



Simposio “Análisis cuantitativo de datos lingüísticos”

Fecha: 9 de Febrero de 2019

Horario: 12:00 a 19:00.

Lugar: Auditorio Jorge Luis Borges de la Biblioteca Nacional (Agüero 2502)

Cronograma

12:00-12:15

Palabras previas

12:15-13:30

Bloque 1: Métodos cuantitativos y estadísticos en ciencias del lenguaje

Métodos cuantitativos y estadísticos en las ciencias del lenguaje: panorama, predicciones y recomendaciones

Damián Blasi
(Max Planck Institute, University of Zurich, Yale University)

Resumen

Esta charla provee un brevísimo panorama de los métodos cuantitativos en las ciencias del lenguaje tal y como se emplean hoy: por qué los necesitamos, cómo se usan en la práctica, qué rol juegan a lo largo de las ciencias del lenguaje y cómo es probable que se vean en un futuro cercano.

Relaciones entre medidas de complejidad lingüística

Germán Coloma
(Universidad del CEMA)

Resumen

Este trabajo sintetiza una serie de conceptos que han sido propuestos por la literatura lingüística para medir la complejidad de los distintos aspectos de los idiomas naturales (fonología, morfología, sintaxis, semántica). Se adentra también

Simposio “Análisis Cuantitativo de Datos Lingüísticos”

en el estudio de las relaciones que pueden establecerse entre dichas medidas de complejidad, y analiza la posible existencia de “efectos de compensación” entre tales medidas (es decir, situaciones en las cuales una característica más compleja de un idioma se ve contrarrestada por otra más simple). Para ello se van introduciendo distintos criterios estadísticos y empíricos, que van desde las correlaciones simples hasta correlaciones parciales, coeficientes de regresión y parámetros surgidos de sistemas de ecuaciones. La mayoría de los conceptos se ilustran apelando a dos bases de datos interlingüísticas que ya han sido utilizadas por el autor en trabajos anteriores.

13:30-14:30

Pausa programada para almuerzo.

14:30-15:45

Bloque 2: Psicolingüística, lingüística computacional y métodos

Cinco dilemas de la metodología en investigación psicolingüística

Carolina Gattei
(IFIBA-CONICET, Laboratorio de Neurociencia - UTDT)

Resumen

Uno de los objetivos de la psicolingüística es explicar, a través del uso paradigmas propios de la psicología experimental, cuáles son los mecanismos neurocognitivos y neurofisiológicos que tienen lugar durante la realización de tareas relacionadas con lenguaje (comprensión, producción, adquisición de lenguaje, etc.). En la presente charla se presentarán algunos de los dilemas metodológicos a los que se enfrentan los investigadores que trabajan en este campo desde el momento de abordaje de su objeto de estudio hasta la integración de los resultados a los modelos psicolingüísticos vigentes. Además, se plantearán posibles soluciones para sortear estos problemas a partir de la aplicación de herramientas propias de otras ramas de la lingüística.

Simposio “Análisis Cuantitativo de Datos Lingüísticos”

Presentación de Grupo de Lingüística Computacional

Macarena Fernández Urquiza
(Facultad de Filosofía y Letras, Universidad de Buenos Aires)
Fernando Schiaffino
(Facultad de Filosofía y Letras, Universidad de Buenos Aires)

Resumen

Nuestro grupo de lingüística computacional está formado por estudiantes y graduados de la carrera de Letras que se han especializado en el área de la Lingüística Computacional, principalmente en el Procesamiento del Lenguaje Natural. Desde el 2017 hemos dictado dos cursos de extensión y nos reunimos regularmente tanto para estudiar diferentes temas relacionados con la lingüística formal, la lingüística matemática, el procesamiento del lenguaje natural y los métodos cuantitativos como para investigar y aprender a usar las herramientas de Procesamiento de lenguaje natural disponibles. En esta presentación vamos a contar algunos de nuestros proyectos, nuestros objetivos e intereses y algunos de los logros obtenidos desde el momento de formación del grupo. Puede encontrarse más información sobre nosotros en <https://sites.google.com/view/grupodepln>

Presentación del “Grupo de Estadística para el Estudio del Lenguaje”

Federico Alvarez
(Facultad de Filosofía y Letras, Universidad de Buenos Aires)
Charo Zacchigna
(Facultad de Filosofía y Letras, Universidad de Buenos Aires)

Resumen

El grupo de estadística para el estudio del lenguaje (GESEL) se formó en 2018 con el propósito de incorporar una perspectiva cuantitativa a la formación clásica en lingüística. En una primera fase nos ocupamos de adquirir una base teórica a fines de facilitar el acceso al material del área. En este mismo proceso generamos material de estudio destinado a compartir el contenido que fuimos abarcando con compañeros con nuestra misma formación de base, como una primera aproximación al área e intentando suplir la falta de material de estudio en castellano y orientado a lingüistas. Tras haber presentado una propuesta de proyecto FiloCyT, presentaremos los lineamientos de la misma y nuestros objetivos para este año. Nuestra página web es <https://gesel.github.io/>

15:45-16:00

Primera pausa programada para descanso

Simposio “Análisis Cuantitativo de Datos Lingüísticos”

16:00-17:15

Bloque 3: Dialectología y lingüística histórica

Lexicografía y dialectología del español argentino: análisis desde las redes sociales

Damián E. Aleman
(Grandata)
Santiago Kalinowski
(Academia Argentina de Letras)

Abstract

The task of detecting regionalisms (expressions or words used in certain regions) has traditionally relied on the use of questionnaires and surveys, and has also heavily depended on the expertise and intuition of the surveyor. The irruption of Social Media and its microblogging services has produced an unprecedented wealth of content, mainly informal text generated by users, opening new opportunities for linguists to extend their studies of language variation. In this work we present three novel metrics based on Information Theory to detect regionalisms on Twitter. Our metrics take into account both the number of occurrences of the word in certain regions and the number of users who mention it. This tool has helped lexicographers discover several unregistered words of Argentinian Spanish, as well as different meanings assigned to registered words.

Lingüística histórica moderna: métodos filogenéticos

Ezequiel Koile
(Max Planck Institute for the Science of Human History)

Resumen

En esta charla introduciremos una serie de métodos modernos en lingüística histórica. Estos surgen de una colaboración interdisciplinaria entre ramas tan diversas como la lingüística, la genética, las ciencias de la computación, la arqueología y la paleogeografía. Discutiremos sus aplicaciones y sus resultados más salientes.

15:45 a 16:00

Segunda Pausa programada para descanso

17:30-18:45

Bloque 4: Procesamiento de lenguaje natural

Simposio “Análisis Cuantitativo de Datos Lingüísticos”

PLN como ventana a la mente y a la sociedad

Edgar Altszyler
(Fundación Sadosky - Laboratorio de Inteligencia Artificial Aplicada, DC, FCEN,
UBA)

Resumen

Existe una gran variedad de técnicas de Procesamiento de Lenguaje Natural que permiten cuantificar similitud semántica entre palabras. Durante la presentación introduciré estas herramientas y contaré sobre algunas aplicaciones de las mismas, que van desde la detección de estereotipos en subtítulos de películas hasta el análisis de contenido en sueños

“PLN, ciencia auxiliar para las humanidades digitales”

Gimena del Rio Riande
(IIBICRIT-HD CAICYT Lab, CONICET)
Nidia Hernández
(HD CAICYT Lab-DILA, CONICET)

Resumen

Las humanidades digitales (HD) constituyen un campo de investigación emergente de carácter interdisciplinario que trabaja y reflexiona en la intersección de las humanidades, las artes y las ciencias sociales con métodos y herramientas computacionales. Un rasgo distintivo de las HD es la relación bidireccional que proponen entre las humanidades y lo digital: se emplea la tecnología digital para la investigación en humanidades pero, a la vez, se interroga a la tecnología y a las humanidades en tanto objetos de estudio. Las HD buscan enriquecer el campo de los estudios literarios, culturales, históricos –cuya aproximación ha sido, por lo general, históricamente cualitativa– a través de un abordaje cuantitativo que descubre potencialidades inéditas en la lectura e interpretación de corpus –principalmente textuales– a partir del uso de medios digitales y aplicaciones informáticas.

El procesamiento del lenguaje natural (PLN) es un método importante para los humanistas digitales, ya que permite a los investigadores buscar relaciones y patrones entre palabras en grandes corpus. Una gran cantidad de investigaciones en HD involucran la aplicación de tareas del PLN, tales como el análisis de sentimientos, el reconocimiento de entidades nombradas o el topic modelling. Esto ha hecho que en los últimos años se haya comenzado a describir al PLN como una ciencia auxiliar para las HD. Esta presentación realizará un acercamiento general a las HD y a su relación con el PLN y la lingüística de corpus desde la época de la Informática Humanística, brindando ejemplos de proyectos y grupos de

Simposio “Análisis Cuantitativo de Datos Lingüísticos”

investigación internacional y nacional, como los de Humanidades Digitales CAICYT Lab, <https://hdcaicyt.github.io/>.

18:45-19:00

Palabras finales

Comité organizador:

Damián Blasi (Max Planck Institute, University of Zurich, Yale University)

Fernando Carranza (Universidad de Buenos Aires, CONICET)

Carolina Gattei (IFIBA-CONICET, Laboratorio de Neurociencia-UTDT)

Ezequiel Koile (Max Planck Institute for the Science of Human History)

**Avalado por el Departamento de Letras de la Facultad de Filosofía y Letras
(Universidad de Buenos Aires)**

**Auspiciado por el Instituto de Filología y Literaturas Hispánicas “Dr. Amado
Alonso” (Universidad de Buenos Aires)**