

Niktelol Palacios, *Corpus orales y lexicografía. A propósito de Presseea-Puebla*, Ciudad de México, El Colegio de México, Centro de Estudios Lingüísticos y Literarios, 2022, 180 pp. ISBN: 978-607-564-388-5.

Julieta Castellanos Armenta

Estancia Posdoctoral CONACYT-UNAM

yecasar@gmail.com

ORCID: 0000-0002-7444-4654

*Corpus orales y lexicografía. A propósito de Presseea-Puebla* es una guía precisa para los interesados en recolectar, preprocesar y procesar datos léxicos con la finalidad de organizarlos en un lecionario y, al mismo tiempo, es un acercamiento panorámico que nos reseña Niktelol Palacios al tratamiento integral del vocabulario del español de Puebla registrado en Presseea-Puebla. El contenido de este libro, además de una sección introductoria, está dividido en 5 capítulos que a continuación desglosaremos.

“Antecedentes: Corpus orales del español de Puebla” es el primer capítulo del texto; se trata de un repaso sobre los proyectos de recolección y documentación de la oralidad del español de Puebla, desde la década de 1980, vinculados con la ahora Benemérita Universidad Autónoma de Puebla (BUAP). Palacios destaca el proyecto Lenguaje, Cultura e Ideología, bajo la dirección de Adrián Gimete-Welsh y el Corpus del Habla de Puebla (CORHPU), cuyos responsables son Jorge Lázaro, Gerardo Sierra y Guadalupe Huerta, desarrollado en el Grupo de Ingeniería Lingüística de la Universidad Nacional Autónoma de México (UNAM), con datos recogidos por estudiantes de la BUAP. La autora reconoce muchas de las limitantes que debieron sortear estos proyectos pioneros para consolidarse y para dar continuidad a sus trabajos; aun así, los considera interesantes aproximaciones para la reconstrucción de la historia lingüística mexicana, de acuerdo a las propuestas teóricas y metodológicas que implementaron. Estos proyectos abrieron caminos y sembraron un interés recurrente por estudiar comunidades de habla de la región, que han llegado a reflejarse en la publicación de este libro, el cual pretende fortalecer los cimientos de la investigación lin-

güística oral y construir lazos comunicativos con proyectos de corte semejante.

El capítulo dos “El Corpus Sociolingüístico de la Ciudad de Puebla y Preseaa-Puebla” inicia con un breve panorama histórico y demográfico del municipio de Puebla que logra mostrar el mosaico cultural que se vive en la localidad y así, tender un puente de acceso al habla regional. La parte medular del capítulo es la descripción del Corpus Sociolingüístico de la Ciudad de Puebla (CSCP) desde sus orígenes, su planeación, su ejecución y particularmente, la descripción de dos de sus módulos de entrevistas: el corpus de léxico de oficios, flora y fauna (2008-2016) y el subcorpus Preseaa-Puebla (2012-2017), integrado por 108 entrevistas, sus correspondientes pruebas lingüísticas y su cuestionario sociológico, de acuerdo al método propuesto por Pedro Martín Butragueño y Yolanda Lastra al desarrollar el Corpus Sociolingüístico de la Ciudad de México (CEMC, 2011-2015). La descripción de la recolección de material está expuesta de tal manera que asegura al lector que el CSCP es una fuente de valiosa información para el estudio empírico del habla de la ciudad de Puebla, ya que proporciona datos para analizar la reconstrucción de la identidad social y lingüística de la comunidad y de los migrantes, y hace posible una acertada descripción de la variación, el contacto y el cambio lingüístico; de igual manera, la preestratificación de la muestra por variables sociales (edad, sexo e instrucción educativa formal) resulta un recurso de gran utilidad para los estudios sociolingüísticos. El capítulo cierra subrayando dos hipótesis importantes, que guiarán las interrogantes de las descripciones a nivel léxico, acerca de la información obtenida de las entrevistas semidirigidas de Preseaa-Puebla: 1) se encontrarían una amplia variedad de registros estilísticos de habla del vocabulario general del español de México y 2) se documentarían el uso de unidades léxicas y fenómenos del habla propios de la oralidad poco comunes en otro tipo de corpus. Ambas hipótesis, más adelante, son exitosamente corroboradas. Este capítulo es una invitación abierta a los interesados en el habla de la comunidad poblana en particular y, de México, en general, a explorar los datos recolectados e idear análisis lingüísticos propositivos.

Los capítulos finales del libro están destinados a aspectos léxicos del español de Puebla. En el tercer capítulo, “Lematización y análisis cuantitativo”, se presenta una descripción del léxico obtenido de las 108 entrevistas semidirigidas de Preseepuebla, con la utilización de herramientas y procesos de análisis computacional; hay dos intenciones explícitas en este apartado: por un lado, se pretende entablar un diálogo entre la aplicación de la ingeniería lingüística y los trabajos lexicológicos, y, por otro, hacer una presentación minuciosa del método de análisis de los textos que se ha aplicado, desde una fase previa al procesamiento de las transcripciones hasta la documentación de vocablos en el leuario del Diccionario del Español de México (DEM) (El Colegio de México, versión en línea).

Palacios da a conocer claramente cada una de las etapas del método de lematización que su equipo de trabajo ha propuesto, con la intención de que sea replicable en cualquier otro corpus oral o regional una vez que se cuenta con textos planos de análisis; la autora explica la importancia de un preprocesamiento de los textos para la extracción automatizada de etiquetas para no alterar los resultados, hace hincapié en la necesidad de encontrar un equilibrio entre la automatización del procesamiento de los datos léxicos a partir del conjunto de transcripciones, mediante herramientas de análisis lingüístico computacional (algunas disponibles de manera gratuita en la web) y sobre todo, insiste en la utilidad del proceso manual del análisis lingüístico para la corrección de distintos tipos de errores en la fase del proceso automatizado. Además, Palacios nos comparte los problemas a los que se enfrentaron durante cada fase del proceso, lo cual podría servir como advertencia a otros equipos de trabajo con objetivos léxicos pero especialmente, para implementar los cuidados previos y evitar caer, en la medida de lo posible, en las mismas dificultades. Algunos de los errores que se reportan en este capítulo se deben a la automatización del proceso por el programa computacional utilizado, por lo que resulta indispensable el análisis manual para hacer las correcciones correspondientes, principalmente en equivocaciones en el etiquetado, erratas en la escritura e inexactitudes en la lematización del propio programa; esta explicación a detalle del funcionamiento de los análisis

automatizados y la forma en que deben suministrarse los datos a estos lematizadores para obtener resultados más eficaces es una orientación de gran interés para el desarrollo teórico y metodológico, por ejemplo, sobre la delimitación y el reconocimiento de la palabra en las lenguas naturales.

Gracias a la lematización y el análisis cuantitativo llevado a cabo por el equipo, se nos presentan los 100 vocablos con mayor frecuencia absoluta de aparición en Preseea-Puebla, se atestigua que el vocabulario recogido es común al español mexicano, ya que alrededor del 90% forma parte del leuario del diccionario integral del español de México (DEM). Se reconoce la poca certeza de que el 10% restante sea parte del español regional o quizá, solo no se encuentra registrado en el DEM por tratarse de innovaciones léxicas, por lo que se plantea la necesidad de futuros estudios para determinar el nivel de lengua que deba asignarse a cada palabra del leuario Preseea-Puebla.

“Rasgos de oralidad en Preseea-Puebla: Reflexiones lexicográficas” es lo concerniente al capítulo cuatro, en el que se desarrolla un acercamiento cualitativo al leuario de Preseea-Puebla, con la firme intención de subrayar la vastedad de los corpus orales como fuente de información para estudios lingüísticos. La autora nos comparte los objetivos planteados para esta parte del análisis y sus logros:

- a) Continuar la descripción de la composición léxica del corpus, lo cual se consiguió mediante una clasificación intuitiva de las palabras que son ejemplo de la creatividad y riqueza de la oralidad, entre los que podemos encontrar *boni*, por *bonito*; *panzar*, como reducción de la locución ‘pasar de panzazo’ y otros juegos de palabras del tipo *buzo caperuzo*, para formar una rima consonante sólo con fines lúdicos.
- b) Proponer posibles tratamientos lexicográficos de voces no definidas en el DEM pero que cuentan con una morfología emparentada con su leuario. Esto debe reconocerse como una aportación relevante de información sobre la tradición lexicográfica y, en especial, acerca de la productividad de ciertos morfemas del español; en este apartado, se nos muestran casos como *antibala*, documentado en Preseea-Puebla, que ejemplifica un proceso directo de derivación entre la entrada

del DEM (*bala*) y el prefijo *anti-* ‘oposición o contrariedad’, y que su uso común lo hace un vocablo candidato a incluirlo, en breve, en el diccionario de la lengua de México. De igual manera, también se hacen propuestas de tratamientos lexicográficos para aquellas palabras ya inmersas en el vocabulario del español mexicano pero que tampoco están registradas por el DEM, ya que no han aparecido en su corpus y que es cuestión de tiempo para que se integren a su lecionario, como es el caso del neologismo *bullying*. Este capítulo finaliza con un listado de vocablos, aunque corto, rico en fenómenos léxicos de estudio, que el equipo de trabajo pone sobre la mesa como una provocación académica que anima a un necesario estudio lexicológico y una documentación a detalle.

El quinto y último capítulo del libro “Notas léxicas del español de Puebla” consta de la descripción y documentación de 11 voces propias de la cultura local poblana obtenidas, en su mayoría, de la última parte de las entrevistas, en la que la conversación se centra en la ciudad y sus costumbres. Se plantea un método para justificar cuándo una palabra es netamente característica del español de Puebla.

Para analizar estas palabras, las han dividido en: a) nombres de prácticas o conocimiento de la cultura local (*huehue, murales, mosaicos, pitufos*), b) vocablos de la gastronomía tradicional de Puebla (*cemita, chalupa, chileatole, molote y tlalitos*) y c) voces que presentan variación sintáctica o semántica en relación con la variedad estándar del español de México (*acusarse y arrimarse*). El propósito principal de esta descripción no dista del objetivo fundamental del libro: explicar la metodología empleada y explicitar la toma de decisiones para que otros equipos de trabajo puedan replicarlo; se hace especial énfasis en una documentación y tratamiento enciclopédico de este tipo de vocabulario puesto que el lexicógrafo debe brindar los datos históricos y socioculturales justos para garantizar la comprensión plena de la realidad que se desconoce más allá de los usuarios locales; la autora exhorta a que las definiciones de esta índole sean descripciones precisas de los objetos para que el hablante local vea representadas sus prácticas cotidianas, ya que se trata de una definición identitaria.

La publicación de *Corpus orales y lexicografía. A propósito de Preseaa-Puebla* y del corpus Preseaa-Puebla cierra una etapa inicial de arduo trabajo de investigación que comenzó en el año 2011. Este primer acercamiento descriptivo del léxico Preseaa-Puebla es una invitación reiterada a la elaboración de corpus lingüísticos orales que puedan seguir metodologías similares a las que se exponen en este libro para asegurar un registro sistemático de las variedades de habla locales, pero es, sobre todo, una contribución innegable al conocimiento del español del Altiplano central mexicano.