

RESEÑAS

GUILLERMO ROJO

Introducción a la lingüística de corpus en español

Routledge. University of Minnesota, USA, 2021, 380 páginas

ISBN: 978-03-6763-584-8

La lingüística del corpus es una de las ramas aplicadas de la ciencia que estudia el lenguaje. Una disciplina con un desarrollo reciente, favorecida especialmente por la rápida evolución de los sistemas computacionales. El volumen *Introducción a la lingüística de corpus en español*, escrito por Guillermo Rojo, expone de forma sencilla, didáctica y descriptiva la situación de la lingüística del corpus (LC, de ahora en adelante) en el ámbito hispánico, específicamente. Para este objetivo, el escritor ha estructurado siete capítulos, los cuales entregan la conceptualización necesaria para comprender la materia tratada en cada capítulo. Cada uno de ellos está precedido por un resumen y finaliza con un apartado de lecturas recomendadas en caso de que el/la lector/a desee profundizar en algún aspecto del capítulo tratado, además de actividades prácticas que permiten poner a prueba la información presentada en cada capítulo. Se incluyen las notas del autor en las últimas páginas de cada capítulo, por lo que la lectura de esta obra requiere una constante revisión de las notas, actividad que, por un lado, complementa la lectura, y por otro, la detiene, para consultar el número de nota y luego continuar leyendo.

El texto posee una serie de funciones discursivas que sirven para cumplir el objetivo central del autor: en primer lugar, la función informativa caracteriza de modo general al volumen, presentando, explicando, detallando y problematizando cada apartado en cada capítulo. En segundo lugar, el autor propone una función práctica, permitiendo a la persona que lee realizar los procedimientos leídos en cada apartado y subapartados. Estos procedimientos se conciben como actividades, preguntas y prácticas.

El primer capítulo es una introducción general a las conceptualizaciones que tendrán uso en todo el volumen, por lo tanto, abundan las definiciones y tipologías sobre lo que se entiende por corpus, sus tipos, usos y también los fundamentos de la LC. La presentación de la información, de modo general, siempre incluye contexto histórico, dado que el tema central es la LC en el mundo hispánico. El autor describe brevemente los panoramas de esta disciplina desde otras tradiciones, como, por ejemplo, la inglesa. El segundo capítulo, al igual que los restantes, retomará aquellos aspectos mencionados sucintamente en el primero para desarrollarlos con mayor detenimiento. En esta sección se presentan los objetivos de la LC, su carácter e intereses en los datos lingüísticos y sus pretensiones de análisis con los textos reales producidos de manera escrita u oral, enfocado en el análisis empírico y exhaustivo de grandes cantidades de este tipo de textos.

El tercer capítulo (y los dos siguientes) comienza a describir y explicar sobre el diseño, creación y usos de los corpus. Por lo tanto, se utiliza abundante

conceptualización nueva, específicamente por la gran variedad tipológica de corpus existentes hasta la fecha. Características centrales, tales como el tamaño, los tipos de textos que contiene, el formato en que se encuentran dichos textos o la utilidad ofrecida por la composición de los distintos corpus señalados son algunos de los parámetros que se detallan en este apartado, entendiendo que la LC requiere un grado de manejo en el lenguaje computacional. También se habla sobre los formatos de texto que pueden componer un corpus, como la codificación HTML, o el lenguaje SGML. Con respecto al uso de los corpus, el cuarto capítulo se enfoca en el proceso de recuperación de información desde corpus textuales, con el fin específico de estudiar fenómenos léxicos. En este sentido, se presenta la conceptualización que se usará en el capítulo y más adelante se explica cómo utilizar un corpus según parámetros de estudios lingüísticos asociados al léxico, como las variaciones diatópica, diacrónica, diafásica y diastrática. El capítulo número cinco trata acerca de la recuperación de la información en diversos corpus con el objetivo de presentar las posibilidades de estudio de fenómenos gramaticales. Se recurre a conceptos tales como la frecuencia general y normalizada, propia de la LC y aspectos gramaticales de interés para estudiar el uso de modos y tiempos verbales, perífrasis verbales, y algunas estructuras sintácticas, por ejemplo, “se los dije”; “ir a + infinitivo”, entre otras.

El sexto capítulo, vuelve a retomar algunas ideas mencionadas de forma general en los capítulos anteriores, con el fin de profundizarlas un poco más, lo que permite comprender un poco más a cabalidad algunos aspectos específicos de la LC como las ventajas y desventajas que presenta el uso de un tipo de corpus, conceptos versátiles de la disciplina, como el tamaño, representatividad y equilibrio de los corpus, entre otros. También se observan proyecciones del autor con respecto al desarrollo de la LC. El séptimo, último capítulo, presenta algunas herramientas de recuperación de datos que no se encuentran en las plataformas de corpus que se explican y presentan en el volumen, el autor considera este capítulo complementario con respecto a los otros. Aquí se encuentran una serie de prácticas, descritas con gran detalle y una clara especificidad, constituyendo un paso a paso para realizar análisis con el sistema Unix, el cual se presenta por medio de comandos computacionales, que el autor explica y usa en las prácticas finales del volumen. Las prácticas finales presentan una configuración más próxima al lenguaje computacional, siempre orientado a analizar fenómenos lingüísticos.

Es interesante observar la función práctica que propone esta obra, puesto que el autor describe literalmente cómo utilizar diversos corpus, a saber, CREA, CORPES, PRESEEA, ESLORA, CdeHist, entre otros. La descripción es un paso a paso, en donde el autor entrega el enlace del corpus y expone qué teclas presionar, en qué orden, qué campos completar y con qué palabras, detallando qué información se presentará a continuación y explicando cómo interpretar dichos resultados. Junto a ello, también se presenta una abundante cantidad de tablas y gráficos, extraídos desde los mismos corpus descritos que sirven para ilustrar todo el proceso descrito. De esta manera, se complementa la exposición de la información con una herramienta visual, siendo este un rasgo didáctico que es transversal en todo el volumen. Esta forma de presentar el uso de los corpus permite entenderlos desde la teoría y la práctica, lo

que disminuye la compleja barrera entre el estudio de los fenómenos lingüísticos y el uso de herramientas computacionales, muchas veces ajenas a los profesionales de la lengua, por lo que esta obra, puede ser utilizada como un referente para quienes deseen aproximarse al mundo de la LC. La cantidad de información, referencias y actividades entregan un panorama general bastante completo para una primera instancia de lectura especializada.

Pablo Quintanilla
Universidad de Chile
pablo.quintanilla@ug.uchile.cl
<http://orcid.org/0000-0002-2707-1576>