

Un ejemplo de la aplicación de las TICs a la investigación humanística: herramientas digitales de análisis textual

Autor: Marina Castilla Ortega,

Licenciada en Historia del Arte.

Profesora Sustituta Interina del Departamento de Historia del Arte de la Universidad de Málaga (España)

(E-mail: mcastilla@uma.es)

INTRODUCCIÓN

Las Humanidades Digitales comprenden la aplicación de las TICs a la investigación en el ámbito humanístico. Favorecen tanto la creación de bases de datos y el acceso inmediato a la información como la aparición de novedades metodológicas y de resultados inéditos hasta la fecha. Datan su aparición en los años cuarenta del siglo XX, con el proyecto del jesuita Roberto Busa, auspiciado por IBM. Tras este trabajo inicial, fueron acogidas con gran fuerza en el ámbito anglosajón, pero no con tanto convencimiento en otras áreas geográficas ni en todas las ramas de las Humanidades por igual. Sin embargo, tras esta fase inicial, actualmente puede afirmarse que las Humanidades Digitales están siendo aceptadas cada vez con más entusiasmo. Entre sus líneas de investigación más recurrentes, se prioriza el análisis de textos, aplicado no sólo al ámbito de la Lingüística, sino también al de disciplinas adscritas a la Historia del Arte, como la Estética o la Teoría del Arte. Por lo tanto, este trabajo expone varias herramientas digitales de libre acceso de análisis léxico-terminológico enfocadas a la investigación histórico-artística.

DESARROLLO

Entre los citados recursos digitales para la investigación léxico-textual, en primer lugar se destaca la plataforma TAPORWARE, *Prototype of Text Analysis Tools* (<http://taporware.ualberta.ca/>), desarrollada gratuitamente por la Universidad de Alberta. Para el procesamiento de una obra completa o de un fragmento de ésta a través de la presente herramienta, será necesaria su digitalización, bien mediante tecleo íntegro, convirtiéndose de formato .doc a texto plano (.txt) o a formato .html, o bien mediante escaneo directo a través de la opción específica OCR. Tras este primer paso, la plataforma provee al usuario de resultados terminológicos cualitativos y cuantitativos, con búsquedas de frecuencias de palabras (el número de veces que se repiten las palabras a lo largo del escrito, pudiéndose obtener las más reiteradas y, por lo tanto, extraer una primera aproximación a qué tipo de idea proyecta prioritariamente su autor, por ejemplo, sobre una obra de arte), concordancias (búsqueda de una palabra en sus distintos contextos de uso o párrafos a lo largo del escrito, para detectar con qué otros vocablos suele relacionarse), coocurrencias (búsqueda de las situaciones en que dos palabras aparecen juntas en un mismo contexto de uso o párrafo, para dilucidar su relación y significado) y comparaciones entre dos textos (ofreciéndose los términos comunes y únicos en cada caso, detectando el grado de riqueza léxica de cada escrito, similitudes y diferencias). A este recurso se suman otros, con los que podría complementarse, como son los corpus on-line de la Real Academia Española (RAE) del Corpus Diacrónico Español (CORDE) (<http://corpus.rae.es/cordenet.html>) y del Corpus de Referencia del Español Actual (CREA) (<http://corpus.rae.es/creanet.html>). La colección textual compendiada en estas bases de datos aporta al usuario un conocimiento exhaustivo de la evolución cronológica de un vocablo, desde los más variados géneros literarios. De este modo, se puede obtener el grado de innovación semántica que un escritor realiza sobre un concepto con respecto al idiolecto colectivo de su contemporaneidad o de otros periodos. A estos recursos de la RAE se añade otro más específico, como su *Diccionario de Autoridades* (<http://www.rae.es/recursos/diccionarios/diccionarios-antiguos-1726-1996/diccionario-de-autoridades>), destinado al vocabulario del siglo XVIII.

CONCLUSIONES

En conclusión, las TICs aplicadas al análisis de textos favorecen la actualización del panorama de la investigación de las Humanidades, abriéndose a futuras posibilidades científicas, más colaborativas e interdisciplinares. Como ejemplo de trabajos precedentes que aplican herramientas digitales de análisis léxico-textual a la investigación de terminología artística, el lector puede remitirse al estudio léxico-estadístico de la Dra. Nuria Rodríguez Ortega publicado en *TREA* (Rodríguez, 2009) sobre dos descripciones del Real Monasterio de El Escorial por parte de los padres Sigüenza (1605) y Santos (1657). En él se emplean una serie de indicadores básicos de léxicoestadística, como *tokens* (frecuencias de una palabra en un texto); *types* (número de palabras distintas empleadas en un escrito); *ratio tokens/type* (intervalo de aparición de una palabra nueva en el texto, lo cual indica su grado de riqueza léxica, por lo que, cuanto más bajo, su corpus será más amplio); *hápax legómena* (palabras con una única aparición en el texto) y *hápax dislegómena* (palabras que aparecen dos veces en un escrito). En segundo lugar, en la publicación previa de la presente autora (Castilla, 2017) se expone un estudio léxico-terminológico, mediante la aplicación de la citada herramienta TAPORWARE, del concepto de `curioso`/`artificio` en dos textos del siglo XVI, mismamente descriptores de la imagen de El Escorial.

REFERENCIAS

Castilla Ortega, M. (2017). Metodologías de la sociedad digital aplicadas al estudio léxicoterminológico en las descripciones escorialenses de la laurentina y la octava maravilla del mundo. *3C TIC: Cuadernos de desarrollo aplicados a las TIC*, pp.11-24. 6 (1), DOI: <<http://dx.doi.org/10.17993/3ctic.2017.55.11-24/>>.

Galina Russel, I. (2011). ¿Qué son las Humanidades Digitales?. *Revista UNAM. Mx (Revista digital universitaria)*. 12 (7). Recuperado de: <http://www.revista.unam.mx/vol.12/num7/art68/>

Berry, David M. (2011). The Computational Turn: Thinking about the Digital Humanities. *Culture Machine*. (12). Londres, Inglaterra: Open Humanities Press.

Rodríguez Ortega, N. (2009). Similitudes y diferencias léxico-textuales en dos descripciones seiscentistas del Real Monasterio de San Lorenzo de El Escorial (I)...*Teoría y Literatura artística en la sociedad digital. Construcción y aplicabilidad de colecciones textuales informatizadas. La experiencia del Proyecto Atenea* (pp. 363-398 y pp. 399-430). Asturias, España: Ediciones TREA.