

Experiencia de lectura de la obra de Fernanda Melchor: análisis de las reseñas en Goodreads

Mariana Reyes-Abundes 
Tecnológico de Monterrey, México
A01550689@tec.mx

Manuel Cebral-Loureda 
Tecnológico de Monterrey, México
manuel.cebral@tec.mx

Recibido: 23/01/2023

Aceptado: 27/09/2023

Resumen

La experiencia de lectura de los usuarios de las redes sociales está influenciada por la cultura digital. En este artículo se analiza la experiencia de lectura de cientos de usuarios de la red social Goodreads, mediante el uso de técnicas de Procesamiento de Lenguaje Natural con programación en R, aplicadas a las reseñas de los libros de la escritora mexicana Fernanda Melchor. Se encontraron las palabras directa e inversamente frecuentes, las emociones vinculadas a las reseñas, los lexemas y verbos más relevantes, y se practicó una clasificación de tópicos. Los resultados muestran cómo los lectores reflexionan, de manera personal, sobre aspectos sociales, como la violencia o la pobreza, dando continuidad a las temáticas que aborda la propia autora.

Palabras clave: Humanidades digitales; promoción de la lectura; reseñas literarias; tecnologías 2.0; minería de textos; Melchor, Fernanda (1982-).

Cómo citar: Reyes-Abundes, M., & Cebral-Loureda, M. (2024). Experiencia de lectura de la obra de Fernanda Melchor: análisis de las reseñas de Goodreads. *Ocnos*, 23(1). https://doi.org/10.18239/ocnos_2024.23.1.363



Reading experience of Fernanda Melchor's work: analysis of reviews on Goodreads

Mariana Reyes-Abundes 
Tecnológico de Monterrey, México
A01550689@tec.mx

Manuel Cebal-Loureda 
Tecnológico de Monterrey, México
manuel.cebral@tec.mx

Received: 23/01/2023

Accepted: 27/09/2023

Abstract

Digital culture influences the reading experience of social network users. This paper analyses the reading experience of hundreds of users of the social network Goodreads by using Natural Language Processing techniques with R programming applied to the book reviews by the Mexican writer Fernanda Melchor. Directly and inversely frequent words were identified, as were emotions linked to the reviews and the most relevant lexemes and verbs, and a classification of topics was performed as well. The results show how readers reflect, in a personal way, on social aspects, such as violence or poverty, giving continuity to the themes addressed by the author herself.

Keywords: Digital humanities; reading promotion; literary reviews; 2.0 technologies; text mining; Melchor, Fernanda (1982-).

How to cite: Reyes-Abundes, M., & Cebal-Loureda, M. (2024). Reading experience of Fernanda Melchor's work: analysis of Goodreads reviews. *Ocnos*, 23(1). https://doi.org/10.18239/ocnos_2024.23.1.363



Introducción

La lectura es una actividad que, aunque regularmente se realiza en solitario, ocurre en un rico contexto social, resultado de la influencia de varias personas y procesos. Los libros permiten varios tipos de interacción social: la selección de un libro puede estar motivada por la recomendación de otros lectores, el trabajo de difusión que realizan editoriales y autores, el proceso de lectura y la recepción e interacción con otros lectores. El proceso de mediación lectora comprende la interacción entre libros y lectores para facilitar el diálogo entre ambos (Cerrillo et al., 2002), por lo tanto, para iniciarse, la lectura necesita del otro. En todos estos procesos participan varias personas y entidades, y suceden tanto en ambientes físicos como en virtuales; ahora bien, las herramientas digitales han permitido que los lectores interactúen de manera directa y con más frecuencia, a veces entre ellos o con otros agentes, gracias al uso de redes sociales y de plataformas de lectura.

Este artículo se propone analizar la recepción de obras literarias discutidas en medios digitales, y el modo en que crean una comunidad que comparte reseñas, valoraciones, opiniones y comentarios de libros, extendiendo el alcance de las temáticas del libro, según el modo en que son asimiladas por sus lectores. Para ello, se estudiará la recepción de la obra de la escritora mexicana Fernanda Melchor, una de las escritoras más premiadas y con altas ventas en los últimos años, a través de un análisis que comprende recolección de reseñas, valoraciones y publicaciones compartidas en Goodreads, plataforma digital dedicada a la promoción de la lectura. Goodreads ofrece la oportunidad de recolectar un gran número de reseñas al ser una de las aplicaciones de lectura más populares del mundo con más de 90 millones de usuarios (Barnett, 2021). El objetivo de revisar esta gran cantidad de reseñas es conocer qué información sobre las prácticas de lectura aportan los lectores, y cómo al aplicar minería de textos se pueden obtener resultados que ayuden a plantear preguntas sobre la experiencia de lectura de cientos de usuarios.

Lectura social y medios digitales

Actualmente, la experiencia de lectura está influenciada, en su gran mayoría, por la cultura digital y las plataformas, a pesar de que se puede dar tanto en espacios virtuales como físicos. Un informe reciente sobre las prácticas de lectura en mujeres mexicanas señala que el 60% de las lectoras se informan de novedades en redes sociales (Quiroga, 2020), desplazando a la librería tradicional como canal de comunicación directo. Este tipo de datos demuestran que los lectores seleccionan sus libros con la ayuda de otros lectores, y que los fenómenos como la valoración de contenidos en internet responden a una realidad sociocultural que supone que leer es una práctica social y hecha en comunidad (Alonso-Arévalo y Cerdón-García, 2014; García-Roca, 2020). El espacio digital y el físico conviven y enriquecen la lectura, de modo que no es difícil imaginar un lector que interactúe con ambos de manera indiscriminada: puede leer una reseña de un libro en una revista en línea, buscar recomendaciones en redes sociales, leer en formato físico o descargar un artículo en PDF y leerlo en el celular, compartir en redes sociales la última lectura, etc. Todas estas actividades se realizan cotidianamente y forman parte de la lectura como actividad social híbrida.

Complementariamente, existen otros espacios que no están dedicados exclusivamente a la promoción de la lectura, donde también se comparten reseñas literarias, retos de lectura y opiniones sobre autores y libros en formato audiovisual. Entre ellos, han ganado relevancia los *booktubers* (usuarios de Youtube) y *booktokers* (usuarios TikTok). Conocidos como *influencers* del libro, son personas que reseñan contenidos, dan su opinión, promueven retos y, en general,

fomentan el hábito de lectura a un público formado en su mayoría por jóvenes y adolescentes usuarios de estas plataformas. Consecuentemente, emergen como mediadores de lectura frente a una audiencia que tiene la posibilidad de interactuar asincrónicamente a partir de las funcionalidades de las plataformas, pero también a través de otras redes sociales (Paladines-Paredes y Aliagas, 2021). Muchos de estos creadores de contenido se han convertido en nuevos vendedores de libros, recomendando textos y escritores según el gusto de sus clientes, con una fuerte incidencia en las prácticas de lectura y la circulación de libros (Saez, 2022). Así, se desafía el modelo de legitimación propio del campo cultural -cuyas instancias tradicionales son la academia, los salones y las editoriales (Altamirano y Sarlo, 1990)- dando cabida a nuevas prácticas digitales de mediación y crítica literaria.

En general, las plataformas de lectura social ofrecen la posibilidad de interactuar con los libros de manera directa, al mismo tiempo que ofrecen una variedad de servicios enfocados a la promoción de la lectura. En Goodreads no solo es posible comentar cualquier libro que esté en su base de datos; además, el usuario puede valorarlos mediante un mecanismo de puntuación, crear sus propias etiquetas del libro, o ver los comentarios y reseñas de otros usuarios. Por ello, Goodreads ofrece una fuente de datos confiable y útil para analizar diferentes prácticas de lectura contemporánea a partir de reseñas en línea, como se ha realizado en distintas investigaciones (Thelwall y Kousha, 2017; Wang, Liu et al., 2019).

La escritura de mujeres: el caso de Fernanda Melchor

En los últimos años se ha publicado y promocionado la lectura de libros escritos por mujeres, cuestionando el canon literario tradicionalmente masculino, dando relevancia y visibilidad a mujeres dentro del circuito editorial (Carroto, 2017; González Arceo, 2017). Este fenómeno editorial, que está impactando de manera gradual pero significativa a los lectores, se presenta como una oportunidad para el análisis, generando datos mensurables, sobre todo en lo que compete a la recepción de sus libros, que es el tema que interesa en este trabajo. Sin embargo, dado que hablar de escritoras mujeres es un tema muy amplio, se ha decidido escoger a Fernanda Melchor como caso de estudio. Melchor destaca por el gran éxito editorial y literario de su novela *Temporada de Huracanes* (2017), publicada en más de veinte idiomas, con varias reimpressiones y una edición de bolsillo (2022). Otros libros de la autora son: la crónica *Aquí no es Miami* (2013), las novelas *Falsa Liebre* (2013) y *Páradais* (2021).

Esta escritora y traductora veracruzana también cuenta con el respaldo de los medios de comunicación y la crítica literaria, y ha sido galardonada con premios como el *Anna Seghers* en 2019 y el *Premio Internacional de Literatura* 2019, entre otros. Sus textos han sido objeto de investigación en distintas publicaciones académicas (Lemus, 2022; Godínez-Rivas y Nieto, 2019), al tiempo que cuenta con una base sólida de seguidores en redes sociales. Por eso, analizar la recepción de su obra permite visualizar y estudiar lógicas y relaciones de un sistema editorial en transformación (Parnell y Driscoll, 2021), siempre teniendo en cuenta que un bestseller literario es un producto cultural de dos caras: llega rápidamente a un gran público, a la vez que gana reconocimiento literario (Pouly, 2016). Melchor se ha convertido en una de las escritoras mexicanas con más reseñas en español en Goodreads. Tan solo *Temporada de huracanes* tiene 2.082 reseñas, número superior a otros libros de escritoras mexicanas contemporáneas como *La hija única* de Guadalupe Nettel con 1.043 reseñas, y muy parecido al de *Nuestra parte de noche* (2019) de la argentina Mariana Enríquez con 2.752.

Fernanda Melchor en Goodreads

A partir del año 2016, la plataforma Goodreads no permite acceder a su contenido de manera abierta al público, ya que no está disponible la aplicación de acceso programático o API (las siglas en inglés de Application Programming Interface) al texto completo de sus reseñas. Así, para recolectar las reseñas se tuvo que utilizar la técnica *web scraping*, que permite extraer datos directamente de la web, en concreto la parte de las reseñas que pueden ser visualizadas por el público general. Goodreads permite visualizar datos en abierto, pero hay otros que solo son accesibles cuando se está activo dentro de la plataforma. En este caso se usó una cuenta de usuario de Goodreads para poder visualizar la mayor cantidad de datos posible.

A pesar de que existe una cantidad de datos superior a tres mil reseñas, Goodreads solamente permite ver 300 de cada libro por cada filtro -hay tres filtros de visualización: más nuevo, más antiguo y más popular; así como el filtro por idioma-. Para acceder a las reseñas en forma de tabla, se utilizó la extensión de Chrome *listly*, que ayudó a recolectar datos públicos para formar una base de datos con 2.048 reseñas en español -575 corresponden a *Páradais*, 1.127 a *Temporada de huracanes*, 292 a *Aquí no es Miami*, y 54 a *Falsa liebre*-. La base de datos está conformada por los campos: usuario, fecha, valoración, reseña, *likes* y comentarios. Es importante destacar que en este estudio solo se realizó el análisis textual usando el campo de reseña, no se mencionan los usuarios para proteger su privacidad y tampoco se toma en cuenta la valoración o comentarios.

Tabla 1. Metadatos recogidos en Goodreads a 21 de julio de 2022 de los libros de Fernanda Melchor

	Año de Publicación	Total de Reseñas	Total de Valoraciones	% de escritura	Rating (1-5)	Longitud Media de Reseñas	Idiomas de Reseñas
Páradais	2021	914	4,738	19.3%	3.75	106	15
Temporada de Huracanes	2017	3,704	17,975	20.6%	4.09	103	20
Aquí no es Miami	2014	312	1,649	18.9%	4.26	84	4
Falsa Liebre	2013	59	271	21.8%	3.89	99.89	1

De acuerdo con los metadatos recogidos, el libro más popular de la autora es *Temporada de Huracanes*, que cuenta con más reseñas escritas y un número mucho más alto de valoraciones, aunque el mejor calificado de todos sus libros es *Aquí no es Miami*, con un promedio de 4.26 estrellas. Es interesante mencionar que todos cuentan con un porcentaje de reseñas muy similar, resultando que aproximadamente el 20% de las personas que leen los libros de Melchor escriben una reseña en la plataforma. La longitud promedio de las reseñas va de 83 palabras la más corta, hasta 106 la más larga, que corresponde a *Páradais*. Este es un dato muy variable, ya que pueden existir reseñas de una palabra y otras de más de 1.000 palabras, pero ayuda a entender cómo se comporta la escritura de reseñas por libro. Otro dato interesante son los idiomas de publicación de reseñas: si bien se ha escogido el español para la recolección de muestras, no hay que olvidar que algunos libros están traducidos a más idiomas, lo cual aumenta el número de reseñas escritas en comparación con los libros que no están traducidos -el caso de *Falsa liebre*-.

Metodología

Para este estudio se han utilizado métodos mixtos. Primeramente, algunos enfoques de las humanidades digitales ofrecen metodologías para cuantificar las historias y predecir el éxito comercial de libros y autores (Toubia et al., 2021; Wang, Yucesoy et al., 2019), así como sobre librerías digitales y literatura electrónica (Pressman et al., 2015). Complementariamente, la analítica cultural se ha dedicado a estudiar redes sociales, procesando grandes volúmenes de imágenes, hashtags y comentarios, demostrando que los métodos cuantitativos basados en datos ofrecen un nuevo marco para describir artefactos, experiencias y dinámicas culturales (Manovich, 2020).

En la presente investigación, este enfoque de analítica cultural es pertinente en la medida en que se analizará una considerable cantidad de comentarios y reseñas recolectadas en las plataformas y redes de lectura y, siguiendo la lógica que plantea Lev Manovich (2020), buscando representar la cultura mediante datos y modelos estadísticos. Además, el objeto de estudio también se entiende desde el enfoque de analítica cultural: se representa la cultura por el modo en el que efectivamente se produce en el presente, poniendo el foco en la baja cultura, o cultura popular, como fuente de legitimación, en vez de la crítica literaria o alta cultura especializada. Finalmente, a diferencia de los métodos tradicionales de estudios literarios, en este trabajo se busca analizar toda la muestra recogida en línea -acotada al caso de estudio-, sin tener que escoger algunos datos más representativos.

Para desarrollar la metodología, se aplicó analítica computacional mediante técnicas de Procesamiento del Lenguaje Natural (NLP por sus siglas en inglés) en RStudio, entorno integrado para manejo de comandos del lenguaje de programación R, realizando las siguientes tareas:

- Análisis de frecuencia directa e inversa de cada libro: mediante el paquete *tidytext* (Silge y Robinson, 2022) se han dividido las reseñas en palabras y posteriormente se han eliminado las palabras vacías o *stopwords* a través de la lista que proporciona el paquete *stopwords* (Benoit et al., 2021). Con las palabras restantes, se contabilizan aquellas que más se repiten para cada documento, en este caso, cada uno de los libros a los que pertenecen las reseñas. Complementariamente, la frecuencia inversa-también conocida como estadística *tf-idf* (*term frequency - inverse document frequency*) se refiere a la frecuencia de un término ajustada a la rareza de su uso en el conjunto de documentos, midiendo la importancia de una palabra en un documento (Silge y Robinson, 2017; Sparck-Jones, 1972).
- Análisis de sentimientos y reconocimiento de emociones a partir de léxicos previos: se trata de una técnica utilizada para clasificar automáticamente un texto de acuerdo con un conjunto de palabras previamente designadas como positivas y negativas, así como otras vinculadas con emociones básicas. Las palabras tienen una puntuación para cada sentimiento y emoción, de tal modo que la frecuencia de esas palabras en los textos analizados se cuenta como el aumento de esos sentimientos y emociones. En el presente estudio se ha utilizado el léxico NRC (Mohammed et al., 2013). Cabe señalar que conocer qué tipo de sentimientos predominan en las reseñas no determina la aceptación o no del libro, lo cual se ve reflejado en las valoraciones de los libros que la misma plataforma ofrece.
- Etiquetado gramatical: se trata de conocer la categoría o clase (sustantivo, verbo, adjetivo, etc.) a la que corresponde una palabra, así como su raíz o lexema, y el posicionamiento sintáctico en una frase. En este caso se usó el paquete *udpipe* (Wijffels, 2022) que dispone de modelos de inteligencia artificial preentrenados para varios idiomas -entre los que se encuentra el español-, los cuales procesan el texto proporcionando las características mencionadas.

- Bigramas y redes de palabras: al separar las palabras por pares, la relación funciona igual que el principio de la cadena de Márkov, es decir, revela el orden secuencial en el que esas palabras se dan en el texto, posibilitando una comprensión más detallada de su uso. A partir de los bigramas, se crearon redes de palabras presentadas como nodos que permiten observar múltiples relaciones. Observando estas redes, y de manera manual, se han formado tópicos, según el modo en que algunos bigramas se unen y forman conjuntos temáticos, mientras otros permanecen más dispersos.

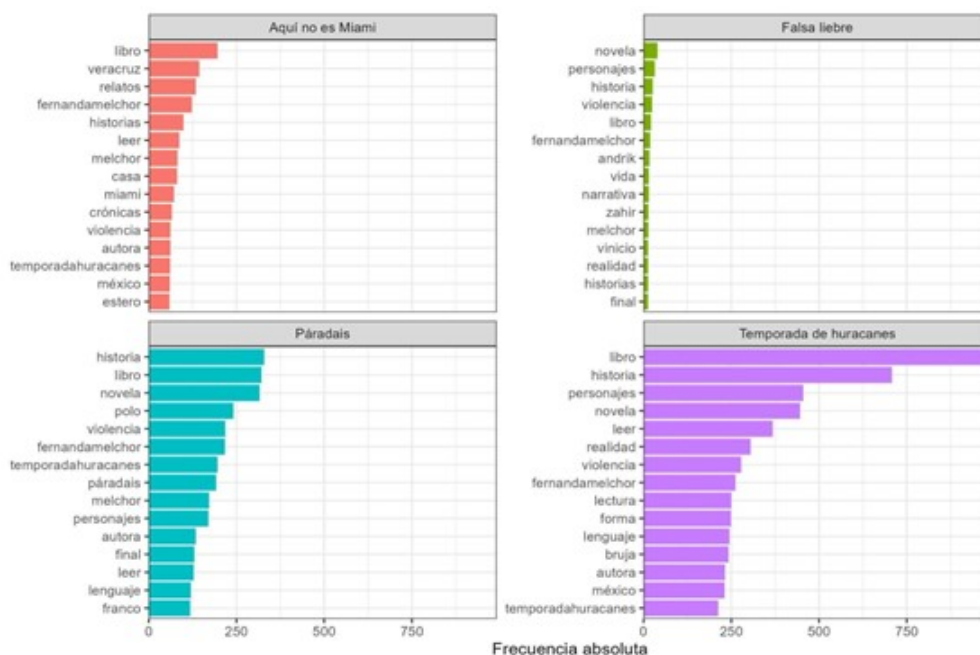
- Modelación por tópicos: se trata de una técnica de *machine learning* que busca construir tópicos o temas con base en la distribución de palabras en un conjunto de textos (Becerra y López-Alurralde, 2021). Para construir los tópicos de la muestra, se usó el modelo Latent Dirichlet Allocation (Blei et al., 2003), que se encuentra en el paquete *topicmodels* (Grün y Hornik, 2021). Esta técnica automática de modelado se ha utilizado de manera complementaria al proceso anterior, que ya arrojó un primer acercamiento al modelado del corpus, esperando obtener un enfoque más amplio. Para configurar el algoritmo, es preciso fijar el valor de k, esto es, el número de tópicos a generar. Dado que este valor influye mucho en el resultado, ha sido calculado mediante la medida denominada perplejidad *-perplexity* (Blei et al., 2003)-, obteniendo que k=4 es un buen valor de configuración.

El uso de métodos digitales y técnicas de minería de textos en Goodreads no es nuevo; ya existen investigaciones que analizan sentimientos para identificar reacciones sobre los libros (Driscoll y Rehberg-Sedo, 2019); que utilizan herramientas como Python y Excel para extracción y procesamiento de datos (Sánchez-García et al, 2021); o que analizan exploratoria y cuantitativamente la participación de usuarios en Goodreads (García-Roca, 2020). Sin embargo, en este trabajo se propone una metodología que haga uso de distintas técnicas para un análisis desde distintas perspectivas.

Resultados

Figura 1. Frecuencia absoluta

Nota. Palabras más frecuentes en las reseñas de Goodreads de las novelas de Fernanda Melchor.

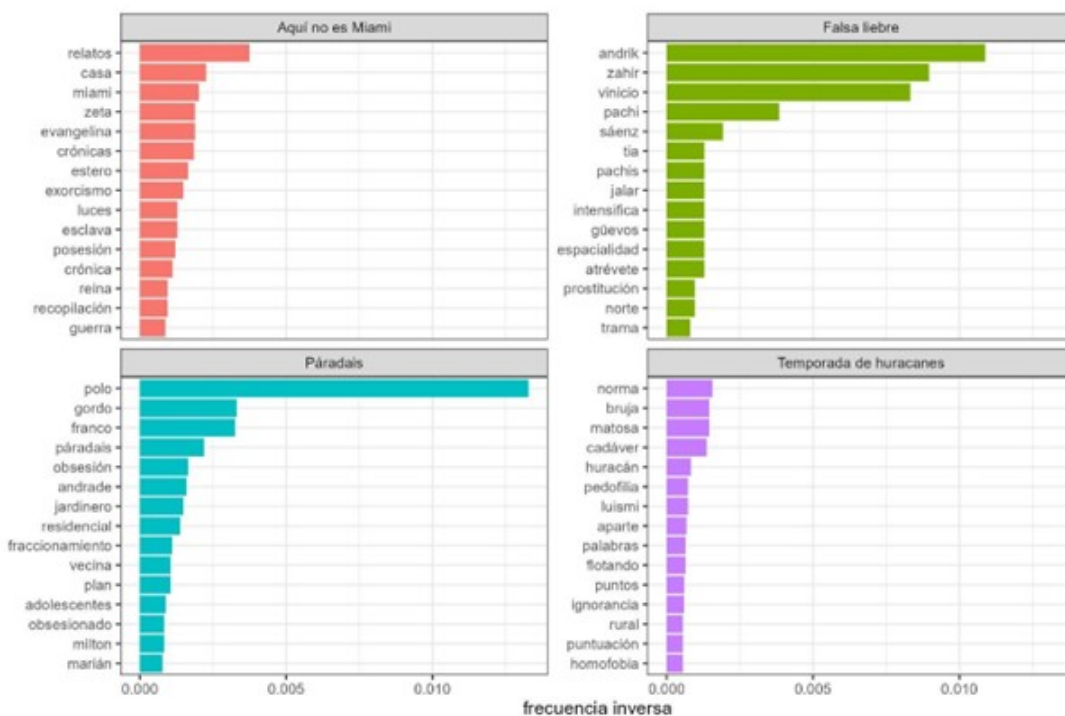


Fuente: Elaboración propia mediante programación en R.

Con respecto a la frecuencia absoluta de las palabras más repetidas en las reseñas de las cuatro novelas, en la [figura 1](#) se muestra la comparativa con las 15 palabras más frecuentes. Puede observarse que las palabras más repetidas son “libro”, “historia” y “novela”, las cuales hacen referencia a la descripción del libro, con una estructura muy parecida en los cuatro conjuntos de reseñas. En el caso de *Temporada de huracanes*, se menciona también “violencia” (278), “lenguaje” (245), “México” (231), y algunos personajes como “bruja” (242). Para *Páradais*, también se hace referencia a “violencia” (218) y “lenguaje” (120), lo que parece ser una constante en los libros de Melchor, aunque también destacan “personajes” (170) como “polo” (241), el “final” (130) de la novela y referencias a la “autora” (134). De igual manera, la mención de “temporada de huracanes” se repite en *Aquí no es Miami* (61) y *Páradais* (196), lo cual refleja que sus lectores han leído también *Temporada de huracanes* y lo comentan en las reseñas de estos libros.

Figura 2. Frecuencia inversa (TF-IDF)

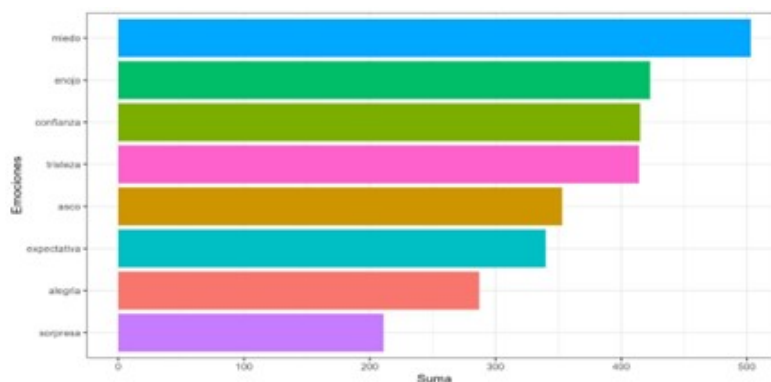
Nota. Quince palabras más frecuentes según el cálculo TF-IDF o frecuencia inversa de documento para las reseñas seleccionadas.



Fuente: Elaboración propia mediante programación en R.

Al hacer la comparación de la frecuencia inversa de palabras, se busca medir la importancia de una palabra en la muestra al disminuir las palabras de uso común y aumentar las relacionadas con las reseñas de cada libro. En este caso, cuanto mayor sea el valor de frecuencia inversa en la gráfica, esa palabra aparecerá menos en los demás conjuntos de reseñas (por libro). En la [figura 2](#), se observa que las palabras con frecuencia inversa mayor están asociadas a personajes y lugares que aparecen en cada libro: “norma”, “bruja”, “matosa”, “luismi” en *Temporada de huracanes*; “polo”, “franco”, “residencial” en *Páradais*; “evangelina”, “miami”, “estero” en *Aquí no es Miami*; “andrik”, “zahir”, “vinicio” en *Falsa liebre*; y otras palabras como “relatos” y “crónicas” en *Aquí no es Miami*, que a diferencia de los otros tres libros no es una novela. Esto es de esperarse, pues en las reseñas se habla de los personajes, los lugares y los elementos del libro que son significativos para cada uno; sin embargo, es más interesante observar cómo también se mencionan palabras que denotan una interpretación del lector sobre temas no tan explícitos de las novelas. En el caso de *Temporada de huracanes* se habla sobre “homofobia”, “pedofilia” e “ignorancia”, lo cual denota una interpretación más profunda del libro, al referir importantes temas que subyacen en la novela. Estos lectores aportan su reflexión e interpretación literaria y personal a la comunidad de Goodreads, estimulando lo que Pierre Lévy (1997) llamaría una inteligencia colectiva, que ayuda a los usuarios a profundizar en el texto de un modo muy diferente al individual ([García-Roca, 2016](#)).

Figura 3. Reconocimiento de emociones



Fuente: Elaboración propia mediante programación en R.

En la figura 3 se observa el conteo de emociones en todas las reseñas de los libros de Melchor. Destaca miedo, seguido de enojo, confianza, tristeza y asco, seguidos por expectativa, alegría y sorpresa. Las palabras que más destacan para la emoción de miedo son “violencia”, “pobreza”, “misericordia”, “duda”, y “bruja”, que hacen referencia no tanto a la percepción de los libros, sino a su temática. En el caso de enojo, los resultados son muy similares, puesto que “violencia”, “bruja” y “asesinato” aportan al puntaje de ambas emociones. Al revisar las palabras que se relacionan con confianza, “real”, “franco” (personaje de una novela) y “fuerte” son las que tienen mayor frecuencia. Por último, cabe señalar que la estructura de emociones es muy similar en las reseñas de todos los libros, predominando palabras como “violencia” y “asesinato”, que definen muy bien la trama común de las novelas de Melchor.

Figura 4. Nube de palabras de sentimientos

Nota. Sentimientos positivos -en azul- y negativos -en rojo- en las reseñas de los libros de Fernanda Melchor

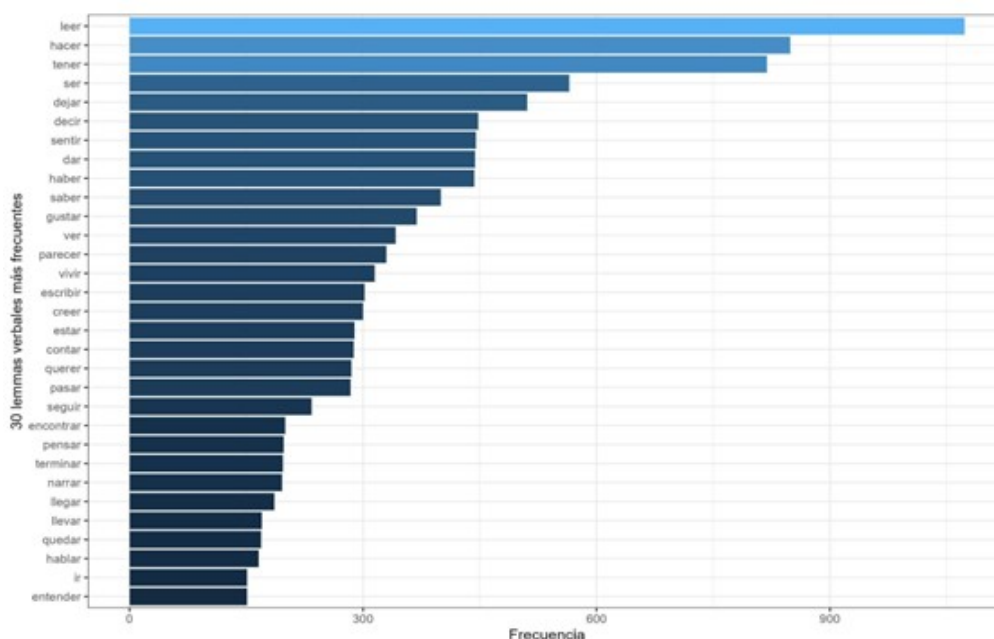


Fuente: Elaboración propia mediante programación en R.

Para visualizar la carga de sentimientos se ha decidido usar una nube de palabras, a diferencia de la información anterior mostrada en tablas, porque se puede apreciar mejor el

impacto de las palabras, como un conjunto de significados, más que de manera jerárquica, siendo las de mayor tamaño las de más frecuencia, las azules corresponden a sentimientos positivos y las rojas a negativos. Como se puede apreciar en la [figura 4](#), la carga de sentimientos negativos es muy fuerte, motivado por la presencia de términos como “violencia”, “pobreza”, “misericordia”, “miedo”, “difícil”, “cruel” y “crudo”. Estas palabras están vinculadas a contenidos y problemáticas de los libros de la autora. Por ejemplo, al hablar sobre violencia y pobreza o, de manera más específica en *Temporada de huracanes*, al referirse al personaje de la bruja, no cabe duda de que al conversar sobre esa temática los comentarios van a tener una carga negativa. Ahora bien, también aparecen en el análisis términos con connotaciones positivas. Curiosamente, son aquellos que no se relacionan tanto con los contenidos, sino mayormente con la experiencia de lectura, encontrando palabras como: “forma”, “estilo”, “lector”, “ritmo”, “sensación”, “leyendo”, “interesante”, “contar”, “perspectiva”.

Figura 5. Lemmas verbales más frecuentes



Fuente: Elaboración propia mediante programación en R.

Se procede a analizar las principales entidades gramaticales usadas en las reseñas, donde destacan los sustantivos -recuérdese que la palabra más frecuente de la muestra es “libro”-, signos de puntuación, determinantes, preposiciones, verbos, pronombres, adjetivos y adverbios. Se ha decidido profundizar en los verbos utilizados, para conocer el contenido de las reseñas respecto al comportamiento de los lectores: ¿qué reacciones y acciones provoca la lectura en los lectores?

Para ello, se han tomado las raíces léxicas de los verbos más frecuentes de las reseñas y se han graficado en la [figura 5](#) de mayor a menor. El verbo “leer” se encuentra en primer lugar, evidenciando que las reseñas son un reflejo de la experiencia de lectura de cada uno de los usuarios. Tras él, se observan verbos como “sentir”, “gustar” y “saber”, que también hablan sobre una percepción personal compartida por los usuarios en las reseñas. En el caso de “pensar”, “mostrar”, “entender” y “encontrar” es posible que refieran a lo que el libro significa para cada lector y cómo la recepción puede trascender a otros aspectos de su vida. La escritura de las reseñas enfatiza lo que los libros hacen sentir a los lectores, reflejando una relación personal

con el texto que se lee, hablando sobre la experiencia propia de cada lector. Así, los verbos más frecuentes son los que usualmente se utilizan para describir impresiones de un libro, hablar sobre el conocimiento previo o adquirido por medio de la lectura, o bien las expectativas de los lectores, en un rico intercambio de experiencias personales y metacognitivas.

Modelado manual y automático de tópicos

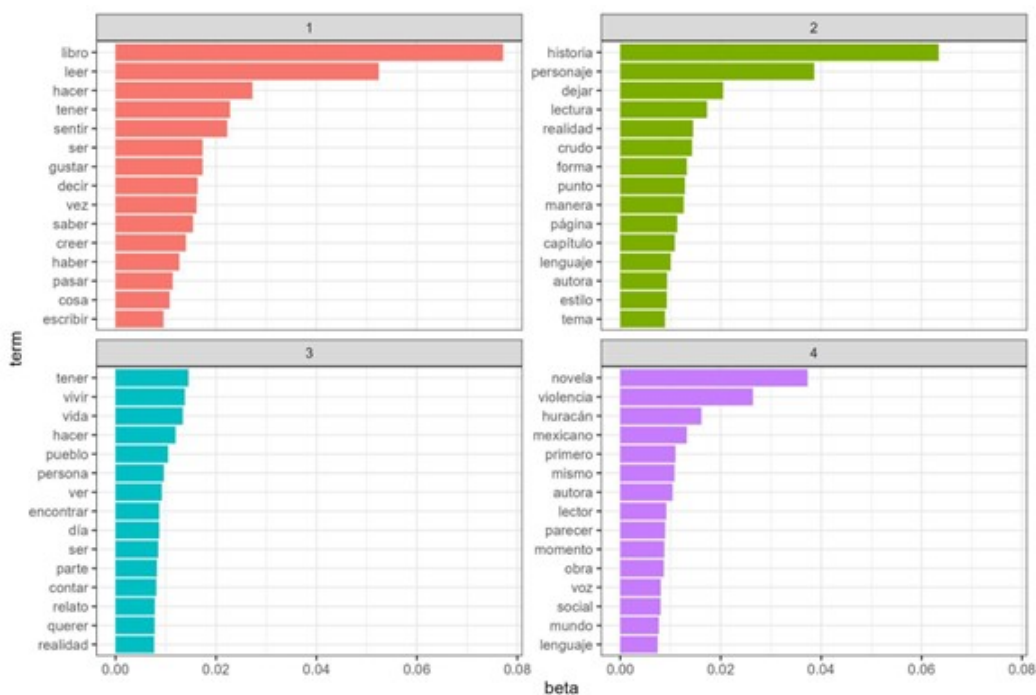
Tabla 2. Aspectos más relevantes conformados por pares de palabras

Aspectos	Bigramas o pares de palabras que lo conforman
<i>Literarios y de escritura del libro</i>	"estructura narrativa", "estilo literario", "voces narrativas", "párrafos interminables", "capítulos largos", "personaje principal", "lenguaje: vulgar, popular, soez, coloquial, crudo, fuerte", "ritmo: trepidante, vertiginoso", "oraciones largas", "forma magistral", "literatura latinoamericana"
<i>Descripción del libro</i>	"libro: incómodo, duro, corto, crudo, fácil, escrito, excelente", "mal libro", "novela: brutal, corta", "historia: gira, contada, narrada, real", "autora: describe, utiliza, logra", "atmósfera asfixiante", "primeras páginas", "modismos mexicanos"
<i>Experiencia de lectura</i>	"sabor amargo", "malas vibras", "sentimientos encontrados", "seguiré leyendo", "quería seguir", "quiero leer", "leí temporada", "recomiendo, disfruté, gustó: mucho", "pude parar", "puedes sentir", "podía dejar", "imposible dejar", "lee rápido", "deja respirar", "3, 4, 5: estrellas", "hubiera gustado", "relato favorito"
<i>Temas sociales</i>	"violencia, abuso: sexual", "tanta: desgracia, violencia", "abuso sexual", "miseria, condición, decadencia: humana", "crimen organizado", "hechos reales", "clases, problemas, diferencias: sociales", "conservadora, clasista", "clase: alta, baja", "desigualdad, contexto, crítica: social", "droga prostitución", "méxico violento", "veracruz actual"
<i>Referencias externas</i>	"gabriel garcía márquez", "daniela aedo", "random house", "pedro páramo", "vargas llosa", "mariana enríquez", "mel gibson"

Nota. Esta tabla es el resultado de la descomposición en bigramas de las reseñas de los libros de Fernanda Melchor.

Los pares de palabras o bigramas más frecuentes ayudan a entender mejor los temas principales que se discuten en las reseñas, proporcionando secuencias léxicas más amplias que los términos aislados. En la [tabla 2](#) estos bigramas fueron agrupados para poder identificar, de manera manual, estos aspectos generales que conforman las reseñas. En primer lugar, están los aspectos literarios y de escritura de los libros, que hacen referencia a la forma, voz, estilo, narrativa, lenguaje y ritmo de los libros, así como a los personajes, párrafos y capítulos, y en general a la literatura latinoamericana. La descripción de la experiencia de lectura es otro tópico frecuente, expresado a través de palabras como "incómodo", "brutal", "crudo" y "fácil", así como mediante la referencia directa a la autora y la manera en la que escribe, la atmósfera que crea, y el uso de modismos. La experiencia de lectura engloba todos aquellos elementos que tienen que ver con la recepción lectora y su interpretación de los libros: qué le pareció ("disfruté mucho"); cómo se sintió ("sabor amargo"); su forma de leerlo ("imposible dejar"); sus expectativas ("hubiera gustado"); o conclusiones propias ("recomiendo mucho"). Otro de los aspectos que destacan de los libros de Melchor, de acuerdo con las reseñas, son los temas sociales que toca: violencia, abuso, crimen, pobreza y desigualdad. Estos elementos son percibidos por los lectores como de gran relevancia, reflejando parte de la "vida cotidiana" y del "méxico violento". Por último, se logran identificar algunas referencias externas a escritores y novelas que los lectores relacionan con los libros de Melchor, como "garcía márquez" y "pedro páramo".

Figura 6. Modelación por tópicos



Fuente: Elaboración propia mediante programación en R.

Al aplicar la modelación por tópicos de manera automática mediante el algoritmo LDA, se obtuvo como resultado un conjunto de tópicos a los cuales se les asigna un número, que no refleja ningún tipo de significado ni prioridad (figura 6). Sin embargo, se han revisado estos resultados para asignar una etiqueta adecuada a cada tópico. Por ejemplo, en el caso del tópico 1, palabras como “leer”, “sentir”, “gustar”, “saber” y “creer”, hacen referencia a la experiencia de lectura que describe cada usuario al leer un libro en particular. En el tópico 2 destacan “historia”, “crudo”, “forma”, “capítulo”, “lenguaje”, “autora”, los cuales remiten a la descripción y a la estructura literaria de cada obra. El tercer tópico es más variado, pues las palabras “pueblo”, “vida”, “relato”, “realidad” y “encontrar” pueden hacer referencia a la percepción de los lectores sobre la temática de los libros. El cuarto tópico, con palabras como “violencia”, “mexicano”, “momento”, “voz”, “social”, “mundo” y “lenguaje”, hace referencia directa a las temáticas y aspectos sociales que trata cada libro. Hasta este punto, los resultados de la modelación por tópicos son bastante similares a los encontrados de manera manual en la separación de bigramas, porque, aunque no son exactamente los mismos, destacan los mismos puntos: experiencia de lectura, aspectos literarios y temática de cada libro (social, violencia, pobreza). La única diferencia significativa sería que en la modelación por tópicos no se mencionan las referencias externas a otros libros o escritores, lo cual puede considerarse un elemento importante para el análisis cualitativo, complementando el análisis automático y computacional.

Conclusiones

Las diferentes técnicas que se usaron para el análisis computacional de las reseñas de los libros de Fernanda Melchor demostraron ser útiles en su aplicación a un corpus extraído de una plataforma socio-digital como Goodreads, aplicando técnicas de procesamiento del lenguaje

natural. Especialmente, debido a la cercanía que las reseñas tienen a los libros que comentan, éstas contienen muchos elementos literarios, como lugares, adjetivos o personajes, combinados con experiencias, sentimientos y emociones de los lectores.

El análisis de frecuencias ha sido el primer acercamiento para el estudio cuantitativo del corpus, dando cuenta de los elementos principales de las reseñas. Aunque el investigador no los haya leído, adquiere una idea general sobre qué tratan. En este caso, las palabras más repetidas son: “libro”, “novela”, “leer” o “historia”, que hacen referencia a aspectos generales de la lectura. Complementariamente, la frecuencia inversa resalta componentes propios de cada libro reseñado, enfatizando su particularidad con respecto al corpus del que forman parte, resaltando temáticas y discusiones propias de cada historia. También da cuenta de la dinámica de esta red social, que genera diferentes lecturas debido a la multiplicidad de usuarios que la conforman, y de esta manera “el proceso de recepción no acaba en el hipotexto, sino que se amplía todo lo que el lector y la comunidad deseen” (García-Roca, 2016, p. 49).

A partir del reconocimiento de sentimientos y emociones se pueden identificar las principales palabras asociadas por los lectores a cada libro. Es importante destacar que, aunque las reseñas de los libros de Melchor están catalogadas como “negativas” o asociadas al “miedo”, no están relacionadas con la percepción o el gusto del lector por ese libro, sino con sus temáticas, que describen aspectos de la violencia, la pobreza y la desigualdad social. Esto se pudo deducir al analizar las palabras destacadas de cada emoción y conociendo el contenido de las novelas. Se podría decir que estos libros no acaban al terminar de leerlos, sino que trascienden en las concepciones de los lectores sobre temas como la violencia, pobreza y desigualdad, y así lo comunican en sus reseñas. Así, se manifiesta una de las funciones de la literatura, despertar y avivar la conciencia social: “allí donde la experiencia literaria del lector entra en el horizonte de expectativas de la práctica de su vida, performa su comprensión del mundo y repercute de ese modo en sus formas de comportamiento social” (Jauss, 2013, p. 200).

El etiquetado gramatical revela cómo se escriben las reseñas y cuál es su sintaxis. En este caso, los sustantivos y adjetivos son muy importantes, puesto que en su mayoría los usuarios hacen una reseña descriptiva de los libros y de diferentes aspectos literarios y de lectura. El análisis de los verbos más usados sirvió para entender la experiencia de lectura de cada usuario, su recepción del libro y lo que éste les proporciona. Se puede constatar que la escritura de reseñas es un acto performativo, supone que la reseña será leída, constituyendo una muestra de identidad social (Driscoll y Rehberg-Sedo, 2019). Los verbos también muestran la riqueza metacognitiva que se da en estos espacios de intercambio de experiencias, puesto que “el acto de leer se multiplica en la colaboración personalizada o anónima de las opiniones y consultas individuales” (García-Canclini, 2015, p. 21). La lectura es social y colaborativa en muchos aspectos, desde el momento en el que participan varios actores, influyendo en la recepción de los libros.

Para el etiquetado por temas se usaron dos métodos: la separación por bigramas para encontrar clústeres (manual) y la modelación por tópicos (automática). En ambos se obtienen resultados similares, compartiendo tres grandes tópicos: experiencia de lectura, aspectos literarios y la temática social de cada libro. La interpelación de estos temas muestra cómo el espacio virtual en el que se comparten las reseñas es, al mismo tiempo, un lugar de comunicación social y expresión personal, donde la lectura se difunde y se comenta de manera íntima, pero también crítica, de modo que “el hecho lector se transforma en conversación social e intercambio común” (Lluch, 2014). Esto sucede especialmente en el caso de Fernanda Melchor, dado el carácter marcadamente social de sus libros.

El presente estudio se ha realizado sobre una autora que consideramos de gran interés, sin embargo, hay otros autores a los que se podría aplicar el mismo estudio, esto nos permitiría comprobar cómo los resultados difieren mientras se mantiene la metodología. Además, un análisis profundo de los tópicos de las reseñas y las percepciones de los lectores podría complementarse mediante técnicas más cualitativas y etnográficas, como entrevistas o *focus group*, lo cual puede ser realizado en próximas investigaciones, profundizando en el alcance de los análisis cuantitativos aquí presentes.

Financiación

Convocatoria *Challenge Based Research Funding Program*. 2022_Project ID: E061-EHE-GI02-D-T3-E.

Referencias

- Alonso-Arévalo, J., & Cerdón-García, J. A. (2014). Lectura social, metadatos y visibilidad de la información. *XLV Jornadas Mexicanas de Bibliotecología*. <http://hdl.handle.net/10760/23095>
- Altamirano, C., & Sarlo, B. (1990). *Conceptos de sociología literaria*. Centro Editor de América Latina.
- Barnett, C. (2021, noviembre 9). Who owns Goodreads? *Book Riot*. <https://bookriot.com/who-owns-goodreads/>
- Becerra, G., & López-Alurralde, J. P. (2021). *Introducción a la ciencia de datos con R. Una guía rápida para científicos sociales*. Bookdown. https://bookdown.org/gaston_becerra/curso-intro-r/
- Benoit, K., Muhr, D., & Watanabe, K. (2021). *Stopwords: Multilingual stopword lists (2.3)*. <https://cran.r-project.org/web/packages/stopwords/stopwords.pdf>
- Blei, D. M., Ng, A., & Michael I., J. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3, 993-1022.
- Carroto, P. (2017, agosto 14). El otro 'boom' latinoamericano es femenino. *El País*. https://elpais.com/cultura/2017/08/13/actualidad/1502641791_807871.html
- Cerrillo, P. C., Larrañaga, E., & Yubero, S. (2002). *Libros, lectores y mediadores: La formación de los hábitos lectores como proceso de aprendizaje*. Ediciones de la Universidad de Castilla-La Mancha.
- Driscoll, B., & Rehberg-Sedo, D. (2019). Faraway, So Close: Seeing the intimacy in Goodreads reviews. *Qualitative Inquiry*, 25(3), 248-259. <https://doi.org/10.1177/1077800418801375>
- García-Canclini, N. (2015). Leer en papel y en pantallas: El giro Antropológico. En N. García-Canclini, V. Gerber-Bicecci, A. López-Ojeda, E. Nivón-Bolán, C. Pérez-Camacho, C. Pinochet-Cobos, & R. Winocur-Iparraguirre, *Hacia una antropología de los lectores* (pp. 1-37). D - Ediciones Culturales Paidós.
- García-Roca, A. (2016). Prácticas lectoras en espacios de afinidad: Formas participativas en la cultura digital. *Ocnos*, 15(1), 42-51. https://doi.org/10.18239/ocnos_2016.15.1.979
- García-Roca, A. (2020). Lectura virtualmente digital: El reto colectivo de interpretación textual. *Cinta de Moebius*, 67, 65-74. <https://doi.org/10.4067/s0717-554x2020000100065>
- Godínez-Rivas, G. L., & Nieto, L. R. (2019). Queers and bewitched: Temporada de huracanes of Fernanda Melchor. *Anclajes*, 23(3), 59-70. <https://doi.org/10.19137/anclajes-2019-2335>
- González Arceo, L. (2017, agosto 19). Libros B4 Tipos: Las chicas que están revolucionando la forma de leer autoras. *Malvestida*. <https://malvestida.com/2017/08/libros-b4-tipos-las-chicas-estan-revolucionando-la-forma-leer-autoras/>

- Grün, B., & Hornik, K. (2021). *Package «Topicmodels»*. *Topicmodels: Topic Models*. <https://cran.r-project.org/web/packages/topicmodels/index.html>
- Jauss, H. R. (2013). *La historia de la literatura como provocación*. Editorial Gredos.
- Lemus, R. (2022). En los márgenes de la nación: Geografías imaginarias en Temporada de huracanes (2017) de Fernanda Melchor. *Bulletin of Hispanic Studies*, 99(2), 163-170. <https://doi.org/10.3828/bhs.2022.12>
- Lévy, P. (1997). *L'intelligence collective: Pour une anthropologie du cyberspace*. La Découverte.
- Lluch, G. (2014). Jóvenes y adolescentes hablan de lectura en la red. *Ocnos*, 11, 7-20. https://doi.org/10.18239/ocnos_2014.11.01
- Manovich, L. (2020). *Cultural Analytics*. MIT Press.
- Mohammad, S. M., Kiritchenko, S., & Zhu, X. (2013). NRC-Canada: Building the state-of-the-art in sentiment analysis of tweets (arXiv:1308.6242). In *Proceedings of the seventh international workshop on Semantic Evaluation Exercises*, June 2013. arXiv. <http://arxiv.org/abs/1308.6242>
- Paladines-Paredes, L., & Aliagas, C. (2021). Videoreseñas de booktubers como espacios de mediación literaria. *Ocnos*, 20(1), 38-49. https://doi.org/10.18239/ocnos_2021.20.1.2489
- Parnell, C., & Driscoll, B. (2021). Institutions, platforms and the production of debut success in contemporary book culture. *Media International Australia*, 1329878X2110361. <https://doi.org/10.1177/1329878X211036192>
- Pouly, M.-P. (2016). Playing both sides of the field: The anatomy of a 'quality' bestseller. *Poetics*, 59, 20-34. <https://doi.org/10.1016/j.poetic.2016.02.002>
- Pressman, J., Marino, M. C., & Douglass, J. (2015). *Reading project: A collaborative analysis of William Poundstone's Project for Tachistoscope (Bottomless Pit)*. University of Iowa Press. <https://doi.org/10.2307/j.ctt20p598m>
- Quiroga, R. (2020). Lectoras mexicanas, el mercado más grande en español. *El Economista*. <https://www.economista.com.mx/arteseideas/Lectoras-mexicanas-el-mercado-mas-grande-en-espanol-20200722-0054.html>
- Saez, V. (2022). De lectores a "influencers". Booktubers, bookstagrammers y booktokers y la circulación de la literatura en redes sociales en Argentina. *Revista Pilquen – sección Ciencias Sociales*, 25(2), 020-046. <https://www.redalyc.org/journal/3475/347572703002/html/>
- Sánchez-García, P., Hernández-Ortega, J., & Rovira-Collado, J. (2021). Leyendo al lector social: Evolución de la literatura infantil y juvenil española en Goodreads. *Ocnos*, 20(1), 7-22. https://doi.org/10.18239/ocnos_2021.20.1.2446
- Silge, J., & Robinson, D. (2017). *Text mining with R: A tidy approach*. O'Reilly.
- Silge, J., & Robinson, D. (2022). *Package 'tidytext'*. *Tidytext: Text Mining using 'dplyr', 'ggplot2', and Other Tidy Tools*. <https://cran.r-project.org/web/packages/tidytext/index.html>
- Sparck-Jones, K. (1972). A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation*, 28(1), 11-21. <https://doi.org/10.1108/eb026526>
- Thelwall, M., & Kousha, K. (2017). Goodreads: A social network site for book readers. *Journal of the Association for Information Science and Technology*, 68(4), 972-983. <https://doi.org/10.1002/asi.23733>
- Toubia, O., Berger, J., & Eliashberg, J. (2021). How quantifying the shape of stories predicts their success. *Proceedings of the National Academy of Sciences of the United States of America*, 118(26). <https://doi.org/10.1073/pnas.2011695118>
- Wang, K., Liu, X., & Han, Y. (2019). Exploring Goodreads reviews for book impact assessment. *Journal of Informetrics*, 13(3), 874-886. <https://doi.org/10.1016/j.joi.2019.07.003>

- Wang, X., Yucesoy, B., Varol, O., Eliassi-Rad, T., & Barabási, A.-L. (2019). Success in books: Predicting book sales before publication. *EPJ Data Science*, 8(1), 31. <https://doi.org/10.1140/epjds/s13688-019-0208-6>
- Wijffels, J. (2022). Package 'udpipe'. *Udpipe: Tokenization, Parts of Speech Tagging, Lemmatization and Dependency Parsing with the 'UDPipe' «NLP» Toolkit*. <https://cran.r-project.org/web/packages/udpipe/udpipe.pdf>