

## **Creación de un corpus previo al registro en la web de anotación: el caso del antillanismo ‘canoa’**

Anabel Mederos-Cedrés & Miguel Ángel Rodríguez-Falcón

*University of Las Palmas de Gran Canaria*

**Palabras clave:** *corpus, web de anotación, americanismos, antillanismos, lexicografía.*

El objetivo de esta investigación es presentar la creación de un corpus previo en la metodología seguida en la creación de una base de datos en abierto en el Proyecto PID2019-104199GB-I00 sobre la vitalidad de los americanismos léxicos en las lenguas española e inglesa documentados en textos de los siglos XVI y XVII-AMERLEX-DATABASE (Proyectos de I+D+i – PGC). Dicho corpus previo se realizó antes del registro de los datos en la web de anotación con los datos extraídos de otros corpus: Léxico Hispanoamericano (1493-1993) (Boyd Bowman 2003); Diccionario de Americanismos (DA 2010); Nuevo diccionario histórico de la lengua española (DHLE 2013); Diccionario de la Lengua Española (DRAE/DLE, distintas ediciones); Corpus Electrónico del Español Colonial Mexicano (COREECOM 2013); y Corpus Diacrónico y Diatópico del Español de América (CORDIAM 2015) de la Academia Mexicana de la Lengua. AMERLEX debía contener ejemplos textuales de los objetivos propuestos. La anotación de corpus o la creación de marcas con información era necesaria para demostrar o probar una determinada teoría lingüística con respecto a los americanismos, y tanto era así que el registro del corpus era fiel reflejo de la teoría que se estaba describiendo o probando (Ide y Pustejovsky 2017)

El propósito de este nuevo corpus es averiguar qué antillanismos, a través del estudio de caso de *canoa*, han pervivido en la evolución histórica de las hablas americanas, y cuál ha sido el grado de difusión panhispánica o regional según las fuentes académicas en la sincronía actual (Quirós García y Ramírez Luengo 2015; Cáceres Lorenzo 2022). La vitalidad de una lexía indígena va unida a otras nociones como adaptación, adopción ortográfica, morfológica y semántica (Alvar 1975) y a la creación de nuevos significados en su proceso diacrónico de transmisión (Sala et álíi 1982). Las conclusiones sobre la incorporación de las voces antillanas al español provienen de investigaciones parciales y dispersas (determinadas crónicas, textos de localidades muy concretas, etc.), pero, al mismo tiempo, en la última década existe una proliferación de nuevos corpus que persiguen obtener resultados acordes a la naturaleza de los escritos indios (Rivarola 2012; Bertolotti y Company 2014) a la incorporación de las variables diacrónica, diastrática, diafásica y diatópica.

La efectividad de un corpus como herramienta de investigación lingüística reside en gran medida en el tipo de búsqueda que permite hacer y en los datos que proporciona (Davies 2009). En nuestro caso, el citado corpus previo tuvo las siguientes funciones: 1) formar parte del marco teórico de la tarea que se quiere resolver y del objetivo del proyecto; 2) registrar datos de la bibliografía; 3) ser útil para crear un modelo de anotación; y 4) ser una información teórica necesaria en la interacción entre ingenieros y lingüistas.

Los resultados obtenidos sobre este estudio de caso, en que *canoa* presenta más de once significados con sus respectivas variantes, supone una evidencia necesaria para la creación de un corpus en el que se informe sobre el reconocimiento de la integración y vitalidad de las voces antillanas en la lengua española.

### **Referencias**

- Alvar, M. 1975. *España y América cara a cara*. León: Bella Época.
- Bertolotti, V. y Company, C. 2014. El corpus diacrónico y diatópico del español de América (CORDIAM). Propuesta de tipología textual. *Cuaderno de la ALFAL*, (6), pp. 130-148.

- Boyd Bowman, P. 2003. Léxico hispanoamericano 1493-1993. En Ray Harris-Northall and John J. Nitti (eds). Nueva York: *Hispanic Seminary of Medieval Studies*. Consultado en [https://textred.spanport.lss.wisc.edu/lexico\\_hispanoamericano/](https://textred.spanport.lss.wisc.edu/lexico_hispanoamericano/).
- Cáceres Lorenzo, M. T. 2022. Hispanic-American dialectology in the 16th century. Penetration of Americanisms in Nicolas Monardes' *Historia Medieval. Dialectologia et Geolinguística*, 30, (1), pp. 115-130.
- [CORDIAM] Academia Mexicana de la Lengua. 2015. Corpus Diacrónico y Diatópico del Español de América. Consultado en <https://www.cordiam.org>.
- [COREECOM] Grupo de Estudio del Español Colonial Mexicano. 2013. Corpus Electrónico del Español Colonial Mexicano. Arias, B. (Coord.). IIFL-UNAM. Consultado en <https://www.iifilologicas.unam.mx/coreecom/index.php?page=inicio&men=1>.
- [DA] Asociación de Academias de la Lengua Española. 2010. *Diccionario de americanismos*. Consultado en <https://www.asale.org/damer/>.
- Davies, M. 2009. Relational databases as a robust architecture for the analysis of word frequency. En Archer, D. (ed.), *What's in a Wordlist? Investigating Word Frequency and Keyword Extraction*. Londres: Ashgate, pp. 53-68.
- [DHLE] Real Academia Española. 2013. *Diccionario histórico de la lengua española*. Consultado en <https://www.rae.es/dhle/>.
- [DLE] Real Academia Española 2014. *Diccionario de la lengua española* (23.<sup>a</sup> ed.). Consultado en <https://dle.rae.es/>.
- Ide, N. y Pustejovsky, J. 2017. *Handbook of Linguistic Annotation*. Springer Netherlands.
- Quirós García, M. y Ramírez Luengo, J. L. 2015. Observaciones sobre el léxico del español de Yucatán (1650-1800). *Revista de filología española*, 95, (1), pp. 183-210.
- Rivarola Rubio, J. L. 2012. Los corpora en el estudio histórico del español de América (un corpus documental del español en el Perú de los siglos XVI y XVII): *Reflexiones y perspectivas. Actas del VIII Congreso Internacional de Historia de la Lengua Española* (pp. 391-396). España: Santiago de Compostela.
- Sala, M., Munteanu, D., Neagu, V. y Sandru-Olteanu, T. 1982. *El español de América*. Bogotá: Instituto Caro y Cuervo.
- Sinclair, J. 1996. *Preliminary Recommendations on Corpus Typology*. EAGLES. Consultado en <http://www.ilc-cnr.it/EAGLES96/corpusstyp/corpusstyp.html>.
-