

Lingüística de corpus y discursos especializados: puntos de mira.

Editado por Giovanni Parodi. Chile: Ediciones Universitarias de Valparaíso, Pontificia Universidad Católica de Valparaíso, 2007. 461 págs.

Viviana Oropeza Gracia

Universidad Nacional Autónoma de México

Por lo general, una de las críticas que reciben los estudios teóricos en las áreas de la Lingüística radica en la falta de confrontación empírica. Esto es, se les exige pruebas del lenguaje natural, ya sean de forma hablada o escrita. Este debate es más frecuente entre los defensores de la Lingüística de Corpus, que tienden a mostrar la importancia de un estudio de corpus para sustentar cualquier análisis teórico.

El libro *Lingüística de corpus y discursos especializados: puntos de mira* surge como un sólido ejemplo del estudio de cualquier fenómeno lingüístico a partir de herramientas metodológicas computacionales para identificar y recuperar material oral o escrito que apoye cualquier investigación teórica.

Los estudios aquí presentados están apoyados por herramientas computacionales desarrolladas por la Pontificia Universidad Católica de Valparaíso. Así, la Pontificia Universidad Católica de Valparaíso se ubica como una institución innovadora que logra, a través de su Escuela de Lingüística, amalgamar dos prácticas lingüísticas que se creían excluyentes.

Giovanni Parodi ha logrado conjuntar dieciocho artículos de investigación empírica, agrupados en seis apartados que, en general, indagan acerca de la comprensión y producción de textos que han sido analizados en diversos contextos gracias al uso de herramientas computacionales. Los apartados con los que se encontrará el lector son:

Parte 1: Lingüística de corpus y recursos computacionales

Parte 2: Análisis multiregistros del español

Parte 3: Comprensión del discurso

Parte 4: Discurso periodístico

Parte 5: Discurso escolar

Parte 6: Análisis semántico latente: Teoría y empiria

Pese a su corte interdisciplinario, este libro está dirigido, principalmente, a los investigadores de la lingüística en general, que deseen evidencia de trabajos apoyados en herramientas computacionales que facilitan la búsqueda de series léxicas, sintácticas y/o discursivas.

Uno de los puntos en común que se encuentran en *Lingüística de corpus y discursos especializados: puntos de mira* es la noción general de unidades léxicas en contexto. Evidentemente recupera la importancia del examen riguroso y el análisis de expresiones lingüísticas dentro de un discurso. Jamás se aparta de la idea de que una unidad lingüística no se encuentra aislada. En este sentido, este libro, editado por Parodi (2007), nos ofrece el resultado de un esfuerzo que combina estudios discursivos apoyados por herramientas de búsqueda computacionales.

A continuación reseño cuatro artículos que abarcan algunas áreas de interés lingüístico, tales como la creación de herramientas de búsqueda, el análisis de comportamiento sintáctico de algunos verbos del español a partir de un estudio de corpus especializado y el estudio del discurso divulgativo-argumentativo científico.

“El Grial: Interfaz computacional para anotación e interrogación de corpus en español” es la reseña de una herramienta computacional que explica el funcionamiento de este motor de búsqueda, el cual está disponible de manera gratuita en el sitio (www.elgrial.cl). Parodi muestra algunas ventajas que nos ofrece esta herramienta para el campo de la investigación, pues entre sus funciones destaca que es posible “anotar morfosintácticamente textos en archivos digitales planos en lengua española y [que es posible también] recuperar esta información en forma de consultas de bases de datos” (32). Además, gracias a las etiquetas que este programa tiene, el investigador puede clasificar gramaticalmente las palabras de un texto y analizar los tipos de estructuras lingüísticas que aparecen en los diferentes textos recopilados por el corpus, una tarea que ahora se ve facilitada por esta herramienta computacional.

Este artículo se complementa con un ejemplo de este tipo de herramientas, pues en “El Manchador de Textos: una herramienta computacional para el análisis de textos”, René Venegas y Julio Silva (53-76) nos explican detalladamente en qué consiste “marcar” estructuras dentro de un texto y cómo esto ayuda al investigador.

Marcar es identificar una estructura a partir de rasgos lingüísticos determinados por el propio investigador, que posteriormente se colorearán dentro del texto para facilitar visualmente el lugar en el que éstas aparecen. Esta herramienta también permite mostrar la frecuencia de aparición de cada secuencia de rasgos,

información, que dentro de un macroestudio resulta relevante a la hora de comparar datos sincrónicamente.

La búsqueda de rasgos lingüísticos está organizada por forma y por lema. Entre los ejemplos de rasgos que se pueden buscar se encuentran: secuencias gramaticales, palabras exactas, partes de palabras (sufijos), flexiones verbales y partes de la oración, entre otros.

Vale la pena mencionar que una de las oportunidades que tiene cualquier estudio de corpus es que el investigador identifica la expresión de su interés, así como sus contextos de aparición y co-ocurrencia. Esta información arroja luz sobre estudios relativos a la variación lingüística que buscan identificar, en algunos casos, sincrónicamente, tendencias de comportamiento de ciertas estructuras lingüísticas.

Un ejemplo de un estudio apoyado por herramientas computacionales se encuentra en el texto preparado por Ornar Sabaj. El autor presenta en “Los esquemas preposicionales en verbos de comunicación”, una discusión teórica y descriptiva con respecto a los complementos argumentales preposicionales de los verbos de comunicación, está fuertemente apoyada por un estudio de corpus. El objetivo de este artículo, en palabras del propio autor, es “describir comparativamente el comportamiento de los esquemas preposicionales en dos tipos de verbos de comunicación en un corpus multiregistro y relacionar ese comportamiento con las características macro-contextuales de los registros estudiados” (187).

Es claro que el objetivo lingüístico parte de una noción teórica que se apoya por la búsqueda en corpus de una secuencia de elementos lingüísticos para identificar la frecuencia de uso y el mismo uso de los esquemas preposicionales de los verbos de comunicación. Es un exitoso trabajo que, gracias al análisis de los registros arrojados por el corpus, logra determinar que los verbos de comunicación, seleccionados por el autor, se presentan en un esquema sintáctico formado por verbos cognitivos + la preposición ‘de’ que presenta cláusulas nominales, estructura que fuertemente favorece la aparición de los verbos en cuestión. Además, fue posible identificar que dichos verbos aparecen en contextos escolares y dentro de discursos científicos.

El último capítulo que reseño es el que da cuenta de la organización argumentativa dentro de los discursos de divulgación científica en la prensa escrita chilena. Se trata de un estudio que busca examinar las estrategias discursivas de los artículos de divulgación científica que presentan información relativa a la

biogenética y que, además, intentan presentar argumentos en favor de la manipulación genética. Como es esperado, dichos textos deben mostrar razones y pruebas contundentes para modificar cualquier sistema de creencias. En “Los textos de divulgación de biogenética en la prensa escrita chilena: Análisis de su organización argumentativa”, Juana Mamkovich y Silvana Ferrari presentan una clara exposición teórica relativa al discurso de divulgación, que acompañan con lo que entienden por “discurso argumentativo”. Para ellas, un discurso de este último tipo debe contar con tres instancias fundamentales: problematización, posicionamiento y prueba, las cuales son ampliamente esclarecidas a lo largo del estudio.

En esta investigación seleccionaron un conjunto de textos referidos a la biogenética, del corpus llamado Divulgación de la Ciencia y Tecnología en la Prensa Escrita. Así, formaron un corpus de investigación que les permitió identificar las estrategias discursivas por las cuales los científicos presentan argumentos y, por otro lado, la prensa divulga el conocimiento científico. De este modo, fue posible concluir que el discurso de los medios impresos que trata sobre la problematización es heterogéneo y mediado por el editorialista. En lo relativo al posicionamiento, los periodistas asumen una postura neutral y apoyan los argumentos que se apoyan en el conocimiento científico, argumentos que muestran las pruebas a favor de la manipulación genética.

Habrà que reconocer que el estudio de corpus, apoyado por herramientas de búsqueda computacionales, es fundamental para el exitoso resultado que presentan las investigaciones recopiladas en *Lingüística de corpus y discursos especializados: puntos de mira*. Evidentemente, resulta esencial conocer las nuevas herramientas que se están generando para apoyar la investigación, pues si bien facilitan el trabajo de búsqueda, permiten también conocer, identificar y explorar contextos de aparición y co-ocurrencia de secuencias lingüísticas, así como su frecuencia de uso. Además, posibilitan las investigaciones que examinan estrategias discursivas dentro de textos especializados, investigaciones inminentemente actuales dentro del área de la Lingüística.