

DOSSIER

Lingüística Computacional

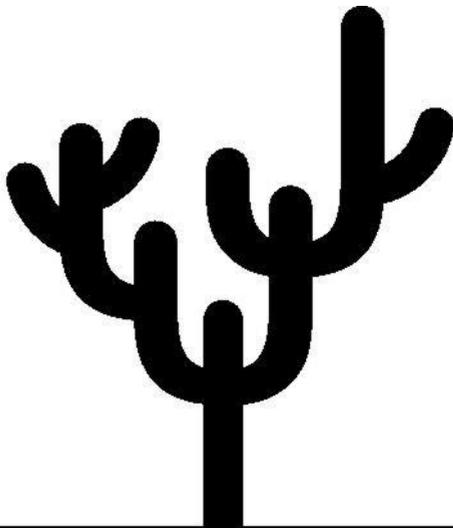
DIRECCIÓN Y COORDINACIÓN

Dr. Rogelio Nazar

 <https://orcid.org/0000-0002-8853-1353>

Instituto de Literatura y Ciencias del Lenguaje
Pontificia Universidad Católica de Valparaíso

 rogelio.nazar@pucv.cl



Introducción

Rogelio Nazar  <https://orcid.org/0000-0002-8853-1353>

Instituto de Literatura y Ciencias del Lenguaje
Pontificia Universidad Católica de Valparaíso, Chile

 rogelio.nazar@pucv.cl

La lingüística computacional es un área de especialidad que ha cobrado gran auge en las últimas décadas con la abundancia de información en formato digital disponible en internet. Con múltiples ramas relacionadas, como la lingüística cuantitativa, la lingüística de corpus o el procesamiento del lenguaje natural, se presenta como un campo interdisciplinario muy atractivo para los interesados en emprendimientos de distinto tipo vinculados al procesamiento de la información textual. Además de la lingüística teórica, se cruza con la informática, la estadística, la inteligencia artificial y todas las áreas en las que este conocimiento puede ser aplicado. También, dentro de la misma lingüística teórica esta especialidad tiene un papel fundamental, puesto que está cambiando la forma de trabajar. Cada vez con mayor frecuencia se exige un respaldo empírico, a través de datos obtenidos de corpus, con el fin de sustentar cualquier formulación teórica.

Si bien es cierto que los avances en el campo han sido extraordinarios, también hay que notar que se han dado de manera desigual, ya que el desarrollo de conocimiento teórico y recursos lingüísticos se ha concentrado en unas pocas lenguas en detrimento de otras. Hasta en castellano, lengua tan extendida en el mundo, se dispone de mucho menos material en comparación con otras lenguas como en el inglés o el francés.

En castellano, a su vez, la mayor parte de lo que existe se ha desarrollado en países como España y México.

Sorprende que el campo se haya desarrollado tan poco en América del Sur y, en particular, en Argentina. El caso de este país resulta paradójico en distintos sentidos, porque tiene una población que desborda en cultura, creatividad e inteligencia, y a pesar de este gran capital humano, no parece conseguir desarrollar su potencial. El ámbito de la lingüística computacional es solamente uno más de los campos en los que resulta incomprensible que no se haya alcanzado un mayor desarrollo. El país tiene todas las condiciones para ser una potencia mundial en el área, pero es poca la gente que se especializa en ella. No es un problema de inversión, porque no se requiere de grandes recursos tecnológicos para poner en marcha una investigación en lingüística computacional. Solo exige tiempo de estudio, al tratarse de una materia compleja.

Las oportunidades, sin embargo, son muchas y muy importantes. En general, quienes trabajan en lingüística computacional están más bien interesados en la lingüística teórica. Esto es, se desarrollan modelos cuantitativos o computacionales, pero es sobre todo para teorizar o avanzar conocimiento del lenguaje mismo. No obstante, además de esta parte teórica, el campo también puede aplicarse a la solución de una serie de problemas concretos.

Es especialmente la parte aplicada de la disciplina la que hace pensar que debería ser una política de Estado el incentivar su desarrollo, ya que esta tiene una importancia estratégica. Disponer de la capacidad del tratamiento masivo de información resulta vital para un país. En particular lo es el poder convertir datos no estructurados (textos) en datos estructurados (bases de datos), es decir, extraer la información que circula en los textos y ordenarla en tablas. Una vez que la información se organiza en tablas, se puede someter a todo tipo de tratamiento estadístico.

Algunos ejemplos concretos de esto pueden ser el análisis de la marcha de la economía a partir de lo que circula en publicaciones, ya sea

especializadas o informales, a través de noticias sobre inversiones, fusiones o adquisiciones de empresas. Otro ámbito puede ser el de las estadísticas de delito a partir de crónicas policiales: dónde se cometen y a qué horas. Otros pueden ser el estudio de comentarios positivos o negativos acerca de un producto, un concepto o un individuo, o las alertas tempranas de alguna amenaza a partir de tendencias de lo que se publica en redes sociales, entre otra diversidad de ejemplos. Los artículos que se presentan en este dossier mencionan ya algunos ejemplos, como la detección de problemas de salud mental o la generación de mensajes publicitarios.

En este contexto, el objetivo principal del dossier es reunir un conjunto de artículos de investigación en el área de la lingüística computacional para difundir líneas de trabajo de la comunidad a nivel local. La publicación de estos artículos contribuirá a la construcción de un espacio para el diálogo, el intercambio de ideas y métodos, así como la reflexión y debate acerca de los desafíos pendientes. Tengo la impresión, además, de que esta publicación es histórica, ya que creo es la primera vez que en Argentina se dedica al tema un número especial de una revista.

El primero de los artículos está a cargo de Franco Luque, de la Universidad Nacional de Córdoba, y presenta un marco general para todo el dossier. En este texto, Franco explica las posibilidades que ofrece la teoría de lenguajes formales para el estudio de las lenguas naturales, centrado en particular en el caso de los fenómenos sintácticos. Además de presentar una revisión histórica de la materia, define cuáles son las potencialidades y los límites de este enfoque.

Ese artículo es seguido por dos que abordan las posibilidades de aplicación de la lingüística computacional al ámbito de la salud mental. El primero de estos está a cargo de María José Garciarena, Leticia Cagnina y Marcelo Errecalde, de la Universidad Nacional de San Luis. Se trata de un tema que ha despertado mucho interés en el ámbito de la lingüística computacional en los últimos años: el de las posibilidades de aplicación de este conocimiento al dominio de las ciencias de la salud, y en este caso en

particular, de la salud mental. El artículo presenta un modelo de detección temprana de síntomas de depresión por medio del análisis de rasgos textuales específicos.

El artículo siguiente está a cargo de la argentina Gabriela Ferraro y su colega catalán Luis Salvador-Carulla, quienes trabajan actualmente para la Australian National University, en Canberra, y sigue una línea similar al anterior. En este caso se aborda el análisis de *posts* en los Grupos de Apoyo en Internet con el objetivo de identificar rasgos textuales que sirvan para detectar publicaciones de individuos con problemas de salud mental que estén en un momento de crisis.

Los últimos dos abordan ámbitos de aplicación diferente. El primero de ellos, a cargo de Fernando Balbachán, de la Universidad de Buenos Aires, y sus colaboradores de la empresa Natural Tech, aborda el tema de la utilización de modelos de *Deep Learning* para las tareas de generación de texto en un problema concreto como la escritura de eslóganes publicitarios.

El último artículo es el que apporto yo, desde la Pontificia Universidad Católica de Valparaíso. En este trabajo presento una descripción del proyecto Kind, que es un algoritmo de inducción automática de taxonomías léxicas. Estas son estructuras de sustantivos mono y poliléxicos organizados por la relación semántica de hiperonimia.

Varios investigadores que habían manifestado interés en participar finalmente no pudieron aportar sus artículos debido a compromisos previamente adquiridos que les impidieron cumplir con las fechas de entrega. Pero albergo la esperanza de que este dossier sea el disparador de nuevas posibilidades, como la publicación de otro número dedicado al tema. Esto podría ser un paso más en la dirección de un mayor desarrollo del campo y la creación de oportunidades laborales en el país.

Nota biográfica

Rogelio Nazar

Rogelio Nazar es Licenciado en Comunicación Social por la Universidad Nacional de Cuyo (Argentina) y Doctor en Ciencias del Lenguaje y Lingüística Aplicada por la Universidad Pompeu Fabra (España). Actualmente es profesor del Instituto de Literatura y Ciencias del Lenguaje de la Pontificia Universidad Católica de Valparaíso (Chile). Su área de investigación es la lingüística computacional y trabaja en lexicología, terminología y análisis computacional del discurso.