

# Un estudio de la prensa latinoamericana entre los siglos XIX y XX con un enfoque en eventos meteorológicos

*Diana Comesaña\**

*Luis M. Vilches-Blázquez\*\**

*Recibido el 15 de agosto de 2018; aceptado el 17 de septiembre de 2018*

## RESUMEN

Uno de los desarrollos más útiles para los historiadores modernos es la digitalización de los archivos de periódicos, ya que éstos se encuentran entre las fuentes más valiosas para los estudiosos que están interesados en comprender cuestiones relacionadas con la política, cultura y sociedad. Este hecho está produciendo una creciente digitalización de estos fondos, lo que a su vez provoca la aparición de “nuevos” desafíos, derivados de la frecuente ausencia de aplicación de tecnologías de reconocimiento de texto a los periódicos históricos digitalizados. Esto dificulta la aplicación de tecnologías que permitan su tratamiento automático y, en consonancia, la extracción de sucesos de la vida cotidiana, entre los que se encuentran los eventos meteorológicos.

Considerando este escenario, en el presente artículo se afrontan estos desafíos para indagar cómo se abordaban y qué lenguaje se utilizaba en las

\* Facultad de Información y Comunicación (FIC), Universidad de la República, Montevideo, Uruguay, correo electrónico: diana.comesana@fic.edu.uy

\*\* Centro de Investigación en Computación, Instituto Politécnico Nacional (IPN), Ciudad de México, correo electrónico: lmvilches@cic.ipn.mx

noticias relacionadas con eventos meteorológicos producidos en la sociedad latinoamericana entre los siglos XIX-XX. Para ello, se seleccionaron periódicos históricos digitalizados, accesibles vía Web, procedentes de las hemerotecas nacionales de Colombia, Ecuador, México y Uruguay; sobre los que se realiza un abordaje de lingüística de *corpus* mediante un proceso de bibliominería, a través de lecturas técnicas y la aplicación de tecnologías (QDA-MinerLite y R), que permite recuperar la terminología de los eventos mencionados en la prensa de los diferentes países considerados.

*Palabras clave: prensa histórica digitalizada, evento meteorológico, recuperación de información, bibliominería, terminología.*

### **A study of the Latin American newspapers from XIX-XX centuries with a focus on meteorological events**

#### **ABSTRACT**

One of the most useful developments for modern historians is the digitization of the newspaper archives, since these ones are appreciated information sources between scholars that want to understand issues associated with politics, culture, and society. This fact is promoting an increasing digitalization of this kind of archives, causing the appearance of “new” challenges, due to the fact lack of text recognition technologies application to digitalized historical newspapers. It makes difficult application of technologies that allow an automatic processing and retrieving events related to daily life, such as meteorological events.

Taking into account this scenario, this paper addresses these challenges for discovering how terminology was used in news related to meteorological events performed in the Latin-American society from the XIX-XX centuries. For that, we gather digitalized historical newspapers, which are available from websites, from National newspaper and periodicals libraries associated with Colombia, Ecuador, México and Uruguay. Once these newspapers have been collected, we carry out a corpus linguistics through a bibliomining process, performing technical readings and the application of technologies (QDA-MinerLite and R), which allows retrieving the terminology about meteorological events related to newspapers from different considered countries.

*Key words: Digitalized historical newspapers, meteorological event, information retrieval, bibliomining, terminology.*

## INTRODUCCIÓN

La mayoría de los historiadores ven el mundo digital como un lugar conveniente para una navegación rápida y eficiente en las ricas fuentes de información disponibles y no como un entorno vital para el análisis histórico. En este sentido, los historiadores que exploran seriamente las nuevas formas de presentación interactiva del conocimiento histórico son todavía escasos (Wijfjes, 2017).

En el contexto actual, incluso para el observador más casual de la profesión histórica, resulta evidente que las prácticas de investigación en esta área se están transformando gradualmente mediante la digitalización de los archivos y las fuentes primarias. Sin embargo, las *Digital Humanities* (Nicholson, 2003) todavía se presentan como un campo experimental de investigación académica y educativa, haciendo conexiones de metodologías tradicionales de humanidades (por ejemplo, hermenéutica histórica) con herramientas que los investigadores pueden utilizar para seleccionar o acceder a colecciones en línea y analizar conjuntos de datos (Wijfjes, 2017). En este sentido, la mayoría de los países no tienen aún una infraestructura accesible y funcional para la investigación de las humanidades digitales.

Aun así, a pesar de los problemas y obstáculos técnicos y metodológicos, las humanidades digitales tienen grandes posibilidades para nuevas investigaciones que se presentan en los territorios como “globales, transhistóricas y transmediales” (Eijnatten, Pieters y Verheul, 2013), por lo que uno de los desarrollos más útiles para los historiadores modernos es la digitalización de los archivos de periódicos (Bingham, 2010).

Los periódicos se encuentran entre las fuentes más valiosas para los estudiosos interesados en investigar la opinión pública y cómo se ha configurado con el transcurso del tiempo. No obstante, entre algunos historiadores hubo cierto escepticismo sobre la utilización de periódicos antiguos como fuente de información, debido a su dudosa precisión y naturaleza efímera (Bingham, 2010), pero, en la actualidad, pocos niegan el valor de su contenido para comprender cuestiones relacionadas con la política, cultura y sociedad; para llegar a conocer cómo los desarrollos e ideas se percibieron y extendieron por diferentes países se revisará lo expuesto por Bingham, 2010; Neudecker y Antonacopoulos, 2016; así como para descubrir detalles de la vida cotidiana, no recogidos en otros documentos.

Por todo ello, los periódicos (digitalizados) facilitan un nuevo punto de acceso a la información que se almacena en periódicos históricos (Smits, 2014), proporcionando a los historiadores acceso a una amplia gama de fuentes de información, lo que deriva en un creciente interés desde las ciencias históricas (Zaagsma, 2013) y en la posibilidad de sacar a la luz una infinidad de *datos menores* que se encuentran ocultos en ellos.

Con esta creciente digitalización aparecen “nuevos” desafíos técnicos, ya que resulta frecuente que a los periódicos históricos digitalizados no se les haya aplicado tecnologías de reconocimiento de texto (Neudecker y Antonacopoulos, 2016). Esto dificulta la aplicación de tecnologías que contribuyan a la extracción de eventos en formas narrativas o a la reconstrucción y análisis de patrones (Wijfjes, 2017) que aporten mayor claridad al contexto histórico recogido en los periódicos.

Considerando este escenario, resulta un desafío acceder a esos datos *ocultos* en los periódicos que pueden proporcionar valiosa información de la vida cotidiana, entre la que se encuentra la relativa a eventos meteorológicos. Entre este tipo de eventos los más comunes son la lluvia o el viento, pero existen otros que sólo se producen en ciertas épocas (por ejemplo, la nieve) o que son más probables en determinadas zonas geográficas. La dimensión e intensidad de estos eventos puede afectar, en mayor o menor medida, a la población o, incluso, poner de manifiesto diversidad de efectos y vulnerabilidades.

En este artículo se afrontan estos desafíos para indagar cómo se abordaban los eventos meteorológicos y sus efectos en la sociedad latinoamericana en periódicos históricos comprendidos entre los siglos XIX-XX. Los eventos meteorológicos producidos en este intervalo de tiempo no fueron recogidos por organizaciones meteorológicas, debido a que se produjeron antes de la conformación de dichas organizaciones,<sup>1</sup> por este motivo los periódicos de la época resultan una de las principales fuentes para conocer lo sucedido.

Para ello, se seleccionan periódicos históricos digitalizados disponibles vía Web, del periodo mencionado, procedentes de las hemerotecas nacionales de Colombia, Ecuador, México y Uruguay. La selección de este lapso se justifica por la consolidación de la prensa escrita (inicios del siglo XIX) y por la mencionada ausencia de registros climatológicos oficiales (primera mitad del siglo XX) en los países que conforman este estudio. Sobre estos periódicos se realiza un abordaje de lingüística de *corpus* (Bolaños, 2015), es decir, un tipo de análisis lingüístico que se basa en herramientas tecnológicas para “reunir, organizar y procesar” datos reales del lenguaje relacionado con agentes/eventos meteorológicos. De esta manera, se trae hacia el presente los hechos que se analizan (Gallegos, 2011) y se recuperan los elementos que conforman la narrativa de estos sucesos desde una perspectiva ajena a los especialistas del área.

<sup>1</sup> En 1951 se fundó la Organización Meteorológica Mundial, por tanto, en este trabajo se asume que, a partir de este momento, resultaría factible hallar datos cuantificados y análisis de los fenómenos meteorológicos.

Esto se lleva a cabo mediante un proceso de bibliominería (Nicholson y Staton, 2003), que se centra en la realización de lecturas técnicas y en la aplicación de tecnologías para la extracción manual y semi-automática de información relacionada con eventos meteorológicos presentes en la prensa digitalizada de las hemerotecas nacionales consideradas. Este proceso permite recuperar y analizar la terminología utilizada para describir estos eventos en el *corpus* de noticias conformado.

Este artículo se organiza de la siguiente manera: la sección 2 recoge los trabajos relacionados presentes en el estado del arte. La parte 3 caracteriza la prensa hispanoamericana entre los siglos XIX y XX. Las diferentes hemerotecas nacionales que conforman este trabajo, así como sus fondos y las principales características asociadas se presentan en la sección 4. La quinta recoge el análisis realizado sobre los periódicos históricos digitalizados que se consideran en este trabajo. Finalmente, se presentan las conclusiones.

## REVISIÓN DE LITERATURA

Existen diferentes propuestas para el análisis y tratamiento de periódicos antiguos que han sido descritas en la literatura. Entre estas propuestas destacan las relacionadas con la construcción de corpus y sus tratamientos. Así, en (Buntinx, Bornet y Kaplan, 2017) se presenta una metodología para analizar los cambios en un *corpus* conformado por cuatro millones de artículos pertenecientes a 200 años de archivos de dos periódicos suizos. En este trabajo los autores realizan una serie de medidas sobre el *corpus* para proporcionar indicadores y visualizaciones de la evolución lingüística. Cortés (2015) presenta una propuesta de lingüística de *corpus* en diferentes secciones de dos periódicos costarricenses. Para ello utiliza las herramientas de análisis cuantitativo AntConc y ATLAS.ti, que le permiten aplicar una serie de técnicas sobre estos periódicos. En Dakić y Trtovac (2014), los autores detectan y procesan analíticamente más de 2 000 artículos de periódicos antiguos serbios, centrando su trabajo en artículos relacionados con la vida cotidiana, que son recuperados a través de los metadatos y las anotaciones que tienen asociados estos periódicos. Además, los autores realizan un trabajo de verificación y evaluación de los resultados de OCR en los periódicos digitalizados.

Otros trabajos han tratado periódicos antiguos (Ahonen y Hyvönen, 2009) y la creación de archivos digitales de periódicos (Castells, Perdrix, Pulido, Rico, Benjamins, Contreras y Lorés, 2004) mediante la utilización de tecnologías de la Web Semántica.

Por otro lado, la literatura también presenta diferentes propuestas relacionadas con el tratamiento y análisis de eventos meteorológicos a partir de diferentes fuentes de estudio. Así, en Añel, Sáenz, Ramírez-González, Polychroniadou, Vidal-Mina, Gimeno y de la Torre (2017) se recopilan notas meteorológicas que aparecen en el periódico español *La integridad*, recogiendo datos de 5 000 periódicos de donde recuperan más de 60 000 medidas individuales de un periodo que abarca desde finales del siglo XIX a principios del XX. Este trabajo de recuperación de diversas variables meteorológicas permite una visión general del tiempo en una determinada zona de estudio.

En Deschamps, Otero y Tonni (2003), se analizan las precipitaciones en el periodo tardo-colonial y comienzos del siglo XIX. Los autores abordan este análisis a través de datos instrumentales continuos de temperatura y precipitación a partir de 1860 para diferentes localidades en Argentina y Uruguay. Este trabajo utiliza diversos acuerdos del Cabildo de Buenos Aires, mapas y crónicas de misioneros jesuitas, así como crónicas y cartas geográficas de varios exploradores como fuentes documentales.

En esta línea de fuentes documentales se enmarcan los trabajos de Barriendos (2005), Pavón (2006), Boyer (1975) y Chávez (1953). En Barriendos (2005) se presenta una caracterización de la variabilidad climática de Cataluña (España) durante los siglos XVIII-XIX, utilizando fuentes agrarias, libros de memorias y crónicas como fuentes documentales. Pavón (2006) presenta un estudio del clima de Colombia durante los siglos XVI-XIX a partir de material histórico. Boyer (1975) se centra en estudios sobre inundaciones coloniales en la Ciudad de México y Chávez (1953) realiza una serie de documentos sobre crisis agrícolas novohispanas.

Adicionalmente, en el trabajo de García (2008) se produce la conexión entre análisis terminológico y el dominio meteorológico, ya que la autora se centra en la descripción y análisis del vocabulario científico que aparece en *Manual de Meteorología popular* de Gumersindo Vicuña (Madrid, 1880), así como en el tratamiento que este léxico recibe en el *Diccionario de la Academia*.

## SÍNTESIS DE LA PRENSA HISPANOAMERICANA

La llegada de la imprenta a América fue consecuencia directa del proceso de conquista y colonización española. El virreinato de Nueva España fue el primer territorio en el continente que contó con una imprenta, la cual fue concesionada por la Corona al impresor alemán Juan Cromberg y al oficial italiano Juan Pablos en 1539 (Fernández, 2010).

Con la imprenta en tierras novohispanas surgieron las hojas volantes, publicaciones que eran papeles sueltos de carácter informativo que en Europa tienen una tradición que arranca desde el siglo XV y que empezaron a aparecer en la Nueva España dos años después de la fundación de la imprenta. Estos impresos recibían indistintamente los nombres de *relaciones*, *nuevas*, *noticias*, *sucesos* o *traslados* y todos los historiadores del periodismo están de acuerdo en considerarlos como germen del periodismo, aunque carezcan de periodicidad (Ruiz, 1995).

Posteriormente, se conforman las gacetas periodísticas, y México se convierte en el país precursor en la región con la *Gaceta de México y Noticias de Nueva España* de 1722. En el contexto de los países que conforman este trabajo, estas gacetas aparecen, conforme a su orden cronológico, en Colombia con la *Gaceta de Santafé* en 1785, en Ecuador ve la luz *Primicias de la Cultura de Quito* en 1772, y en Uruguay, bajo la dominación inglesa, se publica el primer periódico bilingüe, *The Southern Star* o *La Estrella del Sur*, en 1807.

A partir de la segunda mitad del siglo XIX se produce un quiebre en la concepción del periodismo, ya que se abandona la prensa de adoctrinamiento que caracterizaba a América Latina, producto del nacimiento del mercado informativo y de los géneros periodísticos modernos. De esta manera, los nuevos periodistas se ocupan de describir y problematizar los procesos como momentos claves que determinaron la evolución de la prensa latinoamericana. En consecuencia, se puede afirmar que el periodismo nace en este siglo y que los periodistas adquieren durante este tiempo su categoría como profesionales.

Para el estudio de la vida cotidiana y cultural del siglo XIX en América Latina, la prensa escrita se constituye en la fuente por excelencia, no en vano, expresa Emilio Castelar en 1896, citado por Bertrand (1960): “Cuando tomo en mis manos un periódico, cuando recorro sus columnas, cuando considero la diversidad de sus materias y la riqueza de sus noticias, no puedo menos de sentir un raptó de orgullo por mi siglo...” (Castelar, 1896). Y es que en este siglo la prensa escrita se reviste de poder, tomando su lugar en la formación cultural y social de cada uno de los países. En palabras de Mejías y Arias (1998) “El siglo XIX es, por excelencia, el siglo de la prensa escrita”.

Muchas personalidades de la cultura y de diferentes ámbitos científicos y profesionales consideraban un prestigio colaborar en publicaciones periódicas, y es que la prensa hispanoamericana se comprometió, totalmente, con su misión educadora y formadora de opinión. De esta manera, la prensa se constituyó en el escaparate de difusión de proyectos e ideas que llevaron a la conformación de las naciones latinoamericanas, contribuyendo a la formación o destrucción de la imagen de sus líderes.

La prensa se convierte en un medio masivo de comunicación y, poco a poco, aparecen las primeras industrias del mundo de la prensa como Bennet y Pulitzer, transformando a los periódicos en instrumentos de influencia y poder. Esta evolución también conlleva cambios en su estructura, producto de la aparición de secciones bien definidas asociadas a diferentes temáticas. Así, por ejemplo, el periódico uruguayo *El Bien Público*, ilustrado en la Figura 1, a partir de 1930, presenta una sección destinada al informe del tiempo, generalmente ubicada antes de las páginas destinadas a propaganda comercial y poesía. Esto no es impedimento para que, en caso de grandes eventos meteorológicos se registren en primera plana o encabecen las noticias nacionales.



**Figura 1.** Imagen del periódico *El Bien Público* (Uruguay), (14 de febrero de 1936)  
Fuente: Biblioteca Nacional de Uruguay.

Con el transcurrir del siglo XX la prensa compite con otros formatos de divulgación, tales como: la radio, televisión y la aparición de las nuevas tecnologías. Esto provoca que los periódicos se adapten a esta nueva realidad y pasen a editarse tanto en formato impreso como digital.

En lo concerniente a las noticias relacionadas con eventos meteorológicos, éstas no tuvieron cabida en los revueltos tiempos de la prensa doctrinaria, pero el surgimiento de la prensa informativa cambia esta situación, lo que conlleva la progresiva aparición de noticias sobre eventos que, de una forma u otra, afectan la vida y bienes de las personas. Los informes meteorológicos, lejos de las características actuales, comienzan a aparecer en la prensa escrita con escasas alusiones alrededor del año 1900, con menciones del tipo “buen tiempo” o “paró de llover”. Paulatinamente, se comienza a

precisar la información sobre eventos meteorológicos utilizando términos más exactos y vinculados con el dominio, por ejemplo: “tiempo semi-nuboso” o “vientos del NE de suaves a moderados”, consolidándose la presencia de estas noticias o secciones en los periódicos.

### **LAS HEMEROTECAS NACIONALES, SUS FONDOS Y CARACTERÍSTICAS ASOCIADAS**

Las hemerotecas nacionales, con frecuencia adscritas a las bibliotecas nacionales, tienen como objetivo custodiar, organizar, preservar y difundir su acervo, conformado por publicaciones periódicas nacionales y, en ocasiones, extranjeras. Bajo esta misión, los fondos de las hemerotecas nacionales de Colombia, Ecuador, México y Uruguay presentan diferencias y similitudes en cuanto a sus estados de conservación, digitalización y acceso público de los fondos del periodo estudiado. A continuación, se presentan los principales fondos de estas hemerotecas, así como sus características y problemas asociados.

#### ***FONDOS DE LAS HEMEROTECAS NACIONALES***

La Biblioteca Nacional de Colombia fue fundada el 9 de enero de 1777. En lo que respecta a sus fondos hemerográficos, a partir de 1985 esta Biblioteca apuesta por la iniciativa de la microfilmación sistemática de sus fondos, iniciativa que fue suspendida en 1990. En 2004 se produce una verdadera apuesta por la conservación digital de estos fondos, ya que surge un programa para la digitalización de colecciones de las publicaciones periódicas. Este programa ha hecho posible que parte de los fondos existentes estén digitalizados, disponibles en formato PDF, en la página web de la Biblioteca Nacional, donde se puede acceder a periódicos como *El Pabellón Granadino* (1852), *La Voz del Tolima* (1852-1897), *El Constitucional* (1868-1876), *La Concordia* (1899), entre otros.

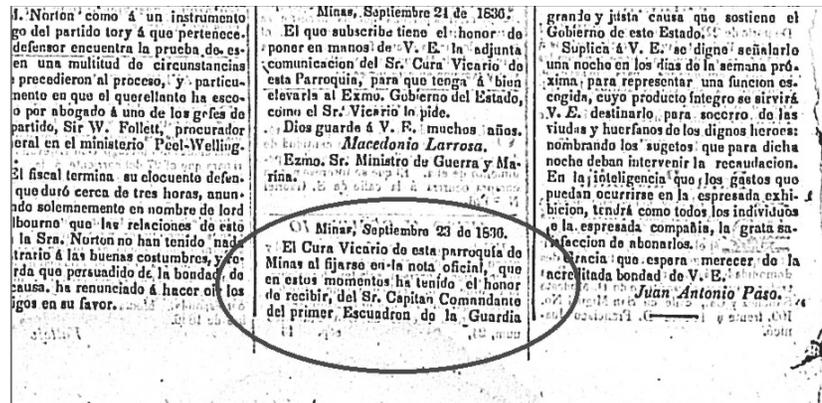
El origen de la Biblioteca Nacional de Ecuador (Biblioteca Nacional “Eugenio Espejo”) se remonta a 1585. Sus fondos hemerográficos están conformados por la colección de periódicos y revistas producidas en el país correspondiente a los siglos XIX y XX hasta 1940. En la actualidad, producto del trabajo de esta biblioteca para rescatar, resguardar y poner en valor el patrimonio bibliográfico y cultural ecuatoriano, se puede acceder a varios títulos digitalizados de prensa de este país, tales como: *El Comercio* (1886), *Los Andes* (1889), *La Nación* (1892-1893), *El Telégrafo* (1901), *La Patria* (1902) o *El Tiempo* (1907).

La Hemeroteca Nacional de México fue fundada el 28 de marzo de 1944 y entre sus fondos hemerográficos se encuentra la colección *Publicaciones periódicas mexicanas 1722-1917*; compuesta por la *Gaceta de México* (1722), *El Monitor Republicano* (1844-1896) y *El Siglo Diez y Nueve* (1841-1846), periódicos de gran trascendencia en el contexto mexicano. Complementan estos fondos el *Diario de los Niños* (1839-1840), *El Álbum Mexicano*, *Semanario de las Señoritas Mexicanas* y *El Mundo Ilustrado*. Es destacable el hecho de que ninguno de estos títulos se encuentre disponible en formato digital en Internet. No obstante, existen fondos hemerográficos disponibles a través del Servicio de Prensa Mexicana (SERPREMEX), que pone a disposición 383 mil registros hemerográficos con resumen de contenido y palabras clave sobre temas relativos a la reforma y transformación del Estado Mexicano.

La Biblioteca Nacional de Uruguay, fundada el 26 de marzo de 1816, cuenta entre sus fondos con el primer periódico uruguayo, *The Southern Star* o *La Estrella del Sur*, publicado en 1807. En cuanto a los fondos digitalizados, esta hemeroteca se nutre de los resultados del Proyecto “Anáforas”, encargado de la digitalización y publicación de los fondos hemerográficos en el portal de la Biblioteca Nacional. Entre los fondos disponibles se encuentran cuatro colecciones: Biblioteca de Autores Uruguayos, Figuras, Publicaciones Periódicas del Uruguay y Sobre la prensa. La colección de Publicaciones Periódicas del Uruguay incluye los primeros impresos periódicos, diarios, otros impresos y semanarios, abarcando desde los años 1810 a 1962. Entre ellos se pueden encontrar títulos como: *El Defensor de las Leyes* (1836-1838), *La Prensa Oriental* (1859; 1861-1862), *Il Propagatore Italiano* (1864), *A Patria* (1880-1888), *The Express* (1888-1890) o *El Bien* (1900-1914). Una característica que diferencia los fondos hemerográficos de esta biblioteca de las anteriores es la existencia de prensa en español, inglés, portugués, italiano y francés, producto de las diferentes dominaciones (británica y luso-brasileña) a las que estuvo sometido y a la presencia de una fuerte colonia italiana y francesa.

#### **CARACTERÍSTICAS Y PROBLEMAS DE LOS FONDOS HEMEROGRÁFICOS**

Los periódicos de las hemerotecas nacionales consideradas presentan una carencia de estructura hasta la segunda mitad del siglo XIX. Por tanto, en estos periódicos se detectan ausencia de secciones, noticias separadas por un guión o incluso, noticias sin título. Un ejemplo de estos diferentes problemas se muestra en la Figura 2.



**Figura 2.** Imagen del periódico *Defensor de las Leyes* (Uruguay, 30 de septiembre de 1836).  
Fuente: Biblioteca Nacional de Uruguay.

Con el transcurso del tiempo, los periódicos del siglo XIX adquieren estructura, esto supuso la aparición de secciones estables que se distinguían conforme al género de los lectores. Así, las noticias de estas nuevas secciones que contenían temas políticos o culturales eran dirigidas a la población masculina, mientras las secciones que recogían noticias sobre moda eran diseñadas para la población femenina.

A pesar de la aparición de estas secciones, sigue resultando difícil identificar datos de la vida cotidiana, como en aquella época podrían considerarse los fenómenos meteorológicos, ya que éstos se asumen como acontecimientos inevitables. Sin embargo, se detectan algunos ejemplos, como el del semanario *El Uruguay* de 1905 (véase Figura 3). Por un lado, nos muestra a su editor como a una personalidad de relevancia en la cultura y la política uruguaya, como lo fue el escritor Javier de Viana y, por otro, se puede encontrar una noticia referida a la producción vitivinícola, en sus páginas 11 y 12. Esta noticia brinda información sobre un revés económico sufrido en el litoral uruguayo y, de forma indirecta, proporciona información sobre eventos meteorológicos poco comunes durante el mes de octubre de 1904, asociados con tormentas que llevaron a la pérdida total de la producción de los viñedos del litoral del país.

Por otro lado, como se mencionaba con anterioridad, la digitalización de los fondos hemerográficos está consiguiendo la posibilidad de un compromiso mucho mayor y más sofisticado para el tratamiento del contenido del periódico. Sin embargo, estas fuentes de información no están exentas de problemas para su tratamiento desde un punto de vista tecnológico, confor-

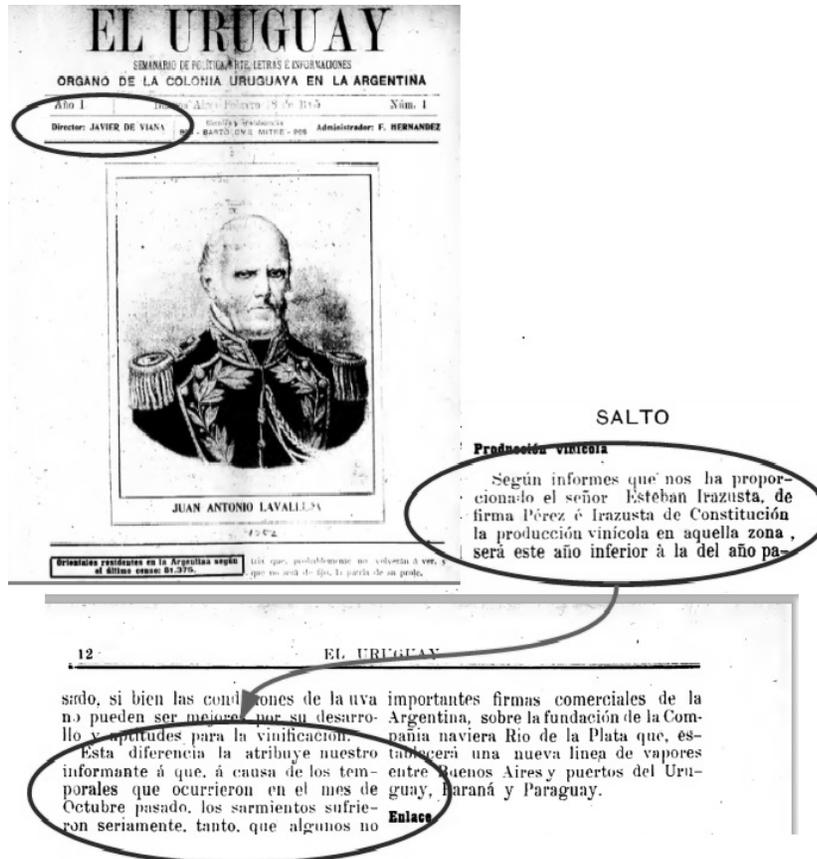


Figura 3. *Semanario El Uruguay* (18 de febrero de 1905).

Fuente: Biblioteca Nacional de Uruguay.

me al movimiento de las Humanidades 3.0 (Bod, 2013). Entre los problemas detectados en los periódicos presentes en los fondos hemerográficos, provenientes de las mencionadas hemerotecas nacionales, se encuentran los siguientes:

- La falta de continuidad de la prensa analizada se debe a las características socio-políticas-económicas que presentaban los países que forman parte de este trabajo durante el periodo de tiempo considerado.
- Las colecciones de las hemerotecas nacionales suelen estar incompletas y en su mayoría presentan importantes problemas de conservación física.

- c) El estado de deterioro de muchos de los originales condiciona y dificulta la calidad de su digitalización.
- d) La ausencia de recursos humanos, tecnológicos o de tiempo deriva en que gran número de periódicos se digitalizaron como imágenes.
- e) La utilización de un software de reconocimiento de óptico de caracteres (OCR, por sus siglas en inglés) resulta escasa y, en los casos donde se aplicó, los resultados presentan “ruido”, es decir, efectos en la imagen digital provocado por manchas reales o grafías imperfectas, o bien por defectos técnicos en la adquisición de la imagen.

### **ANÁLISIS DE INFORMACIÓN**

Desde las primeras décadas del siglo XX, el esfuerzo en el análisis de los fondos hemerográficos se ha centrado en una doble vía. Por un lado, en rescatar artículos relevantes para la cultura o la historia y, por otro, en la descripción de publicaciones periódicas, enumerando su contenido temático principal y los escritores o personajes relevantes de la época que habían plasmado su opinión en ellas. Este foco, como se mencionaba con anterioridad, ha dejado escapar infinidad de datos, referidos a la vida cotidiana, opacados por la importancia de las noticias y artículos relativos a la formación política y social del sentir nacional en los diferentes países de Latinoamérica. Sin embargo, en este trabajo se utilizan los fondos hemerográficos de los países mencionados para indagar en los elementos de la narrativa utilizada para describir los eventos meteorológicos y sus efectos en la sociedad en el periodo que abarca entre los siglos XIX-XX. A continuación, se recogen los principales detalles del proceso seguido.

### **MUESTRA DE PERIÓDICOS CONSIDERADOS**

De los fondos considerados en las diferentes hemerotecas nacionales tratadas, se analiza más de 22 500 documentos digitalizados con el objetivo de identificar noticias relacionadas con eventos meteorológicos en los diferentes países considerados. Este número de documentos se reparte de la siguiente manera: en Colombia se estudiaron 3 867 documentos, en Ecuador se analizaron 1 665 periódicos, en el caso de México se revisaron 54 documentos, y en Uruguay se analizaron más de 5 500 documentos digitalizados (véase Tabla 1). El número de documentos con los que se trabaja está condicionado por la existencia de periódicos digitalizados, por su publicación y disponibilidad en Internet, y por abordar temáticas susceptibles de estar relacionadas con este trabajo.

**Tabla 1**

<i>País</i>	<i>Núm. de documentos</i>	<i>Núm. de noticias</i>	<i>Años</i>
Colombia	3 867	1 805	1813-2000
Ecuador	1 665	191	1860-1922
México	54	16	1907-2000
Uruguay	5 500	725	1810-1956

Fuente: Elaboración propia.



**Figura 4.** Distribución de noticias. Elaboración propia

En este primer análisis del conjunto de los fondos hemerográficos de los diferentes países considerados se identifica que el contenido del 52% de los fondos analizados corresponden a temas culturales, un 28% a temas políticos y un 20% a temas generales, grupo dentro del que se encuentran las noticias (sobre eventos meteorológicos) consideradas en este trabajo, como se muestra en la Figura 4.

Sobre los fondos hemerográficos considerados se lleva a cabo un análisis manual de las noticias para conocer las características que presentan. En este sentido, se encontró que, de las 2.775 noticias analizadas, más del 60% tienen una mala digitalización, consecuencia, como se mencionaba con anterior-

ridad, de la deficiente conservación y del soporte (papel) y la tinta utilizada. Además, se comprueba que un volumen importante de estos periódicos no presenta digitalización a través de un software de reconocimiento óptico de caracteres.

### **SELECCIÓN Y RECOPIACIÓN DE INFORMACIÓN**

Para la selección del *corpus* de noticias sobre eventos meteorológicos del conjunto de fondos hemerográficos considerado resulta necesario un lenguaje o vocabulario primario que permite la mejor identificación de las noticias pertinentes para este trabajo. Para obtener este vocabulario, se realiza un análisis de dominio, basado en el paradigma de Birger Hjørland (Hjørland, 1995). Este paradigma plantea estudiar los dominios del conocimiento como comunidades discursivas, tomando en consideración el contexto psico-social y socio-lingüístico de la sociología del conocimiento y de la ciencia. Este análisis de dominio se realiza a partir de entrevistas a expertos en meteorología y se complementa con el estudio de documentos sugeridos por estos expertos, tales como: el glosario de la Organización Meteorológica Mundial y los diferentes glosarios de los institutos meteorológicos nacionales de cada uno de los países intervinientes. Del conocimiento obtenido se conforma un lenguaje primigenio para la recuperación manual y semiautomática de las noticias objeto de estudio.

Este conjunto de elementos nos permite conformar la terminología, compuesta por los términos recopilados en la prensa considerada y los técnicos o provenientes de expertos, utilizada para la recuperación de las noticias de interés. Así, con esta terminología, la utilización del software de análisis de datos cualitativos QDA-MinerLite y la realización de lecturas técnicas se consigue identificar más de 2 500 resultados positivos, es decir, noticias relacionadas sobre eventos meteorológicos contenidas en los fondos hemerográficos considerados y que pasan a conformar el *corpus* de noticias para proceder a su análisis en detalle.

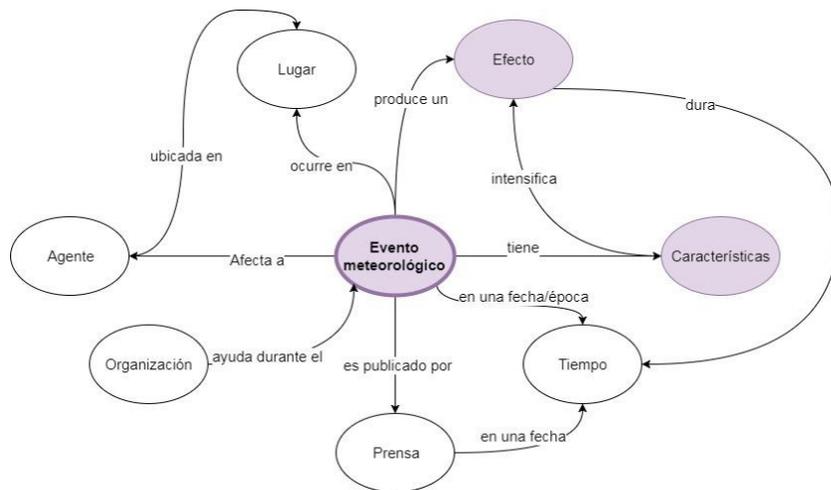
Como se muestra en la Tabla 1, este *corpus* está conformado por 1 805 noticias de Colombia, 191 noticias de Ecuador, 54 noticias extraídas en México y 725 noticias localizadas en los fondos uruguayos.

### **MODELADO DEL CONTENIDO DE LAS NOTICIAS**

En la diversidad de noticias sobre eventos meteorológicos que conforman el *corpus* de este trabajo se identifica una serie de elementos comunes dentro del contenido de este tipo de noticias. A continuación, se muestra un ejemplo de cómo se relata una noticia sobre eventos meteorológicos:

Gran congestión por lluvia. Un fenomenal trancón interrumpió el tráfico automotor por más de dos horas anoche en el norte de Bogotá. La crítica situación comenzó a las cinco de la tarde y a las ocho de la noche aún continuaba y paralizó por completo el sector comprendido entre la carrera 13 entre las calles 57 al Monumento de Los Héroes. Todo el amplio sector estaba totalmente taponado por centenares de automotores que debieron permanecer estacionados por más de dos horas hasta cuando se presentaron patrulleros del Tránsito y agentes de la Policía que poco a poco ordenaron el flujo vehicular.

La identificación de elementos frecuentes permite realizar una abstracción de la noticia y convertirla en un modelo que resulte una base para la recuperación de información de las noticias consideradas. El modelo generado, producto de los elementos que componen la descripción usual de una noticia que recoge información sobre un evento meteorológico, sigue la estructura que se muestra en la Figura 5.



**Figura 5.** Abstracción del contenido de una noticia sobre eventos meteorológicos. Elaboración propia

Un evento meteorológico tiene características como velocidad, grados, cantidad de mililitros, etcétera. Dicho evento ocurre en un lugar geográfico, en un momento (tiempo) determinado y éste puede afectar a un agente (un grupo de personas, edificaciones, autos, etcétera). Además, de acuerdo con las características y efectos asociados al mismo, pueden intervenir organizaciones como bomberos, policía, entre otros. Adicionalmente, la prensa registra los hechos y los relata en una edición determinada.

La conformación de este modelo del contenido de una noticia nos permite dirigir el proceso de recuperación de información (manual, a través de lecturas técnicas, y semi-automática, con la aplicación de QDA-MinerLite) sobre las noticias que recogen información de eventos meteorológicos. Ambos elementos (lecturas técnicas y QDA-MinerLite) permiten recabar información para la descripción temática (y catalogación) de las noticias recogidas. La Figura 6 muestra un ejemplo de las fichas construidas, basadas en el modelo presentado en la Figura 5, para la recuperación de información relevante sobre las diferentes noticias que contienen relación con los eventos meteorológicos.

A	B	C	D	E	F	G	H	I	J	K
Nombre periódico	N°	Fecha	Evento	efecto detectado		URI				
El Siglo	10	1863-02-13	Temporal duro = Temporal	10/02/1863. Fray Bentos Barco-correo p...		<a href="http://bibliotecadigital.bibna.gub.uy/8080/psu/handle/123456">http://bibliotecadigital.bibna.gub.uy/8080/psu/handle/123456</a>				
El Siglo	31	1863-03-12	Temporal = Viento duro	Montevideo, viento y oleaje. daños en el...		<a href="http://bibliotecadigital.bibna.gub.uy/8080/psu/handle/123456">http://bibliotecadigital.bibna.gub.uy/8080/psu/handle/123456</a>				
El Siglo	33	1863-03-14	Inundación	60 días lloviendo copiosa y constante...		<a href="http://bibliotecadigital.bibna.gub.uy/8080/psu/handle/123456">http://bibliotecadigital.bibna.gub.uy/8080/psu/handle/123456</a>				
El Siglo	51	1863-04-08	Temporal = Viento duro	daños en embarcaciones (embarcadas)		<a href="http://bibliotecadigital.bibna.gub.uy/8080/psu/handle/123456">http://bibliotecadigital.bibna.gub.uy/8080/psu/handle/123456</a>				
El Siglo	59	1863-04-16	Sequía	Estragos en Bahía, Brasil		<a href="http://bibliotecadigital.bibna.gub.uy/8080/psu/handle/123456">http://bibliotecadigital.bibna.gub.uy/8080/psu/handle/123456</a>				
El Siglo	111	1863-06-23	Tormenta	Argentina. 13-an-1863. Cabañas, Sple		<a href="http://bibliotecadigital.bibna.gub.uy/8080/psu/handle/123456">http://bibliotecadigital.bibna.gub.uy/8080/psu/handle/123456</a>				
El Siglo	133	1863-07-22	Sequía	Provincia de Buenos Aires, muerte de...		<a href="http://bibliotecadigital.bibna.gub.uy/8080/psu/handle/123456">http://bibliotecadigital.bibna.gub.uy/8080/psu/handle/123456</a>				

B	C	D	E	F	G	H	I	J
N°	Fecha	Evento	coordenadas		efecto detectado			
			latitud	longitud	veloc. Viento	creciente	temperatura	
10	1863-02-13	Temporal duro = Temporal	33°08'00" S	58°18'00" O	89 a 102 Km/hr			10/02/1863. Fray Bentos Barc
31	1863-03-12	Temporal = Viento duro	34°52'01" S	56°10'00" O	62 a 74 km/hr			Montevideo, viento y oleaje. d
33	1863-03-14	Inundación	Provincias Argentinas de Sata, Santiago y Tucumán					60 días lloviendo copiosa y
51	1863-04-08	Temporal = Viento duro	34°52'01" S	56°10'00" O	62 a 74 km/hr			daños en embarcaciones (e
59	1863-04-16	Sequía	12°58'15" S	38°38'39" O				Estragos en Bahía, B
111	1863-06-23	Tormenta	35°02'00" S	58°44'00" O				Argentina. 13-an-1863. Car
133	1863-07-22	Sequía	Sur Argentino					Provincia de Buenos Aires,

Figura 6. Fichas de recopilación de información. Elaboración propia

Asimismo, adoptando las recomendaciones de los expertos y guiados por las diferentes escalas consideradas (Escala Anemométrica de Beaufort y Escala de Douglas), se procede a cuantificar los efectos ocasionados por los distintos eventos meteorológicos, producto de los datos recopilados de forma manual y semi-automática.

Otro aspecto que merece ser destacado producto del proceso de recuperación de información fue la identificación de los lugares donde se produjeron dichos eventos meteorológicos. Considerando la época tratada en este trabajo (siglo XIX-XX), se identifican numerosos topónimos (nombres geográficos) que no existen en la actualidad, cuya presencia resulta un descubrimiento relevante en el contexto de este trabajo, ya que éstos recogen un poco de tradición e historia del país, brindan información sobre la cultura y el modo de vida en general, de quienes nos han precedido (Graur, 1972). Por tanto, considerando que la toponimia se ocupa del estudio del origen, significado y la evolución de los nombres de los lugares en un idioma determina-

do, y también de los nombres geográficos, resulta interesante no sólo para geógrafos e historiadores, sino también para lingüistas y filólogos (Poenaru-Girigan, 2013).

La evolución identificada de los nombres geográficos lleva a consultar diccionarios de nombres geográficos y nomenclatores para proceder a su identificación e, incluso, a una posible asociación con un nombre geográfico actual. Sin embargo, se descubre que no todos los países tienen este tipo de recursos disponibles, por lo que se dificulta llevar a cabo, de manera exhaustiva, la realización de esta labor.

A modo de ejemplo, en el caso de Uruguay, el periódico *The Express*, recoge en su edición del 9 de agosto de 1888, una noticia donde se describe una inundación, acaecida por copiosas lluvias en las quintas de Paso del Molino, la calle Agraciada y la Playa de la Aguada, entre otras zonas. Entre los topónimos extintos se encuentra la Playa de la Aguada, la cual puede verse en la Figura 7 (imagen capturada en el año 1860). Esta playa y su correspondiente topónimo dejaron de existir con la construcción de las instalaciones del puerto de Montevideo, inaugurado el 4 de septiembre de 1909.



**Figura 7.** Bahía de Montevideo vista desde la Aguada. Al fondo a la derecha, el cerro.

Fuente: Intendencia de Montevideo, Centro de Fotografía.



Ante la ausencia de disponibilidad del nomenclátor uruguayo, se recurre a fuentes alternativas para la identificación de este tipo de topónimos y la recopilación de su información asociada. Entre las fuentes consultadas están el *Diccionario Geográfico Argentino* (Latzina, 1891) y el *Diccionario Geográfico del Uruguay* (Araujo, 1900).

### **BIBLIOMINERÍA**

La bibliominería se refiere al uso de técnicas relacionadas con la estadística, inteligencia artificial, aprendizaje automático, etcétera, que permiten sondear las enormes cantidades de datos generados por los fondos digitalizados de las bibliotecas (Nicholson, 2003). Con estas técnicas, en el contexto de este trabajo, se efectúan estudios en los que se combinan el lenguaje natural y la minería aplicada a textos para realizar un abordaje de lingüística de corpus (Bolaños, 2015) que permita descubrir, de forma automática, tokens, frecuencias de palabras, entre otros datos, presentes en las noticias de los fondos digitalizados considerados. En la Figura 8 se recoge el conjunto de actividades llevadas a cabo sobre el corpus de noticias conformado.

Así, en primer lugar, utilizando el software libre R, se procede a la tokenización del corpus de noticias que contienen información sobre eventos meteorológicos. La tokenización (Manning, Raghavan y Schütze, 2008) consiste en realizar un proceso automático para separar las palabras. Esto permite que al tokenizar el texto recogido en una noticia, por ejemplo:

“Gran congestión por lluvia. Un fenomenal trancón interrumpió el tráfico automotor”

Se obtenga la siguiente división de palabras producto de su tokenización:

Gran congestión por lluvia. Un fenomenal trancón interrumpió el tráfico automotor

Tras la realización de la tokenización al *corpus* de noticias, se procede a construir una matriz de términos, utilizando el software R, que permite identificar la frecuencia de repetición de cada palabra en el *corpus* mencionado. Con esta frecuencia se genera un *ranking* de términos que se pueden encontrar en el *corpus* de noticias consideradas. Un ejemplo de este *ranking* de palabras frecuentes se muestra en la Figura 9. En el ranking realizado sobre el conjunto del *corpus* de estudio destacan por su frecuencia palabras como: “personas”, “calle”, “ciudad”, “aguas”, “emergencia”, “bomberos”, “tiempo”, “inundaciones”, “deslizamientos”, “casas”, etc.

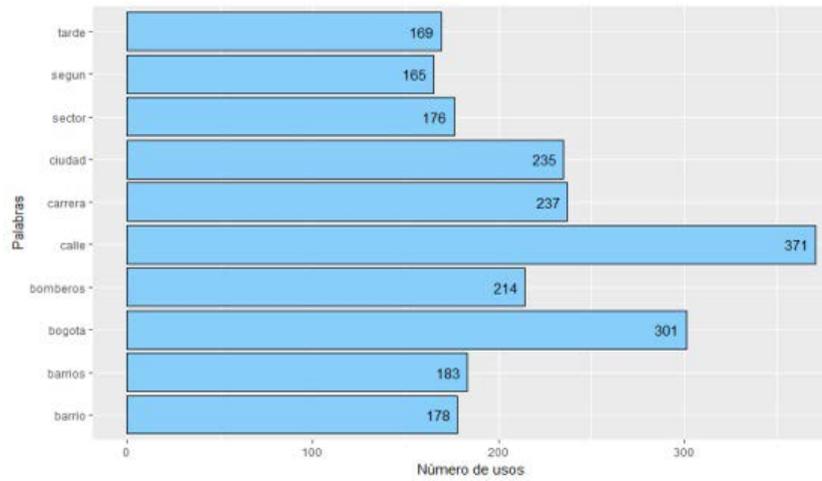


Figura 9. Ejemplo de frecuencia de palabras. Elaboración propia.



Figura 10. Nube de palabras del corpus. Elaboración propia

El proceso de bibliominería concluye con la generación de nubes de palabras, las cuales permiten construir una visualización de las palabras, utilizando el software R, que aparecen en el corpus de noticias reflejando su frecuencia (número de veces que aparecen repetidas en las diferentes noticias consideradas) conforme al tamaño que presentan (a mayor tamaño más veces se encuentran en las noticias y viceversa). Por tanto, la nube de palabras permite tener una idea general de los términos que aparecen en las noticias que conforman el corpus, como se puede apreciar en el ejemplo recogido en la Figura 10.

### CONSTRUCCIÓN TERMINOLÓGICA

El proceso de bibliominería origina un glosario de términos sobre eventos meteorológicos que incluye las palabras y las definiciones dadas en cada uno de los países considerados en este trabajo. De esta forma, se palía el impacto sobre los productos intermedios anteriores, donde el desbalance del número de periódicos considerados en los diferentes países afecta a las visualizaciones y rankings generados.

La Tabla 2 recoge un fragmento del glosario construido en el desarrollo de este trabajo. Este glosario está compuesto por 333 términos divididos en 240 conceptos (por ejemplo: “granizo”, “lluvia”, “evento”, etc.) y 93 instancias (por ejemplo: “Bogotá”, “30 Km/h”, “5 minutos”, etc.) producto del vocabulario extraído del corpus de noticias relacionadas con fenómenos meteorológicos y de las fuentes que presentan vocabulario técnico sobre el dominio de conocimiento en cuestión.

**Tabla 2**  
**Glosario de términos**

<i>Término preferente</i>	<i>Sinónimo</i>	<i>Descripción</i>
Brisa fresca	Fresquito	Olas medianas y alargadas, borreguillos muy abundantes. En tierra hay pequeños movimientos en los árboles y se ondula la superficie de los lagos
Brisa moderada	Bonancible	El mar muestra borreguillos numerosos y olas cada vez más largas. En tierra se levanta polvo y papeles, se agitan las copas de los árboles
Depósito	—	Depósito de partículas acuosas en las superficies
Depresión tropical	—	Perturbación tropical con vientos máximos sostenidos de superficie alcanzando, pero no sobrepasando, los 61km/h (33 nudos). Tiene una o más isobaras cerradas

Continuación Tabla 2

<i>Término preferente</i>	<i>Sinónimo</i>	<i>Descripción</i>
Evento meteorológico	Fenómeno meteorológico	Los fenómenos meteorológicos incluyen todos los acontecimientos que se pueden producir ligados a la dinámica de la atmósfera
Precipitación continua	—	Se dice que la precipitación es continua, cuando su intensidad aumenta o disminuye gradualmente
Presión atmosférica	—	Es la presión que ejerce la atmósfera en un punto específico como consecuencia de la acción de la fuerza de gravedad sobre la columna de aire que se encuentra encima de este punto

**Fuente:** Elaboración propia.

**Tabla 3**  
**Sinónimos por país**

<i>Término preferente*</i>	<i>Escala de Beaufort</i>	<i>Colombia</i>	<i>Ecuador</i>	<i>México</i>	<i>Uruguay</i>
Atardecer					Puesta de sol /atardecer
Brisa débil	Flojo				Flojo
Brisa fresca	Fresquito				Fresquito
Brisa fuerte	Fresco				Fresco
Brisa moderada	Bonancible				Brisa moderada
Brisa muy débil	Flojito				Flojito
Calima				Bruma/ calima	Bruma/ neblina
Canícula				Canícula	Veranillo
Cellisca					Agua nieve
Chubasco				Chubasco	Aguacero/ chubasco
Cumulonimbos					Nube de tormenta

Continuación Tabla 3

<i>Término preferente*</i>	<i>Escala de Beaufort</i>	<i>Colombia</i>	<i>Ecuador</i>	<i>México</i>	<i>Uruguay</i>
Despejado					<i>Few</i>
Escarcha				Cencellada blanca/escarcha	Escarcha
<i>Fohen</i>					<i>Föhn/Fohen</i>
Frente			Zona frontal		Frente
Galerma	Viento duro	Mar de leva			Tempora/mar de leva
<i>Graupel</i>		Granizo		Granizo	Granizo menudo
Helada		Helada		Cencellada transparente /Hielo liso / Helada	Helada
Helada dura					Helada negra
Huracán	Temporal huracanado	Huracán		Huracán	Temporal huracanado
Nublado					Cielo nuboso/nublado
Ola de calor				Onda de calor/ Ola de calor	Ola de calor
Onda tropical			Onda tropical	Onda del Este	
Precipitación ligera			Precipitación ligera		Día con precipitación
<i>Scattered</i>					Nubes dispersas
Sistema de alta presión					Anticiclón/ Alta presión

Continuación Tabla 3

<i>Término preferente*</i>	<i>Escala de Beaufort</i>	<i>Colombia</i>	<i>Ecuador</i>	<i>México</i>	<i>Uruguay</i>
Sistema de baja presión		Depresión tropical	Zona de baja presión	Baja presión /Ciclón/ Depresión tropical/ Perturbación tropical	Ciclón/Baja presión
Temporal fuerte	Viento muy duro	Marejada de tempestad			Temporal fuerte
Temporal muy duro	Borrasca			Borrasca	Borrasca
Vendaval		Vendaval		Ventarrón	Vendaval
Viento fuerte	Frescachón				

\* Término proveniente de la Organización Mundial de Meteorología. Elaboración propia.

También se han detectado, en un menor número, algunos casos de polisemia entre la terminología resultante de este trabajo. El detalle de los casos identificados se recoge en la Tabla 4.

Si se observan estos casos (Tabla 4), la definición de “Alerta” que se emplea en Colombia, resulta un caso particular a la definición que recoge Uruguay. Asimismo, la definición de “Despejado” que se emplea en Uruguay resulta un caso particular al compararla con la utilizada en Colombia, Ecuador y México.

Además, se identifican términos aparentemente provenientes del lenguaje coloquial como, por ejemplo; “fresquito”, “frescachón”, “flojo”, “flojito”, “ventolina”. Estos términos provienen de la escala anemométrica de Beaufort, base del lenguaje de la disciplina, y son presentados como sinónimos en la misma. Esta escala, creada en 1830 por Sir Francis Beaufort para la Marina Británica, fue adaptada al uso no naval en la década de 1850 y continúa en uso en la actualidad, expresando en un lenguaje sencillo, los efectos del viento detectables en tierra y mar.

Finalmente, se identifica que el lenguaje recuperado que conforma este glosario aún resulta en uso; esto puede ser debido a que este lenguaje parte

de observaciones empíricas, resulta especializado, incluso, aunque no sea manejado por especialistas, y se utiliza en la narrativa periodística, propiciando la continuidad del mismo. Este hecho justifica la ausencia de variaciones temporales en la terminología que compone este glosario, a lo que se une que ésta se conforma en países con antecedentes comunes (colonización, lengua y, en algunos casos, proximidad geográfica, entre otros aspectos), lo que permite sustentar la evidencia de la escasa variación de los términos en los diferentes países que conforman este estudio.

**Tabla 4**  
**Polisemia**

<i>Término preferente</i>	<i>Colombia</i>	<i>Uruguay</i>
ALERTA	Advierte a los sistemas de prevención sobre el riesgo de inundaciones	Advierte a los sistemas de prevención y atención de desastres sobre la amenaza que puede ocasionar un fenómeno con efectos adversos para la población y que requiere atención inmediata por parte de la población y de organismos de prevención. Se emite una alerta cuando un evento indica probabilidad de amenaza inminente
DESPEJADO	Las nubes cubren un máximo de 2/8 partes del cielo	Cielo totalmente sin nubes

**Fuente:** Elaboración propia.

## CONCLUSIONES

Es un hecho que los periódicos antiguos se encuentran entre las fuentes más valiosas para los estudiosos que están interesados en investigar diversas cuestiones relacionadas con la política, cultura y sociedad, y que éstos proporcionan a los historiadores acceso a una amplia gama de fuentes de información. Sin embargo, como ha quedado evidenciado en la revisión de los fondos hemerográficos de los países objeto de estudio (Colombia, Ecuador, México y Uruguay), son diversos los problemas que presentan este tipo de periódicos al ser digitalizados.

Partiendo de estos problemas, en este artículo se crea un corpus de noticias sobre eventos meteorológicos, conformado por más de 2 500 noticias, que abarcan un periodo de tiempo comprendido entre los siglos XIX-XX. Sobre este *corpus* se identificó una serie de problemas comunes, asociados con elementos propios de las fuentes y tecnológicos. Entre los primeros aparecen la falta de continuidad, completitud de los fondos, problemas de conservación física y de ausencia de estructuración. Por otro lado, en cuanto a los problemas tecnológicos se comprueba que un gran número de documentos digitalizados presentan baja calidad y no se ha utilizado tecnología de reconocimiento de caracteres (OCR), lo que genera excesivo “ruido” en la obtención automática de datos.

Sobre este *corpus* se realiza un abordaje de lingüística de corpus que permite recopilar agentes/eventos meteorológicos a través de un proceso de bibliominería, realizado con lecturas técnicas y la aplicación de diversas técnicas para el tratamiento automático de la información. Como resultado de ello se obtuvieron rankings con las frecuencias de palabras que aparecen en las noticias, visualizaciones de estos rankings mediante nubes de palabras y un glosario de términos que recoge la terminología con la que se abordaban las noticias sobre eventos meteorológicos en el periodo de tiempo tratado.

El proceso de bibliominería llevó a comprobar que la terminología utilizada en el área de las noticias sobre eventos meteorológicos no presenta grandes diferencias ni al considerar factores geográficos ni temporales. Asimismo, el análisis realizado sobre el corpus permitió detectar una evolución en los nombres geográficos, por lo que fue necesario consultar diccionarios de nombres geográficos y nomenclátors que recogen nombres geográficos históricos en los diferentes países. No obstante, se identificó que existen importantes carencias en cuanto a la disponibilidad de este tipo de recursos (diccionarios y nomenclátors) entre los países considerados en este trabajo.

### *AGRADECIMIENTOS*

Este trabajo fue desarrollado en el marco del proyecto “Compartiendo la Historia Escondida del cambio climático en Latinoamérica a través de las TIC”, auspiciado por el Instituto Panamericano de Geografía e Historia (IPGH), enmarcado en el contexto de los Proyectos Panamericanos de Asistencia Técnica (PAT), 2017. Además, agradecemos el trabajo realizado por la estudiante de maestría Lorena de Jesús Arrieta Moreno de la Pontificia Universidad Javeriana, Bogotá, Colombia.

## BIBLIOGRAFÍA

- Ahonen, E. y Hyvönen, E., "Publishing historical texts on the Semantic Web - A case study", IEEE International Conference on Semantic Computing, 2009.
- Añel, J.A.; Sáenz, G.; Ramírez-González, I.A.; Polychroniadou, E.; Vidal-Mina, R.; Gimeno, L. y de la Torre, L., "Obtaining meteorological data from historical newspapers: La Integridad", *Weather*, no. 72, 2017, pp. 366-371. DOI:10.1002/wea.2841.
- Araujo, O., *Diccionario Geográfico del Uruguay*, Imprenta Artística de Dornaleche y Reyes, Montevideo, Uruguay, 1900.
- Barité, M., *Método de la Grilla. Terminología*, Universidad de la República, Montevideo, Uruguay, 2013.
- Barriados Vallvé, M., "Variabilidad climática y riesgos climáticos en perspectiva histórica. El caso de Catalunya en los siglos XVIII-XIX", *Revista de historia moderna*, núm. 23, 2005, pp. 11-34.
- Bertrand, O., *Oratoria, política y oradores del ochocientos*, Instituto de Humanidades, Universidad del Sur, Buenos Aires, 1960, p. 31.
- Biblioteca Nacional de Colombia, *Depósito legal. Normas*, <[http://camlibro.com.co/wp-content/uploads/PDF/deposito\\_legal.pdf](http://camlibro.com.co/wp-content/uploads/PDF/deposito_legal.pdf)>, consultado el 9 de junio de 2018.
- Biblioteca Nacional de Colombia. Ver Biblioteca Nacional de Colombia, Hemeroteca, <[http://catalogoenlinea.bibliotecanacional.gov.co/client/es\\_ES/bd/?rm=PRENSA+COLOMBIO%7C%7C%7C1%7C%7C%7C3%7C%7C%7Ctrue](http://catalogoenlinea.bibliotecanacional.gov.co/client/es_ES/bd/?rm=PRENSA+COLOMBIO%7C%7C%7C1%7C%7C%7C3%7C%7C%7Ctrue)>, consultado el 9 de junio de 2018.
- Biblioteca Nacional de Colombia. Ver Biblioteca Nacional de Colombia. ¿Quiénes somos? Historia, <<http://bibliotecanacional.gov.co/es-co/Footer/bibliotecanacional-de-colombia/quienes-somos/historia>>, consultado el 9 de junio de 2018.
- Biblioteca Nacional de Ecuador. Ver Biblioteca Nacional de Maestros. Historia y Bibliotecas, Biblioteca Nacional de Ecuador, <<http://www.bnm.me.gov.ar/novedades/?p=16872>>, consultado el 9 de junio de 2018.
- Biblioteca Nacional de Uruguay. Ver. Biblioteca Nacional, Quiénes somos, <<http://www.bibna.gub.uy/>>, consultado el 9 de junio de 2018.
- Biblioteca Nacional Eugenio Espejo. Ver. Biblioteca Nacional Eugenio Espejo, Acerca de la Biblioteca Nacional <[http://www.casadelacultura.gob.ec/index.php?ar\\_id=4&ge\\_id=7](http://www.casadelacultura.gob.ec/index.php?ar_id=4&ge_id=7)>, consultado el 9 de junio de 2018.
- Bingham, A., "The Digitization of Newspaper Archives: Opportunities and Challenges for Historians", *Twentieth Century British History*, vol. 21, núm. 2, 2010, pp. 225-231, 226.

- Bod, R., "Who's Afraid of Patterns? The Particular versus the Universal and the Meaning of Humanities 3.0". *BMGN-Low Countries Historical Review*, vol. 128, núm. 4, 2013, pp. 171-180.
- Bolaños, S., "La lingüística de corpus: perspectivas para la investigación lingüística contemporánea", *Forma y función*, vol. 28, núm. 1, pp. 31-54. DOI: 10.15446/fyf.v28n1.51970. 2015.
- Boyer, R.E., "La gran inundación. Vida y sociedad en México (1629-1638)", *SEP-Setentas* 218, SEP, México, 1975.
- Buntinx, V., Bornet, C., Kaplan, F., "Studying Linguistic Changes over 200 Years of Newspapers through Resilient Words Analysis", *Frontiers in Digital Humanities*, vol. 4, no. 2. DOI: 10.3389/fdigh.2017.00002. 2017.
- Castelar, E., "Las Noticias", VII Jornadas E.U.B.D., Universidad Complutense de Madrid, 1896.
- Castells, P.; Perdrix, F.; Pulido, E.; Rico, M.; Benjamins, R.; Contreras, J. y Lorés, J., "Neptuno: Semantic Web Technologies for a Digital Newspaper Archive", *The Semantic Web: Research and Applications. LNCS:3053*, 2004.
- Centro de Fotografía. Véase Centro de fotografías de Montevideo, Catálogo de fotografías, <<http://www.cdf.montevideo.gub.uy/catalogo/foto/0164fmhb>>, consultado el 9 de junio de 2018.
- Chávez Orozco, L., "La crisis agrícola novohispana de 1784-1785", Publicaciones del Banco Nacional de Crédito Agrícola y Ganadero, México, 1953.
- Cortés Kandler, M., "Diferencias en el léxico entre los periódicos *La Nación* y *Diario Extra* desde la lingüística de corpus: aplicación de una metodología", *Revista Káñina*, vol. XXXIX, núm. 1, pp. 201-212. ISSN: 0378-0473. 2015.
- Dakić, N. y Trtovac, A., "Historical Newspapers Content as a Base for Scientific Research- Europeana Newspapers Project", IFLA 2014, 13-14, August, Geneva, Switzerland, 2014.
- Deschamps, J.R.; Otero, O. y Tonni, E.P., "Cambio climático en la pampa bonaerense: las precipitaciones desde los siglos XVIII al XX". Documento de Trabajo núm. 109, Universidad de Belgrano, 2003 <<http://repositorio.ub.edu.ar/handle/123456789/1580>>, consultado 1 de octubre de 2018.
- Eijnatten, J. van; Pieters, T. y Verheul, J., "Big Data for Global History: The Transformative Promise of Digital Humanities", *BMGN-Low Countries Historical Review*, vol. 128, núm. 4, pp. 55-77, 2013. DOI: <http://doi.org/10.18352/bmgn-lchr.9350>.
- Escala Anemométrica de Beaufort. Ventanas al Universo. Escala de Beaufort para Estimar la velocidad del viento, <[https://www.windows2universe.org/earth/Atmosphere/tornado/beaufort\\_scale.html&edu=high&lang=sp](https://www.windows2universe.org/earth/Atmosphere/tornado/beaufort_scale.html&edu=high&lang=sp)>, consultado el 9 de junio de 2018.
- Escala de Douglas, Diccionario Náutico, <<https://diccionario-nautico.com.ar/escala-de-douglas/>>, consultado el 9 de junio de 2018.

- Fernández Fernández, I., "Un recorrido por la historia de la prensa en México. De sus orígenes al año 1857", *Documentación de las Ciencias de la Información*, vol. 33, pp. 69-89.
- Gallegos, C., "Abordaje metodológico de prensa escrita: el semanario Cuba Libre", *Latinoamérica*, núm. 53, 2011, <[http://www.scielo.org.mx/scielo.php?script=sci\\_arttext&pid=S1665-85742011000200006](http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1665-85742011000200006)>, consultado el 9 de junio de 2018.
- García Aranda, M.A., "El léxico de especialidad en el siglo XIX: el manual de meteorología popular de Gumersindo Vicuña (Madrid, 1880)", *ELUA*, vol. 22, núm. 91-110, 2008.
- Graur, A. *Nume de locuri [Place Names]*, Editura Științifică, Bucarest. 1972.
- Hemeroteca Nacional de México. Véase Hemeroteca Nacional de México, <<http://www.hnm.unam.mx/>>, consultado el 9 de junio de 2018.
- Hjorland, B., "Toward a new horizon in information science: Domain-analysis", *Journal of the Association for Information Science and Technology*, vol. 46, núm. 6, pp. 400-425, 1995.
- Latzina, F., *Diccionario Geográfico Argentino*, Espasa Calpe, Buenos Aires, 1891, <<http://cdi.mecon.gov.ar/bases/librosantiguos/diccionario%20geografico%20argentino.pdf>>, consultado el 9 de junio de 2018.
- Manning, C.D.; Raghavan, P. y Schütze, H., *Introduction to Information Retrieval*, Cambridge University Press, 2008.
- Mejías, A. y Arias, A., "La prensa del siglo XIX como medio de difusión de la literatura hispanoamericana", *Revista General de Información y Documentación*, vol. 8, núm. 2, 1998, pp. 241-257.
- Neudecker, C. y Antonacopoulos, A., *Making Europe's Historical Newspapers Searchable*, 12th IAPR Workshop on Document Analysis Systems (DAS), pp. 405-410. DOI: 10.1109/DAS.2016.83, 2016.
- Nicholson, S., "The Bibliomining Process: Data Warehousing and Data Mining for Library Decision-Making", *Information Technology and Libraries*, vol. 22, núm. 4, 2003.
- Nicholson, S. y Stanton, J., "Gaining strategic advantage through bibliomining: data mining for management decisions in corporate, special, digital, and traditional libraries", in Nemati, H. & Barko, C. (eds.), *Organizational data mining: Leveraging enterprise data resources for optimal performance*, pp. 247-262, 2003.
- Organización Meteorológica Mundial - WMO. Véase METEOTERM. WMO Official terminology database.
- Pavón Caicedo, J.D., "El clima de Colombia durante los siglos XVI-XIX a partir de material histórico. Parte I: inventario de fuentes de información", *Cuadernos Geográficos Revista Colombiana de Geografía*, núm. 15, pp. 75-92, 2006.
- Poenaru-Girigan, O.M., *The Relationship between Toponymy and Linguistics*, BDD-A17728. Editura Universității din Suceava, 2013,

- <<http://www.diacronia.ro/ro/indexing/details/A17728/pdf>>, consultado el 9 de junio de 2018.
- Proyecto Anáforas, UDELAR, Facultad de Información y Comunicación. Véase UDELAR, FIC, <<http://anaforas.fic.edu.uy/jspui/>>, consultado el 9 de junio de 2018.
- QDA-MinerLite. Véase PROVALIS RESEARCH, <<https://provalisresearch.com/products/qualitative-data-analysis-software/freeware/>>, consultado el 24 de septiembre de 2018.
- R. Ver. The R Project for Statistical Computing, <<https://www.r-project.org/>>, consultado el 24 de septiembre de 2018.
- Ruiz Castañeda, M.C., “Capítulo III. Periodismo colonial. Las Hojas Volantes (1541-1700)”, en Ruiz Castañeda, María del Carmen y Luis Reed Torres, *El periodismo en México: 500 años de historia*, 2a. edición, México, EDAMEX/Lotería Nacional para la Asistencia Pública, 1995.
- SERPEMEX, UNAM. Véase Servicio de Prensa Mexicana. <<http://serpemex.hnm.unam.mx/>>, consultado el 9 de junio de 2018.
- Smits, T., “Problems and possibilities of digital newspaper and periodical archives”, *Tijdschrift voor Tijdschriftstudies*, núm. 36, pp. 139-146, 2014. DOI: 10.18352/ts.317.
- Wijffes, H. “Digital Humanities and historical newspaper research”, *Tijdschrift voor Mediageschiedenis*, vol. 20, núm. 1, pp. 4-24, 2017, <<http://hdl.handle.net/11245.1/d02ef470-22b1-4b7a-bc0a-de73e9cacb6c>>, consultado el 9 de junio de 2018.
- Zaagsma, G., “On Digital History”, *BMGN-Low Countries Historical Review*, vol. 128, núm. 4, pp. 3-29, 2013.