

<https://artnodes.uoc.edu>

ARTÍCULO

NODO «ARTES EN TIEMPOS DE PANDEMIA»

Análisis cuantitativo de la revista *Índice Literario* (1932-1936)

Juana María González

Universidad Internacional de la Rioja

Fecha de presentación: septiembre de 2020

Fecha de aceptación: noviembre de 2020

Fecha de publicación: enero de 2021

Cita recomendada

González, Juana María. 2021. «Análisis cuantitativo de la revista *Índice Literario* (1932-1936)». En: Benítez, Laura; Berger, Erich (coord.) «Artes en tiempos de pandemia». *Artnodes*, núm. 27: 1-11. UOC. [Consulta: dd/mm/aa]. <http://doi.org/10.7238/a.v0i27.374373>



Los textos publicados en esta revista están sujetos –si no se indica lo contrario– a una licencia de Reconocimiento 4.0 Internacional de Creative Commons. La licencia completa se puede consultar en https://creativecommons.org/licenses/by/4.0/deed.es_ES.

Resumen

En este artículo aplicaremos diferentes análisis estadísticos con el lenguaje de programación R (frecuencia de palabras, bigramas, coocurrencia de palabras y *Topic Modeling*) a las reseñas bibliográficas de 1932-1933 de la revista española *Índice Literario* (1932-1936) para aproximarnos (i) a la visión de los colaboradores de *Índice Literario* del contexto general de la literatura escrita en los años treinta en España y, particularmente, de la obra literaria escrita por mujeres; y (ii) a la situación social y política española en estos años reflejada en la obra de los autores reseñados y en las reseñas de la revista. *Índice Literario* dependía de la sección «Archivos de Literatura Española Contemporánea» del Centro de Estudios Históricos (1910-1939) y estuvo dirigida por el poeta español Pedro Salinas (1891-1951) y un grupo de colaboradores. La revista fue una clara apuesta del Centro de Estudios Históricos por abrirse a la actualidad literaria y supone una fuente de información indispensable para conocer la realidad cultural de los años republicanos en España. El artículo concluye que el análisis cuantitativo (descubrimiento de patrones, tales como tendencias en el uso de palabras o la identificación de los tópicos más relevantes en las reseñas incluidas en la revista) complementa y enriquece de forma satisfactoria el análisis cualitativo de la publicación.

Palabras clave

Edad de Plata, literatura, *Text Mining*, *Topic Modeling*, R.

*Quantitative analysis of the journal Índice Literario (1932-1936)***Abstract**

In this article we will apply various statistical analyses with the R programming language (word frequency, bigrams, word co-occurrence and Topic Modelling) to the bibliographical reviews of 1932-1933 from the Spanish journal Índice Literario (1932-1936) in order to bring us closer (i) to the vision of the collaborators on Índice Literario in the general context of the literature written in Spain in the 1930s and, particularly, in the literary work written by women and (ii) to the social and political situation in Spain during these years which is reflected in the work of the authors reviewed and in the journal reviews. Índice Literario was part of the section “Archivos de Literatura Española Contemporánea” (“Contemporary Spanish Literature Archives”) in the Centre for Historical Studies (1910-1939) and was managed by the Spanish poet Pedro Salinas (1891-1951) and a group of collaborators. The journal was a clear attempt by the Centre for Historical Studies to open itself up to the literary scene of the time and provides an indispensable source of information for getting to know the cultural reality during the years of the Second Spanish Republic. The article concludes that the quantitative analysis (the discovery of patterns, such as trends in the use of words or the identification of the most important subjects in the reviews included in the journal) complements and enriches the qualitative analysis of the publication in a satisfactory way.

Keywords

Silver Age, literature, *Text Mining*, *Topic Modelling*, R

1. Introducción y metodología

En este artículo aplicaremos metodologías de análisis cuantitativo a la revista *Índice Literario* (1932-1936) para valorar en qué medida estas técnicas nos ayudan a conocer mejor la visión que los colaboradores de *Índice Literario* tenían sobre de la literatura escrita en los años treinta en España, la obra literaria escrita por mujeres, así como la situación social y política española en aquellos años.

Concretamente, se realizarán diferentes análisis estadísticos con el lenguaje de programación R (frecuencia de palabras, bigramas, coocurrencia de palabras y *Topic Modeling*), cuyos resultados contrastaremos con los de diversos trabajos e investigaciones de tipo cualitativo sobre la publicación.

En este estudio se analizan los números de los primeros dos años de la revista (1932 y 1933), que constituyen un total de 600 páginas aproximadamente y que, a priori, suponen un corpus suficiente para poder valorar la utilidad y el alcance de dichos métodos cuantitativos.

2. La revista *Índice Literario* (1932-1936): origen, características y estructura

La revista *Índice Literario* (1932-1936) fue el principal resultado de la sección del Centro de Estudios Históricos¹ «Archivos de Literatura Española Contemporánea», creada en 1932. Estuvo dirigida por el poeta Pedro Salinas (1891-1951) con la colaboración del también escritor José María Quiroga Plá (1902-1955), el poeta y crítico literario Guillermo de Torre (1900-1971), el historiador Vicente Llorens (1906-1979), la archivera y bibliotecaria María Galvarriato, y la filóloga y escritora María Josefa Canellada (1913-1995). No se trata de una publicación estrictamente bibliográfica ni crítica; lo que se presenta en *Índice Literario* es una serie de reseñas anónimas de libros de reciente edición donde se intenta adoptar un tono objetivo y exento de crítica personal. A las reseñas anónimas elaboradas por los colaboradores de la sección se añaden, ocasionalmente, fragmentos de reseñas de

1. El Centro de Estudios Históricos fue una institución española creada en 1910 a instancias de la Junta para Ampliación de Estudios e Investigaciones Científicas (1907-1939) presidida por el Premio Nobel de Medicina Santiago Ramón y Cajal desde su fundación hasta su muerte en 1934. El Centro de Estudios Históricos estuvo dirigido por el importante filólogo, historiador y medievalista Ramón Menéndez Pidal (1869-1968) y tuvo como objetivo principal el fomento y desarrollo de la cultura en España (López, 2006, y López-Ocón 2007)

críticos relevantes que se seleccionan de la prensa diaria en función de la importancia del título reseñado.

La publicación tenía una frecuencia mensual (a excepción de los meses de julio y septiembre, en que no se publicaba). Cada número de *Índice Literario* está organizado de la siguiente manera:

1. Un artículo de redacción o de fondo, consagrado al estudio especial de uno de los libros entre los de mayor valor y significación dentro de los publicados en el mes.
2. Reseña bibliográfica y de contenido de la producción literaria.
3. Pasajes de crítica seleccionados en la prensa (que se adjuntan a determinadas reseñas).
4. Estrenos teatrales.
5. Bibliografía de aquellos estudios o artículos críticos referentes a las obras más importantes reseñadas en anteriores números de *Índice Literario* y aparecidos con posterioridad a la publicación de estos.

Las reseñas bibliográficas están organizadas por secciones que, básicamente, se refieren al género literario de la obra reseñada. Estas son, fundamentalmente: antologías, dramática (estrenos y obras publicadas), ensayos literarios, literatura histórica (biografías, enseñanza de la historia, historia, historia de España, memorias, narraciones, narraciones históricas, temas contemporáneos), literatura para niños, novela y narraciones, poesía y viajes. Las secciones suelen mantenerse de modo general en todos los números de la revista, pero algunos presentan variaciones.

La revista *Índice Literario* y la sección «Archivos de Literatura Contemporánea» fueron una clara apuesta del Centro de Estudios Históricos por abrirse a la actualidad literaria. Ambas iniciativas coincidieron en el tiempo con la instauración de la Segunda República en España (1931) y suponen un muestrario de la visión de su director y colaboradores, reputados escritores y críticos del momento, sobre la realidad cultural y literaria de los años treinta en España.

3. La presencia femenina en la revista *Índice Literario*

Además de ser la primera revista del Centro de Estudios Históricos dedicada exclusivamente a la realidad cultural contemporánea del ámbito hispánico en los años treinta, *Índice Literario* destaca por incluir en sus páginas reseñas sobre obras escritas por mujeres intelectuales y escritoras relevantes del momento y por contar en el elenco de sus colaboradores con dos mujeres (las ya mencionadas María Josefa Canellada y María Galvarriato) (González 2018). Esto convierte a esta revista, a la sección «Archivos de Literatura Española Contemporánea», de la que dependía, y al Centro de Estudios Históricos –vinculado a la Institución Libre de Enseñanza (1876-1936) y a su nueva y modernizadora visión de la mujer (Vázquez, 2012)– en

fuentes de información importantes para estudiar la progresiva mejora de la situación de la mujer en España, de su acceso al espacio público y a la vida cultural, especialmente durante el periodo republicano (1931-1939) (Gómez, 2009: 211-240).

Hasta finales del siglo XIX el acceso de la mujer a la educación y a la vida pública en España había sido muy limitado. A partir de 1910 las mujeres empiezan a poder acceder libremente a la universidad (Scanlon, 1986: 56). En los años siguientes, las mujeres van dando pasos en su participación en la vida cultural del momento y avanzan en la conquista de derechos y libertades. Esto encontrará su más clara expresión en la obtención del derecho al voto en 1931.

Índice Literario surge coincidiendo con el impulso modernizador republicano en favor de la educación y el progreso de las mujeres, y animado por los principios de defensa de la mujer que desde años antes promovía la Institución Libre de Enseñanza. Esto la convierte en una fuente de información interesante acerca de la visión sobre la obra literaria de las mujeres que tenían los colaboradores de la revista y que, en consecuencia, transmitió el propio Centro de Estudios Históricos.

La revista *Índice Literario* se publicó durante cinco años y contiene aproximadamente unas mil reseñas diferentes sobre multitud de obras de carácter muy heterogéneo. Se trata, por tanto, de un corpus particularmente adecuado para ser analizado con metodologías cuantitativas, ya que uno de los principales objetivos de estas es ayudar a resumir e interpretar grandes cantidades de información.

4. R como herramienta para realizar análisis cuantitativo sobre textos

R es un lenguaje de programación y un entorno computacional cuya principal característica es ser un lenguaje de programación creado con el único propósito de hacer estadística (Mendoza, s.f.: 1).

En la actualidad, R es uno de los lenguajes de programación más populares para el análisis de datos y es usado por numerosas multinacionales de primer nivel como son Google, Facebook, Twitter, Microsoft, Uber o AirBnB (Bhalla, 2017).

Gran parte de la potencia de R viene dada por sus numerosas rutinas y librerías (paquetes). Se define paquete como «una colección de funciones, datos y código que se almacenan en una carpeta conforme a una estructura bien definida, fácilmente accesible para R» (Santana y Hernández, s.f., párr. 2). A fecha de mayo de 2020 existen 15 639 paquetes diferentes en R (CRAN, 2020).

Uno de los paquetes de R más relevantes para el análisis de textos es el {tidytext}, desarrollado por Julia Silge y David Robinson en 2016 conforme a los principios de los datos limpios (en inglés: *tidy data*).

Estos fueron desarrollados por Hadley Wickham (2014: 4) al proponer una manera estándar de organizar la información siguiendo tres principios:

1. Cada variable forma una columna.
2. Cada observación forma un renglón.
3. Cada tipo de unidad observacional forma una tabla.

El formato `{tidytext}`, por tanto, consiste en tablas en las que cada fila contiene un *token*, siendo *token* aquella unidad de texto –por ejemplo, palabras– que queremos analizar y siendo la tokenización el proceso de separar el texto en *tokens*.

Definir un estándar para organizar la información según los principios de los datos limpios (*tidy*) permite crear un amplio catálogo de funciones y librerías compatibles entre sí: es el llamado universo *tidy* (`{tidyverse}`), que incluye librerías como `{dplyr}`, `{tidyr}`, `{ggplot2}` y `{broom}` (Silge y Robinson, 2020).

Por minería de textos entendemos «el proceso o práctica de examinar grandes colecciones de escritos con el objetivo de generar nueva información» (Universidad de Alcalá, 2019: 2) o, dicho de otra forma, «el conjunto de procesos necesarios para transformar documentos o recursos de texto desestructurados en información significativa y estructurada» (Jetley, 2019: 3).

Por su parte, el modelado de tópicos es una aplicación de la minería de textos relacionada con el Procesamiento del Lenguaje Natural (PLN) y el aprendizaje automático (*Machine Learning*) que consiste en identificar aquellos grupos de palabras que tienen mayor tendencia a aparecer de forma concurrente a lo largo del texto. Esto se logra detectando patrones en una colección de documentos llamada corpus y agrupando las palabras que aparecen en dichos documentos en tópicos (Hammoe, 2018: 4).

En lo referente a la minería de textos y al modelado de tópicos, el entorno de R posee potentes paquetes entre los que destacan el `{tm}` (del inglés *Text Mining*) y el `{topicmodels}` (para modelado de tópicos).

Una característica relevante del paquete `{tm}` es que incluye la clase DTM (*Document-term matrix*), una forma muy popular de representar textos para su tratamiento posterior por parte del ordenador y cuya característica principal es que obvia el orden de aparición de las palabras (Feinerer, Hornik y Meyer, 2008: 10).

Por su parte, el paquete `{topicmodels}` incluye, entre otras, las funciones necesarias para poder aplicar los algoritmos del modelo LDA (Latent Dirichlet Allocation) a un texto, ampliando y complementando las funcionalidades de minería de textos del paquete `{tm}`. El modelo LDA es uno de los algoritmos que permiten descomponer un texto o conjunto de textos en tópicos (Blei, 2012).

Así, el paquete `{tm}` nos ofrece la funcionalidad básica de minería de textos, incluida la transformación de un corpus de documentos a matrices de tipo DTM, y el paquete `{topicmodels}` nos permite aplicar distintos algoritmos matemáticos de modelado de tópicos, sobre la información contenida en esas matrices DTM (Grün y Hornik, 2011).

En resumen, R es un lenguaje de programación diseñado para análisis estadísticos y sobre el que se ha creado un potente entorno gracias al diseño de paquetes por parte de sus usuarios. En este

proyecto usaremos especialmente los paquetes `{tidytext}`, `{tidyverse}`, `{tm}` y `{topicmodels}`.

5. Análisis cuantitativo de la revista *Índice Literario*

5.1. Cosechar los textos para su análisis

La revista *Índice Literario* está accesible de forma digital en formato JPG y PDF en dos instituciones diferentes: (i) la Universidad de Florida (prácticamente completa, JPG) y (ii) la Biblioteca Virtual Cervantes (solo 14 números en total, los que se conservan en la biblioteca Tomás Navarro Tomás (CSIC), PDF).

En el análisis se utilizaron, cuando fue posible, los números de la revista que se encuentran en Biblioteca Virtual Cervantes, ya que cuenta con PDF de mayor calidad. En el resto de los casos se emplearon los números de la Universidad de Florida. Se descargaron todas las imágenes de los números publicados entre 1932 y 1933, un total de 600 páginas, aproximadamente. Después se transformaron dichas imágenes a PDF y se reunieron en un solo archivo. A continuación, se aplicó a los PDF el reconocimiento óptico de caracteres (OCR) de ABBY Fine Reader para transformarlos a texto plano con extensión .TXT.

5.2. Preparar los textos para su análisis

Tras la transformación del OCR, se revisó el texto para corregir erratas, errores de puntuación, tildes y otros. A continuación, se procedió a preparar el texto para su análisis con R:

1. Se actualizó la ortografía.
2. Se unieron palabras partidas al final de línea, también las coincidentes con final de página.
3. Se eliminaron los números de página a fin de evitar párrafos partidos.
4. Se señalaron con almohadilla algunas líneas del texto como las que contienen el número de la reseña entre corchetes, números sueltos, datos bibliográficos de las reseñas complementarias, a fin de que el sistema no las tuviese en cuenta.
5. Se añadieron indicadores entre almohadillas para señalar el comienzo y fin de las reseñas y los datos bibliográficos de estas.
6. Se unió el nombre de algunos autores y lugares a fin de que el sistema pudiera identificarlos correctamente.

5.2. Preparar los textos para su análisis

Para comenzar el análisis, se extrajeron las reseñas incluidas en los números de la revista. Para ello fue necesario reunir todos los números de la revista en un solo archivo TXT. Este está disponible junto con el código para su análisis en un repositorio en Github (<https://github.com/NGG-UNIR/Indice-Literario>).

Se agruparon algunas de las secciones/géneros que estaban excesivamente subdivididas dentro de la revista para obtener así una mayor base estadística. Concretamente, se agrupó la sección «Dramática», que está subdividida en *Índice Literario* en «obras estrenadas» y «obras publicadas», y la sección «Literatura histórica», que estaba subdividida en subsecciones como «biografías», «memorias», «temas contemporáneos», etc.

Posteriormente se contaron las palabras absolutas/*token* de las diferentes reseñas, agrupándolas por el género literario/sección de la revista en la que se enmarca la reseña, y por el sexo del autor que está siendo reseñado (hombre o mujer), obteniéndose los siguientes datos (tabla 1).

Sorprende la poca representación de las autoras en las reseñas, en particular en la sección de «Literatura histórica» (biografías, memorias, temas contemporáneos, etc.). Esta sección fue una de las más extensas de *Índice Literario* entre 1932 y 1933 y, sin embargo, solamente una reseña hacía referencia a una obra escrita por una mujer, Concha Peña, autora del ensayo *El comunismo*.

Estos datos corroboran que «la producción de bienes literarios o filosóficos por parte de las mujeres resultó extremadamente complicada y estuvo llena de limitaciones y paradojas» (Alonso, 2016: 97). Según Alonso (2016: 103) esto se debe a que a las mujeres escritoras de aquellos años «se las ha recordado siempre por sus relaciones personales con otros escritores o artistas del 27». «Solamente se les permitía introducirse en esos círculos si se las percibía al lado de un hombre, pero únicamente se las reconoce y recuerda por su relación con ese hombre» (Alonso, 2016: 103).

Así, a pesar de que en sus números de 1932 y 1933 la revista *Índice Literario* se hace eco de autoras y obras importantes escritas por mujeres del momento como Concha Espina (1869-1955), Elena

Fortún (1886-1952) o Margarita Ferreras (1900-1964), la mayor parte de las reseñas se dedican a autoras de novela rosa como María Mercedes Ortoll (¿?-¿?) o Concha Linares (1910-2009) o a autoras de comedias de corte conservador como Pilar Millán Astray (1879-1949). No hay referencias, por ejemplo, a textos importantes escritos por mujeres publicados estos años como los de la poeta Concha Méndez, *Vida a vida* (1932), la dramaturga María Teresa León, *Huelga en el puerto* (1933), o la periodista Josefina Carabias, *Por nacer tarde* (1932).

También llama la atención la escasa referencia en estos números al tema del voto femenino, que se obtuvo por primera vez en España a finales de 1931, y sobre el que trata únicamente la reseña sobre el libro de Ángel Ossorio, *Cartas a una señora sobre temas de derecho político* (noviembre de 1932). Esta obra es básicamente un manual de educación política para uso de la mujer, a la que el autor no considera capacitada para ejercer su derecho al voto recientemente adquirido.

Estos datos quizás podrían interpretarse en la línea de las afirmaciones de Navas (2009: 79) cuando señala que los críticos y los historiadores de la literatura han sido en su mayoría hombres y que esta «condición sexuada» de los discursos críticos, aparentemente neutrales, deriva en estrategias para silenciar la producción literaria de las mujeres. En opinión de Quance (1987), esto ha implicado «el predominio de lecturas sesgadas, impregnadas de tópicos misóginos» (Navas, 2009: 80) de la obra literaria de las mujeres, hasta el punto de considerarla de tono o calidad menor, marginal o excéntrica y, por tanto, de menor relevancia.

A continuación, se estudió la frecuencia de aparición de determinadas palabras. Para ello se eliminaron del análisis las denominadas palabras vacías, que son aquellas cuyo contenido no es relevante

Sección de la Revista	Número de reseñas en el Corpus	Número de palabras <i>token</i> en el Corpus	Número de reseñas de obras escritas por mujeres	Número de palabras <i>token</i> de obras escritas por mujeres
Estudio introductorio	15	26 311	0	0
Antologías	3	262	0	0
Dramática	66	21 273	4	398
Ensayos literarios	32	12 725	1	100
Literatura histórica	116	43 951	1	294
Literatura para niños	9	2 163	1	105
Novela y narraciones	103	47 677	13	4 143
Poesía	28	6 519	5	728
Viajes	2	945	0	0
Total	374	161 826	25	5 768

Tabla 1. Número de reseñas y número de palabras *token* de las reseñas agrupadas por sección y sexo

para el análisis. Las palabras vacías empleadas en este trabajo son: (i) las llamadas stopwords ya predefinidas en el paquete {tidytext} (fundamentalmente conjunciones, preposiciones, determinantes, etc.) y (ii) palabras incluidas en listas de elaboración propia que se encuentran a disposición del lector en nuestro repositorio en Github <https://github.com/NGG-UNIR/Indice-Literario> (palabras que, según el contexto de análisis, no aportan información relevante, por ejemplo: «vida» en la sección de «biografías», etc.). Posteriormente, se realizó un conteo simple de la frecuencia absoluta de aparición de lemas en las reseñas de las principales secciones de la revista, obteniéndose los siguientes resultados (figura 1).

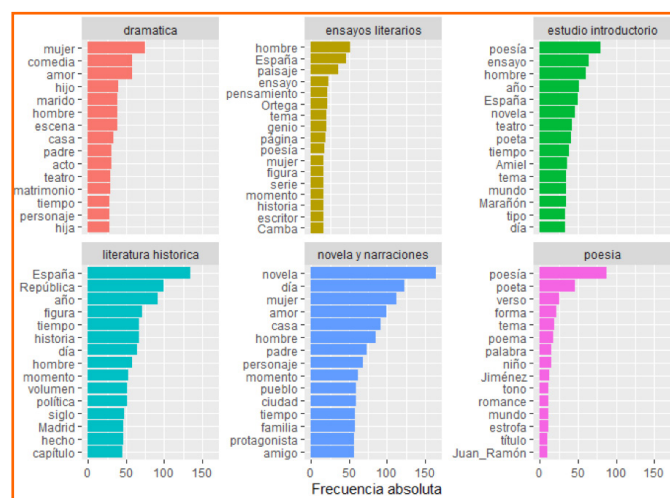


Figura 1. Lemas más frecuentes en las reseñas de las principales secciones de *Índice Literario* (nombres comunes y propios)

Resulta interesante observar el empleo frecuente de palabras como «amor», «mujer», «casa», «hombre», «hijo», «marido», «matrimonio» y «comedia» en las reseñas de obras de la sección «Dramática» y en las reseñas de la sección «Novela y narraciones».

En lo que respecta a la sección de «Literatura histórica», sin embargo, los reseñistas de *Índice* destacan el contenido político de las obras que comentan: entre las palabras más frecuentes están «República», «España», «política», «Madrid» o «historia». Muchas de estas obras están centradas en la reciente instauración de la Segunda República o en cuestiones políticas.

Destaca también la aparición de tres de los grandes referentes intelectuales de la Edad de Plata (1898-1939) de la cultura española: José Ortega y Gasset (1883-1955), padre del intelectualismo literario e impulsor del concepto de Arte Nuevo en la época (Lasaga, 2006); Gregorio Marañón (1887-1960), figura indiscutible en el mundo intelectual de estos años, particularmente, en el ámbito de la medicina (López, 2011); y Juan Ramón Jiménez (1881-1958), poeta, maestro de los poetas españoles de la Generación del 27 y Premio Nobel de Literatura (Blasco y Piedra, 2006).

5.4. Análisis de bigramas y coocurrencia de palabras

El análisis cuantitativo del texto continúa con el análisis de los bigramas y las coocurrencias de palabras más frecuentes en el corpus. En este artículo el concepto bigrama hace referencia a un grupo de dos palabras que aparecen una inmediatamente después de la otra, mientras que el concepto coocurrencia hace referencia a las palabras que tienden a coincidir juntas en una misma reseña, pero no necesariamente seguidas.

5.4.1. Análisis de bigramas

Tras analizar con R los bigramas de palabras más frecuentes en nuestro corpus agrupándolos por secciones de la revista, obtenemos gráfica de la figura 2.

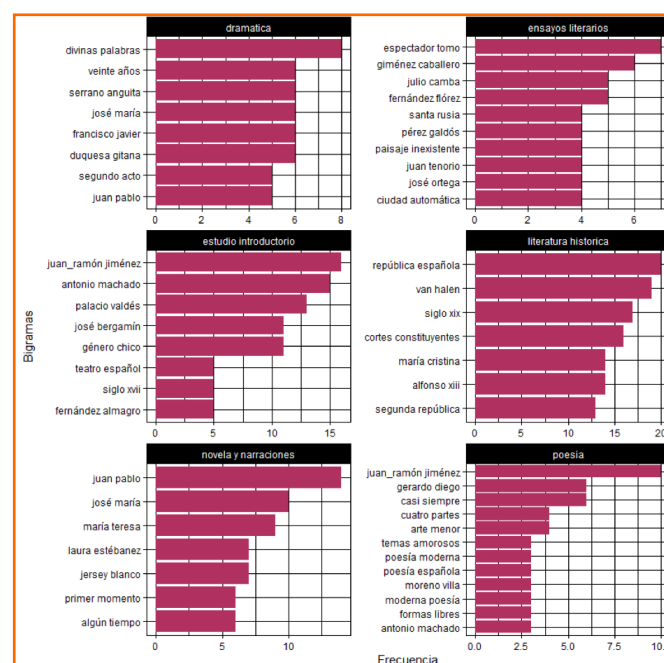


Figura 2. Bigramas más frecuentes en cada género/sección de *Índice Literario*

En la sección «Ensayos literarios», junto a «Pérez Galdós» y «José Ortega» [y Gasset], aparecen nuevas referencias como «Fernández Flórez», en clara alusión al escritor y periodista Wenceslao Fernández Flórez (1885-1964), quien había recibido el Premio Nacional de Literatura en 1926 por su novela *Las siete columnas*, o [Ernesto] «Giménez Caballero» (1899-1988), relevante escritor y crítico literario en la época, impulsor de la revista cultural *La Gaceta Literaria* (1927-1932), principal plataforma del vanguardismo español y la llamada Generación del 27, cuyo secretario fue en sus primeros meses el colaborador de *Índice Literario* Guillermo de Torre.

En la sección de «Estudio introductorio», el análisis de bigramas permite identificar a los principales autores objeto de las reseñas de dicha sección: «Palacio Valdés» [Armando Palacio Valdés (1853-

1938]), escritor y crítico literario español, «Juan Ramón Jiménez» o «Antonio Machado». La frecuencia en esta sección de bigramas como «género chico» o «teatro español» nos hace pensar que los colaboradores de *Índice Literario* dieron un espacio importante en esta sección a los textos dramáticos.

La sección de «Literatura histórica» parece centrada, en cambio, en la actualidad política. Entre los bigramas más frecuentes se encuentran «república española», «Alfonso XIII», o «Cortes Constituyentes».

En la sección «Dramática» destaca «divinas palabras», referida a una de las obras teatrales de Ramón del Valle Inclán.

Finalmente, en la sección de «Poesía» destacan bigramas que ponen de relieve a las figuras del panorama lírico más importantes en este momento, los mencionados «Juan Ramón Jiménez», «Antonio Machado» y «Gerardo Diego».

5.4.2. Análisis de coocurrencia de palabras

Respecto al análisis de coocurrencias, se ha aplicado este análisis a un determinado grupo de autores que aparece con asiduidad en los números de 1932 y 1933 de la publicación, a fin de poder determinar las palabras que más frecuentemente tienden a aparecer junto a sus nombres. Los resultados se pueden ver en las gráficas de la figura 3.

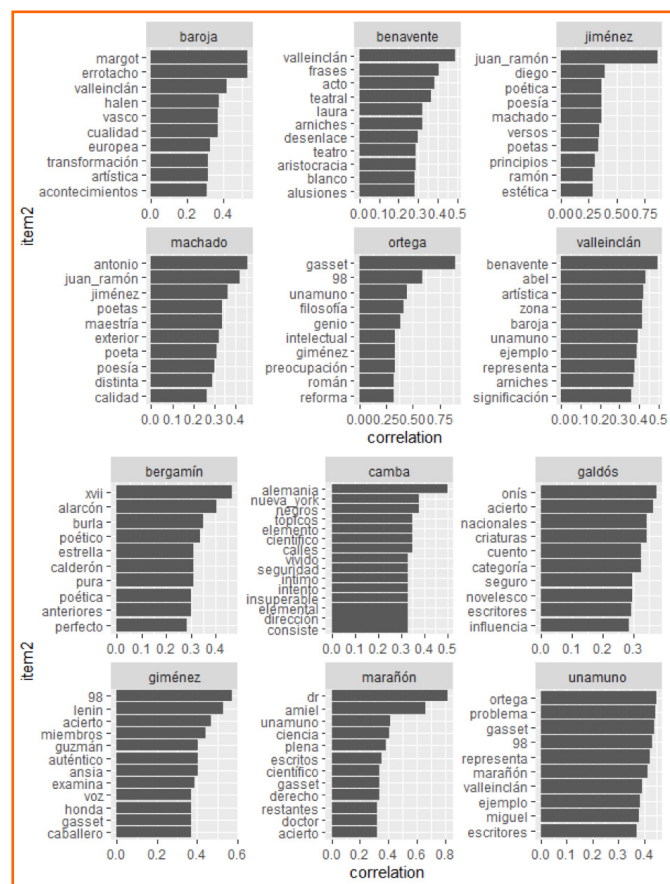


Figura 3. Términos más correlacionados con algunos escritores

A partir de este análisis pueden verse asociaciones interesantes como, por ejemplo, las de José Ortega y Gasset a términos como «intelectual», «filosofía» o «Unamuno» –Miguel de Unamuno (1864-1936)–. A este respecto, también Unamuno es asociado a «Ortega», como puede verse en la figura 3, y al término «problema». Esto nos hace pensar que ambos autores son percibidos por los colaboradores de *Índice Literario* como ideólogos, cuyas reflexiones conectan por su alto contenido filosófico.

Juan Ramón Jiménez se asocia, en cambio, a los conceptos de «poesía» y «estética». Machado es asociado también al poeta moguerño, quizás por ser considerados por los colaboradores de *Índice Literario* como dos de las figuras clave en la culminación de la transformación de la poesía española desde el Modernismo. Gregorio Marañón es identificado claramente como científico.

En este punto, también nos resultó interesante analizar las coocurrencias asociadas a los términos de «hombre» y «mujer» en las distintas secciones de la revista, siendo este análisis particularmente interesante en la sección «Dramática», como se puede ver en las gráficas de la figura 4.

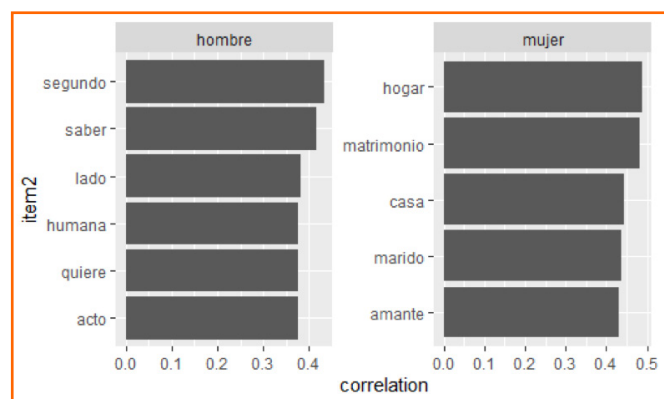


Figura 4. Términos más correlacionados con «hombre» y «mujer» en la sección «Dramática»

En esta sección, parece que los términos «saber» y «querer» suelen asociarse con mayor probabilidad a «hombre», mientras que los términos «matrimonio», «hogar» o «casa» suelen asociarse con mayor probabilidad a «mujer».

Algo similar sucede en el resto de secciones de la revista: en la sección de «Novela y narraciones», los términos «alma», «relación», «soledad» y «deseos» se asocian con mayor probabilidad a «hombre», mientras que «amante» y «marido» se asocian típicamente a «mujer».

En la sección de «Estudio introductorio», emergen las asociaciones «sentimiento de la mujer», «compasión de la mujer» y «tarea de la mujer», apareciendo también la palabra «amor» frecuentemente asociada al término «mujer».

Y en la sección de «Ensayos literarios» el término «hombre» es asociado frecuentemente a «intelectual» e «inteligencia», mientras que el de «mujer» es asociado frecuentemente a «musa».

5.5. Topic Modeling

También se realizó un modelado de seis tópicos de los números de 1932 y 1933 de la revista *Índice Literario*. Recordamos aquí que un tópico es uno o varios grupos de palabras que tienen mayor tendencia a aparecer de forma concurrente a lo largo de un texto. Tras procesar nuestro corpus, se obtienen los siguientes tópicos:



Figura 5. Términos más frecuentes en cada uno de los tópicos presentes en *Índice Literario*

En el tópico 4 predominan los conceptos asociados a la poesía, los tópicos 1 y 6 estarían asociados a novelas y narraciones rosas, centrados en tramas amorosas y familiares, y en el tópico 3 predominan palabras relacionadas con la república y la política.

Los tópicos dominantes de cada una de las reseñas, agrupados por sección, son los que se pueden ver en la figura 6.

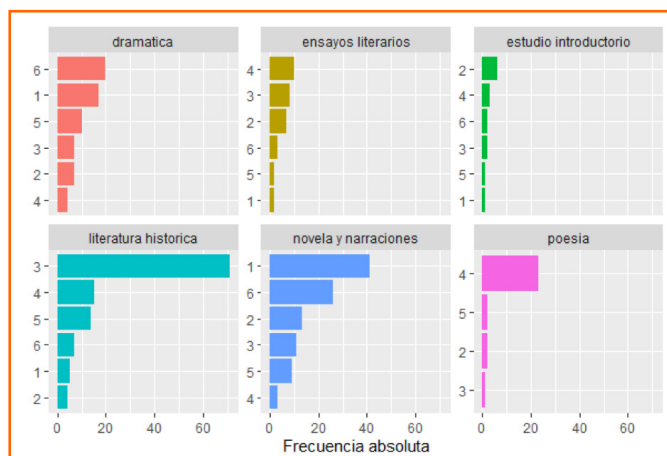


Figura 6. Relevancia de cada tópico en las principales secciones de *Índice Literario*

Como podía esperarse, el tópico 3 (que hemos asociado a temas políticos) destaca muy claramente dentro de la sección «Literatura histórica», mientras que el tópico 4 (que hemos asociado a poesía) domina con claridad la sección «Poesía». Los tópicos 1 y 6 (que hemos

asociado a novela rosa y temas amorosos) destacan notablemente en las secciones «Novela y narraciones» y «Dramática».

Este análisis presenta conclusiones interesantes. En primer lugar, que un número muy elevado de publicaciones se centraron en cuestiones de tipo históricopolítico, probablemente debido a la especial situación política del momento y a la adhesión pública de muchos de los colaboradores de *Índice Literario* (Pedro Salinas, Jose María Quiroga Plá y Guillermo de Torre) a la Segunda República.

En segundo lugar, que un importante número de dramas, novelas y narraciones de la época consistían en historias de amor o narraciones de tipo rosa con final feliz, lo cual no resulta sorprendente, porque es este tipo de literatura la que predominó entre el público general en estos años (Eguidazu, 2008). Lo llamativo es, no obstante, que *Índice Literario*, una revista en principio orientada a un público culto, se hiciera eco mayormente de este tipo de literatura más popular.

Adicionalmente, no es infrecuente encontrar en la revista reseñas de novelas y obras dramáticas que se centran en tramas de tipo amoroso donde los problemas que surgen entre la pareja se deben a cuestiones de tipo político o ideológico y que, frecuentemente, se resuelven a favor del amor.

Por último, analizamos si existen diferencias en los tópicos de las reseñas de obras de autores hombres y mujeres. En primer lugar, centraremos el análisis en el género/sección «Novela y narraciones», ya que es el único género para el que se cuenta con un mínimo número de reseñas de autoras (en concreto, 13):

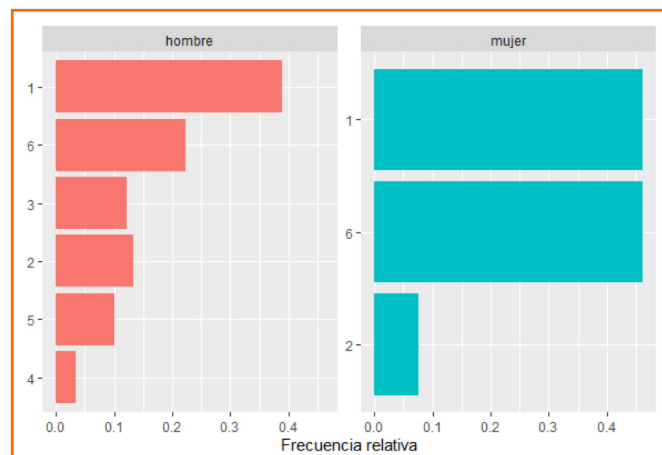


Figura 7. Relevancia de cada tópico en la sección «Novela y narraciones» en función del sexo del autor de la obra reseñada

Del análisis anterior resulta llamativo observar que en prácticamente todas las reseñas de obras escritas por mujeres los tópicos dominantes son los relacionados con poesía, novela rosa y enredos amorosos, mientras que en las reseñas de obras escritas por hombres hay más diversidad y no es infrecuente encontrar reseñas en las que predominen otros tópicos de carácter más político o intelectual.

Este mismo efecto se produce si analizamos la diferencia de tópicos en las reseñas de obras escritas por hombres y por mujeres en toda la revista:

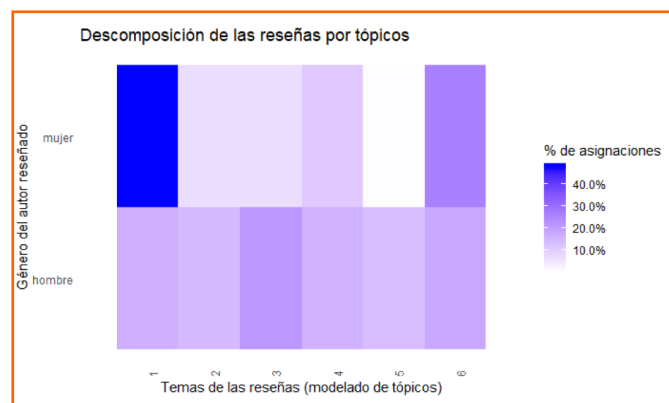


Figura 8. Tópicos más presentes en las reseñas de *Índice Literario* agrupados por el sexo del autor reseñado

6. Conclusiones

El análisis cuantitativo de los números de 1932-1933 de la revista *Índice Literario* (1932-1936) arroja unos resultados que complementan y enriquecen de forma satisfactoria el análisis cualitativo de la publicación. Mediante una extracción completa, rigurosa y estructurada de los datos contenidos en el texto y la identificación de patrones, este análisis nos permite aproximarnos a la visión de los colaboradores de *Índice Literario* del contexto general de la literatura escrita en los años treinta en España y, particularmente, de la obra literaria escrita por mujeres, así como a la situación social y política del momento.

De esta forma, se comprueba que la revista *Índice Literario* dio mucha importancia a las obras relacionadas con la «Literatura histórica», «Novela y narraciones» y «Dramática», y que en estas dos últimas secciones predominan los temas de amor y enredo, es decir, la publicación dio mucho espacio en sus páginas a obras de carácter popular, sin duda las más solicitadas por el público de la época. Asimismo, los nombres de autores que aparecen con más frecuencia en la publicación son hombres, la mayor parte de ellos conocidos referentes intelectuales para la Generación del 27.

La presencia femenina en la publicación es, sin embargo, muy escasa. La mayor parte de las reseñas sobre sus obras se encuentran en las secciones de «Poesía», «Novela y narraciones» y «Literatura para niños». En este sentido, a pesar de que *Índice Literario* y el Centro de Estudios Históricos dieron un paso adelante en la consideración de la obra intelectual y artística de las mujeres, coincidente con el impulso modernizador republicano (Vázquez, 2012), la representación femenina en la revista es insuficiente y desigual. Finalmente, se ha de señalar que la mayor parte de los tópicos asociados a estas

reseñas de obras escritas por mujeres son temas relacionados con el amor y la novela rosa.

En cuanto a la situación política y social, el análisis muestra que la política es un tema muy presente en las obras literarias del momento y que la publicación se muestra afectada a los ideales republicanos. Los temas relacionados con la mujer (derecho al voto, feminismo, divorcio) no se tratan apenas. Las mujeres aparecen en las reseñas generalmente asociadas al ámbito doméstico, al matrimonio, al amor y al sentimiento.

A la vista de estos resultados, parece oportuno extender estos análisis cuantitativos al resto de la revista *Índice Literario* y plantear un posible análisis relacional-comparativo con otras publicaciones de la época.

Referencias bibliográficas

- Alonso, Encarna. 2016. *Machismo y vanguardia. Escritoras y artistas en la España de preguerra*. Madrid: Devenir el otro.
- Bhalla, Deepanshu. 2017. *List of Companies using R* [Post]. <https://www.datasciencecentral.com/profiles/blogs/list-of-companies-using-r>.
- Blasco, Javier, y Antonio Piedra (Com.). 2006. *Juan Ramón Jiménez, Premio Nobel 1956* [Catálogo de exposición]. Madrid: Sociedad Estatal de Conmemoraciones Culturales/Publicaciones de la Residencia de Estudiantes.
- Blei, David M. 2012. «Probabilistic Topic Models». *Communications of the ACM* 55(4): 77-84. <http://www.cs.columbia.edu/~blei/papers/Blei2012.pdf>. DOI: <https://doi.org/10.1145/2133806.2133826>
- CRAN. 2020. *Contributed Packages of R*. <https://cran.r-project.org/web/packages/>.
- Egudazu, Fernando. 2008. *Del folletín al bolsilibro: 50 años de novela popular en España (1900-1950)*. Guadalajara: Silente Ciencia Ficción.
- Feinerer, Ingo, Kurt Hornik, y David Meyer. 2008. «Text Mining Infrastructure». *R. Journal of Statistical Software* 25(5): 1-54. Recuperado el 9 de mayo de 2020 de <https://www.jstatsoft.org/article/view/v025i05>. DOI: <http://dx.doi.org/10.18637/jss.v025.i05>
- Gómez, Mercedes. 2009. *Modernas y vanguardistas. Mujer y democracia en la II República*. Madrid: Laberinto.
- González, Juana María. 2018. «La sección de Archivos de Literatura Española Contemporánea (1932-1936) y su labor difusora de la obra creativa de la mujer en la Edad de Plata». *Bulletin of Hispanic Studies* 95: 959-975. DOI: <https://doi.org/10.3828/bhs.2018.56>
- Grün, Bettina y Kurt Hornik. 2011. «Topicmodels: an R package for fitting Topic Models». *Journal of Statistical Software* 40(13): 1-30. <https://epub.wu.ac.at/3987/1/topicmodels.pdf>. DOI: <http://dx.doi.org/10.18637/jss.v040.i13>
- Hammoe, Luciano. 2018. *Detección de tópicos utilizando el modelo LDA* [Trabajo final presentado para la obtención del título de especialista

- en ciencia de datos]. Buenos Aires: Instituto Tecnológico de Buenos Aires. https://ri.itba.edu.ar/bitstream/handle/123456789/1250/TFI_Hammoe.pdf?sequence=1&isAllowed=y.
- Jetley, Raoul. 2019. «Minería de textos». *ABB Review* 4, 2018. <https://new.abb.com/news/es/detail/14828/mineria-de-textos-ABB-Review-4/2018>
- Lasaga, José (Ed.). 2006. *El Madrid de Ortega*. Madrid: Publicaciones de la Residencia de Estudiantes.
- López, Antonio. 2011. *Gregorio Marañón: radiografía de un liberal*. Barcelona: Taurus.
- López, José María. 2006. *Heterodoxos españoles: el Centro de Estudios Históricos, 1910-1936*. Madrid: Marcial Pons Historia.
- López-Ocón, Leoncio. 2007. «El cultivo de las ciencias humanas en el Centro de Estudios Históricos». *Revista Complutense de educación*, 18(1): 59-76. <https://revistas.ucm.es/index.php/RCED/article/view/RCED0707120059A/15710>.
- Mendoza, Juan Bosco. s.f. *R para principiantes*. <https://bookdown.org/jboscomendoza/r-principiantes4/introduccion-que-es-r-y-para-que-es-usado.html>.
- Navas, Isabel. 2009. *La literatura española y la crítica feminista*. Madrid: Editorial Fundamentos.
- Quance, Roberta. 1987. «Entre líneas: Posturas críticas ante la poesía escrita por mujeres». *La Balsa de la Medusa* 4: 73-96.
- Santana, Angelo, y Carmen Nieves Hernández. s.f. *Librerías en R*. <http://www.estadistica-dma.ulpgc.es/cursoR4ULPGC/5-librerias.html>.
- Scanlon, Geraldine M. 1986. *La polémica feminista en la España contemporánea 1868-1974*. Madrid: Akal.
- Silge, Julia, y David Robinson. 2020. *Text Mining with R. A Tidy Approach*. <https://www.tidytextmining.com/>.
- Universidad de Alcalá. 2019. *Text Mining vs Natural Language Processing* [Post]. <https://www.master-data-scientist.com/text-mining-natural-language-processing/>.
- Vázquez, Raquel. 2012. *Mujeres y educación en la España contemporánea. La Institución Libre de Enseñanza y la Residencia de Señoritas de Madrid*. Madrid: Akal.
- Wickham, Hadley. 2014. «Tidy Data». *Journal of Statistical Software* 59(10): 1-23. <http://vita.had.co.nz/papers/tidy-data.pdf>. <http://hdl.handle.net/10.18637/jss.v059.i10>. DOI: <https://doi.org/10.18637/jss.v059.i10>

CV

**Juana María González**

(juanamaria.gonzalez@unir.net)

Universidad Internacional de La Rioja

Doctora en Literatura Española (Universidad de Granada), Máster en Humanidades Digitales (UNED) y acreditada como Profesor Contratado Doctor. Profesora de Literatura Española en la Universidad Internacional de La Rioja, se ha dedicado, fundamentalmente, al estudio de la literatura de la Edad de Plata y las humanidades digitales. Es colaboradora de proyectos de investigación del Instituto de Historia del Consejo Superior de Investigaciones Científicas, del Grupo de Estudios del Exilio Literario de la Universidad Autónoma de Barcelona y del grupo La Otra Edad de Plata de la Universidad Complutense de Madrid. Ha publicado varias ediciones de epistolarios de la Edad de Plata centrándose en figuras como León Sánchez Cuesta, Pedro Salinas, Guillermo de Torre, Juan Ramón Jiménez o José María Quiroga Plá y artículos en revistas especializadas como *Arbor*, *Rilce* o *Bulletin of Hispanic Studies*.