

Prensa digitalizada: Herramientas y métodos digitales para una investigación a escala

Digitized periodical press: Digital tools and methods for a research at scale

Amelia SANZ CABRERIZO

Universidad Complutense de Madrid

amsanz@ucm.es

Dirección

Clara Martínez

Cantón

Gimena del Río

Riande

Ernesto Priani

Secretaría

Romina De León

ABSTRACT

This article presents the advantages and the challenges of a research based on digitized periodical press and digital tools and methodologies for the Humanities, in general, and for Literary Studies, and Women Studies, in particular. It deals with the new conditions for research as a result of massive digitization processes, and the epistemological problems that come along with them. The author reviews some of the most innovative research projects on the basis of digitized periodical press and makes some proposals for a very close future.

KEYWORDS

Digitized Periodical Press, Digital Tools, Digital Methods, Literature.

RESUMEN

Este artículo presenta las ventajas y desafíos de una investigación realizada a partir de la prensa digitalizada con métodos y herramientas digitales, tanto para las ciencias humanas en general, como para los estudios literarios y de la mujer en particular. Plantea la transformación de las condiciones de investigación que ha traído consigo la digitalización masiva de la prensa periódica y los problemas epistemológicos que ello acarrea. Presenta un conjunto de proyectos de investigación novedosos en curso y propone líneas posibles de desarrollo para un futuro que ya es muy inmediato.

PALABRAS CLAVE

Prensa digitalizada, herramientas digitales, métodos digitales, literatura.

RHD 2 (2018)

ISSN

2531-1786



1. PORQUE CAMBIAN LAS CONDICIONES

Nos hemos hecho digitales: desde que nos levantamos y miramos el teléfono, hasta que llegamos al trabajo y encendemos el ordenador. El cambio fundamental se perfila en torno a 2004: *Google Print* vio la luz en 2004, *Google Earth* nació cuando Google compró *Keyhole EarthViewer* en 2004 y fue liberado por la compañía en junio de 2005; Facebook fue creado en febrero de 2004. Es una fecha simbólica, porque, en realidad, todo sucede entre el primer ordenador individual con ratón de Macintosh en 1984 y el primer *smartphone* de Apple en 2007.

El hecho es que las condiciones mismas de nuestra vida intelectual han cambiado, aunque sabemos bien que no en todas partes de la misma manera, ni al mismo ritmo: basta con echar un vistazo a las estadísticas sobre la penetración de Internet según continentes y países en Internet World Stats¹, o a las desigualdades en los procesos de digitalización en la propia Europa según Enumerate². Las diferencias son particularmente claras en la prensa europea digitalizada y disponible en The European Library, que pasará pronto a Europeana³: frente a la gran cantidad de resultados provenientes de las bibliotecas alemanas, el British Newspapers Archive es un archivo con suscripción de pago y no está integrado⁴. Y es que los procesos de digitalización y su estandarización, el cuidado de los datos, su agregación y su preservación requieren mayor intervención de bibliotecarios y documentalistas de lo que creemos y no parece ser una prioridad presupuestaria para la Europa comunitaria.

Entre tanto, el centro del trabajo del investigador ha dejado de ser el edificio de la biblioteca (del archivo o del museo), ni siquiera el de su despacho, sino la pantalla de su ordenador que funciona como un punto de acceso global, aunque restringido (debido a que no toda la ciencia se ofrece en acceso abierto): no es un lugar físico (aquí), sino un espacio virtual (allí). La función de una biblioteca ya no solo consiste en guardar y custodiar según una lógica territorial: su valor no reside tanto en lo que posee, sino en lo que comparte y cómo lo hace⁵.

Ello es importante porque la cantidad de masa crítica accesible (o no) ha modificado las condiciones de la cientificidad y, desde luego, el archivo aumentado de las literaturas que leemos, enseñamos, estudiamos. Esa transformación epistemológica no es solo cuantitativa, sino que es de naturaleza hermenéutica. Hemos de entender que si ahora podemos y queremos leer más es porque partimos de una definición plural de la cultura como culturas y esas definiciones

¹ Accesible desde: <https://www.internetworldstats.com/>.

² Accesible desde: <http://www.enumerate.eu/>.

³ The European Library. *Newspapers Home. Browse by country*. Accesible desde: <https://goo.gl/YkXMNY>, materiales que estarán pronto disponibles en Europeana. *Europeana Collections*, accesible desde: <https://www.europeana.eu>.

⁴ British Library. *British Newspapers Archive*. Accesible desde: <https://www.britishnewspaperarchive.co.uk/>.

⁵ Véase el cambio de estrategia de la biblioteca y el reposicionamiento del usuario en la Universidad Complutense de Madrid en el trabajo de García García y Tardón (2018), titulado "Cooperación global e innovación en el siglo XXI: la apuesta estratégica de la Biblioteca Complutense".

posibles son fruto de la explosión teórica de los años 60 y 70, 80 y 90, es decir, de los estructuralismos, los estudios culturales y post-coloniales, los psicoanálisis y los marxismos, los feminismos, las antropologías, etc.

Desde esa concepción plural de los sistemas culturales, los periódicos son archivos privilegiados de voces distintas y distantes. Así, sin el papel continuo (químico) que dio la palabra (escrita) a tantos (y se la negó a tantas), no podríamos concebir las revoluciones del siglo XIX, las modernidades del XX ni sus contiendas.

Esa pluralidad está ahora mucho más al alcance para su estudio, gracias a la digitalización masiva de la prensa. Las dificultades de acceso a las bibliotecas han sido tradicionalmente un obstáculo mayor: había que desplazarse a una y a otra, pasar las hojas de los ejemplares en busca de citas, copiar y acopiar para una comparación posterior, posible o imposible y siempre necesariamente limitada.

En este contexto, cabe preguntarse en qué medida el número de periódicos digitalizados en España resulta ya relevante y útil para una investigación. La *Hemeroteca Digital* contaba en 2014 con 1.154 títulos y en 2018 con 2.060 (Sanz, 2016; BNE, 2017; Rodríguez, 2018) y la *Biblioteca Virtual de Prensa Histórica* con 2.181 cabeceras en 2014 y 2.410 hoy⁶, a los que han de añadirse los 387 de *Arxiu de Revistes Catalanes Antiques (ARCA)*⁷.

Con todo, la prensa local no ha sido suficientemente digitalizada: de los 373 publicados en Jaén en el siglo XIX nos han llegado 78 y sólo 2 aparecen digitalizados en la Hemeroteca Nacional (HN) (Contreras, 2009). Sabemos que existieron 200 periódicos en Palma de Mallorca de los cuales 22 aparecen disponibles en la (Biblioteca Virtual de Prensa Histórica (BVPN) (Díaz y Moll, 1982). Tampoco están completos los repertorios de Lleida y solo 4 títulos están disponibles en la HN y en BVPH en formato electrónico (Giné y Domínguez, 2004 y 2016 y pp. 129-166)⁸. Las ausencias llegan a ser graves, por ejemplo en el caso de las revistas de mujeres: de las 129 publicaciones periódicas dirigidas a las mujeres de las que tenemos constancia (Simón Palmer, 1975; Perinat y Marrades, 1980; Simón Palmer, 1993; Palomo, 2014)⁹, solo han sido digitalizadas 44. Llama la atención que 54 publicaciones del total fueran realizadas en Barcelona, de las cuales sólo 2 están disponibles en ARCA.

⁶ Biblioteca Virtual de Prensa Histórica. *Estadísticas*. Accesible desde: <http://prensahistorica.mcu.es/es/consulta/estadistica.cmd>.

⁷ Accesible desde: <http://www.bnc.cat/digital/arca/>. Véase Navarro (2018). A ello habría que añadir e la Càtedra Marius Conde, el *Corpus Literari Digital. Hemeroteca*. Accesible desde <http://www.catedramariustorres.udl.cat/materials/hemeroteca/diccio.php>. Tampoco podemos olvidar las sorpresas que puede depararnos Hispana. *Directorio de Colecciones Digitales*. Accesible desde <http://hispana.mcu.es/es/comunidades/directorio.cmd>, aunque en la mayoría de las ocasiones no permita buscar en modo texto.

⁸ En el Arxiu Municipal de Lleida, accesible desde: <http://premsadigital.paeria.es/cgi-bin/pandora.exe>, se conservan unos 250 títulos desde el año 1854 hasta nuestros días, de los cuales solo 14 (en su mayoría del siglo XX) han sido digitalizados. En ARCA solo encontramos una cabecera editada en Lleida en el siglo XX.

⁹ También en la *Spanish Periodicals and Newspapers: Women's Magazines Digital Collection*, accesible desde: <http://doddcenter.uconn.edu/asc/collections/spanwomen.htm>.

Es cierto que las bibliotecas optaron, en primer lugar, por las cabeceras más demandadas, las colecciones más completas, las que corrían más peligro por la calidad del papel químico, las de mayor circulación (Rodríguez, 2012). Cabe preguntarse, sin embargo: ahora, ¿qué se está digitalizando?, ¿qué nuevo canon se está formando al socaire de la digitalización? Es importante hacerse esta pregunta porque sabemos (partiendo de las ideas en otros ámbitos de Michel Foucault) que el archivo electrónico selecciona, modeliza colecciones, etiqueta así lo que era invisible y lo trae a la existencia. La digitalización, como cualquier otra memoria escrita, connota.

Lo que es, quizás, indiscutible, es el éxito de la digitalización, que muestra ese cambio en las condiciones del que al principio hablábamos. Queda patente en las estadísticas de la Hemeroteca Digital de la Biblioteca Nacional de España (BNE, 2015; BNE, 2016; BNE, 2017):

	2014	2015	2016	2017
Documentos descargados en la H. D.	2.771.576	3.413.956	3.964.458	3.897.665

2. PORQUE CAMBIAN LAS ESCALAS

Siempre podemos preguntarnos cuántos granos de arena necesitamos para constituir un montón o cuántas publicaciones debemos estudiar para extraer una generalización. Hasta hace poco, el vaciado y estudio de una sola publicación periódica bien valía un artículo. Hoy, ¿cuántas debemos tratar para que nuestra investigación entre en el paradigma científico al uso? Pues dependerá de lo arraigado de las tradiciones epistemológicas e institucionales de las disciplinas, regiones y países, porque algunos aceptarán aún un modelo a partir de las citas extraídas de una sola publicación y se resistirán a aceptar modelizaciones realizadas con herramientas electrónicas a partir de cientos de publicaciones. Trabajamos con y para los modelos, con metodologías inductivas o deductivas, siempre sometidas a hermenéuticas: esos modelos han de ser sometidos a pruebas, validados, contrastados, a partir de uno o mil ítems en un corpus.

Lo cierto es que la digitalización masiva va a obligar y permitir revisar (una y otra vez según vaya creciendo) muchos trabajos: eso es la ciencia. Existen multitud de brillantes trabajos de investigación sobre publicaciones locales que no han sido aún digitalizadas¹⁰. Pero buena parte de esa investigación realizada deberá ser completada cuando esos títulos estén

¹⁰ Pongamos como ejemplo Sánchez Agustí (2002), sobre el periódico republicano *La Libertad*. Se trata de una publicación local, de Valladolid, pero con una clara proyección iberoamericana, pues muestra los precios para Cuba y Puerto Rico, Filipinas y países de ultramar, como se ve en la portada del número 1. O también cualquiera de los capítulos en Medina (2009).

disponibles en formato electrónico y puedan ser estudiados con metodologías electrónicas: cuando se trata simplemente de identificar y extraer cadenas de letras, los programas informáticos lo hacen y lo harán mejor que el ojo humano.

De hecho, empieza a ser posible una investigación transnacional, transgénica, hipermedial e hipertextual, aunque, vistos los ejemplos que vamos a revisar a continuación, nos tememos que se está desarrollando en su mayoría en el primer mundo.

El proyecto *Oceanic Exchanges* (2017) investiga desde 2012, en un marco relacionado con la Ciencia Abierta, los modelos de información que cruzan fronteras nacionales y lingüísticas, esto es: gracias a un análisis computacional a gran escala, los límites que separan los corpus de periódicos digitalizados aislados se borran para ilustrar conexiones globales entre publicaciones del siglo XIX. En esta línea, el proyecto *Viral Texts*¹¹ persigue aquellos rasgos textuales y temáticos que convierten una historia, un cuento, un poema, una información relatada, en viral; muchas veces se trata de potentes híbridos por su capacidad de circulación, como es el caso de las viñetas en los periódicos.

Efectivamente, las herramientas computacionales permiten estudiar los procesos culturales transnacionales, como es el caso de los de mediadores culturales hispánicos en la modernidad internacional durante el primer tercio del siglo XX, estudiado en el proyecto *Cartografía de la modernidad hispánica. Redes literarias transnacionales y mediadores culturales (España-Latinoamérica, 1908-1939)*¹².

Y es que estos métodos rompen fronteras nacionales, culturales, sociales, genéricas. El proyecto *Borderlands Archives Cartography* se dedica a visualizar los periódicos de la fluida frontera México-Estados Unidos y, de esta forma, produce otras historias alternativas a la oficial que representen la frontera desde las propias comunidades de esas fronteras y visualiza los archivos de periódicos en la plataforma digital¹³.

Un proyecto como *Recovering the U.S. Hispanic Literary Heritage Program*¹⁴ propone recuperar el legado escrito hispano de los Estados Unidos a partir de los más de 1.400 periódicos publicados por latinas/os en los Estados Unidos desde 1808 a 1960, mientras presta especial atención a las conjeturas que delimitan la producción, la distribución y el consumo, y rata de contestar así a las preguntas quién habla, de quién se habla, por quién se habla. Es la oportunidad para que las comunidades puedan crear sus propias representaciones en forma de

¹¹ *Viral Texts: Mapping Networks of Reprinting in 19th-Century Newspapers and Magazines*. Accesible desde: <http://viraltxts.org>.

¹² *Cartografía de la modernidad hispánica. Redes literarias transnacionales y mediadores culturales (España-Latinoamérica, 1908-1939)*. Accesible desde: <https://mapmodern.wordpress.com/cartografia-modernidad/>.

¹³ *Borderlands Archives Cartography*. Accesible desde: <https://www.bacartography.org/>.

¹⁴ *Recovering the U.S. Hispanic Literary Heritage Program*. Accesible desde: https://touch.facebook.com/RecoveringUSHispanicHeritage/?_tn=~H-R.

archivo digital: el *cuarto archivo* del que habla Cook (2013), el de un *ethos* social y político de una comunidad, no de una sociedad.

Son trabajos que cuestionan y reconfiguran, porque revelan y visualizan, las identidades, como hace *La traduction littéraire sous l'Occupation - France, Belgique, 1940-1944*¹⁵ que muestra la intensa cobertura mediática de las obras traducidas en la llamada *prensa de la Ocupación*, símbolo de una colaboración intelectual exitosa y una enorme apuesta para imponer la línea ideológica del ocupante.

No solo abren identidades, sino que las colocan en el mapa cultural como es el caso del proyecto *Digitization of Rare Philippine Periodicals*¹⁶, con una vocación y una metodología didáctica que cuenta con los estudiantes y para ellos. En este mismo sentido también pone en el mapa Luis Andrade la revista portuguesa *A_Aguia*, de principios del siglo XX en Portugal, por el hecho de ponerla en línea y descifrarla¹⁷.

El proyecto del Austrian Centre for Digital Humanities apuesta por una perspectiva y una metodología centrada en el usuario, de forma que sea la comunidad la encargada de poblar de anotaciones una publicación del siglo XVIII, como es el *Wien[n]erisches Diarium*¹⁸.

Se trata de sacar el mayor partido a la digitalización de las publicaciones periódicas mediante el uso de herramientas y tecnologías digitales, por ejemplo, mediante análisis de redes de contactos, como propone el portal *Revistas Culturales 2.0*¹⁹ de la Universidad de Tübingen, que está diseñado como un entorno virtual para investigadores de revistas culturales históricas en lengua española. Se puede sacar también partido mediante una estilometría digital que busque constelaciones léxicas y estilísticas en las reseñas y comentarios, junto con visualizaciones propias de la llamada *geo-crítica*, como en el proyecto *The reception of the English novel in the Italian literary press between 1700 and 1830*²⁰. También cabe la construcción de extensas bases de datos con parámetros diversos que permitan inventariar categorías de traducción en los periódicos británicos del Romanticismo (Hacke, 2018).

¹⁵ *La traduction littéraire sous l'Occupation - France, Belgique, 1940-1944*. (2014-18). Accesible desde: <http://www.tsocc.univ-nantes.fr>.

¹⁶ *Digitization of Rare Philippine Periodicals and Training in DH*. Accesible desde: <https://www.uantwerpen.be/en/research-groups/digitalhumanities/about/projects/vlir-uos/>. Véase también Ortuño Casanova (2018).

¹⁷ Accesible desde: http://ric.slii.pt/A_Aguia/um_voo_singular_e_longo.

¹⁸ Austrian Centre for Digital Humanities. *Das Wien[n]erisches Diarium. A Digital Data Treasure for the Humanities*. Accesible desde: <https://www.oead.ac.at/acdh/projects/wienerisches-diarium-digital>. Cualquiera puede contribuir a partir de la aplicación web que permite anotaciones a texto completo, identificación de entidades nombradas, correcciones textuales y anotaciones semánticas o estructurales.

¹⁹ Accesible desde: <https://www.revistas-culturales.de/es>.

²⁰ *The reception of the English novel in the Italian literary press between 1700 and 1830: a transcultural enquiry into the early shaping of the modern Italian literary and cultural identity*. Accesible desde: [https://cris.vub.be/en/projects/the-reception-of-the-english-novel-in-the-italian-literary-press-between-1700-and-1830-a-transcultural-enquiry-into-the-early-shaping-of-the-modern-italian-literary-and-cultural-identity\(113a4336-9aca-402f-bc69-3f4aa55b90b1\).html](https://cris.vub.be/en/projects/the-reception-of-the-english-novel-in-the-italian-literary-press-between-1700-and-1830-a-transcultural-enquiry-into-the-early-shaping-of-the-modern-italian-literary-and-cultural-identity(113a4336-9aca-402f-bc69-3f4aa55b90b1).html).

Cada vez que cambiamos de escala, modificamos la granularidad: las preguntas son otras y las respuestas también.

3. PORQUE CAMBIAN LAS POSICIONES DE OBSERVACIÓN

Para el estudio de las literaturas y de sus lecturas, la prensa puede ser considerada un punto de encuentro entre un autor que escribe y un lector que lee, de modo que éste resulte tan importante como aquel. Los periódicos anuncian las respuestas predecibles de los lectores: editores y redactores se adelantan a las suposiciones compartidas sobre modelos que se ven así reforzados por los estereotipos en anuncios y reseñas, críticas y traducciones, alusiones o comentarios (Charnon-Deutsch, 2000). Nos interesan esos jirones de frases aparentemente desconectados porque remiten a la selección que la literatura de llegada hace de los textos de origen. Más aún, aportan instrucciones de lectura, modelan normas que, en última instancia, rigen el funcionamiento de un sistema literario dado²¹.

Particularmente en el siglo XIX, la prensa es una práctica social fundamental para la construcción del modelo occidental de nación por cuanto que favorece la territorialización, la serialización, la estandarización del lenguaje periodístico²² al tiempo que, paradójicamente, incorpora una heterogeneidad de voces e internacionaliza los discursos de las construcciones nacionales (Thiesse, 1999).

En el caso de la circulación de la literatura de mujeres, la prensa resulta especialmente relevante por cuanto que, frente a la progresiva exclusión de las mujeres en la vida pública después de la Revolución Francesa (Goodman, 1994), la escritura y la lectura se mantienen como formas de actividad aristocrática y burguesa propia del espacio doméstico. Estudiar la circulación de obras escritas por mujeres, traducidas o comentadas en los periódicos y revistas, es una manera de afrontar la historia (literaria) desde una perspectiva de género. Tomaremos la posición de la recepción en ese encuentro como punto de arranque de la investigación para entender el papel de las mujeres en la historia literaria: la literatura escrita y leída por mujeres puede ganar legitimidad y visibilidad si estudiamos su circulación (Sanz, Scott y Van Dijk, 2014; Sanz, 2016). Por eso, cualquier evidencia de que un texto/autora ha sido leído, tratado, traído, es una prueba (una huella) de la circulación de esas mujeres escritoras, como se pone de manifiesto en el proyecto *Women Writers Networks*²³.

Nos han faltado fuentes para una historia de la circulación de la escritura y de la lectura femeninas, a pesar de la gran investigación empírica realizada desde los años 70. Esas fuentes son variadas, están dispersas, son escasas porque las mujeres han dejado pocas huellas en los

²¹ En el sentido de los *Descriptive Translation Studies* y la teoría de los polisistemas de Toury (1995) y de las teorías post-Bourdieu (Sanz Roig, 2014).

²² Son los principios propuestos en la trilogía de Benedict Anderson (1983, 1990 y 1998).

²³ Accesible desde: <http://www.womenwriters.nl>.

archivos tradicionales para el historiador (Farge, 1979; Thébaut, 2007). Necesitamos la prensa, porque las voces femeninas sí que llegaban a la prensa, aunque refractadas.

La digitalización de la prensa permite ampliar la escala de la investigación, pero, además, las herramientas digitales nos llevan a otras preguntas de investigación. Tenemos ya una profusión de publicaciones sobre casos aislados de mujeres, en forma de pluri-biografías o de síntesis (Planté, 2003), que se dedican a la semblanza biográfica (o bio-bibliográfica) de una sola autora en una sola revista en un determinado periodo, de forma que se subraya la excepcionalidad de la figura considerada única: es una historiografía de monumentos. Pero la digitalización, junto con herramientas y métodos digitales, nos permite buscar cuadros de conjunto más exhaustivos y transnacionales, encontrar relaciones y dibujar circulaciones, poner en valor grandes cantidades y sus pluralidades: una historiografía de documentos. Así, si la biografía es un modelo (una reducción) de una vida basada en la diferencia social (*La question biographique en littérature*, 2008), otras modelizaciones son posibles, basadas en las repeticiones, los contactos, las semejanzas, los encuentros.

Es por eso que hemos lanzado diferentes proyectos de investigación con estudiantes de grado y de máster en busca de referencias a Madame de Staël (Sanz y Menéndez, 2015) en la prensa española digitalizada, o de las 452 mujeres que, nacidas entre 1780 y 1900, escriben en francés y aparecen en la base de datos NEWW²⁴ (Sanz, 2018).

En una primera etapa, los estudiantes descubren las grandes hemerotecas digitales en España y buscan, de forma automática, las referencias a la(s) escritora(s), transcriben las citas corrigiendo los errores del OCR y las comparten en una wiki en Campus Virtual. En una segunda etapa, descubren herramientas como *Voyant Tools* o *Sketch Engine*²⁵ que nos permiten visualizar y representar frecuencias, solidaridades, ocurrencias y concordancias. En una tercera etapa, se contextualizan y analizan citas que parecen particularmente interesantes porque contradicen o puntualizan los resultados estadísticos que arrojan los algoritmos de las herramientas. Al final, publicamos en la colección de *Eprints* de la Universidad Complutense los resultados en bruto, debidamente referenciados (Sanz et al., 2015 y 2018).

Nos instalamos así en diferentes topografías de la lectura: desde más arriba o desde más abajo, desde lo cuantitativo a lo cualitativo, lectura con microscopios o con telescopios electrónicos (motores de búsqueda, wikis, programas para análisis de corpus, herramientas de visualización, etc.)²⁶, en un intento de aunar actividades de investigación y de aprendizaje, de mejorar las competencias digitales, informacionales y literarias sin solución de continuidad.

²⁴ Accesible desde: <http://resources.huygens.knaw.nl/womenwriters>.

²⁵ *Voyant Tools*, accesible desde: <https://voyant-tools.org/>, y *Sketch Engine*, accesible desde: <https://www.sketchengine.eu/>, todas herramientas abiertas, amigables, orientadas hacia el usuario y, en consecuencia, al alcance de los estudiantes que pueden así desarrollar sus habilidades y competencias.

²⁶ Es la lectura con recursos o cuarto tipo de lectura propuesta por Bode y Dixon (2009) como *resourceful reading*.

Practicamos una hermenéutica de los datos y una hermenéutica de los textos: una metodología con granularidades diferentes a partir de unas premisas teóricas particulares.

4. PORQUE TIENEN QUE CAMBIAR LAS CONDICIONES

4.1. Extensiones y distensiones en los procesos de digitalización

Es cierto que muchas bibliotecas y archivos se encuentran en la difícil tesitura de poner en cola (o aplazar *sine die*) una buena parte de su patrimonio para su digitalización, al tiempo que han de producir contenidos digitales para satisfacer la demanda de millones de usuarios y mientras los formatos que se utilizaron hace no más de diez años ya son presa de la obsolescencia.

Para la investigación con herramientas digitales no basta cualquier tipo de digitalización: para realizar una lectura *hábil*, las bibliotecas deben garantizar búsquedas transversales gracias a un OCR con un porcentaje medio correcto de fiabilidad (Rodríguez Fuentes, 2012). Una mera fotografía de las páginas de una publicación periódica (en formato jpg o pdf) puede ampliar el acceso, pero no modificará el paradigma de investigación. Tampoco nos sirve un acceso local (solo en el espacio físico de una biblioteca) a ciertas cabeceras digitalizadas por ser de actualidad y con derechos²⁷. Las colecciones digitales tienen que respetar los protocolos estándares que permiten la interoperabilidad entre diferentes entornos (Subdirección General de Coordinación Bibliotecaria, 2016), la interrogación a gran escala, la exportación y la descarga de datos. No nos sirven las bibliotecas digitales que son meros escaparates de imágenes (ahí reside su fracaso), necesitamos hacer cosas con los documentos: construir conocimiento, no (ad)mirarlo. Más aún: según corren los tiempos y sus generaciones, los paradigmas y sus modas, los materiales no digitalizados/no digitales pueden desaparecer del campo de visión y del canon. Si el paradigma científico dominante resulta ser digital y del Norte, los investigadores y la investigación del Sur Global (en el que podemos incluirnos por razones geopolíticas) nos arriesgamos a quedar excluidos. Y es lo que ya está sucediendo.

4.2. Epistemologías de las herramientas y métodos digitales

Reconozcamos que, frente a la sobre-teorización de la historia literaria y de las ciencias humanas en general, las herramientas y métodos digitales están sub-teorizados, como si fueran neutros, pero no lo son. Tampoco son transparentes: han nacido digitales de la mano (en su gran mayoría) de ingenieros occidentales, blancos y masculinos, y se han querido hacer globales

²⁷ A la Hemeroteca Digital de la BNE se están incorporando grandes cabeceras de prensa española a un ritmo vertiginoso pero, por desgracia, este material tiene derechos de propiedad intelectual y no está accesible desde fuera de las instalaciones de la biblioteca.

para usos locales sobre la base de una única idea de modernidad concebida como progreso técnico.

En Humanidades sabemos mucho de los conflictos hermenéuticos de las interpretaciones. Sabemos que no hay neutralidad posible ni en la posición de observación ni en la elección de la escala de observación: todas son reductivas y cada una hace de ese acto de simplificación su propio *ethos*. Y las Ciencias Humanas también saben contar: desde la llamada *Escuela de los Annales* y mucho antes. Hoy los ordenadores nos permiten contar más y cosas diferentes. Las herramientas digitales pueden funcionar como reflectores de un radar o sistemas de radiolocalización capaces de señalar la presencia y la distancia de masas de información. Desde ahí, tendremos que aceptar una premisa de los métodos digitales: el carácter no lineal de los modelos y relaciones que los análisis computacionales arrojan como principios organizativos (alternativos) para un documento único o para miles, pero que no pueden ser detectados por el lector humano por sí solo.

Necesitamos esos emisores capaces de lanzar ondas potentes (las búsquedas), en la dirección precisa que solicita el investigador; necesitamos unos receptores capaces de eliminar el ruido que no nos interesa (los filtros de la búsqueda) y, finalmente, un buen sistema de tratamiento de la señal recibida (los registros electrónicos) para clasificar y tratar la información recibida por el investigador que necesariamente ha de interpretar el resultado.

Para dar sentido a esos datos a gran escala (digital) podemos echar mano de las teorías sistémicas y los *Cultural Studies*, de la historia del libro y los *Translation Studies*, de la sociología de Bourdieu y post-Bourdieu, de los análisis del discurso social y los Estudios de la Memoria, de la lingüística del corpus y de la lingüística cognitiva. Más aún, desde esos modelos interpretativos, extraeremos nuestros propios *capta* (Drucker, 2011).

En este momento, herramientas como las ya citadas *Voyant Tools* o *Sketch Engine* deberían ser ya moneda corriente entre cualquier estudiante de Grado en Letras; las librerías de Python y sus herramientas NLTK (Natural Language Toolkit) (Perkins, 2014), incluso las técnicas de *topic modeling* (Rhody, 2012), deberían estar al alcance de cualquier doctorando. Es la estilometría digital la que puede comparar constelaciones léxicas, semánticas, sintácticas en diferentes prácticas discursivas de periódicos y periodistas. Herramientas de GIS (*Geographical Information System*) como *TimeMapper*²⁸ o *ArcGIS*²⁹ son utilizables desde los primeros ciclos universitarios; las visualizaciones y representaciones de estructuras con *Gephi*³⁰ o *Nodegoat*³¹ son herramientas indispensables para muchos trabajos de investigación.

²⁸ Accesible desde: <http://timemapper.okfnlabs.org/>.

²⁹ Accesible desde: <https://www.arcgis.com/home/index.html>.

³⁰ Accesible desde: <https://gephi.org/>.

³¹ Accesible desde: <https://nodegoat.net/>.

4.3. Competencias digitales de todos para todos

Lo vemos todos los días hasta en el cine (piénsese en *Yo, Daniel Blake*, la película de Ken Loach en 2016): la falta de competencia digital genera exclusión social. Lo prueban los economistas: la mejora de la cualificación laboral contribuye al crecimiento mucho más de lo que se pensaba, lo cual reduce el papel mecánico del mero progreso técnico³².

Si queremos que las hermosas colecciones de periódicos digitalizados en España y en Europa con dinero público sean nuestras (porque son nuestro patrimonio), necesitamos unas Humanidades Digitales para nosotros, que no tienen que ser exactamente *Digital Humanities*, porque no designan al mismo referente. Necesitamos herramientas y métodos digitales a nuestro alcance para formular preguntas de investigación y respuestas culturales que hasta ahora no habían sido posibles, esto es, algo más que colecciones de revistas en fotos. Necesitamos que los ciudadanos puedan apoderarse de ellas y empoderarse con ellas, entre otras cosas, para legitimar la inversión de los poderes públicos que (no lo olvidemos) nosotros elegimos.

La democratización del conocimiento no reside en la *democracia de los clicks* que propone Google, porque oculta las condiciones materiales de las tecnologías y los espacios en los que se interacciona con ellas. Aprender a usar un producto o servicio digital implica tener las condiciones suficientes, los conocimientos adecuados, el tiempo social requerido, de forma que las hemerotecas digitales no sean repositorios estáticos sino espacios activos de producción, como muestran bien Pons (2014), Rissam (2017), Meza (2017). Pero echamos de menos más experiencias de aprendizaje-investigación con publicaciones periódicas digitalizadas³³.

Porque solo desde la pregunta *para quién*, podremos responder a las necesidades de dónde, cómo y para qué las herramientas y las habilidades digitales.

REFERENCIAS BIBLIOGRÁFICAS

- Anderson, B. (1983). *Imagined Communities*. New York: Verso.
- _____ (1990). *Language and Power: Exploring Political Cultures in Indonesia*. Ithaca, New York: Cornell University Press.
- _____ (1998). *The Spectre of Comparisons*. New York: Verso.
- Bakker, G., Crafts, N. y Woltjer, P. (2017). The Sources of Growth in a Technologically Progressive Economy. The United States, 1899-1941. *Economic History Working Papers*, 269, The London School of Economics and Political Science. Recuperado de <http://eprints.lse.ac.uk/85081/1/WP269.pdf> el 10/10/20118.

³² Por ejemplo, en un muy comentado artículo de Bakker, Crafts y Woltjer (2017).

³³ Son muy limitadas, véase por ejemplo: Institución Libre de Enseñanza. Jornada *Las revistas culturales como materiales para la educación*. Accesible desde: <https://goo.gl/KsyxFo>. También, la International Conference and Workshop, *Teaching Women Writers, Exploring NEWW Virtual Research Environment Possibilities*, Ljubljana 16th and 17th November 2017. Accesible desde: <https://goo.gl/oD5WPF>.

- Biblioteca Nacional de España (2015). *Informe de Indicadores. Biblioteca Nacional de España 2015*. Recuperado de <https://goo.gl/SmKfoo> el 10/10/20118. (=BNE)
- _____ (2016). *Informe de Indicadores. Biblioteca Nacional de España 2016*. Recuperado de <https://goo.gl/1Q8RiV> el 10/10/20118. (=BNE)
- _____ (2017). *Informe de Indicadores. Biblioteca Nacional de España 2017*. Recuperado de http://www.bne.es/export/sites/BNWEB1/webdocs/LaBNE/Transparencia/Institucional/Informe_indicadores_2017.pdf el 10/10/20118. (=BNE)
- Bode, K. y Dixon, R. (2009). *Resourceful Reading. The New Empirism, eResearch, and the Australian Literary Culture*. Sidney: Sidney University Press.
- Charnon-Deutsch, L. (2000). *Fictions of the Feminine in the Nineteenth-Century Spanish Press*. Pennsylvania: The Pennsylvania State University Press.
- Contreras Gila, S. (2009). Francia en la prensa giennense del siglo XIX. En E. Medina (Coord.), *La prensa/La Presse* (pp. 205-221). Jaén: Diputación de Jaén.
- Cook, T. (2013). Evidence, Memory, Identity, and Community: Four Shifting Archival Paradigms. *Archival Science*, 13(2-3), 95-120. doi:10.1007/s10502-012-9180-7.
- Díaz de Castro, F. J. y Moll Blanes, I. (1982). Prensa conservadora a finales del siglo XIX. El Ancora de Mallorca. En B. Barrière et al. (Coords.), *Metodología de la historia de la prensa española* (pp. 58-80). Madrid: Siglo XXI.
- Drucker, J. (2011). Humanities Approaches to Graphical Display. *Digital Humanities Quarterly*, 5(1). Recuperado de <https://goo.gl/YSDvZW> el 10/10/2018.
- Farge, A. (1979). L'histoire ébruitée. En Ch. Dufrancastel et al. (Coords.), *L'histoire sans qualités* (pp. 20). Paris: Galilée.
- García García, J., Tardón, E. (2018). Cooperación global e innovación en el siglo XXI: la apuesta estratégica de la Biblioteca Complutense. *E-Prints Complutense*, abril 2018. Recuperado de <https://eprints.ucm.es/47137/> el 10/10/2018.
- Giné, M. y Domínguez, Y. (2004) *Prensa hispànica i literatura francesa al segle XIX. Petites i grans ciutats*. Lleida: Ed. de la Universitat de Lleida.
- Goodman, D. (1994). *The Republic of Letters: a Cultural History of the Enlightenment*. New York: Ithaca.
- Hacke, M. (2018). Metadata Mining: Translation in British Romantic Review Periodicals, 1809-1827. En *Translation in Periodical Publications. Theoretical Problems and Methodological Challenges for a Transnational Study on a Large Scale. International Conference, U.O.C., 13-14 septiembre 2018*. Recuperado de: <https://goo.gl/9DRvM2> el 10/10/2018.
- Medina, E. (Coord.). (2009). *La prensa/La Presse. Coloquio Hispano-francés "Provincia de Jaén" de Estudios del siglo XIX*. Jaén: Diputación de Jaén.

- Meza, A. (2017). Decolonizar las humanidades digitales: cómo diseñar un repositorio digital de sur a norte. *Intervenciones en estudios culturales*, 4, 109-131. Recuperado de <https://goo.gl/btYvKK> el 10/10/2018.
- Navarro, P. (2018). ARCA (Archivo de Revistas Catalanas Antiguas): un momento para el cambio. En Biblioteca de Galicia, *Prensa histórica e xeneracion de coñecemento: do papel ao dixital. Xornada técnica*. Recuperado de <https://goo.gl/KEj6wd> el 10/10/2018.
- La question biographique en littérature*. (2008). "Présentation du dossier". Número spécial *Contextes*, 3. Recuperado de <http://contextes.revues.org/2653> el 10/10/2018.
- Oceanic Exchanges Project Team. (2017). Oceanic Exchanges: Tracing Global Information Networks In Historical Newspaper Repositories, 1840-1914. doi:10.17605/OSF.IO/WA94S.
- Ortuño Casanova, R. (2018). Philippines at the Crossroads: Enhancing Research on Philippine Periodicals and Finding Transnational Attitudes in Them. En *DH18.Mexico*. Recuperado de <https://dh2018.adho.org/en/perspectivas-digitales-y-a-gran-escala-en-el-estudio-de-revistas-culturales-de-los-espacios-hispanico-y-lusofono/> el 10/10/2018.
- Palomo Vázquez, M. P. (2014). Las revistas femeninas españolas del siglo XIX. Reivindicación, literatura y moda. *Arbor*, 190(767). doi:10.3989/arbor.2014.767n3001.
- Perinat, A. y Marrades, M.I. (1980). *Mujer, prensa y sociedad en España 1800-1939*. Madrid: Centro de Investigaciones Sociológicas.
- Perkins, J. (2014). *Python 3 Text Processing with NLTK 3 Cookbook*. Birmingham, UK: Packt Publishing.
- Planté, C. (2003). La place des femmes dans l'histoire littéraire : annexe, ou point de départ d'une relecture critique? *Revue d'histoire littéraire de la France*, 103(3), 655-668.
- Pons, A. (2014). *El desorden digital: guía para historiadores y humanistas*, Madrid: Siglo XXI de España.
- Rissam, R. (2017). Breaking and Building: The Case of Postcolonial Digital Humanities. En J. G. Singh y D. D. Kim (Eds.), *The Postcolonial World* (pp. 345-362). Londres: Routledge.
- Rhody, L. (2012). Topic Modeling and Figurative Language. *Journal of Digital Humanities*, 2(1). Recuperado de <https://goo.gl/m7vMF5> el 10/10/2018.
- Rodríguez Fuentes, D. (2012). El reto de la digitalización de prensa: Hemeroteca Digital de la BNE. *Boletín de ANABAD*, 4 (oct.-dic.). Recuperado de <https://goo.gl/pe879D> el 10/10/2018.
- _____. (2018). El renacer de la prensa histórica a través de la hemeroteca digital. En *Prensa histórica e xeneracion de coñecemento: do papel ao dixital. Xornada técnica*. Recuperado de http://ficheiros-web.xunta.gal/descargas/bibgal/renacer_prensa_historica.pdf el 10/10/2018.
- Sánchez Agustí, M. (2002). *La educación española a finales del XIX: una mirada a través del periódico republicano La Libertad*. Madrid: Milenio y Ministerio de Educación.

- Sanz, A. (2016). Lentes electrónicas sobre la prensa española del s. XIX: en busca de escritoras que vienen de Francia. En Y. Bürki y H. Partzsch (Eds.). *Redes de comunicación. Estudios sobre la prensa en el mundo hispanohablante* (pp. 351-379). Berlin: Frank and Timme.
- Sanz, A. y Menéndez, A. (2015). Resourceful Reading Experiences with Undergraduates: On Madame De Staël in the British Newspapers Archive (1800-1830). En *E-Prints-UCM*. Recuperado de <http://eprints.ucm.es/32725/> el 10/10/2018.
- Sanz, A., Menéndez de la Cuesta González, A., Carrillo, J. M., Constans, C., Coué, E., Domínguez, C., Esteban, A., Hernández, C., Kochan, D., Lacambra, J., Mazzariello, F., Montedeoca, C., Mvogo, A., Orliac, J., Paz, L., Pérez, I., Prieto, C., Salomón, J., Samoletnikova, A., Veloso, I. D. y Villanueva, L. (2015). Resourceful Reading Experiences with Postgraduate Students: on Madame de Staël in the Spanish Digitized Press. En *E-Prints-UCM*. Recuperado de <http://eprints.ucm.es/32726/> el 10/10/2018.
- Sanz A. et al. (2018). 19th French Women Writers in the Spanish Digitized Press. En *E-Prints-UCM* (en prensa).
- Sanz, A., Scott, F. y Van Dijk, S. (Eds.). (2014). *Women Telling Nations*. Amsterdam: Rodopi.
- Sanz Roig, D. (2014). *Bourdieu después de Bourdieu*. Madrid: Arco.
- Simón Palmer, C. (1975). Revistas españolas femeninas del siglo XIX. *Homenaje a D. Agustín Millares Carlo*. Tomo I (pp. 401-445). Las Palmas: Cajas de Ahorro.
- _____ (1993). *Revistas femeninas madrileñas*. Madrid: Artes Gráficas Municipales.
- Subdirección General de Coordinación Bibliotecaria. (2016). *Requisitos técnicos de los proyectos de digitalización de patrimonio bibliográfico y de prensa histórica de la SGCB*. Madrid: Ministerio de Educación, Cultura y Deporte. Recuperado de <https://goo.gl/t4wQeD> el 10/10/2018.
- Thébaud, F. (Ed.). (2007). *Ecrire l'histoire des femmes et du genre*. Paris: ENS Editions.
- Thiesse, A. M. (1999). *La création des identités nationales. Europe XVIIIe-XXe siècles*. Paris: Seuil.
- Toury, G. (1995). *Descriptive Translation Studies and Beyond*. Amsterdam: John Benjamins.