

Mar Cruz Piñol. *Lingüística de corpus y enseñanza del español como 2/L*. Madrid: Arco/Libros, 2017. 189 págs.

Beatriz G. Granda
Universidad Nacional Autónoma de México
Centro de Enseñanza para Extranjeros

El libro de Piñol, *Lingüística de corpus y enseñanza del español como 2/L*, tiene el objetivo de proporcionar la información necesaria que permita el uso de datos de corpus informáticos, con el fin de que el maestro de español como segunda lengua pueda usarlos en sus trabajos de investigación y en su aplicación en materiales para el aula de clases.

El uso de datos empíricos que apoyen cualquier investigación o elaboración de materiales didácticos es hoy un requisito indispensable para otorgar validez a la información. Por esta razón es indiscutible el interés creciente en la lingüística de corpus (LC), pero también es un hecho, como menciona la autora, que la mayoría de los especialistas en enseñanza de español como 2/L no conocen los alcances que puede tener el uso de un corpus textual, ni cuentan con las habilidades necesarias para consultarlo de acuerdo con sus intereses particulares.

Con el fin de facilitar a los profesores de español como 2/L la incorporación de los corpus textuales en sus clases, la autora presenta la LC de una manera didáctica y amable para el lector y distribuye el contenido en dos secciones: una que aborda aspectos relacionados con las características generales de la disciplina, en la que se refiere fundamentalmente a definir términos y explicar la metodología específica para la elaboración de los corpus, y otra que, a partir de esos conocimientos, se enfoca en describir y ejemplificar sus posibles usos en las clases de español como 2/L.

Piñol abre la primera sección de la obra con la definición de LC de Sinclair (2004), que toma como marco teórico para su trabajo y que, en su opinión, sintetiza los aspectos fundamentales de un corpus lingüístico, entre los que destaca: la recopilación de textos o fragmentos de texto de acuerdo con criterios objetivos de selección que representen, tanto como sea posible, la lengua que es objeto de estudio.

La autora retoma los antecedentes de esta disciplina refiriéndose a trabajos de análisis de corpus en los que aún no se hablaba en términos de LC ni se usaban computadoras para procesar los datos. Entre otros, menciona el trabajo de lingüistas de la primera mitad del siglo xx que se interesaron por datos empíricos para analizar las lenguas, como Boas para las lenguas amerindias, o acervos de muestras de lengua más formales como las listas de concordancia (aparición de palabras en contexto) del *Index Thomisticus*, elaborado entre los años cuarenta y noventa del siglo pasado, y que fue informatizado en 2005.

Es a mediados del siglo xx cuando se desarrollan importantes corpus textuales de lengua inglesa, como el *Survey of English Usage* (SEU) que fue la base para tres corpus de referencia de estudios lingüísticos angloamericanos con los que se inicia la que hoy se denomina lingüística de corpus, que se caracteriza por el uso de técnicas automatizadas a partir de computadoras.

Piñol hace un breve pero imprescindible recuento de las etapas por las que pasó la LC, vinculadas con diversas corrientes lingüísticas; esto resulta muy apropiado para esclarecer la relación entre las concepciones sobre la lengua y sus implicaciones en el desarrollo del pensamiento lingüístico, incluida la importancia del uso de datos tomados de contextos de uso de la lengua. Así, un primer auge en el interés por el uso de datos empíricos se da en el marco del estructuralismo, etapa en la que la observación de datos reales (de la *parole*) era el método apropiado para el análisis de la lengua. El innegable interés por el pensamiento que Chomsky desarrolla en una etapa posterior, centrado en dar cuenta de la competencia del hablante, de su capacidad para generar un número infinito de oraciones y de las reglas subyacentes que permiten esta capacidad humana, tuvo como consecuencia un énfasis en el análisis de aspectos universales y abstractos de la lengua. Más tarde, el pensamiento lingüístico se centró en la función social del lenguaje y en las condiciones contextuales de la producción del mismo, lo que generó un renovado interés por el uso de corpus como fuente de información, sumado a los avances informáticos que han permitido procesar grandes volúmenes de datos y ampliar la extensión de muestras de lengua que podemos estudiar.

Es en este contexto, a finales del siglo, en el que los corpus lingüísticos pasan a ser una necesidad para la comunidad académica; la razón más importante de su uso es la necesidad de contar con datos obtenidos del uso real y coti-

diano de la lengua, que permitan al investigador verificar sus teorías de un modo más objetivo, transparente y válido. A partir de este momento y hasta la actualidad se han producido corpus en muy distintas lenguas y con muy diferentes finalidades, a los que la autora hace referencia.

Un concepto clave que es necesario comprender para un mejor uso de corpus textuales es su *representatividad*, “la piedra angular de la LC” de acuerdo con Piñol, quien sostiene que un corpus no puede abarcar la totalidad de una lengua, sino que es una muestra representativa de la misma; representatividad que dependerá de las decisiones y los criterios que se tomen en el proceso de creación y compilación de un corpus. En cuanto a las decisiones involucradas en la representatividad del español, la autora menciona el trabajo de la Real Academia Española (RAE) en la conversión del Corpus de Referencia del Español Actual (CREA), que modificará la representación de las áreas geográficas, de modo que América pasará a tener un 70% y España un 30% (Rojo, 2008), en lugar del 50/50 que este corpus contemplaba anteriormente. Sin duda, menciona la autora, se trata de un paso importante hacia la representatividad del CREA.

En el apartado titulado “Extracción de información: *concordancias* y *frecuencias*” se presentan las dos maneras básicas de visualizar la información de un corpus: los *listados de frecuencias* y la consulta de *concordancias*. El primer método nos permite recabar datos sobre las palabras que son más frecuentes en la lengua de manera descontextualizada, mientras que el segundo presenta las palabras buscadas en su contexto. Ambas maneras de acceder a la información son útiles para diferentes propósitos.

Cabe mencionar que, para facilitar la comprensión de lo que se va exponiendo a lo largo del texto, se incluyen capturas de imágenes de algunos corpus, a manera de ejemplo, para que el lector pueda visualizar la forma en que se presenta la información (la obra incluye 74 imágenes que ilustran tanto consultas como aplicaciones para la enseñanza); también se incluyen los enlaces correspondientes a cada uno de los corpus para facilitar su consulta.

Resulta muy esclarecedor un ejemplo tomado de Rojo (2001), en el que se presenta una tabla con los verbos más frecuentes del español en tres corpus (Base de Datos Sintácticos del Español Actual, BDS; Léxico Informatizado del Español, LexEsp, y Diccionario de frecuencias del español de Juilland-Chang) para ilustrar cómo se pueden obtener resultados muy diferentes de acuerdo con

el proceso de selección de textos y procedimientos de análisis lingüístico a que se haya sometido el corpus. Así, según interpreta Rojo:

Es fácil observar las discrepancias entre las tres listas y suponer su causa. *Haber e ir* en la BDS están mucho más abajo que en las otras dos listas precisamente porque no se han tenido en cuenta más que sus usos como verbos plenos. En sentido contrario, creo que se puede pensar que *decir* ocupa en la BDS una posición más alta que en las otras listas como consecuencia del peso, quizás excesivo que los textos narrativos tienen en ella (2008: 265).

En esta sección, la autora nos ofrece una diversidad de ejemplos de listados de frecuencias, de líneas de concordancias y de concordancias y combinaciones de palabras; ejemplos de corpus en español en los que se pueden consultar tanto concordancias y colocaciones como programas para obtener concordancias, colocaciones y frecuencias; también aborda una descripción de la web como corpus.

En el capítulo 5, titulado “Lematización y análisis de corpus”, se describen los corpus que incluyen algún tipo de análisis (por ejemplo “categoría gramatical”) o información metalingüística. Al respecto, la autora destaca el hecho de que el valor de un corpus no se mide por el número de palabras que contiene sino por el nivel de análisis a que haya sido sometido: un mayor nivel de análisis aumenta enormemente la utilidad del corpus. *Lematizar* un corpus, en palabras de la autora, es relacionar cada *forma* (la palabra tal como aparece en el texto) con su *lema* (la entrada bajo la que buscaríamos esa palabra en un diccionario). Como ejemplo para realizar una consulta en un corpus lematizado, se muestran los resultados de la búsqueda del verbo *ser* en el Corpus Técnico del Institut Universitari de Lingüística Aplicada (CT-IULA). Con solo introducir el infinitivo (el lema) se obtiene el listado de concordancias de *todas las formas* de este verbo que aparece en el corpus. En un corpus no lematizado, en cambio, solo se obtendría la forma *ser*, y si se quisiera buscar todas las formas flexionadas del paradigma, habría que hacerlo una por una.

Además, la autora hace una breve referencia a programas lematizadores en línea, afirmando que más allá de entender el funcionamiento técnico de estos programas, su consulta puede ser de utilidad para la enseñanza de español 2/L ya que estos aportan información sobre la estructura morfológica de la lengua.

Se muestran dos ejemplos correspondientes a la búsqueda de la palabra *casa*; el primero se trata de una consulta de manera aislada en la que se obtienen los posibles lemas de *casa* (como sustantivo y como verbo) y, el segundo, una búsqueda en su contexto oracional (*vivo en una casa grande*) en la que se obtiene el lema que corresponde al contexto de la oración con la que se hace la consulta. En esta última búsqueda el programa reconoce cuando *casa* es un sustantivo y cuando es una forma flexionada del verbo *casar*.

Además de relacionar cada forma con su lema, un corpus se puede procesar de modo que se indique la *categoría gramatical* correspondiente a cada forma, la *función sintáctica* de cada palabra en su contexto y también información *semántica* o *pragmática*; cuanto mayor es el nivel de análisis, más posibilidades de utilización tendrá el corpus.

A fin de motivar al lector a explorar algunos corpus en español se ofrece un listado para hacer consultas por lemas, sobre categorías gramaticales de las palabras, sobre funciones o relaciones sintácticas y sobre relaciones y valores semánticos.

Para concluir con la primera sección del libro, Piñol presenta una tipología de corpus —de la que muestra ejemplos— elaborada con base en el canal de producción, el grado de representatividad y la especificidad de los textos. También hace una breve presentación de aplicaciones de la LC a la gramática general (lexicografía, estudios diacrónicos, etcétera) antes de adentrarse, en la segunda parte del libro, en la aplicación de esta disciplina a la enseñanza de E/2L.

En la segunda sección, la autora nos muestra los beneficios de incorporar la LC en la docencia de 2/L, para lo cual hace un análisis de las principales aplicaciones que se pueden llevar a cabo a partir de la consulta de corpus.

Piñol comienza con las aplicaciones de *ocurrencias* y *concordancias*. La búsqueda de ocurrencias es la forma más sencilla de utilizar un corpus, generalmente para resolver dudas lingüísticas, ya que muestra las construcciones que realmente ocurren en la lengua. De hecho, advierte sobre la importancia de que la búsqueda en Google, por ejemplo, que ya es muy habitual entre docentes y alumnos, requiere del conocimiento de los usuarios o de la guía de un maestro experimentado que pueda evaluar con qué corpus es conveniente trabajar y para qué propósitos.

Para la obtención de palabras clave en su contexto de uso, se utilizan las búsquedas de *concordancias* que nos permiten obtener en segundos contextos reales “típicos, recurrentes y observables repetidamente” (Alonso, 2007: 12). Piñol muestra un listado de posibilidades de uso de *concordancias* en la enseñanza de lenguas para diferentes actividades centradas en la gramática, el vocabulario o el uso.

Como bien menciona Humblé (2001), citado por la autora, los diccionarios y las gramáticas resultan muchas veces insuficientes para resolver los problemas que los alumnos encuentran a la hora de producir la lengua, y el uso de corpus resulta práctico en estas situaciones. Algunos ejemplos de estas son las búsquedas de preposiciones que acompañan ciertos verbos o las de falsos amigos que se aclaran por el contexto, entre otros.

En este sentido, resulta ilustrativa una actividad que la autora expone sobre su propia experiencia y que surge de la duda de sus alumnos sobre la diferencia entre *chance* o *chanza*. La búsqueda de las dos palabras en el CREA les permitió comprobar a los estudiantes que las dos son posibles, conocer cuáles son sus contextos de aparición y cuál es la distribución de ambas según ámbitos geográficos (esta última información relacionada con el interés de los alumnos por contrastar el español de México con el de España); al final se presentaron también sus definiciones en el diccionario de la Real Academia Española (DRAE). Toda esta información sirvió no solo para aclarar dudas, sino también para desarrollar la autonomía de los alumnos a partir de la experiencia de trabajar en procedimientos que les permitirán resolver sus dudas en el futuro.

La consulta de concordancias le permite también al alumno extraer de modo inductivo reglas gramaticales a través de la observación repetida de muestras de la lengua en uso. Esto último se ejemplifica con la consulta de *ser/estar + adjetivo*, en la que el adjetivo tiene un significado diferente según el verbo que se utilice, lo que posibilita al alumno deducir el valor semántico de cada combinación.

Por su parte, la búsqueda de concordancias en *corpus paralelos* (versiones de un mismo texto en varias lenguas) son especialmente útiles en estudios de traducción, pero también se utilizan, cada vez más, en la enseñanza de lengua, sobre todo por alumnos de nivel inicial para comparar la lengua meta con la L1.

En el capítulo 10, “Aplicaciones de las *listas de frecuencia*”, Piñol aborda la rama con más tradición de la LC, que estuvo dirigida a la elaboración de glo-

sarios y diccionarios. En cuanto a su aplicación en ELE, hay autores que abogan por su beneficio, aunque también hay quienes cuestionan sus limitaciones bajo el argumento de que el léxico no puede desvincularse de la gramática y del discurso, por lo tanto la frecuencia de uso no puede tener un valor por sí mismo para la enseñanza de determinadas formas, cuyo aprendizaje dependerá de otros factores, como bien menciona Alvar Ezquerro (2004), citado también por la autora.

En el capítulo 11, “Aplicaciones de los *corpus de aprendices*”, se incluyen muestras de textos producidas por alumnos que están adquiriendo la lengua y se ofrecen ejemplos de corpus tanto de L1 como de L2, orales y escritos. En cuanto a las aplicaciones de este tipo de corpus, se ejemplifican con las posibilidades de uso del corpus como el Corpus Escrito del Español L2 (CEDEL2), que recoge muestras de alumnos de todos los niveles y que puede “ser utilizado como fuente de datos por los investigadores de español 2L y como fuente de ejemplos para profesionales de ELE y aprendices del español” (Lozano, 2009: 207).

En el capítulo 12, “Aplicación de los corpus lematizados y de los corpus que incluyen información sintáctico-semántica”, la autora se refiere a corpus que ofrecen información más amplia que otros (relaciones sintácticas e información semántica). Estos corpus son sometidos a un nivel de análisis más detallado y, en consecuencia, brindan más posibilidades de realizar búsquedas precisas. La autora nos ofrece ejemplos de consulta de ambos tipos de corpus.

En los capítulos 13, “Sobre las aplicaciones de la LC a EFE (Enseñanza de 2/L para fines específicos)”, y 14, “Aplicaciones de la LC a la edición de materiales didácticos”, Piñol presenta aplicaciones de la LC a la enseñanza de español para fines específicos, área cuya demanda es cada vez más creciente y nos muestra sus posibilidades de aplicación en la edición de materiales didácticos. La aplicación de la LC en la edición de materiales es, en palabras de la autora “un ámbito profesional-productivo fundamental en la enseñanza de lenguas” (p. 151). En estos capítulos se ejemplifica, de manera particular, la aplicación de la LC en la confección de diccionarios, de libros de texto y de ejercicios interactivos en línea.

Cabe mencionar que la obra incluye un cuestionario de síntesis al final de cada una de sus secciones, mediante el cual el lector puede evaluar su comprensión de los temas tratados. También ofrece las respuestas respectivas, así como sugerencias de actividades para enriquecer la información y la experiencia de operar con corpus lingüísticos.

En mi experiencia como docente e investigadora he manejado corpus de aprendices, generalmente recabados en las clases de lengua, para estudios sobre adquisición de ciertos usos lingüísticos que requerían textos con información muy específica. Si bien he explorado corpus en búsqueda de información para mis investigaciones, mi experiencia en su uso en realidad ha sido limitada, lo que evaluó con base en las posibilidades de explotación de este recurso que descubro a partir de la lectura de este libro y de cierta exploración que llevé a cabo, guiada por las sugerencias que la autora hace a lo largo del texto.

El libro de Mar Cruz Piñol cumple con creces el propósito para el que fue concebido: proporcionar a los profesores de español 2/L el conocimiento y las herramientas básicas que les permitan usar corpus informáticos. El docente de español 2/L, como resultado de la lectura de este libro, tendrá un panorama exhaustivo de los corpus más importantes en lengua española, información sobre cómo acceder a ellos, cuáles son sus características, cómo y para qué propósitos pueden consultarse y cómo usar los datos obtenidos en aplicaciones que enriquecerán, sin duda, su ejercicio profesional.

Referencias

- ALONSO PÉREZ-ÁVILA, ELENA (2007). El corpus lingüístico en la didáctica del léxico del español como LE. *Boletín de Asele*, 37, 11–32.
- ALVAR EZQUERRA, MANUEL (2004). La frecuencia léxica y su utilidad en la enseñanza del español como lengua extranjera. En María Auxiliadora Castillo Carballo, Olga Cruz Moya, Juan Manuel García Platero & Juan Pablo Mora Gutiérrez (Coords.), *Las gramáticas y los diccionarios en la enseñanza del español como segunda lengua: deseo y realidad. Actas del XV Congreso Internacional de Asele* (pp. 19–39). Sevilla: Universidad de Sevilla. Recuperado de https://cvc.cervantes.es/ensenanza/biblioteca_ele/asele/pdf/15/15_0017.pdf
- HUMBLÉ, PHILIPPE (2001). O uso de corpora no ensino de línguas. Alguns exemplos do português e do espanhol. En Loni Grimm Cabral, Pedro de Souza, Ruth E. Vasconcelos Lopes & Emilio Gozze Pagotto (Orgs.), *Lingüística e ensino: novas e tecnologias* (pp. 157–180). Blumenau: Nova Letra.
- LOZANO, CRISTÓBAL (2009). CEDEL2: corpus escrito del español como L2. En Carmen M. Bretones Callejas, José Francisco Fernández Sánchez, José Ramón Ibáñez Ibáñez, María Elena García Sánchez, M. Enriqueta Cortés de los Ríos, Sagrario Salaberri Ramiro, María Soledad Cruz

- Martínez, Nobel Perdú Honeyman & Blasina Cantizano Márquez (Eds.), *Applied linguistics now: Understanding language and mind. La lingüística aplicada hoy: comprendiendo el lenguaje y la mente* (pp. 197–212). Almería: Universidad de Almería. Recuperado de <https://viancep2012.files.wordpress.com/2012/10/actasalmeria.pdf>
- ROJO, GUILLERMO (2001). La explotación de la Base de Datos Sintácticos del Español Actual (BDS). En Josse de Kock (Ed.), *Lingüística con corpus: catorce aplicaciones sobre el español* (pp. 255–286). Salamanca: Universidad de Salamanca. Recuperado de https://gramatica.usc.es/~grojo/Publicaciones/Explotacion_BDS.pdf
- ROJO, GUILLERMO (agosto, 2008). Lingüística de corpus y lingüística del español. Ponencia plenaria en el XV Congreso de Asociación de Lingüística y Filología de América Latina, Montevideo, Uruguay. Recuperado de https://gramatica.usc.es/~grojo/Publicaciones/Lgca_corpus_lgca_espanol.pdf
- SINCLAIR, JOHN M. (2004). Corpus and texts – basic principles. En Martin Wynne (Ed.), *Developing linguistic corpora: A guide to good practice* (pp. 1–16). Oxford: Oxbow Books. Recuperado de <http://ota.ox.ac.uk/documents/creating/dLC/chapter1.htm>