación translingüística y adaptación a diferentes dominios. Estado del arte y vigencia de los algoritmos de *deep learning* en múltiples tareas estándares del campo disciplinar.

- 1.- ¿Qué es una red neuronal?
 - 1.1 El concepto de back-propagation
 - 1.2 Arquitecturas recientes en *deep learning*
- 2.- Diferencias epistemológicas y plausibilidad psicolingüística para los modelos formales no transformacionales. Parsing y Procesamiento de Lenguaje Natural dentro de la Lingüística Computacional

d. Bibliografía específica (subrayada la bibliografía obligatoria)

Unidad I

- **Chomsky, N. 1957. Estructuras sintácticas. México, SigloXXI. (hay diversas ediciones en español)**
- **Eguren, L. y Fernández Soriano, O. 2004. Introducción a una sintaxis minimalista. Madrid, Gredos. Cap.1**
- Grune, D. y Jacobs, C. *Parsing Techniques. A practical guide*. Amsterdam, Springer, 2008
- **Jurafsky, D y Martin, J. 2000. Speech and language processing. New Jersey. Prentice-Hall. caps. 2 y 9**
- Partee, B., Meulen, A. y Wall, R. 1993. *Mathematical Methods in Linguistics*. Dordrecht: Kluwer Academic Publishers.

- Savich, W. Bach, E., Marsh, W. y Safran-Naveh, G. 1987 *The Formal Complexity of Natural Languages* Dordrecht: Kluwer Academic Publishers.
- Sag, I. And Wasow, Th. 1999. *Syntactic Theory. A Formal Introduction*. Stanford, California: CSLI Publications. Apéndice B.
- **Sandoval Moreno, Antonio. 1998. *Linguística Computacional*. Madrid. Editorial Síntesis**
- **Hierro S. Pescador, José. 1980. *Principios de Filosofía del Lenguaje I. Teoría de los signos y de la Gramática*. Madrid. Alianza Editorial cap.4**

Unidades II y III

- Alsina, A. 1996. *The Role of Argument Structure in Grammar. Evidence from Romance*. Stanford, California: CSLI Publications.
- **Jurafsky, D y Martin, J. 2000. *Speech and Language Processing*. New Jersey. Prentice-Hall. cap. 11**
- Radford, A. 1997. *Syntactic Theory and The Structure of English. A Minimalist Approach*. Cambridge: Cambridge University Press.
- Radford, A. 1997. *Syntax. A Minimalist Introduction*. Cambridge: Cambridge University Press.
- Pollard, C. y Sag, I. 1987. *Information-Based Syntax and Semantics*. Stanford, CA: CSLI Publications.
- Pollard, C. y Sag, I. 1994. *Head-Driven Phrase Structure Grammar*. Stanford, CA: CSLI Publications.
- Sag, I. y Wasow, T. 1999. *Syntactic Theory. A Formal Introduction*. Stanford, California: CSLI Publications.
- Sells, P. 1989 *Teorías sintácticas actuales (GB, GPSG, LFG)* Barcelona, Ed. Teide.
- **Russell S. y Norvig P. 1995. *Inteligencia Artificial: un enfoque moderno*. New Jersey. Prentice Hall caps. 22 y 23**
- **Sandoval Moreno, Antonio. 2001. *Gramáticas de Unificación y Rasgos*. Madrid .Ed. Antonio Machado**
- Shieber, S. 1986. *An Introduction to Unification-Based Approaches to Grammar*. Stanford, CA: CSLI Publications.

Unidad IV

- Abney, S. 1996. "Statistical methods and linguistics". En: Klavans J. y Resnik, P. (eds.), *The balancing act*. The MIT Press, Cambridge, MA.
- **Balbachan, F. 2014. *Técnicas de clustering para inducción de categorías sintácticas en un corpus de español*. Tesis de doctorado (UBA-FFvL).**

- **Balbachan, F. y Dell’Era, D. 2010. "Inducción de constituyentes sintácticos en español con técnicas de clustering y filtrado por información mutua". En *Linguamática* n°2, pp.39-57**
- Chater, N. y Manning, C. 2006. “Probabilistic models of language processing and acquisition”. En *TRENDS in Cognitive Sciences*, vol.10 n°7, pp.335-344
- **Clark, A. 2001. *Unsupervised language acquisition: theory and practice*. Sussex. School of Cognitive and Computing Sciences, University of Sussex Press.**
- **Jurafsky, D y Martin, J. 2000. *Speech and Language Processing*. New Jersey. Prentice-Hall. caps. 1, 10, 15, 21**
- **Manning, C. y Schütze H. 1999. *Foundations of Statistical Natural Language Processing*. The MIT Press. Cambridge (Massachusetts). cap. 1**
- **Redington M. et al. 1998. "Distributional information: a powerful cue for acquiring syntactic categories". En *Cognitive Science* Vol 22 (4) 1998. pp. 425-469**

Unidad V

- **Flechas, N. y Balbachan, F. (en prensa). Modelos de Deep Learning para NLU: estado de la cuestión y evolución del paradigma.**
- Peacocke, Ch. 1989. "When is a grammar psychologically real?" en George, A. (comp.) *Reflections on Chomsky*. Cambridge: Basil Blackwell.
- **Pinker, S. 1994. *El instinto del lenguaje*. Versión española de José Manuel Igoa González. Madrid: Alianza, 1995.**
- Savich, W. Bach, E., Marsh, W. Y Safran-Naveh, G. 1987. *The Formal Complexity of Natural Languages* Dordrecht: Kluwer Academic Publishers.
- **Shieber, S. 1985. “Evidence against the context-freeness of natural language”. En *Linguistics and Philosophy* (8) pp.333-343**
- Soames, S. 1984. "Linguistics and Psychology" en *Linguistics and Philosophy* 7.
- Soames, S. 1985. "Semantics and Psychology" en Katz, J. (comp.) *The Philosophy of Linguistics*. Cambridge: Oxford University Press.

Bibliografía general

- **Balbachan, F. (editor) Dell’Era, D., Berros J., Cruces, M. Haimovici, S. 2009. *Cuadernillo de ejercicios y apuntes*. Serie Fichas de Cátedra. (FFyL-UBA)**
- Bolshakov, I. y Gelbukh, A. 2004. *Computacional Linguistics: model, resources, applications*. México. UNAM
- Grune, D. y Jacobs, C. 2007. *Parsing techniques: a practical guide*. Amsterdam. Springer

e. Organización del dictado de la materia:

La materia se dicta en modalidad virtual mientras duren las restricciones establecidas por el Aislamiento Social Preventivo y Obligatorio definido por el gobierno nacional (DNU 297/2020). 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 96 horas

Modalidad de trabajo

La materia estará dividida en clases teóricas y prácticas. En las primeras, se realizarán exposiciones acerca de los contenidos expuestos en el ítem anterior. En las segundas, se resolverán ejercicios relacionados con la parte teórica. Oportunamente se repartirá una guía de ejercicios para cada trabajo práctico a través del campus virtual de la Facultad de Filosofía y Letras.

f. Organización de la evaluación

La materia se dicta bajo el régimen de promoción con EXAMEN FINAL (EF) establecido en el Reglamento Académico (Res. (CD) N° 4428/17) e incorpora las modificaciones establecidas en la Res. (D) N° 732/20 para su adecuación a la modalidad virtual de manera excepcional.

-Regularización de la materia:

Es condición para alcanzar la regularidad de la materia aprobar 2 (dos) instancias de evaluación parcial (o sus respectivos recuperatorios) con un mínimo de 4 (cuatro) puntos en cada instancia.

Quienes no alcancen las condiciones establecidas para el régimen con EXAMEN FINAL deberán reinscribirse u optar por rendir la materia en calidad de libre.

-Aprobación de la materia:

La aprobación de la materia se realizará mediante un EXAMEN FINAL presencial en el que deberá obtenerse una nota mínima de 4 (cuatro) puntos. La evaluación podrá llevarse a cabo cuando las condiciones sanitarias lo permitan.

Se dispondrá de UN (1) RECUPERATORIO para aquellos/as estudiantes que:

- hayan estado ausentes en una o más instancias de examen parcial;
- hayan desaprobado una instancia de examen parcial.

La desaprobación de más de una instancia de parcial constituye la pérdida de la regularidad y el/la estudiante deberá volver a cursar la materia.

Cumplido el recuperatorio, de no obtener una calificación de aprobado (mínimo de 4 puntos), el/la estudiante deberá volver a inscribirse en la asignatura o rendir examen en calidad de libre. La nota del recuperatorio reemplaza a la nota del parcial original desaprobado o no rendido.

La corrección de las evaluaciones y trabajos prácticos escritos deberá efectuarse y ser puesta a disposición del/la estudiante en un plazo máximo de 3 (tres) semanas a partir de su realización o entrega.

VIGENCIA DE LA REGULARIDAD:

Durante la vigencia de la regularidad de la cursada de una materia, el/la estudiante podrá presentarse a examen final en 3 (tres) mesas examinadoras en 3 (tres) turnos alternativos no necesariamente consecutivos. Si no alcanzara la promoción en ninguna de ellas deberá volver a inscribirse y cursar la asignatura o rendirla en calidad de libre. En la tercera presentación el/la estudiante podrá optar por la prueba escrita u oral.

A los fines de la instancia de EXAMEN FINAL, la vigencia de la regularidad de la materia será de 4 (cuatro) años. Cumplido este plazo el/la estudiante deberá volver a inscribirse para cursar o rendir en condición de libre.

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 la cátedra.



Miguel Vedda
Director del Departamento de Letras