

/03/

UN EJEMPLO DE LA APLICACIÓN DE LAS TICS A LA INVESTIGACIÓN DE LA HISTORIA DEL ARTE: HERRAMIENTAS DIGITALES DE ANÁLISIS TEXTUAL AN EXAMPLE OF THE APPLICATION OF THE ICTS TO THE HISTORY OF ART RESEARCH: DIGITAL TOOLS OF TEXTUAL ANALYSIS

Marina Castilla Ortega

Licenciada en Historia del Arte. Profesora Sustituta Interina,
Departamento de Historia del Arte de la Universidad de Málaga (España).
E-mail: mcastilla@uma.es ORCID: <https://orcid.org/0000-0002-8058-6202>

Recepción: 17/12/2018. **Aceptación:** 16/01/2019. **Publicación:** 29/03/2019

Citación sugerida:

Castilla Ortega, M. (2019). Un ejemplo de la aplicación de las TIC's a la investigación de la Historia del Arte: Herramientas digitales de análisis textual. *3C TIC. Cuadernos de desarrollo aplicados a las TIC*, 8(1), pp. 58-73. doi:<http://dx.doi.org/10.17993/3ctic.2019.81.58-73>

RESUMEN

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, bajo el auspicio de 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. Debido a que entre sus líneas de investigación destaca el análisis textual, se exponen algunos recursos digitales para el estudio léxico-terminológico, aplicados a un texto de interés para la Historia del Arte. Una de estas herramientas es la plataforma TAPORWARE (<http://taporware.ualberta.ca/>), la cual propicia resultados terminológicos cualitativos y cuantitativos (frecuencias de palabras, concordancias y coocurrencias). Por otra parte, se estudian algunas herramientas de la RAE, como el CORDE (<http://corpus.rae.es/cordenet.html>) y el CREA (<http://corpus.rae.es/creanet.html>). Desde variados géneros literarios, el corpus textual de estas bases de datos facilita el análisis de la evolución cronológica de los términos y el esclarecimiento de su significado.

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.

ABSTRACT

The Digital Humanities comprise the application of ICTs to research in the humanistic field. They favor both the creation of databases and immediate access to information and the appearance of new methodologies and unpublished results to date. They appear in the 1940s with the project of the Jesuit Roberto Busa, under the auspices of IBM. After this initial work, they were received with great force in the Anglo-Saxon area, but not with so much conviction in

other geographical areas or in all branches of the Humanities equally. However, after this initial phase, it can now be affirmed that the Digital Humanities are being accepted with increasing enthusiasm. Due to the fact that among its lines of research the textual analysis stands out, some digital resources are exposed for the lexical-terminological study, applied to a text of interest for the History of Art. One of these tools is the TAPORWARE platform (<http://taporware.ualberta.ca/>), which provides qualitative and quantitative terminological results (word frequencies, concordance and co-occurrences). On the other hand, some tools of the RAE are studied, such as CORDE (<http://corpus.rae.es/cordenet.html>) and CREA (<http://corpus.rae.es/creanet.html>). From varied literary genres, the textual corpus of these databases facilitates the analysis of the chronological evolution of the terms and the elucidation of their meaning. In conclusion, the TICs applied to the analysis of texts favor the updating of the panorama of the Humanities research, opening up to future scientific possibilities, more collaborative and interdisciplinary.

PALABRAS CLAVE

Humanidades Digitales, TICs, Análisis léxico-terminológico, TAPORWARE, CORDE y CREA (RAE).

KEYWORDS

Digital Humanities, ICTs, Lexical-terminological analysis, TAPORWARE, CORDE and CREA (RAE).

1. INTRODUCCIÓN

Según las palabras de Galina Russel, las Humanidades Digitales engloban un “nuevo campo interdisciplinario que busca entender el impacto y la relación de las tecnologías de cómputo en el quehacer de los investigadores en las Humanidades”. En relación a ello, esta disciplina recibe distintas denominaciones, como “Recursos Digitales para las Humanidades, Cómputo para las Humanidades, Cómputo en las Humanidades, Informática Digital y Cultural e Informática para las Humanidades” (Galina, 2011).

Por lo tanto, la utilidad de las Humanidades Digitales, como su nombre indica, radica en la aplicación de los beneficios de las nuevas tecnologías de la información y la comunicación (TICs) a la comunidad científica del ámbito humanístico. Berry (2011) acuña el término “Giro Computacional” dentro del ámbito de las Humanidades, para referirse al momento en que los recursos digitales son absorbidos en todos los ámbitos de las disciplinas.

Los orígenes más antiguos de las Humanidades Digitales se remontan a la década de los años cuarenta del siglo XX en Italia, a partir de los trabajos a ordenador por parte del considerado “fundador” de la disciplina, el jesuita Roberto Busa. Su intención era la creación de un índice de concordancias (lista de todas las palabras de un texto, sus frecuencias o veces que aparecen en el escrito y sus relaciones en su contexto de uso o párrafos)¹ y la “lematización” de las obras completas de Santo Tomás de Aquino y autores relacionados. La “lematización” en el etiquetado de textos (*POS tagging*), así como en la disciplina de la Lingüística, consiste en otorgar una categoría gramatical a cada palabra de un escrito, determinada por una forma flexiva concreta, tal y como que se adjudicarían en un diccionario (ej. *-ar* para los infinitivos, *-ado* o masculino singular para los sustantivos, etc) (Bassi y Polanco, 2001). Como la amplitud del corpus de dichas obras resultaba desmesurado, el Padre Busa se valió del uso de una máquina llamada “computadora”, para lo cual recibió el apoyo de *International Business Machines Corporation (IBM)* de Estados Unidos. El corpus

¹ Previamente a la aparición de los ordenadores, se realizaban pocas concordancias de obras completas debido al gran esfuerzo y tiempo que dicho trabajo requería.

textual se trasladó a tarjetas perforadas para, acto seguido, desarrollar un programa para realizar concordancias. “Se requería una vagoneta para transportar todas las tarjetas perforadas” (Galina, 2011, citando a Hockey, 2004). Finalmente, en 1974 se publicaron los primeros tomos del Índice *Thomasticus*, con un corpus de más de 11.000 palabras en latín medieval.

Tras este trabajo, las Humanidades Digitales tuvieron una gran acogida en el ámbito anglosajón (Pino y Fiorimonte, 2018). No obstante, fuera de esta área geográfica, su aceptación no fue tarea fácil, sino una ardua carrera de obstáculos. De este modo, uno de los primeros inconvenientes fue la falta de medios materiales o de infraestructuras suficientes para su desarrollo (desde aulas de informática hasta *hardwares* y *softwares*) o la carencia de una “formación específica de Informática para humanistas” (López, 2015). Junto a estas trabas se añadieron otras de carácter inmaterial o “psicológico”, como la fobia hacia lo tecnológico o la conciencia arraigada de una incompatibilidad absoluta entre las Humanidades y las nuevas tecnologías. Ésta idea se vio fomentada porque la investigación humanística tradicionalmente se cimenta en un cariz crítico y especulativo. De ahí que, en un principio, el fenómeno de la incursión de las TICs en el territorio de las Humanidades recibiese nombres relacionados con la Informática, como “Informática Humanística”, modificándose posteriormente a una denominación más adecuada, como es la de “Humanidades Digitales”. Esta resulta más conveniente para ilustrar su verdadero cometido (“...queremos pensar que este nombre refleja que ahora lo sustantivo es la disciplina humanística y lo adjetivo es la tecnología”) (López, 2015). A estos prejuicios, se suma una tradición científica individual y hermética, de mérito propio, ajena a la interdisciplinariedad y al trabajo colaborativo inherente a la web 2.0.

Uno de los primeros inconvenientes fue la falta de medios materiales o de infraestructuras suficientes para su desarrollo (desde aulas de informática hasta *hardwares* y *softwares*) o la carencia de una “formación específica de Informática para humanistas” (López, 2015).

Pero, afortunadamente, puede afirmarse que, en la actualidad, en líneas generales, las Humanidades Digitales están siendo aceptadas con cada vez más entusiasmo, proponiendo como buen ejemplo de ello la asociación de las Humanidades Digitales Hispánicas (<http://humanidadesdigitales.org/>). Enfocada

a la investigación de la Literatura Hispánica, supone un máximo ejemplo de trabajo colaborativo e interdisciplinar, donde las ramas humanísticas comparten su investigación con la Informática, la Biblioteconomía y la Documentación. Con tan sólo tres años de existencia, se componía de un total de setenta socios provenientes, primordialmente, de la Literatura, pero también de la Lengua, la Historia, la Historia del Arte, la Teoría de la Literatura, los Estudios Culturales, la Biblioteconomía y Documentación, la Lingüística, la Comunicación, la Filología Clásica, la Filosofía y la Informática (López, 2015).

Hoy en día, entre las líneas de investigación más recurrentes de las Humanidades Digitales, se prioriza el análisis de textos. Aunque, aparentemente, el análisis textual se aplica en el ámbito de la Lingüística, también es una línea de estudio asociada a disciplinas adscritas a la Historia del Arte, como la Estética o la Teoría del Arte, de talante teórico y filosófico. De este modo, este trabajo expone varias herramientas digitales de libre acceso de análisis de textos (léxico-terminológico) como ejemplo de la incursión de las Humanidades Digitales aplicadas a la Historia del Arte y de la consiguiente modernización de esta área de las Humanidades. Para ilustrar la aplicación de estas herramientas, se expondrán los resultados del estudio del texto de la obra del Padre fray José de Sigüenza, *Historia de la Orden de San Jerónimo* (1605), concretamente de los pasajes de la tercera parte sobre la descripción del Real Monasterio de El Escorial.

2. DESARROLLO

2.1. LA HERRAMIENTA DE ANÁLISIS LÉXICO-TERMINOLÓGICO TAPORWARE: UN EJEMPLO DE LA APLICACIÓN DE LAS TICS A LA INVESTIGACIÓN HUMANÍSTICA

En primer lugar, se expone el estudio de las posibilidades de la herramienta digital TAPORWARE, *Prototype of Text Analysis Tools* (<http://taporware.ualberta.ca/>), desarrollada gratuitamente por la Universidad de Alberta.

Como punto de partida, para el procesamiento de las obras o textos en dicho sistema informático, será necesaria su completa digitalización. Ésta podrá realizarse 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 de reconocimiento óptico de caracteres (OCR; del inglés *Optical Character Recognition*)². Una vez llevada a cabo la digitalización textual, esta herramienta nos ofrece la posibilidad de desarrollar estudios léxicos cuantitativos y cualitativos, obteniéndose resultados más refinados y variados sobre el uso del vocabulario de su autor.

De este modo, se posibilitan diferentes tipos de búsquedas imposibles de realizar mediante el simple rastreo del ojo humano. Estas búsquedas, que se expondrán a continuación, permiten concluir qué palabras componen el idiolecto particular del escritor (búsqueda de frecuencias de palabras) junto con sus contextos de uso (búsquedas de concordancias y co-ocurrencias). Por lo tanto, con TAPORWARE se posibilitan dos tipos de análisis: una primera fase de análisis más general y cuantitativo, seguido de una segunda fase de análisis más minucioso y cualitativo. Para ilustrar ambas fases de este procedimiento, se presentarán los resultados obtenidos del análisis del pasaje concerniente a la descripción del Real Monasterio de El Escorial de la *Historia de la Orden de San Jerónimo* del Padre Sigüenza (1605) (pp. 525-899), el cual se compone de un total de 229.823 palabras. El texto digitalizado está disponible en la Plataforma digital ATENEA-TTC (proyecto I+D HAR2009-07068) (www.proyectoatenea.es), destinada a la consulta on-line de textos españoles de los siglos XVI-XVIII y que provee al usuario de herramientas informáticas para su análisis lingüístico-terminológico. Su transcripción, para su digitalización y procesamiento, fue realizada por José Luis Vega Loeches (Universidad Complutense de Madrid) y Nuria Rodríguez Ortega (Universidad de Málaga).

² Sin embargo, en muchos casos se impondrá la necesidad de un tecleo manual íntegro, como en el caso de las obras de una tipografía anterior al siglo XX, ya que los programas OCR de determinados escáneres no están habilitados para registrarla en su totalidad.

Primera fase

El análisis de frecuencias de palabras pone a nuestra disposición una selección de los términos más reiterados en el escrito, ordenados de mayor a menor afluencia y seguidos del número de sus repeticiones entre paréntesis (p.ej. tabla 1). En esta misma fase inicial, ya podríamos obtener conclusiones iniciales, puesto que los vocablos de mayor número son indicadores previos (aunque no del todo concluyentes) para una primera aproximación a qué tipo de idea se proyecta prioritariamente en un texto, por ejemplo, vinculado a un monumento u obra de arte. De este modo, si en un escrito predominan los vocablos relacionados con elementos arquitectónicos, decorativos y la riqueza, connotará que, a simple vista, para ese escritor la magnificencia, la suntuosidad y la abundancia de partes y elementos, con su inherente variedad, serían los elementos más destacables o prioritarios en su visión frente a otros conceptos.

No obstante, es indispensable una revisión posterior del listado de frecuencias proporcionado por TAPORWARE, para la eliminación de aquellos vocablos puramente gramaticales, como artículos, conjunciones, preposiciones, etc., así como otras tipologías de palabras (pronombres, etc.) que, en definitiva, no constituyen un punto de interés para la investigación de la terminología artística. A su vez, habría que considerar las diferentes formas flexivas de cada término: ej.: `ingen-ios/-ioso(s)/-iosa´, `gracioso/-iossa/-iosamente´, etc., aunque algunas de estas variaciones no aparezcan dentro del marco de las palabras seleccionadas como más recurrentes. Así mismo, también resultaría imprescindible una revisión de los vocablos restantes, a pesar de que no sobrepasen el umbral de los términos más reiterados, pues podrían llegar a ser concluyentes por su significación.

Posteriormente, estas palabras podrían ser clasificadas por el investigador en grupos específicos o categorías de vocabularios, según su uso y significado en el texto (ej.: “vocabulario descriptivo-crítico o estético”, “vocabulario técnico de las artes”, “vocabulario religioso”, etc.), según el ámbito del estudio pertinente. Por ejemplo, del vocabulario descriptivo-crítico del citado texto del Padre Sigüenza, los vocablos más reiterados son el `orden´ (308) y lo `hermoso´ (203) (Tabla 1).

Segunda fase

En la siguiente fase, de estudio cualitativo, se afinarán estos resultados, puesto que se analizará la significación de las palabras en relación con otros términos del texto y su definición dentro de su contexto real (párrafos o frases). Por ejemplo, éste sería el caso de conceptos ambiguos que podrían responder a varias acepciones semánticas.

Para llevar a cabo este estudio cualitativo, se desarrollan el resto de análisis que nos ofrece la plataforma: concordancias (búsqueda de la palabra en sus distintos contextos de uso o párrafos a lo largo del texto, para detectar con qué otros vocablos suelen relacionarse) y co-ocurrencias (búsqueda de las situaciones en que dos palabras aparecen juntas en un mismo contexto de uso o párrafo, para dilucidar su relación y significado). Con todas estas posibilidades, se esclarecerían los significados de los términos dentro del contexto textual en el que se inscriben, así como las palabras con las que suelen relacionarse. Para ilustrar estas posibilidades de búsquedas, se ha tomado como ejemplo uno de los conceptos más frecuentes del juicio estético del citado monje: `maravilla`/lo `maravilloso` (Tabla 2).

Se pueden esclarecer los significados de los términos dentro del contexto textual en el que se inscriben, así como las palabras con las que suelen relacionarse.

Comparación con otros textos

Por último, TAPORWARE aumenta todavía más nuestro ámbito de actuación con la opción *comparator*, por la que sería posible la comparación entre dos textos. Como resultado, se obtienen las palabras comunes en ambos escritos y las propias en cada caso. De este modo, se detectan las similitudes y divergencias léxico-terminológicas entre ambos textos, así como su grado de riqueza textual.

Tabla 1. Análisis cuantitativo: frecuencias de palabras del vocabulario descriptivo-crítico estético en la tercera parte de la *Historia de la Orden de San Jerónimo* (1605) de fray José de Sigüenza, sobre la descripción del Real Monasterio de El Escorial.

Orden(es)(308), *hermos-o(s)/-a(s)/-ura* (203), *razón/raçón* (147), *arte(s)* (146), *manera* (128), *natural(es)* (119), *mal* (76), *admìr-aciòn/-able* (73), *excelente(s)/-encia* (126), *ingenio (s)*(88), *antigu-os/-as/-edad* (80), *verdad* (65), *proporciòn* (62), *gracia* (61), *precios-o/-as* (61), *clar-os/-idad* (55) *juicio/juizio* (54), *alegr-es/-ía* (53), *gusto* (47), *medida(s)* (47) *primor* (42), *sentimiento* (40), *invenciòn* (40), *Roma* (37), *correspondencia* (34), *estima* (34), *modo* (31), *grieg-a/os* (39), *maravill-a(s)/-o/-an/-osas/-ándose/-aba/-ado/-ávase/-arse/-áronse/-óse/-arán/-éme* (34), *notable* (28), *artificio* (21), *dórico* (20), *compostura* (19), *enemigo* (19), *malizia* (14), *ignorancia* (14), *hereges* (16), *curios-o (s)/-a (s)/-idad (es)* (16), *Vitruvio* (15), *grutescos* (14).

Se exponen los términos más numerosos de este fragmento, de un total de 229.823 palabras, con su número de repeticiones totales entre paréntesis. Además, en estas cantidades se han tenido en cuenta otros conceptos y declinaciones de palabras que se hallaron fuera de este umbral y se han descartado vocablos puramente gramaticales, carentes de sentido (artículos, pronombres, preposiciones, conjunciones, etc). Entre los términos más reiterados destacan, en primera y segunda posición, el `orden´ (308) y lo `hermoso´ (203), convirtiéndose, a simple vista, en los principios más sobresalientes de la estética del Monasterio.

Tabla 2. Análisis cualitativo de `maravilla´/lo `maravilloso´ (concordancias y coocurrencias) en la tercera parte de la *Historia de la Orden de San Jerónimo* de fray José de Sigüenza (1605).

`Siete maravillas del mundo´, `octava´, `prodigalidad´, `ostentación y locura´, `santa´, `santos fines´, `desbaratadas´, `tanta profanidad´, pp.870 y 871; `lícita´, `santa´, `aprovada de Dios´, `maravillas y milagros continuos del Cielo´, `David y Salomón´, `templo tan admirable, para que reposase allí el arca del Señor´, `santos´, pp. 661 y 662; `Virgen´, `ángel´, `hermosos´, `suma honestidad y compostura´, `temor y reverencia´, `muchas maravillas´, `devoción´, p. 837; `ha sido maravilla´, `milagro´, `siervo de Dios´, p. 897; `maravillosos escorços y perspectivas´, `diversas partes´, `varían´, `singular deleite´, p. 757; `fábrica cumplida y excelente´, `se excede a sí misma´, `aventaja´, `maravilla y recrea´, p. 829; `fue maravilla´, `Dios´, p.587; `no les haga, pues, maravilla´, `altares de Dios´, `tantos ornamentos´, `cajones´, `mudas de paños y de colores´, `tanto cuidado´, `Rey tan pío y tan poderoso´, p.815; `no es maravilla´, `tan incomparable tesoro´, `Príncipe tan grande como el de Filipo II´, `gran devoción´, `reliquias de los santos´, pp. 828 y 829; `romanos´, `gente pagana y bárbara´, `tiranos´, `no es maravilla´, `perecido´, `vanidad soberbia y tyrana´, p.549.

Se muestran aquellas palabras con las que `maravilla`/lo `maravilloso`, uno de los vocablos más reiterados por el Fraile, suele relacionarse en los sucesivos párrafos del texto y que, por lo tanto, esclarecen el significado del término (véase tabla I: `maravill-a(s)/-o/-an/-osas/-ándose/-aba/-ado/-ávase/-arse/-áronse/-óse/-arán/-éme`, índice de frecuencias 34). De acuerdo a estas asociaciones terminológicas, `maravilla`/lo `maravilloso` se vincula al tópico de El Escorial como `Octava Maravilla del Mundo`, superior a las siete restantes. Además, `maravilla`/lo `maravilloso` se traduciría bajo las ideas de perfección y excelencia, tanto material (traducida por el lujo de sus `ornamentos`, la variedad y cantidad de elementos y `diversas partes` y la pericia técnica de sus artistas que, como resultado, crearon una `fábrica cumplida y excelente`) como sobre todo espiritual (como templo consagrado a `Dios`, vinculado a una belleza metafísica, unitiva a él, y a la alta `devoción` de su fundador, `Felipe II`). Su supremacía superará, incluso, al mítico Templo de Salomón (`David y Salomón`). Todos estos conceptos hacen del Monasterio una obra única en el mundo, frente a la `soberbia`, la `profanidad`, la barbarie (`bárbaros`) y la tiranía (`tiranos`) del resto de Maravillas (los vocablos que inspiran juicios peyorativos o negativos irían atribuidos a éstas). Además, en la presente tabla se especifican los números de páginas donde localizar todas estas relaciones terminológicas.

2.2. OTROS RECURSOS ON-LINE COMPLEMENTARIOS A LA INVESTIGACIÓN LÉXICO-TEXTUAL DE TAPORWARE: ALGUNAS HERRAMIENTAS DE LA RAE

Finalmente, para discernir todavía más el grado de originalidad o aportación del texto objeto de estudio, podrían desarrollarse otros análisis comparativos con respecto al lenguaje o idiolecto colectivo de la época en la que fue redactado. De este modo, la RAE pone a nuestra disposición, y de libre acceso, el Corpus Diacrónico Español (CORDE) (<http://corpus.rae.es/cordenet.html>) y el Corpus de Referencia del Español Actual (CREA) (<http://corpus.rae.es/creanet.html>), junto a otros medios más específicos, como su *Diccionario de Autoridades* (<http://www.rae.es/recursos/diccionarios/diccionarios-antteriores-1726-1996/diccionario-de-autoridades>), destinado al vocabulario del siglo XVIII.

El banco de datos del CORDE es una plataforma permite el acceso a un amplio corpus textual para la identificación de las posibles acepciones de los conceptos, su grado de antigüedad, vigencia

y desuso. Esta colección de textos abarca los comienzos de nuestra lengua hasta el año de 1974, en épocas y enclaves en los que se habló español. Al introducir el término de interés, la plataforma devuelve numerosos registros en los que el vocablo ha sido utilizado (párrafos y fragmentos de textos). Por ejemplo, del estudio comparativo del término `maravilla` /lo `maravilloso` en la *Historia de la Orden* de Sigüenza (uno de los conceptos más recurrentes de su juicio estético, véase Tabla 1) con la jerga popular de su tiempo (siglo XVII) a través del recurso del CORDE, se revela la coincidencia de que lo `maravilloso` podría poseer un trasfondo trascendental o divino (`celestiales`). Por lo tanto, esta acepción de lo `maravilloso` en la *Historia* no resultaría una innovación terminológica con respecto a su época (véase Tabla 2):

“...Santidad esta su Universidad de Salamanca, suplicó a V. Beatitud se dignase de insertar en el Catálogo de los Santos a la Bienaventurada Madre Teresa de Jesús, moviéndonos a tan religiosa súplica una no vulgar certeza, que fuera de toda duda nos persuade, que todo el período de su vida fué un agregado de verdaderas virtudes, y un maravilloso dechado de celestiales recibos...” (Anónimo, 1611, CORDE).

Por otro lado, la plataforma CREA recopila textos escritos y orales, desde 1975 hasta 2004, producidos en todos los países de habla hispana y procedentes de diversos géneros y medios, como periódicos, revistas y libros.

Por último, el *Diccionario de Autoridades* es definido como el “*Diccionario de la lengua castellana*, en el que se explica el verdadero sentido de las voces, su naturaleza y calidad, con las *phrases* o modos de hablar, los proverbios o refranes, y otras cosas convenientes al uso de la lengua [...]”, constituyendo la base de las sucesivas ediciones de los diccionarios de la RAE. Su segunda edición data de 1770, por lo que su uso sería, como bien se ha indicado, específico para el estudio de vocablos dieciochescos.

Con todos estos recursos, podrían llevarse a cabo estudios comparativos con el uso coloquial de los conceptos, a través de textos que comprendan los mismos periodos y procedan de ámbitos diversos (escritos notariales, jurídicos, literarios, de prensa, revistas, etc.), representando, por lo tanto, todas las variedades relevantes de nuestra lengua. Además, con estas herramientas podrían realizarse

reflexiones en torno a la evolución cronológica de un vocablo en concreto, desde los más variados géneros literarios. Así, se obtiene el grado de innovación semántica que un escritor realiza sobre un concepto con respecto a la jerga popular de su contemporaneidad o de otros periodos.

La plataforma CREA recopila textos escritos y orales, desde 1975 hasta 2004, producidos en todos los países de habla hispana y procedentes de diversos géneros y medios, como periódicos, revistas y libros.

3. CONCLUSIONES

Las Humanidades Digitales aplicadas al análisis de textos amplían su campo de acción a través de nuevas herramientas de análisis y de rastreo de información, cuya búsqueda sería inconmensurable al ojo humano. En base a ello, fomentan resultados inéditos y múltiples posibilidades de estudios, con la consiguiente apertura de futuras líneas de investigación.

Como ejemplo de trabajos precedentes que aplican herramientas digitales de análisis léxico-textual al estudio de terminología artística, el lector puede remitirse a la investigación léxico-estadística de la Dra. Rodríguez Ortega (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 del Real Monasterio de El Escorial (*Laurentina* (ca. 1580) de Luis Cabrera de Córdoba y la *Octava Maravilla del Mundo* (ca. 1595) de Juan Alonso de Almela).

4. REFERENCIAS BIBLIOGRÁFICAS

Anónimo. (1611). *Nueva instancia de la universidad de Salamanca sobre la beatificación de la santa...* España. Santa Teresa, S. de, (1915-1923), 397 (p. 1.). Monte Carmelo, Burgos: España. Tema: documentos notariales. Real Academia Española-CORDE. Recuperado de: <http://corpus.rae.es/cordenet.html> y <http://corpus.rae.es/creanet.html>.

Bassi Acuña, A. y Polanco, X. (2001). Lematización basada en análisis no supervisado de corpus. *Segundo taller internacional de procesamiento computacional del español y tecnologías del lenguaje (SLPLT2)*, Jaén, España. Recuperado de: <https://users.dcc.uchile.cl/~abassi/ecos/lema.html>

Berry, D. M. (2011). The Computational Turn: Thinking about the Digital Humanities. *Culture Machine*, (12).

Castilla Ortega, M. (2017). Metodologías de la sociedad digital aplicadas al estudio léxicoterminológico en las descripciones escurialenses de *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/>

Hockey, S. (2004). The History of Humanities Computing. Schreibman, S., Ray S., Unsworth, J. (Eds.). (2004). *Companion to Digital Humanities*. Oxford, Inglaterra: Blackwell.

López Poza, S. (2015). Humanidades digitales y Literaturas Hispánicas: presente y futuro. *Ínsula. Revista Bibliográfica de Ciencias y Letras*, (822), p. 3.

Marcos Marín, F. (2009). Historia humana de la lengua española y su computación. *Studies in Hispanic and Lusophone Linguistics*, 2(2), pp. 387-415.

Pino Díaz, J y Fiormonte, D. (2018). La geopolítica de las humanidades digitales: un caso de estudio de DH2017 Montreal. *RIUMA*. Recuperado de: https://riuma.uma.es/xmlui/bitstream/handle/10630/16329/DH2018_Pino-Fiormonte_sex.pdf?sequence=4

Rodríguez Ortega, N. (2009). Similitudes y diferencias léxico-textuales en dos descripciones seiscentistas del Real Monasterio de San Lorenzo de El Escorial (I). Ejemplo de exploración lingüístico-computacional de textos digitales y Similitudes y diferencias léxico-semánticas en dos descripciones seiscentistas del Real Monasterio de San Lorenzo de El Escorial (II). Consideraciones en torno al devenir del vocabulario descriptivo crítico de las artes. Rodríguez Ortega, N. (dir.). *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.

Sigüenza, fray J. de (1605). Tercera Parte de la Historia de la Orden de San Jerónimo, Doctor de la Iglesia... dirigida al Rey Nuestro Señor, Don Philippe III Por fray Iosep de Siguença, de la misma Orden. *Historia de la Orden de San Jerónimo* (pp. 525-899). Madrid, España: Imprenta Real, (por Iuan Flamenco). M-RAH, 5/832 M (manuscrito original). BNE, 2/47849.

Humanidades Digitales Hispánicas. Recuperado de: <http://humanidadesdigitales.org/>

Plataforma digital ATENEA-TTC (proyecto I+D HAR2009-07068). Recuperado de: www.proyectoatenea.es

Real Academia Española-CORDE. Recuperado de: <http://corpus.rae.es/cordenet.html>

Real Academia Española-CREA. Recuperado de: <http://corpus.rae.es/creanet.html>

Real Academia Española-Diccionario de Autoridades. Recuperado de: <http://www.rae.es/recursos/diccionarios/diccionarios-anteriores-1726-1996/diccionario-de-autoridades>

TAPORWARE. *Text analysis portal for research (Prototype of text Analysis Tools)*. Recuperado de: <http://taporware.ualberta.ca/>