

El corpus textual DITECA: Análisis y resultados

Pilar Carrasco / Diana Esteba Ramos

Universidad de Málaga

pcarrasco@uma.es

desteba@uma.es

El *Diccionario de Textos Concejiles de Andalucía (DITECA)* se ha elaborado a partir de un corpus documental de ordenanzas municipales andaluzas redactadas entre los siglos XIII y XVIII. Los textos, aunque pertenecientes a la misma tradición discursiva, presentan variaciones tanto de contenido como de forma: junto al estilo más formal y al tema más abstracto de la regulación de la vida ciudadana pueden coexistir ciertos estilos de lengua más cercanos a la propia variedad lingüística vernacular. Numerosas voces de la lengua vulgar aparecen en estos textos para dar expresión a los contenidos más populares o más relacionados con las costumbres u oficios tradicionales. Nombres autóctonos de plantas, peces, aperos de labranza, oficios artesanales aparecen por doquier junto a los términos más especializados del léxico administrativo y jurídico, amén del léxico general.

El diccionario se compone de 4497 entradas ordenadas alfabéticamente y de 17.151 variantes analizadas semánticamente. Se trata, por tanto, de un amplio corpus en el que ofrecemos no solo los textos (135.896 citas textuales) sino también el significado y etimología de los términos.

Nos proponemos en esta comunicación analizar la macroestructura del diccionario formada por nombres propios, topónimos, sustantivos, verbos, adverbios, locuciones, lexemas complejos, etc. con la aplicación de los métodos de la estadística léxica con el objeto de generar índices de las palabras y de las estructuras sintácticas más frecuentes, o de colocaciones y otros indicadores léxicos. La variabilidad de frecuencias debe servir para la reconstrucción de los campos léxicos o semánticos que servirán de base para la interpretación histórica de la lengua.

Pretendemos poner de manifiesto el interés que para la historia del léxico español tiene la confección de diccionarios a partir de corpus especializados, de pequeñas dimensiones y formados con textos pertenecientes a una misma tradición discursiva y a un área lingüística determinada, lo que conlleva poder documentar históricamente términos no recogidos en fuentes lexicográficas debido en parte a su carácter dialectal o poco formal. Por ejemplo, algunas de las voces se han definido a

partir de los contextos específicos en que se usan y con el apoyo de los estudios históricos medievales y modernos que han abordado la organización social de los reinos andaluces.

Bibliografía

- Carrasco, Pilar (2002): «El Corpus de Textos Jurídicos Andaluces. El léxico de las Instituciones municipales», en M^a Teresa Echenique y Juan Sánchez (eds.), *Actas del V CIHLE*, Madrid: Gredos, II, 1937-1946.
- _____ (2003): «El léxico de los oficios municipales en las ordenanzas andaluzas», en José Luis Girón et al. (eds.), *Homenaje al Prof. J.J. de Bustos Tovar*, Madrid: Universidad Complutense, I, 497- 509
- Corominas, Joan/ José Antonio Pascual (1991-1997): *Diccionario crítico etimológico castellano e hispánico*. Madrid: Gredos.
- Nieto Jiménez, Lidio / Manuel Alvar Ezquerro (2007): *Nuevo tesoro lexicográfico del español (s. XIV-1726)*, Madrid: Arco/Libros.
- Real Academia Española (2001). *Nuevo tesoro lexicográfico de la lengua española*. Madrid: Espasa Calpe.
- Rojo, Guillermo (2003), “El empleo de corpus textuales en la investigación lingüística”, En *Actas del VII Simposio de actualización científica y didáctica de Lengua española y Literatura*, Málaga, 107-119.