

## ¿Cómo leemos en la era digital?

Romina De León

Nuestro proceso cognitivo de lectura es sumamente complejo y nos habilita a decodificar símbolos que nos permiten inferir y/o construir significados. Es por ello, que cuando leemos se produce una compleja interacción con el texto, sumada a conocimientos y experiencias previas que favorecen la comprensión. A lo que se añade, las formas de lecturas y sus muy variadas formas, pero solo mencionaremos la que realizamos en los ámbitos académicos superiores, que incluye citas, notas al pie, etc.

En el contexto de las Humanidades Digitales (HD), la introducción de los textos en soporte digital ha ampliado aún más las maneras de leer. La aparición del hipertexto, desde hace más de 50 años, ha profundizado las formas no secuenciales de lectura, enriqueciendo los textos con referencias cruzadas, hipervínculos, imágenes, vídeos y diversas fuentes conectadas por medio de enlaces asociativos y redes sociales<sup>1</sup>. Este cambio en la forma de acceder y consumir información digital ha impactado significativamente en cómo entendemos y analizamos textos en el entorno digital dentro del campo de las HD. Asimismo, cabe destacar que las complementariedades hipertextuales no son exclusivas de las HD, sino que es parte del modo en que la era digital ha permeado en nuestras rutinas diarias, incluyendo la forma en que consumimos y procesamos la información, como hemos visto en la unidad N° 1.

Por lo cual, en el contexto de este módulo y sus áreas afines, podemos distinguir entre **big data** y datos inteligentes (**smart data**), estos últimos definidos por Schöch (2013)<sup>2</sup>. Como vimos en la unidad anterior, los datos masivos, o big data, son aquellos

<sup>1</sup> El término hipertexto fue utilizado por primera vez en 1965 por Ted Nelson, puede acceder a más información desde: <http://www.hyperfiction.org/texts/whatHypertextIs.pdf>. Asimismo, la idea original fue desarrollada por Vannevar Bush, en 1945, en su artículo "As we may think", donde describe el dispositivo MEMEX. Explicando que "un individuo almacena sus libros, anotaciones, registros y comunicaciones, y esta colección de información es mecanizada de forma que puede consultarse con alta velocidad y mucha flexibilidad". Es decir que, característica esencial de MEMEX es su habilidad de "atar" o asociar dos ítems. Se puede acceder a dicho trabajo desde: <https://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/>.

<sup>2</sup> Puede accederse a información sobre el tema en Schöch, C. (2013). Big? Smart? Clean? Messy? Data in the Humanities Journal of Digital Humanities. *Journal of Digital Humanities*, 2(3).

no estructurados, crudos, de diversos formatos y en grandes volúmenes, que requieren organizarse para su análisis. Por otro lado, los datos inteligentes o **smart data**, son semiestructurados, enriquecidos, “limpios” y se presentan en cantidades más pequeñas, puesto que facilitan la intervención humana en su análisis y comprensión.

También hemos reconocido que los datos masivos, por sí solos, aunque poseen valor, es mediante el análisis crítico que pueden generar beneficios y reflexiones significativas. Por ello, Espinosa Menéndez (2022) enfatiza la necesidad de realizar ese tipo de estudios de los datos digitales, que implica formular preguntas sobre su naturaleza y cómo se construyen. Este tipo de enfoque se alinea con lo discutido en la última clase sincrónica, donde se resaltó la importancia de cuestionar y comprender la procedencia y calidad de los datos, así como el valor de estos.

En ese tenor, la misma autora destaca que los datos masivos como fuentes valiosas para investigaciones y para la creación de nuevos artefactos, deben contar con la intervención de la tecnología para desarrollar metodologías y técnicas para aprovechar plenamente su valor, y, además, entender, transformar y reutilizarlos de manera efectiva.

Con lo anterior, podemos entender que con el paso del tiempo se fueron construyendo y evolucionando las prácticas de lectura y eso incluyó la adaptación a nuevos dispositivos. Sumado a ello, el crecimiento exponencial de la cantidad de información y la variación de los soportes de lectura, como dispositivos de producción, almacenamiento y circulación textual provocaron que las maneras de leer fueran totalmente disímiles al de la imprenta, y con la intervención de la tecnología, podemos diferenciar dos tipos de **nuevas** lecturas, la distante y la cercana.

## Lectura cercana y lectura distante

Hemos trabajado en la unidad anterior, lo importante de la mirada crítica al analizar datos, sobre todo cuando se producen de manera cuantiosa. Sin embargo, la mirada crítica humana puede estar limitada, porque es imposible poder leer todo lo que se

---

<http://journalofdigitalhumanities.org/2-3/big-smart-clean-messy-data-in-the-humanities/>. Así como en: Menéndez, M. J. E. (2022, marzo 23). ¿Qué son los datos digitales para las Humanidades? [Billet]. *Amontonamos las palabras: Blog de la Biblioteca de El Colegio de México*. <https://doi.org/10.58079/lwfl>.

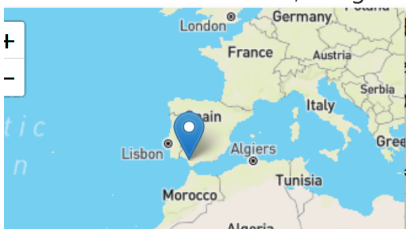
produce o produjo sobre un tema de interés. En ese punto, Franco Moretti (2004a), explica que:

un canon de doscientas novelas [...] parece muy grande para la Gran Bretaña del siglo XIX (y es mucho mayor que el actual), pero sigue siendo menos del 1[%] de las novelas que de hecho se publicaron [número que nadie] sabe realmente; y la lectura [*cercana*] no nos va a ayudar aquí, pues a novela por día, todos los días del año, necesitaríamos aproximadamente un siglo” (pp. 60-61).

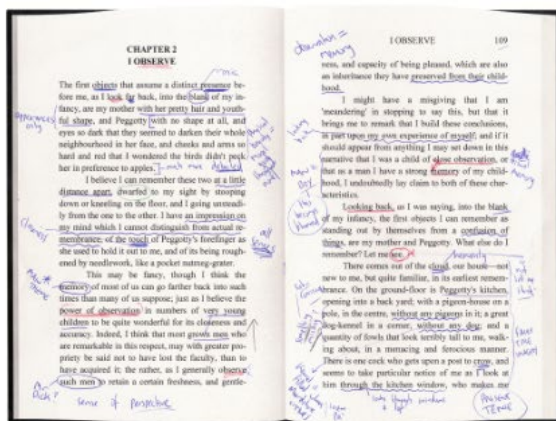
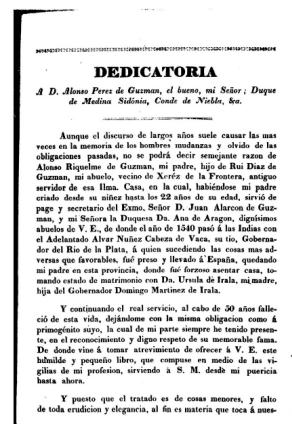
Por lo cual, con lectura cercana entendemos, nuestra práctica de análisis que prescinde, como regla general, de cualquier tipo de información contextual, y se concentra exclusivamente en el texto (Moretti, 2004a). A continuación, podemos observar ejemplos de lectura cercana enriquecida y convertida en hipertexto.

A don Alonso Pérez de Guzmán, el bueno, mi Señor; Duque de Medina Sidonia, Conde de Niebla, etcétera

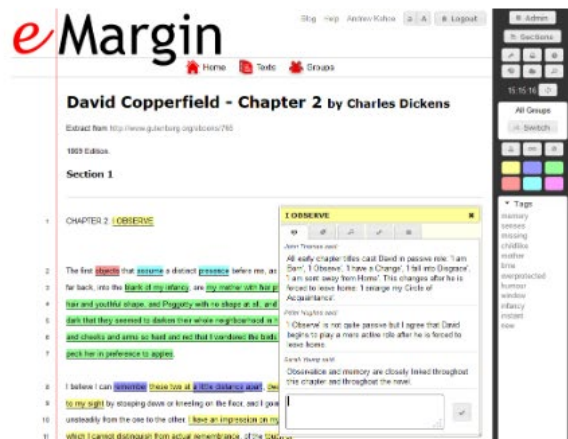
Aunque el discurso de largos años suele causar las más veces en la memoria de los hombres mudanzas y olvido de las obligaciones pasadas, no se podrá decir semejante razón de Alonso Riquelme de Guzmán, mi padre, hijo de Rui Díaz de Guzmán, mi abuelo, vecino de Jerez de la Frontera, antiguo servidor de esa ilustrísima casa, en la cual, desde su niñez hasta los 22 años de su edad, el Excelentísimo señor don Juan Alarcón de la Puquesa doña Ana de Aragón, dignísimos de donde el año de 1540 pasó a las Indias por la cabeza de Vaca, su tío, gobernador del Perú, sucediendo las cosas más adversas que



Fuente: [https://hdlab.space/La-Argentina-Manuscrita/argentina\\_manuscrita\\_pages/pg\\_0001](https://hdlab.space/La-Argentina-Manuscrita/argentina_manuscrita_pages/pg_0001)



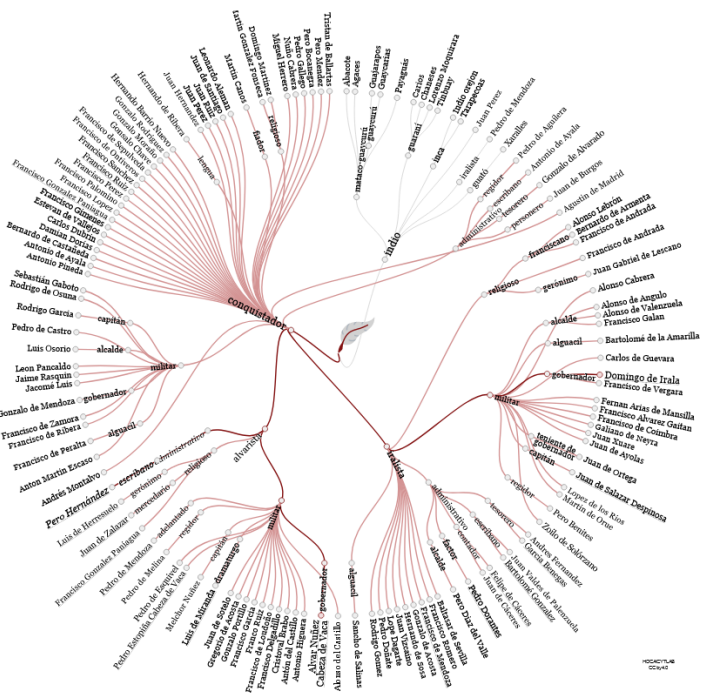
(a) Traditional close reading.



(b) Digital close reading with eMargin.

Figure 1: Examples of close reading of the second chapter of Charles Dickens' *David Copperfield* (Figures reproduced with permission from Kehoe et al. [KG13]).

Fuente: Jänicke, S., Franzini, G., Cheema, M. F., & Scheuermann, G. (2015).

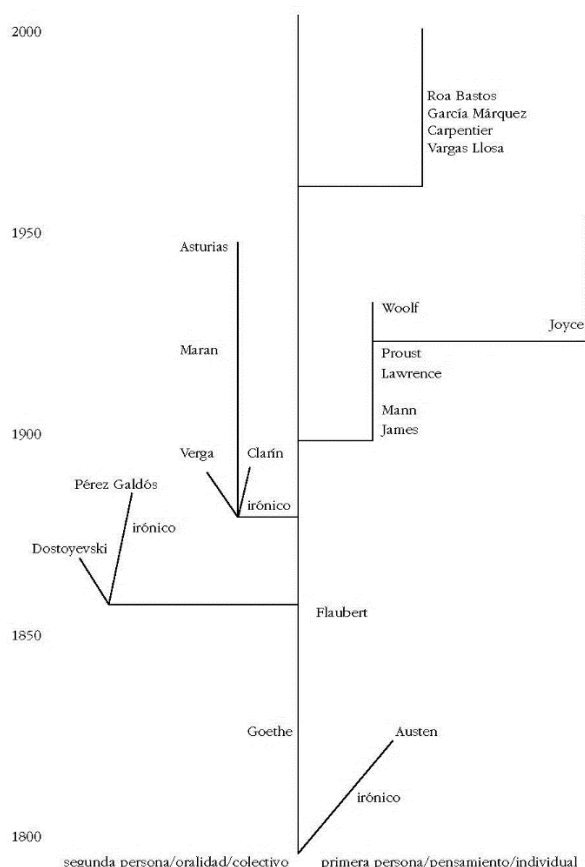
[illegible]

Con las propuestas anteriores, se comprende que mediante la **lectura distante**, los pequeños detalles no son tenidos en cuenta, sino que se busca delimitar ciertas características o ciertos patrones que se cotejarán a través del procesamiento



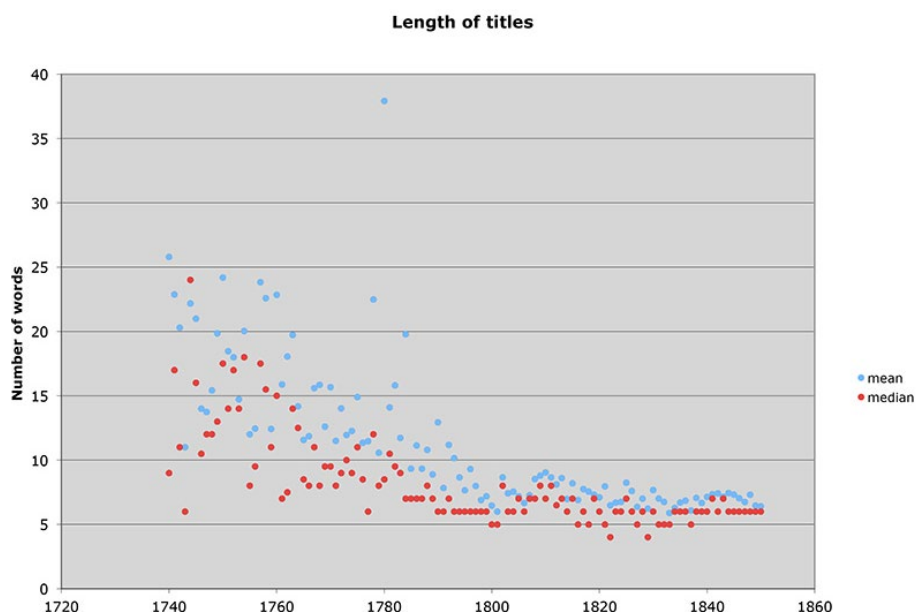
computacional. Otro enfoque analítico realizado por el crítico literario italiano fue sobre las narrativas contemporáneas desde 1800 al 2000 haciendo foco en el estilo indirecto libre (Moretti, 2004c, pp. 30-32).

FIGURA 6. *Estilo indirecto libre en la narrativa contemporánea, 1800-2000*



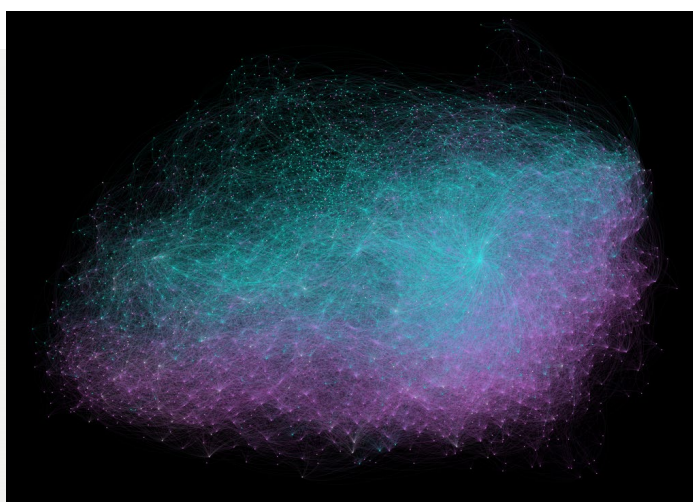
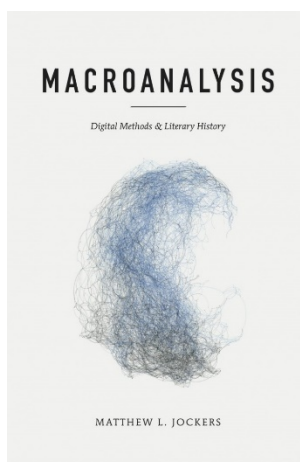
Fuente: Moretti, 2004c, p. 31. “Esta figura refleja la obra en progreso, y, por consiguiente, es bastante tentativa, especialmente en el caso de las literaturas no europeas, y en el del lapso temporal diacrónico de las diversas ramas”.

En esta representación se pueden considerar las comparaciones y análisis a gran escala, en este caso considerando un período de tiempo extenso, que proporciona la evolución de estilos en diferentes contextos y períodos de la historia literaria. De esta manera, se puede apreciar la identificación de patrones y tendencias que no serían evidentes en la lectura cercana. Otro ejemplo que podemos observar se publicó en 2013 en el libro “Distant Reading” del mismo autor, y que también implicó labor computacional. Moretti trabajó con los títulos de novelas, en un lapso de más de cien años, demostrando cómo estos fueron acortándose desde fines del siglo XVIII, particularmente por la relación con la novela de folletín y el emergente mercado editorial.



Fuente: Moretti, 2013, p. 183, representación de número de palabras en los títulos de novelas entre 1720 y 1860.

Otro humanista digital, Matthew Jockers (2013) ha preferido hablar de *macroanálisis* para describir métodos cuantitativos y computacionales similares, encontrando a este término más intuitivo. En su obra, explora cómo el análisis computacional y estadístico puede proporcionar aspectos novedosos y revelar estándares en la literatura que escapan a la lectura tradicional.



Fuente: M. Jockers (2013), "Macroanalysis: Digital Methods and Literary History". Gráfico a la derecha representa una red de las novelas del siglo XIX, donde se resaltó con colores las diferencias por género, en lo que respecta a estilo y temas (figura 9.4).

En síntesis, estas nuevas formas de evaluación de grandes cantidades de datos sólo son proporcionadas cuando las máquinas leen por nosotros, luego de estos procesamientos queda volver sobre los datos y darle sentido con nuestro trabajo humano.

## Voyant Tools

Con relación a la reflexión y el desarrollo sobre grandes datos, las metodologías cuantitativas y cualitativas están vinculadas a la visualización de datos, pues, en numerosas ocasiones, este tipo de representación es la técnica escogida para la creación de modelos y la determinación de patrones de análisis.

Por ello, ahora trabajaremos este marco teórico mediado por la plataforma Voyant Tools. Esta realizará una lectura distante de un gran corpus y generará automáticamente visualizaciones que luego nos permitirá volver sobre los datos desde la lectura cercana<sup>3</sup>.

Este software libre fue creado por Stéfán Sinclair y Geoffrey Rockwell. Alojado en una web abierta, permite el análisis de grandes cantidades de textos, cuenta con funciones como la comparación de corpus, cuantificación de frecuencia de palabras, análisis de ocurrencias, interpretación de términos clave, etc. No requiere instalación y es compatible con la mayoría de los sistemas operativos, además puede accederse desde <https://voyant-tools.org/><sup>4</sup>.



Figura 1: Interfaz de Voyant Tools.

<sup>3</sup> Les comparto un ejemplo de lectura distante, <https://www.theatlantic.com/politics/archive/2015/01/the-language-of-the-state-of-the-union/384575/>. Se pueden apreciar las palabras más utilizadas en los discursos de presidentes norteamericanos, para un analista político pueden ser de gran utilidad al analizar las políticas de cada gobierno.

<sup>4</sup> En el siguiente link encontrarán un tutorial, que utiliza un ejemplo un poco más complejo, realizado por De La Torre, S. G. (2019) en <https://programminghistorian.org/es/lecciones/analisis-voyant-tools>. Además, pueden acceder al tutorial de Voyant: <https://voyant-tools.org/docs/#!/guide/about>.

Síntesis de las características de Voyant para iniciar la lectura distante.

## Carga de documentos

Existen cuatro opciones:

- pegar directamente en el cuadro de inicio un texto copiado
- pegar direcciones web –separadas por comas– de los sitios en donde se encuentren los textos que quieres analizar
- hacer clic en “Abrir” y seleccionar algunos de los corpus precargados (las obras de Shakespeare o las novelas de Austen, ambos en inglés)
- cargar directamente nuestros documentos. Los textos por subir deben ser en texto plano, o sea, \*.txt y codificado en UTF-8.

El ejercicio que presentaré a continuación se realizó con los datos obtenidos desde <https://covid.dh.miami.edu/get/> con los tuits realizados en Argentina desde el 25 de abril hasta la fecha (24/11/2020)<sup>5</sup>.

Una primera visualización general:

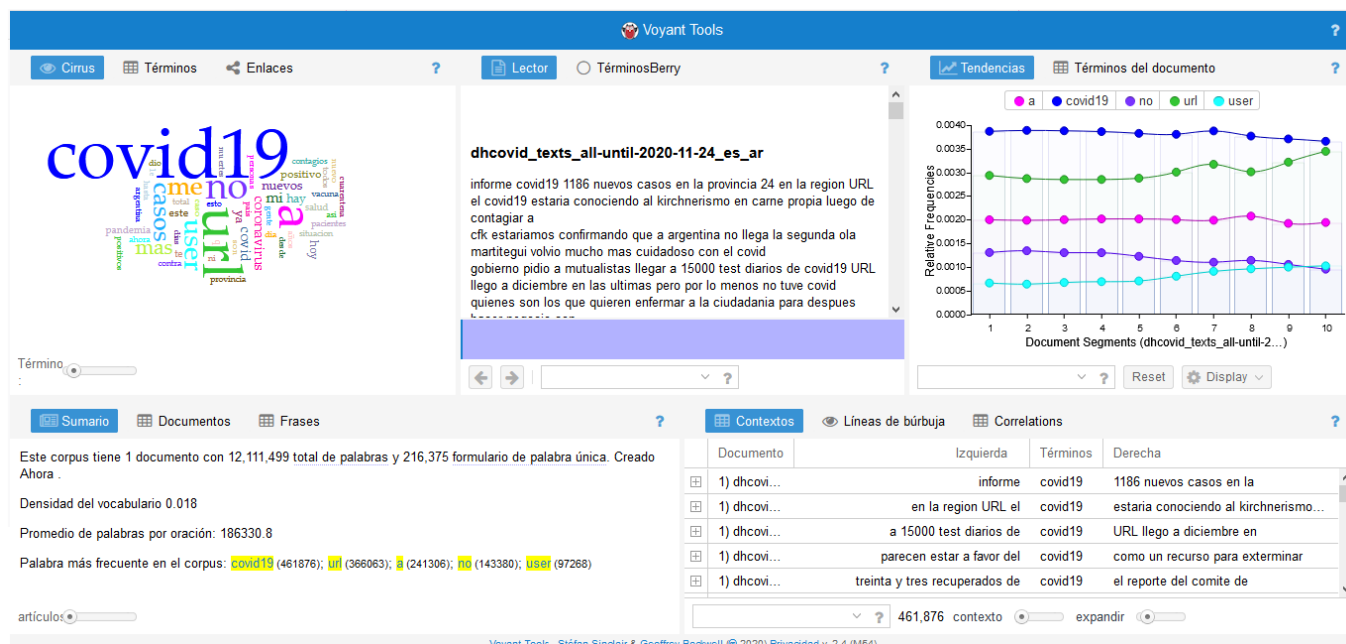


Figura 2: Primeras visualizaciones con los datos cargados.

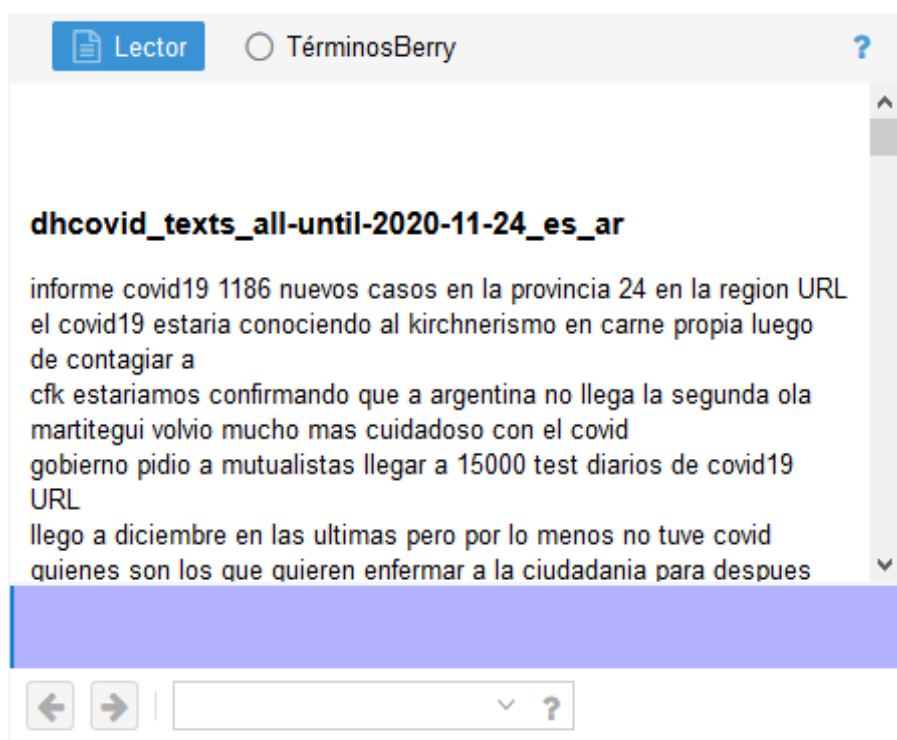
<sup>5</sup> Este proyecto se mencionó en la clase sincrónica de la Unidad N° 1, es sólo de carácter informativo y los datos son sólo para poder realizar una aproximación a la herramienta. No será necesario descargar los datos ni repetir la ejercitación.



- **Cirrus:** nube de palabras que muestra los términos más **frecuentes**



- **Lector:** espacio para la revisión y lectura de los textos completos con una gráfica de barras que indica la cantidad de texto que tiene cada documento



<sup>6</sup> Limpieza y filtrado serán explicados en el próximo apartado.

- **Tendencias:** gráfico de distribución que muestra los términos en todo el corpus (o dentro de un documento cuando sólo se carga uno)

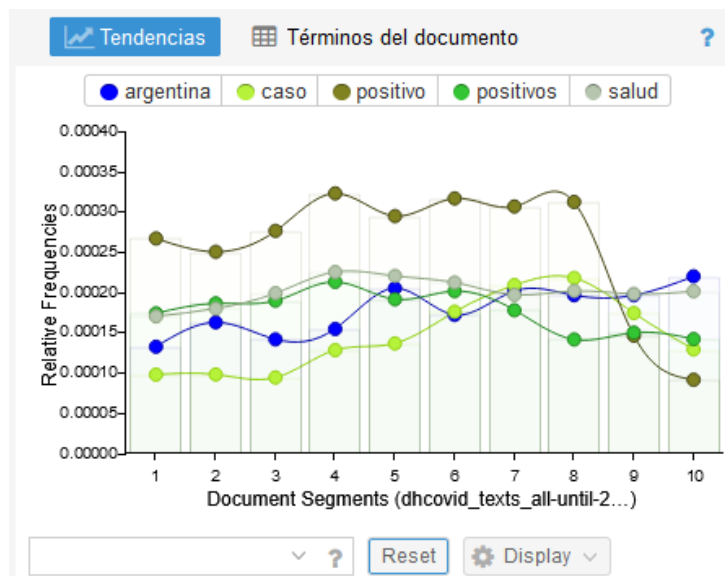


Figura 5. Vista de Tendencias, se puede realizar el mismo filtro de palabras como en el resto de las visualizaciones de los otros campos.

- **Sumario:** proporciona una visión general de ciertas estadísticas textuales del corpus actual, como ser el número total de palabras, si se trata de un corpus el número total de textos, las palabras únicas, densidad del vocabulario (divide el número de palabras únicas con el número total de palabras), palabras por oración (que en nuestro corpus no fue posible debido a la eliminación de signos de puntuación en la limpieza previa de los tuits)

**Este corpus tiene 1 documento con 12,111,499 total de palabras y 216,375 formulario de palabra única.**

**Densidad del vocabulario 0.018** (*Entre más cercano a 1 es el índice de densidad, el vocabulario tiene mayor variedad de palabras, es decir, que es más denso*)<sup>7</sup>

**Promedio de palabras por oración: 186330.8**

**Palabra más frecuente en el corpus: salud (24230); caso (17639); vacuna (16834); muertes (13397); contagios (13311)**

<sup>7</sup> Este sencillo análisis de textos tiene en cuenta la cantidad de palabras que se repiten en un corpus, cuanto más denso es mayor variedad de palabras. También se puede destacar que cuando es menos denso, implica que las palabras son más comunes, como preposiciones, conjunciones, artículos, adverbios, pronombres, algunos verbos, etc., estas son denominadas **stopwords** o palabras vacías. Estas unidades léxicas gramaticales se encuentran presentes en los textos en una proporción de alrededor del 30-40% en el español, y en la lectura distante no aportarían información, pero sin ellas en la lectura cercana no encontraríamos coherencias gramatical ni naturalidad.

- **Contextos:** concordancia que muestra cada ocurrencia de una palabra clave con un poco de contexto circundante.

Contextos	Líneas de burbuja	Correlations	
Documento	Izquierda	Términos	Derecha
1) dhcovi...	informe	covid19	1186 nuevos casos en la
1) dhcovi...	en la region URL el	covid19	estaria conociendo al kirchnerismo en
1) dhcovi...	a 15000 test diarios de	covid19	URL llevo a diciembre en
1) dhcovi...	parecen estar a favor del	covid19	como un recurso para exterminar
1) dhcovi...	treinta y tres recuperados de	covid19	el reporte del comite de
1) dhcovi...	fallecidos URL creo que tengo	covid19	me esta faltando el aire
1) dhcovi...	uuuu le estoy rozando al	covid19	de una maneraaaa el covid19
1) dhcovi...	covid19 de una maneraaaa el	covid19	es a uruguay lo que

461,876 contexto expandir

byant Tools , Stéfán Sinclair & Geoffrey Rockwell (© 2020) Privacidad v. 2.4 (M54)

Figura 6. Vista de Contextos, se puede realizar el mismo filtro de palabras como en el resto de las visualizaciones de los otros campos.

Una vez realizada esta primera **lectura** general, podemos acceder a la **frecuencia bruta** (en la ventana Cirrus, y al pasar el mouse sobre las palabras para obtener sus frecuencias), esta es obtenida antes de realizar la limpieza de palabras vacías, es decir, una enorme cantidad de palabras que no son útiles para el análisis, por lo cual podemos filtrar las palabras vacías: artículos, preposiciones, interjecciones, pronombres, etc., por lo que debemos realizar una **limpieza**.

## Limpieza

Voyant nos ofrece dos opciones, utilizar la lista que viene por default o editarla agregando nuestras propias palabras vacías. En particular, para nuestro análisis sumamos palabras como **covid**, **url**, **cuarentena**, **positivos**, **positivo**, algunos lugares (como nombres de provincias, ciudades), meses, números, entre otros. Este paso es determinante para avanzar en el análisis de las narrativas que provienen de Twitter, pues buscamos puntualizar en los contextos lingüísticos y humanísticos. **Cuidado:** por defecto está seleccionada la opción “Aplicar a todo”; si ésta se deja seleccionada el filtrado de palabras afectará las métricas de todas las otras herramientas. **Es muy importante que documentes tus decisiones, esto servirá para explicar cada paso para el análisis.** Una buena práctica es guardar la lista de palabras vacías en un archivo de texto (.txt).

Por otro lado, también señalaremos que para nuestro proyecto la lista de palabras vacías incluyó las unidades léxicas antes mencionadas. Asimismo a los tuits al recolectarlos de Twitter se les realizó un primer proceso de curado, donde se filtraban signos de puntuación, URL, nombre de usuarios, emojis, entre otros. Las decisiones tomadas fueron para facilitar el análisis, mantener la privacidad de quienes escribieron en la red social, y ayudarnos a llevar adelante una mayor comprensión de las narrativas que forman nuestro corpus.



Figura 7. Vista general con las palabras vacías filtradas. Con el círculo naranja se indica donde realizar la edición de la lista de palabras vacías.

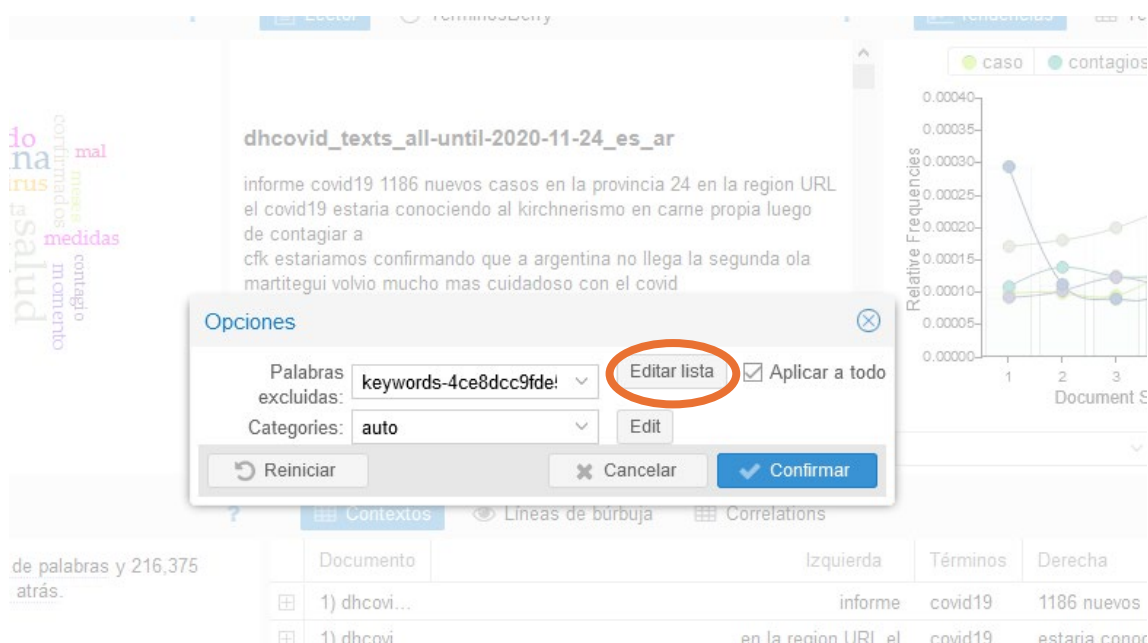
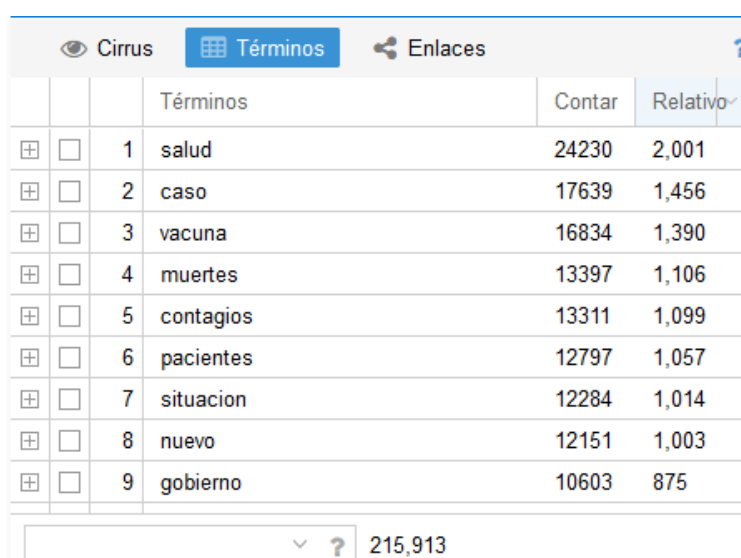


Figura 8. Opciones para el filtrado de las palabras vacías. Con el círculo naranja se indica donde realizar la edición de la lista de palabras vacías, y agregar una debajo de la otra.

De esta forma, para el ejercicio que les propondré, les compartí capturas de pantalla. Sin embargo, estas pueden diferir considerablemente dependiendo del texto que se “lea” o de las palabras vacías que se incluyan o no.

## Un poco más de lectura distante

En Cirrus, damos clic sobre “Terms” o “Términos”, abrirá una tabla que por defecto tiene tres columnas: Términos (con la lista de palabras en los documentos, sin las filtradas), Contar (con la “frecuencia bruta o neta” de cada término) y Tendencia (con una gráfica de la distribución de una palabra tomando su frecuencia relativa). Para obtener la frecuencia relativa de un término, en la barra de los nombres de columna, en el extremo derecho, se da clic sobre el triángulo que ofrece más opciones y en **Columnas** se selecciona la opción **Relativo**:



			Términos	Contar	Relativo
+	<input type="checkbox"/>	1	salud	24230	2,001
+	<input type="checkbox"/>	2	caso	17639	1,456
+	<input type="checkbox"/>	3	vacuna	16834	1,390
+	<input type="checkbox"/>	4	muerres	13397	1,106
+	<input type="checkbox"/>	5	contagios	13311	1,099
+	<input type="checkbox"/>	6	pacientes	12797	1,057
+	<input type="checkbox"/>	7	situacion	12284	1,014
+	<input type="checkbox"/>	8	nuevo	12151	1,003
+	<input type="checkbox"/>	9	gobierno	10603	875

215,913

Figura 9. Vista de Cirrus y Términos, donde se pueden seleccionar las columnas antes mencionadas.

El orden de la frecuencia bruta (**Contar**) y el de la frecuencia relativa (Relativo) es el mismo. Pero será de gran utilidad cuando comparemos diferentes corpus. Además, señala los datos que por sí solos no nos dicen mucho, que necesitan que los leamos su contexto y las distintas tramas para que adquieran significado.





Figura 10. Vista de Enlaces en Cirrus, ofrece la visualización de los enlaces de las palabras más frecuentes.

## Otras vistas (y faltan muchas más)

Desde el cuadro Tendencias se puede acceder a los términos del documento y desde allí visualizarla con el contexto de las palabras más frecuentes y su contexto. A continuación, se presentan figuras de otras vistas, que no repasaremos con detalle.

Voyant Tools				
Términos del documento		Colocaciones		
#	Términos	Contar	Relativo	Tendencia
<input type="checkbox"/>	1 salud	24230	2,001	
<input type="checkbox"/>	1 vacuna	16834	1,390	
<input type="checkbox"/>	1 muertes	13397	1,106	
<input type="checkbox"/>	1 contagios	13311	1,099	
<input type="checkbox"/>	1 pacientes	12797	1,057	
<input type="checkbox"/>	1 situacion	12284	1,014	
<input type="checkbox"/>	1 nuevo	12151	1,003	
<input type="checkbox"/>	1 gobierno	10603	875	
<input type="checkbox"/>	1 virus	10580	874	
<input type="checkbox"/>	1 ciudad	10472	865	
<input type="checkbox"/>	1 mundo	10384	857	
<input type="checkbox"/>	1 muertos	9890	817	
<input type="checkbox"/>	1 casa	9559	789	
<input type="checkbox"/>	1 hospital	9484	783	
<input type="checkbox"/>	1 informe	8800	727	
		215,907		

Figura 11. Vista de Tendencias, al lado en Términos de Documentos, se puede seleccionar Términos y las columnas antes mencionadas, así como la Tendencia, que grafica la distribución de una palabra tomando su frecuencia relativa.

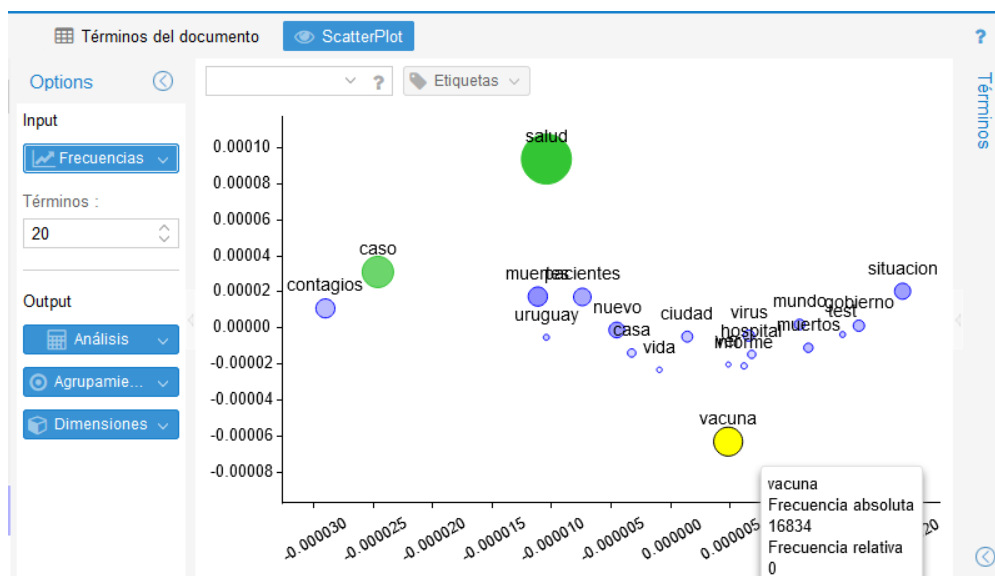


Figura 12. Vista ScatterPlot, provee una gráfica de visualización de cómo las palabras se agrupan en un corpus uso de similitud de documentos, el análisis de correspondencias o análisis de componentes principales.

## Descarga de datos

Por último, para poder conservar los datos de los análisis que resultan de cada herramienta se pueden descargar y compartir de diferentes formas, dependiendo de cada campo, como archivos de imágenes para las visualizaciones, como texto plano para luego hacer otros análisis o como código HTML para embeberlo en una página. De igual manera, se puede exportar los datos haciendo clic en el cuadro con flecha que aparece cuando pasas el cursor sobre la esquina derecha de “Contextos”, y seleccionar la opción “Exportar datos actuales”, donde se extienden diferentes opciones de descarga según tus necesidades.

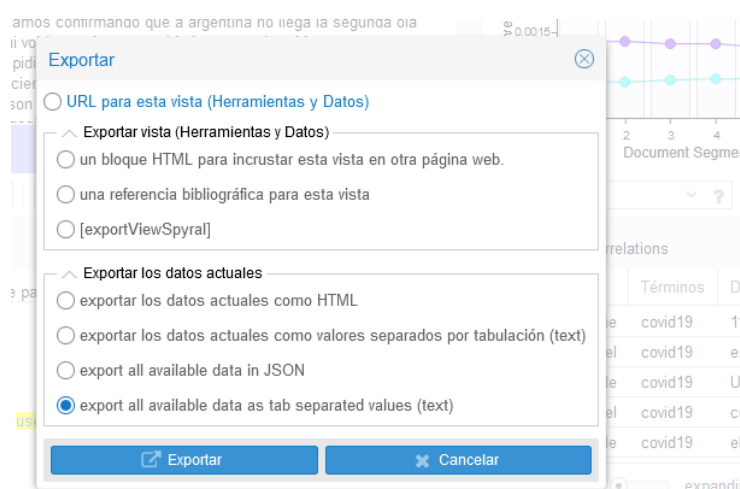


Figura 13. Vista con opciones de Descarga.

## Actividad:

- 1) Seleccioná un libro desde <https://www.gutenberg.org/browse/languages/es>. Este sitio es una librería online gratuita. Este proyecto fue desarrollado por Michael Hart en 1971 siendo la primera biblioteca digital. Pueden bucear y descargar un libro o varios, o copiar el url del archivo del libro. Por ejemplo, <https://www.gutenberg.org/cache/epub/26404/pg26404-images.html> o <https://www.gutenberg.org/cache/epub/26404/pg26404.txt>. Importante, tener en cuenta que si Voyant no devuelve nada cuando se trata de usar la dirección de un sitio es porque no es correcto el enlace para que pueda procesarlo.
- 2) ¿Qué palabras son las más frecuentes en el corpus? Antes y después de realizar la limpieza de palabras vacías.
- 3) ¿Qué nos dicen estas palabras del corpus?, ¿son significativas todas?
- 4) ¿Qué palabras vacías están en la nube de palabras?
- 5) ¿Cuáles eliminarías y por qué?
- 6) ¿Qué ventaja ofrece la visualización en Cirrus con respecto a la información que nos dan las palabras más frecuentes en la sección Sumario? ¿Cómo se complementan?
- 7) La actividad deberá ser subida al campus en Actividad de Unidad N° 2.

## Bibliografía recomendada (NO obligatoria)

- Bush, V. (1945, Julio). As we may think. *The Atlantic*. <https://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/>
- del Rio Riande, G., De León, R., & Hernández, N. (2021). Relación de las cosas sucedidas. *HDLAB*. <https://hdlab.space/Relacion-de-las-cosas-sucedidas/>
- De La Torre, S. G. (2019). *Análisis de corpus con Voyant Tools*. Programming Historian. <https://programminghistorian.org/es/lecciones/analisis-voyant-tools>
- Jänicke, S., Franzini, G., Cheema, M. F., & Scheuermann, G. (2015). On Close and Distant Reading in Digital Humanities: A survey and future challenges. *Eurographics Conference on Visualization (EuroVis) - STARs*, 83-103. <https://doi.org/10.2312/eurovisstar.20151113>
- Jockers, M. (2013). *Macroanalysis: Digital Methods & Literary History*. University of

Illinois Press. <https://www.matthewjockers.net/macroanalysisbook/>

Menéndez, M. J. E. (2022, marzo 23). ¿Qué son los datos digitales para las Humanidades? [Billet]. *Amontonamos las palabras: Blog de la Biblioteca de El Colegio de México*. <https://doi.org/10.58079/lwfl>.

Moretti, F. (2000). Conjeturas sobre la literatura mundial. *New Left Review*, 3, jul-ago. <https://newleftreview.es/issues/3/articles/franco-moretti-conjeturas-sobre-la-literatura-mundial.pdf>

Moretti, F. (2004a). Gráficos, mapas, árboles - 1. *New Left Review*, 24, ene-feb. <https://newleftreview.es/issues/24/articles/franco-moretti-graficos-mapas-arboles-1.pdf>

Moretti, F. (2004b). Gráficos, mapas, árboles - 2. *New Left Review*, 26, may-jun. <https://newleftreview.es/issues/26/articles/franco-moretti-graficos-mapas-arboles-2.pdf>

Moretti, F. (2004c). Gráficos, mapas, árboles - 3. *New Left Review*, 28, sep-oct. <https://newleftreview.es/issues/28/articles/franco-moretti-graficos-mapas-arboles-3.pdf>

Moretti, F. (2013). *Distant Reading*. London: Verso.

Nelson, T. H. (1965, Agosto). Complex Information Processing: A File Structure for The Complex, The Changing and The Indeterminate. *Proceedings of the 1965 20th national conference*. <https://dl.acm.org/doi/abs/10.1145/800197.806036>

Schöch, C. (2013). Big? Smart? Clean? Messy? Data in the Humanities. *Journal of Digital Humanities*, 2 (3). <http://journalofdigitalhumanities.org/2-3/big-smart-clean-messy-data-in-the-humanities/>

Wardrip-Fruin, N. (2004). What Hypertext Is. *HYPERTEXT '04: Proceedings of the Fifteenth ACM Conference on Hypertext and Hypermedia*, 04, 126–127. <https://doi.org/10.1145/1012807.1012844>

Weber, M. (1982). *Ensayos sobre metodología sociológica*. Amorrortu Editores.