



Ctrl + F

Texto, archivo

e interpretación

en entornos digitales

{ Maria José Afanador Llach,
Nicolás Vaughan };
Edición académica

Ctrl + F

Texto, archivo e interpretación en entornos digitales

Ctrl + F

Texto, archivo

e interpretación

en entornos digitales

{ Maria José Afanador Llach,
Nicolás Vaughan };
Edición académica

Nombre: Afanador-Llach, María José, edición académica. | Vaughan, Nicolás, edición académica.

Título: Ctrl + F : texto, archivo e interpretación en entornos digitales / María José Afanador Llach, Nicolás Vaughan, edición académica.

Descripción: Bogotá: Universidad de los Andes, Facultad de Artes y Humanidades, Ediciones Uniandes, 2026. | 430 páginas: ilustraciones; | (Colección Humanidades Digitales; 2)

Identificadores: ISBN 978-628-7903-08-1 (rústica) | 978-628-7903-09-8 (e-book) | 978-628-7903-11-1 (e-pub)

Materias: Humanidades digitales | Preservación digital | Medios digitales | Edición digital

Clasificación: CDD 001.30285--dc23

SBUA

Primera edición: abril del 2026

© María José Afanador Llach y
Nicolás Vaughan, autores compiladores

© Universidad de los Andes, Facultad de
Artes y Humanidades, Departamento de
Narrativas Digitales, Departamento
de Literatura

Ediciones Uniandes
Carrera 1.ª n.º 18A-12, bloque Tm
Bogotá, D. C., Colombia
Teléfono: 601 339 4949, ext. 2133
<http://ediciones.uniandes.edu.co>
ediciones@uniandes.edu.co

ISBN: 978-628-7903-08-1
ISBN *e-book*: 978-628-7903-09-8
ISBN *epub*: 978-628-7903-11-1
DOI: <http://doi.org/10.51573/Andes.9786287903081.9786287903098.9786287903111>

CORRECCIÓN DE ESTILO:

Laura Porras Montenegro

DISEÑO DE COLECCIÓN Y DIAGRAMACIÓN:

Boga Visual

**Universidad de los Andes |
Vigilada Mineducación.**

Reconocimiento como universidad: Decreto 1297 del 30 de mayo de 1964. Reconocimiento de personería jurídica: Resolución 28 del 23 de febrero de 1949, Minjusticia. Acreditación institucional de alta calidad, 10 años: Resolución 000194 del 16 de enero del 2025, Mineducación.

Todos los derechos reservados. Esta publicación no puede ser reproducida ni en su todo ni en sus partes, ni registrada en o transmitida por un sistema de recuperación de información, en ninguna forma ni por ningún medio, sea mecánico, fotoquímico, electrónico, magnético, electro-óptico, por fotocopia o cualquier otro, sin el permiso previo por escrito de la editorial.

CONTENIDO

Introducción

MARIA JOSÉ AFANADOR LLACH
NICOLÁS VAUGHAN

< pág. 10 />

{ Capítulo 1

Interpretación de textos por humanistas digitales
usando R: una aproximación metodológica

SILVIA MARGARITA BALDIRIS NAVARRO
CARLOS AMÓRTEGUI ÚSUGA
EDURNE SÁENZ

< pág. 24 />

Capítulo 2

Ser vendida a él, así comenzó nuestra historia de amor: un estudio discursivo asistido por corpus de una tendencia problemática en los fanfictions románticos en Wattpad

ANA MARÍA BERMEO

< pág. 56 />

Capítulo 3

Trabajar con ChatGPT para el etiquetado de textos
con TEI: un caso de estudio

INMACULADA UREÑA ASENSIO

< pág. 88 />

Capítulo 4

Edición académica digital mínima: entre la crítica
textual mínima y la edición académica a escala como
coordenadas de formación en humanidades digitales

CRISTIÁN SUÁREZ-GIRALDO

< pág. 166 />

Capítulo 5

CICLCOI: retos en la concepción y adaptación en el
modelamiento de una base de datos relacional

GEMMA BERNADÓ FERRER

< pág. 196 />

Capítulo 6

Padrones digitales históricos: más allá de una edición digital de textos

JORGE CACHIOTIS SALAZAR

PILAR LÓPEZ BEJARANO

< pág. 226 />

Capítulo 7

Aplicación de *Hand Written Text Recognition* a la transcripción de impresos antiguos y manuscritos

ANDRÉS FELIPE ECHAVARRÍA PELÁEZ

< pág. 268 />

Capítulo 8

Las huellas de la represión: recuperación y procesamiento de archivos militares producidos durante la dictadura uruguaya (1973-1985)

ELINA GÓMEZ BONAGLIA

< pág. 314 />

Capítulo 9

Colecciones fugaces: análisis de *ephemera* digital a través de Tropy

LUIS ARAMIS GARCÍA BENDEZÚ

< pág. 342 />

Capítulo 10

Archivos sonoros, inteligencia artificial y computación: una mirada a los aportes de la tecnología en la catalogación y creación de metadatos para historias orales

LILIANA PARÍS BORDA

< pág. 380 />

};

Sobre los autores y editores académicos

< pág. 422 />

Introducción

MARIA JOSÉ AFANADOR LLACH

NICOLÁS VAUGHAN

f Las bibliotecas y los archivos han sido las instituciones encargadas de gestionar, recolectar, preservar y difundir el conocimiento humano y la cultura. A partir de la segunda década del siglo xx, las instituciones de la memoria adoptaron los desarrollos de la computación, el internet, la World Wide Web y las tecnologías de digitalización, lo que propició un proceso de transformación digital en la gestión de la información. Desde la década de 1960, diferentes bibliotecas y organizaciones por todo el mundo comenzaron a crear índices bibliográficos digitales y a crear las primeras bases de datos para almacenar y recuperar información bibliográfica (Falk, 1982). En 1971, se estableció la primera biblioteca digital sobre la base de un proceso de digitalización de miles de libros en el Proyecto Gutenberg, en Estados Unidos (Cruz Quintana, 2022). De ahí en adelante, los flujos de digitalización de documentos se han convertido en una política de archivos y bibliotecas en todo el mundo. En el contexto hispanohablante, uno de los proyectos más tempranos de digitalización fue el del Archivo General de Indias

de Sevilla, que entre 1986 y 1992 digitalizó miles de documentos sobre la América española (García González y Más Bleda, 2010).

En los primeros proyectos de digitalización en archivos y bibliotecas, uno de los principales beneficios de crear representaciones digitales de los documentos fue la posibilidad de conservarlos en formato electrónico y así limitar su manipulación física. Con los proyectos de digitalización llegaron, además, los retos de construir repositorios digitales para ser consultados en terminales y más adelante vía internet. Las infraestructuras digitales para acceder a estos materiales inicialmente permitieron almacenar y recuperar registros bibliográficos y objetos digitalizados con sus metadatos. Otras posibilidades de automatización y uso de la computación para el trabajo con documentos comenzarían a desarrollarse desde las humanidades. Un hito bien conocido en el surgimiento del campo de las humanidades digitales en Estados Unidos es el proyecto del padre Roberto Busa. Busa es considerado pionero en el campo de las humanidades digitales, al utilizar computadores para la digitalización del extenso corpus de la obra de Tomás de Aquino, en la década de 1950.

Posteriormente, diferentes centros, asociaciones y publicaciones académicas comenzaron a formarse, articulando las posibilidades de la computación, en el estudio de la experiencia y cultura humanas. En 1962, se fundó la Association for Computational Linguistics y, en 1964, el Literary and Linguistic Computing Center en la Universidad de Cambridge en el Reino Unido. En 1978, se funda en Estados Unidos la Association for Computers and Humanities; en 1986, la revista académica *Literary and Linguistic Computing*, hoy llamada *Digital Scholarship in the Humanities*.

Al ser el análisis y la interpretación de textos un aspecto central de la investigación en humanidades, los investigadores de las humanidades, la lingüística y la computación comenzaron a pensar cómo los computadores podían ayudar en su estudio. A finales de la década de 1980, un grupo de investigadores de estas áreas, pertenecientes a las asociaciones mencionadas arriba, se reunieron para desarrollar la primera edición del estándar de la denominada *Text Encoding Initiative* (TEI). TEI es tanto un estándar de codificación y representación de textos en formato digital¹ como el consorcio que lo regula. La codificación de textos implica crear textos procesables por computador (lo que en inglés denominan *machine readable code*) para apoyar la investigación en humanidades (Alcaraz-Martínez y Vázquez, 2016). Más exactamente, implica indicar por medio de marcas o etiquetas qué significados tienen ciertas cadenas de texto plano (Vaughan, 2021). En el contexto de TEI, desde finales de la década de 1990 hasta el presente se ha producido un gran número de proyectos digitales en las humanidades.

Recientemente, en este campo se han explorado formas en que los corpus marcados en TEI pueden facilitar tareas asociadas al procesamiento del lenguaje natural (NLP, por sus siglas en inglés), que consiste en construir herramientas y algoritmos para el procesamiento automático del lenguaje. Asociado con el NLP está el uso de lenguajes de programación como R y Python, para tareas asociadas con el procesamiento, la transformación y el análisis de

1 En sus primeras ediciones, la codificación prescrita por TEI estaba construida sobre el lenguaje de marcado GML; luego, desde la edición P4, en el 2002, la codificación fue migrada a XML. TEI está hoy en su edición P5 (versión 4.9.0). Consulte el siguiente enlace <https://tei-c.org>

texto. La programación se ha convertido en una práctica central en las humanidades. En la actualidad, el uso de herramientas de inteligencia artificial (IA) ha permitido procesos automatizados como la indexación de textos, su digitalización y marcado, la transcripción automática y los modelos predictivos para apoyar diferentes procesos relacionados con la gestión de archivos digitalizados, la creación de metadatos, entre otros.

Sobre este volumen

Los investigadores de las disciplinas de las humanidades han construido sus interpretaciones sobre el pasado, la lengua, la literatura, el arte, el pensamiento y la experiencia humana a partir del estudio de documentos. La digitalización de documentos de diferentes épocas de la historia (o su nacimiento en digital), su procesamiento y publicación en línea han facilitado que los investigadores anoten, estudien y representen sus objetos de estudio con el apoyo de flujos de trabajo computacional, metodologías y herramientas digitales. El segundo volumen de la Colección Humanidades Digitales de la Facultad de Artes y Humanidades de la Universidad de los Andes pretende acercarse a proyectos de investigación en humanidades que exploran problemas y posibilidades de digitalización y procesamiento de archivos, creación de ediciones críticas digitales y análisis de textos computacionales.

El presente volumen reúne diez contribuciones de investigadoras e investigadores de América Latina y España. Los capítulos abordan enfoques pedagógicos para el uso de lenguajes de programación en el análisis de textos y describen proyectos de edición

digital con anotación digital de XML-TEI y HTML, en cursos de formación universitaria a nivel de pregrado. Las contribuciones incluyen reflexiones sobre proyectos de modelado de bases de datos de corpus de incunables, el uso del reconocimiento automatizado de letra manuscrita (*handwritten text recognition*), para la transcripción de impresos antiguos y manuscritos complejos, y la edición digital de textos del siglo XVIII. El volumen incluye el aporte del uso de metodologías de lingüística de corpus para el análisis literario en plataformas digitales y el uso de sistemas de IA para el trabajo con marcado TEI. En efecto, los capítulos de este volumen exploran temas de vanguardia en la investigación con texto en humanidades digitales. Por último, el volumen incluye reflexiones teóricas, epistemológicas y metodológicas sobre el trabajo de catalogación y análisis de *ephemera* digitales², archivos militares digitalizados y archivos sonoros de memoria.

El capítulo primero, titulado “Interpretación de textos por humanistas digitales usando R: una aproximación metodológica”, escrito por Silvia Margarita Baldiris Navarro, Carlos Amórtegui Úsuga y Edurne Sáenz, aborda el análisis automatizado de textos a través del lenguaje de programación R desde un enfoque pedagógico. Mediante dos ejemplos concretos —uno basado en el libro *Hannibal the Carthaginian* y el otro en la tragedia *Hamlet* de Shakespeare—, los autores demuestran cómo un estudiante lego puede iniciar el procesamiento de un texto desde cero y, con el

2 Los *ephemera* digitales son colecciones de datos que están en la web, cuya naturaleza misma los hace susceptibles a la desaparición. Entre estos se encuentran, por ejemplo, entradas de *blogs*, correos electrónicos o páginas web de la desaparecida Geocities. Internet Archive (<https://archive.org>) ha sido un pionero en la preservación de estas colecciones.

uso de las herramientas estadísticas de R, llegar a extraer conclusiones significativas sobre su contenido.

El capítulo segundo, titulado “*Ser vendida a él, así comenzó nuestra historia de amor: un estudio discursivo asistido por corpus de una tendencia problemática en los fanfictions románticos en Wattpad*”, de Ana María Bermeo, explora el tema desde el prisma de la lingüística de corpus. El estudio se centra en un corpus de textos extraídos de Wattpad, una plataforma digital que alberga miles de ejemplos de lo que hoy se conoce como *fanfiction*. Los textos seleccionados para el estudio son los denominados “romances captor-captiva”, en los que la protagonista se enamora de su secuestrador. Utilizando la herramienta digital de análisis de corpus AntConc, la autora demuestra cómo esta metodología estadística de lectura a distancia permite concluir que la mayoría de los textos del corpus tienden a objetivar a la protagonista, dejándola a merced de la voluntad de su captor, que suele estar vinculado a una mafia.

Los dos siguientes capítulos discuten el marcado o etiquetado semántico de textos con TEI-XML. El capítulo tercero, escrito por Inmaculada Ureña, se titula “Trabajando con ChatGPT para el etiquetado de textos con TEI: un caso de estudio”. Tras hacer una rápida descripción de la sintaxis del lenguaje TEI —cómo se marcan las divisiones de sección, los títulos de estas, etc.—, se discute cómo podría usarse un *chatbot* basado en un modelo de lenguaje grande (LLM, por sus siglas en inglés), tal como ChatGPT, para automatizar una parte del marcado semántico en TEI de un texto. Si bien la autora muestra la no predictibilidad (o irregularidad) del marcado resultante, señalan que estas debilidades no superan las ventajas de la utilización de la herramienta. Finalmente, la autora

discute y problematiza la posibilidad de usar dichos LLM para alinear diferentes versiones de un mismo texto (en el caso estudiado, un original latino y dos traducciones, en español e inglés); esto es, la posibilidad de automatizar la identificación de unidades sintácticas (oraciones, párrafos, etc.) en función de su significado.

El capítulo cuarto, “Edición académica digital mínima: entre la crítica textual mínima y la edición académica a escala como coordenadas de formación en humanidades digitales”, de Cristian Suárez-Giraldo, se aproxima al asunto desde una tendencia teórica reciente denominada *minimal computing*. En este, el autor busca responder a la pregunta por cómo enseñar la creación de una “edición mínima digital” —por ejemplo, una que haga uso de herramientas digitales de código y acceso abierto, y que permita la publicación de los resultados en una plataforma web que no requiera de *frameworks JS*— a estudiantes de pregrado. Partiendo de la selección del material a editar, se discuten paso a paso las dificultades y oportunidades de la enseñanza del proceso de edición digital: escaneo y digitalización, marcado semántico y publicación del documento HTML final. El autor resalta de este modo la utilidad del *minimal computing* tanto para el proceso de edición como para, lo que quizás es más importante, su aprendizaje.

El capítulo quinto, titulado “CICLCOI: retos en la concepción y adaptación en el modelamiento de una base de datos relacional”, de Gemma Bernadó Ferrer, discute los retos y dificultades de la creación de una base de datos bibliográfica, correctamente estructurada, de las ediciones incunables existentes en Colombia. La autora parte del trabajo realizado por la Biblioteca de Ediciones de Clásicos Latinos en el Renacimiento (BECLAR), para describir la creación e implementación del Corpus de Incunables de autores

Clásicos Latinos en Colombia (CICLCO), discutiendo las dificultades en el modelaje y la estructuración de la base de datos para el caso idiosincrásico de los incunables en Colombia.

El capítulo sexto, titulado “Padrones digitales históricos: más allá de una edición digital del texto”, de Pilar López Bejarano y Jorge Cachiotis Salazar, reflexiona sobre la edición digital de una serie de centros históricos de población de finales del siglo XVIII. Los autores presentan y reflexionan sobre el proceso metodológico para las transcripciones y su codificación mediante marcado TEI/XML para la construcción de una base de datos de consulta. El capítulo va más allá de presentar un proceso metodológico, interesándose por reflexionar sobre la investigación en la disciplina de la historia. Los autores se preguntan por lo que implica presentar el archivo histórico como datos y cuestionan hasta qué punto los asuntos técnicos son al mismo tiempo problemas epistemológicos.

El capítulo séptimo, titulado “Aplicación del *Hand Written Text Recognition* (HTR) a la transcripción de impresos antiguos y manuscritos complejos”, de Andrés Felipe Echavarría Peláez, se ocupa del uso de la computación para la transcripción de manuscritos antiguos. Ante la complejidad de los manuscritos —la caligrafía, la calidad de la tinta y los tipos de fuentes, entre otros—, el autor muestra que las tecnologías digitales pueden ser utilizadas para asistir un proceso que, con todo, no se puede automatizar por completo. Para esto, el capítulo aborda los conceptos básicos de los modelos de transcripción como HTR y presenta un ejemplo de un modelo de transcripción con el programa *Transkribus*. Además, explica los procesos de obtención y la caracterización del corpus, el establecimiento de normas de transcripción, la selección de un

modelo de transcripción, el proceso de segmentación del texto y la aplicación y evaluación del modelo de HTR.

Los últimos tres capítulos del volumen se ocupan de problemas relacionados con gestión, catalogación y análisis de distintos tipos de archivos. El capítulo octavo, “Las huellas de la represión. Recuperación y procesamiento de archivos militares producidos durante la dictadura uruguaya (1973-1985)”, de Elina Gómez Bonaglia, se ocupa del problema de la recuperación de documentación relacionada con los órganos de inteligencia militar en el Cono Sur. Parte de una indagación teórica para iluminar la discusión epistemológica y técnica, el capítulo discute la dimensión política del archivo e incluye una mirada a los desafíos interpretativos sobre los archivos militares, en su mayoría textuales. Los contextos de producción y de comprensión del archivo, plantea la autora, pasan por problemas relacionados con los principios de verdad y literalidad. La propuesta, en relación con el caso de estudio propuesto, pretende pensar las herramientas tecnológicas como formas de asistir el procesamiento y lo que la autora llama el “des-silenciamiento” de los archivos históricos.

El capítulo continúa develando el proceso llevado a cabo con los archivos militares, definiendo un flujo de trabajo que parte desde la digitalización, la catalogación de los materiales, la anotación, los modelos de transcripción automática hasta las herramientas de extracción de información. Finalmente, el capítulo reflexiona sobre la noción de *dato*, el procesamiento de lenguaje natural y la dimensión política en la noción del *archivo*, y ubica la discusión en una dimensión epistemológica, relevante y necesaria en la discusión sobre historia y humanidades digitales.

El capítulo noveno, titulado “Colecciones fugaces: catalogando *ephemera* digital a través de Tropy”, de Luis Aramis García Ben-dezú, presenta una estrategia metodológica con dicha herramienta digital para organizar y analizar contenidos digitales efímeros en Instagram en relación con su valor documental, su naturaleza archivística y sus rasgos narrativos. Al partir de contenidos digitales producidos por instituciones de la memoria, el autor recopiló, archivó y catalogó los contenidos, con el fin de trabajarlos como *ephemera* nacidos digitales con una aplicación de código abierto. El autor se pregunta por la relevancia de estos objetos para las ciencias de la información, la investigación histórica o cualquier campo interesado en estudiar el registro cultural digital. Además, la reflexión incluye preguntas por los desafíos técnicos, éticos y legales de los objetos digitales y de las características narrativas de las historias de Instagram, para presentar una propuesta de catalogación de contenidos multimodales.

Por último, el capítulo décimo, titulado “Archivos sonoros, inteligencia artificial y computación: una mirada a los aportes de la tecnología en la catalogación y creación de metadatos para historias orales”, de Liliana París Borda, explora el problema de la creación de metadatos en archivos sonoros y las posibilidades de utilizar métodos computacionales e IA para este problema. La autora presenta un análisis utilizando diferentes herramientas de código abierto para evaluar su utilidad en la catalogación y generación de metadatos sobre el contenido de archivos sonoros, específicamente entrevistas. Partiendo de una contextualización sobre la IA y sus usos, el capítulo describe un caso de estudio que comienza con los procesos de transcripción, el procesamiento y el reconocimiento de emociones en la oralidad. Finalmente, la

autora reflexiona sobre las implicaciones y oportunidades del uso de IA en la catalogación y creación de archivos.

En el contexto hispanohablante, aún es necesario investigar genealogías propias de la aplicación de la computación y el uso de los medios digitales para las humanidades. En los departamentos de literatura, lingüística e historia en América Latina y España, con seguridad podremos encontrar iniciativas e investigaciones que han explorado caminos metodológicos que podríamos situar en la historia de las humanidades digitales no anglófonas. A su vez, en las instituciones de la memoria como los archivos, los museos y las bibliotecas, podemos rastrear las historias de digitalización y procesamiento de archivos. Mientras investigamos antecedentes y contextos que iluminen los desarrollos de las humanidades digitales hispanohablantes del pasado, este volumen presenta una mirada actual sobre investigaciones y proyectos digitales que articulan el problema del texto, el archivo y la interpretación en entornos digitales.

Nota sobre la edición

A menos que se indique lo contrario, todas las gráficas, las imágenes, las tablas y los listados de código son de elaboración propia de los autores.

Agradecimientos

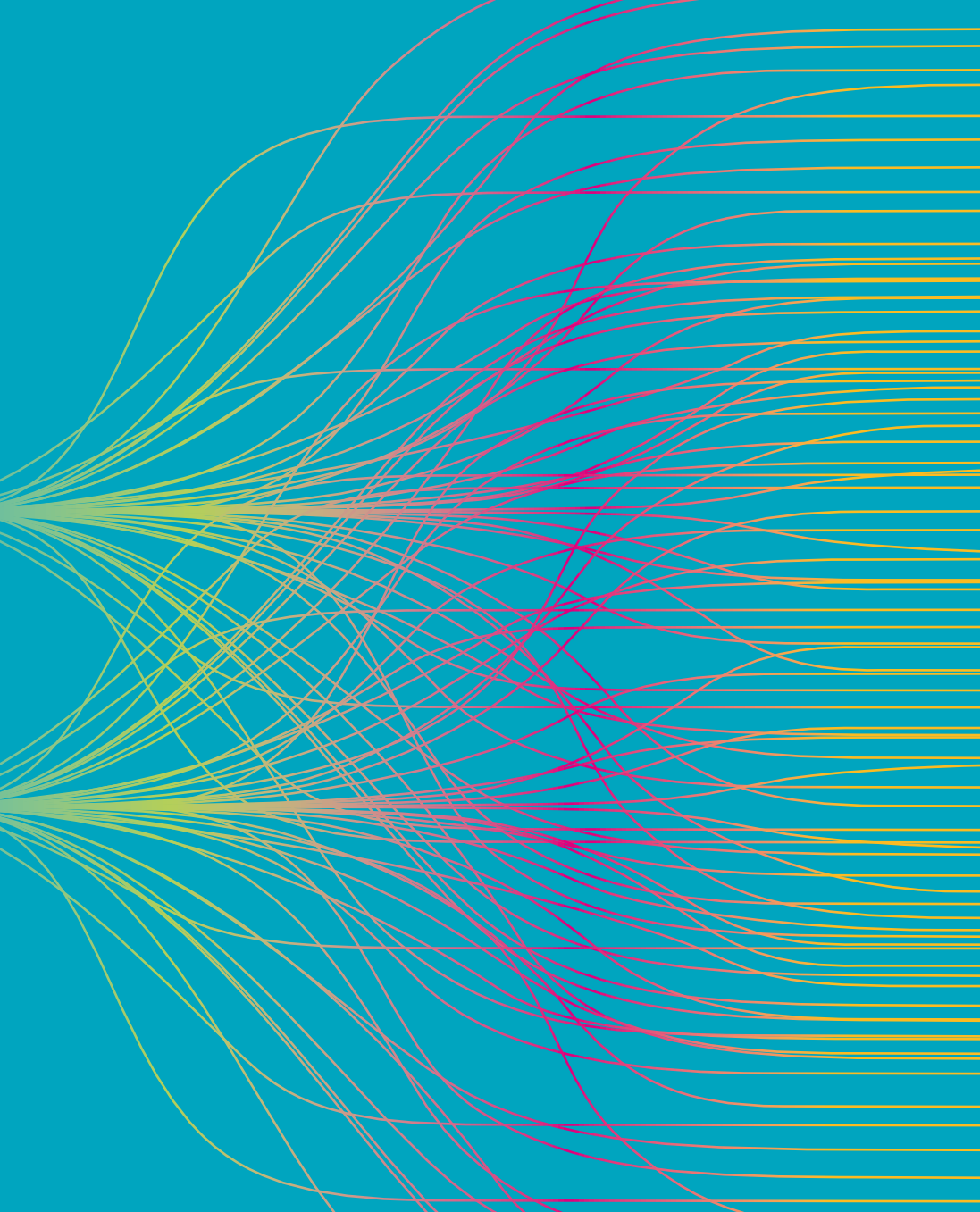
Queremos agradecer al equipo de Ediciones Uniandes, especialmente a María Ortiz y Carolina Mazo, por el minucioso trabajo en la revisión, edición y publicación del presente volumen, así como al editor general de la Universidad de los Andes, Juan Camilo González, por su apoyo y acompañamiento en la gestación de la Colección de Humanidades Digitales. También queremos agradecer al equipo de Boga Visual, responsables del diseño de la maqueta de la Colección y la diagramación de este volumen. Finalmente, agradecemos a los pares evaluadores del volumen y a la Facultad de Artes y Humanidades de la Universidad de los Andes.



Referencias

- Alcaraz-Martínez, R. y Vázquez, E. (2016, 1.º de diciembre). TEI: un estándar para codificar textos en el ámbito de las humanidades digitales. *BiD: Textos Universitaris de Biblioteconomia i Documentació*.
- Cruz Quintana, F. (2022). *Panorama histórico del libro y la edición digital*. Universidad de los Andes, Universidad de Guadalajara, Universidad Nacional de Villa María y Pontificia Universidad Católica de Chile.
- Falk, J. D. (1982). In search of history: The bibliographic databases. *The History Teacher*, 15(4), 523-544. <https://doi.org/10.2307/493442>

- García González, M. y Más Bleda, A. (2010). Acceso a los fondos del Archivo General de Indias a través de la plataforma “PARES” (Portal de Archivos Españoles). *Naveg@mérica. Revista Electrónica Editada por la Asociación Española de Americanistas*, (5).
- Vaughan, N. (2021). Introducción a la codificación de textos en TEI (parte 1). *Programming Historian en Español*, 5. <https://doi.org/10.46430/phes0053>



CAPÍTULO 1

Interpretación de textos por humanistas digitales usando R: una aproximación metodológica

Para citar este capítulo:

<http://doi.org/10.51573/Andes.9786287903081.9786287903098.9786287903111.1>

SILVIA MARGARITA BALDIRIS NAVARRO

CARLOS AMÓRTEGUI ÚSUGA

EDURNE SÁENZ

Resumen

Las humanidades digitales emergen como un campo de investigación en rápido crecimiento. Este campo ofrece una amplia gama de posibilidades como lo es el análisis de textos, proceso que permite acceder a información en una vasta cantidad de literatura de diferentes épocas, contextos y autores. Sin embargo, esta técnica no es tarea fácil para los humanistas, ya que requiere habilidades técnicas específicas. Dentro del escenario de la enseñanza, un *software* muy utilizado para ejecutar dicha acción por parte de programadores y no programadores es RStudio, un *software* estadístico que representa un desafío en el binomio educador/estudiante por su complejidad. En el presente capítulo se propone una solución innovadora al problema que, en lugar de centrarse en la programación en sí, plantea un enfoque didáctico enfocado en la definición de preguntas sobre los textos analizados y en la interpretación de los resultados del análisis usando R, con el propósito de acercar al estudiante al análisis de texto. Se ofrece una evaluación cualitativa del enfoque que arroja resultados prometedores.

PALABRAS CLAVE: R, análisis de textos, humanidades digitales, enseñar R.



Introducción

f Las humanidades digitales se han convertido en un área de investigación en constante crecimiento, con una amplia gama de posibilidades de realización, entre ellas, el análisis literario o el musical, la conservación del patrimonio, la estilometría, el análisis geoespacial o la realidad aumentada y virtual (Schreibman *et al.*, 2016).

Como una rama de las humanidades digitales, el análisis de textos se convierte en una técnica novedosa para los humanistas, ya que brinda la posibilidad de descubrir conocimientos en una enorme cantidad de literatura de diferentes épocas, contextos o autores. Sin embargo, el análisis de textos exige habilidades técnicas específicas en los humanistas para poder tomarlo en sus propias manos y explotarlo adecuadamente (Wiedemann y Niekler, 2017). En este sentido, enseñar herramientas como los lenguajes de programación a no programadores, como es el caso de los humanistas digitales, es realmente desafiante.

En particular, aunque R (RStudio, s. f.; The R Foundation, s. f.)¹ es ampliamente conocido como un *software* estadístico y de visualización efectivo, porque es potente, flexible y de código abierto, su enseñanza no es una tarea fácil. De hecho, al reflexionar sobre sus prácticas docentes, los educadores que enseñan análisis de textos utilizando R argumentan con frecuencia sobre las siguientes dificultades en su enseñanza: (1) R tiene una curva de aprendizaje más pronunciada en comparación con otro *software* disponible comercialmente, como IBM SPSS Statistics (SPSS), STATA o Microsoft Excel (Excel), y (2) R no se considera un lenguaje amigable para principiantes, ya que se les dificultará más a quienes no tienen conocimientos en lógica de programación, siendo esta una de las razones más importantes para no enseñar R a los estudiantes de pregrado (Dae-HeeKim, 2019).

En este capítulo presentamos un enfoque didáctico práctico y eficaz para apoyar el aprendizaje de análisis de textos por parte de estudiantes de humanidades digitales utilizando R. El proceso metodológico propuesto se enfoca en la interpretación de los resultados de los análisis literarios logrados, más que en el uso de R como herramienta de programación, lo que permite lograr una apropiación real de los conceptos y procedimientos de análisis de

1 RStudio, desarrollado por J. J. Allaire, es un entorno integrado de desarrollo (IDE, por sus siglas en inglés) para desarrollar programas en el lenguaje de programación R. Es gratuito y de código abierto, aunque tiene una versión comercial. Cf. <https://posit.co/> Por su parte, R es un lenguaje de programación, también gratuito y de código abierto, especialmente diseñado para análisis estadístico, si bien tiene una variedad de otras aplicaciones. Cf. <https://www.r-project.org/>

datos, tales que estos puedan ser transferidos fácilmente a nuevas situaciones (Fries *et al.*, 2021; Son *et al.*, 2021).

El presente capítulo está organizado de la siguiente manera: en la sección 1 se describen las tendencias en el área de la pedagogía y didáctica del análisis de texto en el contexto de las humanidades digitales; la sección 2 describe la aproximación metodológica diseñada en este estudio; la sección 3 describe la metodología de investigación desarrollada, en la sección 4 se detallan los resultados. Finalmente, se concluye el capítulo y se destacan los trabajos futuros de relevancia.

Tendencias en el área de la pedagogía y didáctica del análisis de texto en el contexto de las humanidades digitales

Libros como *A new companion to digital humanities* reflejan la gran cantidad de ejes temáticos que representan el área de las humanidades digitales (Schreibman *et al.*, 2016). Sin embargo, la brecha entre la literatura disponible sobre pedagogía e investigación en el terreno, paradójicamente, se está ampliando (Hirsch, 2012). En el ámbito internacional han emergido diversos centros e institutos para el desarrollo de las humanidades digitales, y desde estos se lleva a cabo gran diversidad de proyectos orientados a fortalecer el uso de las tecnologías en el campo, así como a la creación de nuevas tecnologías en áreas no exploradas anteriormente. En este contexto, se hace importante también fortalecer las prácticas pedagógicas y didácticas que habilitan la

generación de humanistas digitales con fortalezas para promover la transformación del campo.

Una primera cuestión importante sobre la que se tendría que reflexionar es si el proceso de enseñanza/aprendizaje de un humanista debería buscar fortalecer habilidades técnicas u ofrecer a los alumnos aproximaciones metodológicas que le faciliten la exploración del campo.

En este sentido, Mahony y Pierazzo (2012) argumentan que el currículo debe centrarse en enseñar a los estudiantes nuevos enfoques y nuevas formas de pensar sobre las humanidades y, para lograrlo, se hace necesario enseñar enfoques metodológicos y no simplemente habilidades tecnológicas.

Asimismo, los autores recomiendan que los humanistas reciban una instrucción que fortalezca sus habilidades de trabajo en comunidades de práctica y fortalezcan su alfabetización informacional y mediática. Esta medida se ve justificada por el hecho de que el desarrollo de este campo sin duda se lleva a cabo en colaboración con otras disciplinas, lo que implica que, en la actual sociedad de la información, sea cada vez más relevante estimular el sentido crítico del alumno. Los mismos autores recomiendan incluir en la práctica docente estrategias como el uso de ejemplos relevantes y el uso de ejercicios contextualizados para situar a los participantes.

El enfoque didáctico descrito en el presente capítulo asume la postura de los autores, ofreciendo al alumno una metodología para hacer análisis de texto, el cual podrá implementar al hacer uso de cualquier herramienta de programación y que se apoya en el aprendizaje basado en ejemplos y ejercicios previamente contruidos que guiarán su trabajo.

En el ámbito particular de la enseñanza del análisis de texto, Sinclair y Rockwell (2012) resaltan lo intimidante que puede resultar para un estudiante de humanidades digitales el análisis de texto, pero también lo importante que esta área de las humanidades se vuelve cada día. En particular, actualmente con la incidencia de la inteligencia artificial generativa, el análisis de texto y en general el tratamiento del lenguaje natural se ha vuelto un tema de interés creciente.

Sinclair y Rockwell proponen dos diferentes modelos que podrían guiar el proceso de enseñanza del análisis de texto. El primero de estos es elegir una herramienta que incluya corpus predefinidos con los cuales el estudiante pueda interactuar, eliminando la necesidad de que sea el alumno, quien realice el proceso de búsqueda e importación de los corpus en las herramientas de análisis. Esta aproximación centra la atención en el proceso de análisis del texto más que en procesos de soporte ETL (*extract, transform, load*; en español: extraer, transformar y cargar datos). Los autores proponen guiar también el análisis del texto, sugiriendo a los alumnos las actividades a realizar con los corpus disponibles, como análisis de frecuencias de palabras, temas emergentes en el texto, entre otros. El segundo modelo sugiere que sean los propios estudiantes, quienes definan un corpus a analizar de acuerdo con sus propios intereses. Esta aproximación exige a los alumnos un conocimiento de los procesos de ETL que les permitan apropiarse de los corpus a trabajar. Una recomendación final por resaltar por parte de los autores es la manera en la que se presenta el análisis de texto a los alumnos, la cual debe ser muy cercana a la interpretación, se sugiere el uso de “recetas” o historias que cuenten posibles tipos de análisis a realizar a partir de textos preseleccionados.

Este enfoque se alinea más con el diseño metodológico propuesto en este capítulo.

Los estudios revisados nos han permitido identificar buenas prácticas en la enseñanza de análisis de texto a humanistas digitales. En la siguiente sección se describe el enfoque didáctico diseñado para facilitar el aprendizaje de análisis de textos por humanistas.

Enfoque didáctico para apoyar el aprendizaje de análisis de textos por humanistas

El enfoque didáctico que hemos diseñado para facilitar el aprendizaje de análisis de texto a humanistas digitales se basa en el modelo de ciencia de datos denominado modelo del epiciclo (Peng y Matsui, 2017), que se muestra en la figura 1.1.

El modelo del epiciclo ha sido diseñado para guiar los procesos de descubrimiento de conocimiento en ciencia de datos. El epiciclo central plantea tres procesos que iteran a lo largo de todas las etapas del modelo, el desarrollo de expectativas (*develop expectations*), la recolección de datos (*collect data*) y el establecimiento de la relación entre expectativas y cómo pueden estas ser alcanzadas a través de los datos disponibles (*match expectations with data*).

Estos procesos del epiciclo son analizados en cada una de las etapas del modelo, a saber: definición de preguntas (*stating the question*), el análisis exploratorio de datos (*exploratory data analysis*), la construcción del modelo (*model building*), la interpretación (*interpret*) y la comunicación de resultados (*communicate*).

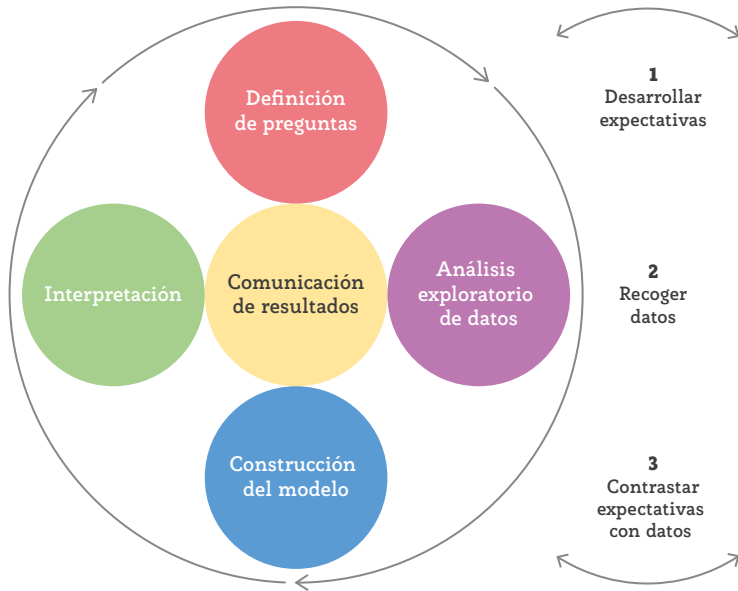


FIGURA 1.1. Modelo del epiciclo

FUENTE: elaboración propia a partir de Peng y Matsui (2017)².

Para ejemplificar cómo actúa el epiciclo, la etapa de Definición de preguntas permite claramente *formular la expectativa a través de preguntas*, es decir qué comprensión se espera lograr luego del análisis textual, verificando claramente que los *datos de que se dispone contribuyan a dar respuesta a la pregunta o ajustando la pregunta (expectativa)* en caso contrario.

2 La figura es una representación del modelo del epiciclo, en el que se observa un epiciclo central que opera sobre las etapas del modelo.



FIGURA 1.2. Enfoque didáctico para apoyar el aprendizaje de análisis de textos

Sobre la base del modelo del epiciclo se plantea el enfoque didáctico para apoyar el aprendizaje de análisis de textos por humanistas, cuyo esquema se define en la figura 1.2.

A continuación, se describen cada una de las etapas del modelo.

1. *Etapa de aprendizaje situado de las funcionalidades de la herramienta seleccionada utilizando ejemplos.* En esta se sitúa al alumno en el uso de la herramienta de análisis seleccionada por el facilitador de los aprendizajes. Es importante

que esta primera aproximación de los alumnos a un proceso de análisis se realice sobre la base de ejemplos concretos que los estudiantes estén en capacidad de comprender y apropiar. Se recomienda analizar las competencias digitales iniciales de los alumnos, así como seleccionar para este primer ejercicio un texto fuente que sea de conocimiento de los estudiantes. De esta manera, las preguntas emergerán con mayor claridad y prontitud.

2. *Identificar un texto de interés para el alumno.* En el proceso de análisis textual es muy importante captar la atención temprana de los estudiantes, no solo porque realizar un análisis textual lleva el acercamiento del humanista a áreas —como la programación— que son nuevas y podrían ser difíciles de abordar, sino porque la motivación guía el logro de un aprendizaje exitoso (Meyer *et al.*, 2014). En este sentido, que el humanista se acerque al análisis en el contexto de un texto de interés resulta relevante. Es por esto recomendable que el texto elegido por el alumno sea un texto conocido, preferiblemente en profundidad, o que haya sido trabajado por el estudiante desde otra perspectiva.
3. *Plantear preguntas que orienten el proceso de interpretación a partir del análisis de texto.* Dado un texto conocido, el estudiante está entonces en la capacidad de formular preguntas orientadoras para el desarrollo del análisis, discutir sobre ellas buscando alinear sus expectativas y analizar cómo a partir de la obra, como insumo textual, estas preguntas pueden ser resueltas.
4. *Apropiar el texto a analizar en un formato apropiado.* Teniendo claras las preguntas a responder, un asunto

relevante es la búsqueda de la fuente del dato, es decir, el lugar en el que se encuentra el texto que se desea analizar. Los textos pueden provenir de diversas fuentes como repositorios, bases de datos o sitios web de proyectos que alojan diversos tipos de fuentes. En este sentido, lo importante es que la fuente sea fiable, tal que se garanticen la integridad, la completitud y la veracidad del dato.

5. *Importar y explorar el texto a analizar preprocesándolo.* Una manera de conocer el texto a analizar es preprocesándolo, llevando a cabo acciones como crear listas de las palabras más frecuentes, nubes de palabras, identificación de la densidad de palabras, entre otras acciones. Esto permite al alumno familiarizarse con el texto que analizará.
6. *Guiar el análisis textual tal que se logre dar respuesta a las preguntas formuladas por los alumnos.* En esta etapa el profesor, como guía, orienta a los estudiantes en técnicas de análisis que puedan serles de utilidad para la búsqueda de las respuestas a las preguntas del alumno. En este punto, proveer pequeños trozos de código en el lenguaje escogido puede ser de utilidad para orientar el trabajo de los alumnos. Es importante que el estudiante se sienta acompañado en esta fase, por lo que es recomendable crear una estrategia de comunicación que permita reducir el estrés y la ansiedad de los alumnos.
7. *Interpretar los resultados logrados del análisis.* La interpretación debe emerger del análisis de los resultados obtenidos por los estudiantes. Es importante que la interpretación asegure dar respuestas a las preguntas formuladas por los alumnos. En este punto, el profesor, que guía los aprendizajes de

los estudiantes, debe asegurarse de que los análisis desarrollados por los alumnos permitan dar respuesta efectiva a las preguntas planteadas, promoviendo el análisis crítico de los resultados logrados y la reorientación o mejora de las técnicas de análisis, en caso de que las desarrolladas antes no permitieran responder las preguntas planteadas.

8. *Realizar una revisión por pares del trabajo de análisis e interpretación desarrollado por los alumnos.* La visión de un par, que se encuentra en una situación similar de aprendizaje, que comparte los mismos intereses, puede provocar la reflexión. La revisión por pares es un espacio interesante de enriquecimiento de las experiencias de los alumnos por el contraste de opiniones, experiencias y realidades conceptuales. Esta revisión debe ser guiada y las rúbricas deben ser un buen instrumento para facilitar la revisión que se realice entre pares. Se sugiere que la revisión de un par del mismo nivel sea el primer momento de comunicación del trabajo de los alumnos.
9. *Retroalimentar la revisión por pares, entre pares y también por parte del docente.* Asimismo, retroalimentar el trabajo de revisión es fundamental, permite identificar nuevas vías de enriquecimiento de las experiencias, mejora de los trabajos realizados por los estudiantes y promover el aprendizaje en comunidades de práctica.
10. *Concluir el trabajo realizado en plenaria.* Finalizar el trabajo de análisis en plenaria brinda a los alumnos la experiencia de comunicar en público, no solo sus hallazgos, sino las conclusiones de la revisión preliminar de pares y su retroalimentación. Enriquece las habilidades comunicativas de

los alumnos y provoca que emerjan ideas de futuros trabajos conjuntos.

Tras analizar críticamente la metodología planteada antes de iniciar el proceso de evaluación, es preciso señalar que, aunque algunos autores se muestran contrarios al uso de metodologías cuantitativas para analizar objetos literarios, por considerarlas, entre otras cosas, imprecisas, reduccionistas y propensas a errores (Da, 2019), también hay muchos partidarios de las posibilidades que ofrece este análisis. Estamos frente a una realidad cada vez más orientada hacia un enfoque cuantitativo, por lo que, tal y como indica Rodríguez-Ortega (2018), es crucial combinar aspectos cuantitativos, estadísticos y objetivos con la incertidumbre, subjetividad y vaguedad propias de las disciplinas humanísticas. Por su parte, Moretti (2013) lo defiende argumentando que gracias al uso del análisis cuantitativo obtendremos hechos, es decir, distribuciones, frecuencias o patrones que serán más o menos idénticos para todos los investigadores, pero cuya interpretación semántica no es un hecho por sí misma, sino una interpretación que se consigue observando el texto de manera atomística (Hackler y Kirsten, 2016).

Evaluación

Concepto de la evaluación

El enfoque didáctico propuesto fue validado en el contexto del curso Herramientas para el Almacenamiento, la Extracción y la

Visualización de Datos en Humanidades (Baldiris, 2022), con un número de 17 participantes, siguiendo un diseño cualitativo basado en casos de estudio.

Atendiendo a la diversidad de disciplinas que aúnan las humanidades digitales, se han seleccionado dos ejemplos que representan los diferentes orígenes académicos de los alumnos del mencionado curso.

En primer lugar, se presenta un texto histórico que narra la vida y obra de uno de los guerreros antiguos más destacados de la historia, reconocido por sus innovadoras estrategias de guerra. En segundo lugar, se aborda una obra literaria ampliamente estudiada en el ámbito de la Filología Inglesa, la cual, al igual que el texto anterior, ha sido analizada en su lengua original, el inglés, debido a su riqueza en detalles descriptivos. Los casos de estudio que se presentan en el presente capítulo como objetos de evaluación son los siguientes:

- Análisis cuantitativo del contenido de la obra *Hannibal the Carthaginian: Makers of history series* (Abbot, 1849), caso en el cual se logran resolver preguntas como ¿quién es el personaje principal?, ¿quién es el enemigo del protagonista?, y ¿qué otro personaje tiene relevancia en la obra?
- Análisis cuantitativo del contenido de la obra *Romeo y Julieta* (Shakespeare, 1597), caso en el cual se analiza cuál es el peso que cada uno de los dos protagonistas tiene en las conversaciones del resto de personajes.

Metodología

Cada caso de estudio se analizó con respecto al seguimiento del proceso metodológico planteado en el enfoque didáctico. En este sentido, lo que se buscaba corroborar es que el proceso hecho por el alumno le permitió llevar a cabo el análisis de un texto para dar respuesta a preguntas concretas previamente definidas, realizando una adecuada interpretación de los resultados del análisis. A continuación, se presentan los dos casos de estudio, un documento histórico y uno literario, una muestra interesante que permite evidenciar las posibilidades del análisis textual.

Para el caso particular de los casos de estudio, se utilizó R como herramienta para el análisis textual, específicamente el ambiente de RStudio.

Resultados

Caso 1. Hannibal the Carthaginian

El libro *Hannibal the Carthaginian* (Abbot, 1849) es un texto que hace parte del seriado biográfico *Makers of History Series*, producido por los hermanos Jacob y John S. C. Abbott, quienes se dedicaron a estudiar las vidas de las personas más influyentes del mundo antiguo.

En esta ocasión logran recopilar las principales hazañas del militar y gobernante cartaginés, Hannibal (247-183 a. C.)³. Gracias a sus acciones, pasa a ser recordado en la historia universal como uno de los estrategas de guerra más eficientes de todos, hasta tal

3 En español, Aníbal Barca. Nos referiremos a los personajes con sus nombres en inglés, ya que los análisis estadísticos se hacen sobre el texto escrito en esa lengua.

punto de ser considerado como uno de los pocos en repeler al Imperio romano en su propio territorio.

Empero, con la selección y definición del insumo literario a abordar, se continuó con el análisis propiamente dicho. El punto de partida fue la definición de preguntas problematizadoras como ¿quién es el personaje principal?, ¿quién es el enemigo del protagonista? y ¿qué otro personaje tiene relevancia?

Para dar respuesta a las preguntas definidas se desarrolló un conjunto de líneas algorítmicas que se concretaron en el despliegue de tres gráficos que sirvieron como respaldo analítico y visual para apoyar las respuestas a cada inquietud. A continuación, se describen las respuestas a las preguntas planteadas con el soporte del análisis logrado.

Pregunta 1: “¿Quién es el personaje principal?”

La respuesta a esta pregunta es Hannibal.

El análisis realizado para dar respuesta a esta pregunta fue la generación de la frecuencia de aparición de cada uno de los personajes en la obra. Las gráficas generadas permiten ver en el eje horizontal la evolución temporal del texto analizado e incluyen una línea vertical en cada punto de la obra en la que aparece el personaje buscado.

En el presente estudio, las gráficas revelaron que los personajes de *Hannibal*, Cornelius Scipio y Hanno exhiben una frecuencia de aparición significativamente superior en comparación con aquellos nombres citados en una o dos ocasiones como máximo a lo largo del texto. De estos tres personajes, Hannibal recibe el mayor número de menciones y, por consiguiente, se postula como

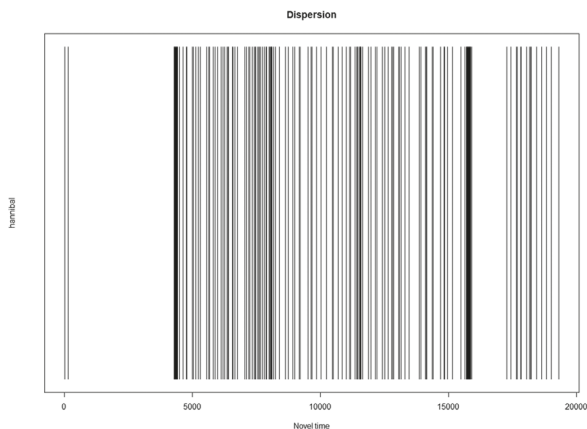


FIGURA 1.3. Vista de la frecuencia de aparición de la palabra “Hannibal”

el protagonista. La figura 1.3, generada mediante el *software* R Studio a través de la opción de creación de gráficas, ilustra la frecuencia de aparición de “Hannibal” en la obra:

En el análisis de la biografía de Hannibal, se observa una peculiaridad que merece atención: el autor hace referencia al personaje en una cantidad mínima al inicio de la obra, omitiéndolo por un extenso periodo que abarca casi un cuarto del texto. Según los principios de la interpretación, este fenómeno se explica por la generación de la contextualización geográfica del Imperio romano y la región de Cartago, así como la explicación sociopolítica que desencadenó las numerosas batallas entre ambos territorios. Este enfoque permite obtener una radiografía social detallada del estado de cada lugar y de sus habitantes, al abarcar aspectos como su origen étnico, su vestimenta, sus prácticas culturales y su sistema social.

Por otro lado, se puede apreciar al final la cantidad de veces que se repite la palabra “Hannibal”, entendiendo el gesto como el

desenlace de la historia que plantearon los autores en la trama. Básicamente, allí se explican las diferentes teorías que existen sobre por qué el protagonista decidió finalmente no continuar batallando contra Roma tras una década de interminables encuentros bélicos.

Pregunta 2: “¿Quién es el enemigo del protagonista?”

La respuesta a esta pregunta es Cornelius Scipius⁴.

Las figuras 1.4 y 1.5 muestran el comportamiento de aparición de este personaje en la obra.

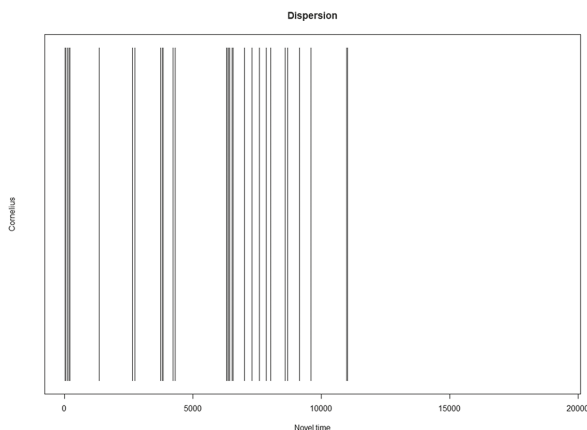


FIGURA 1.4. Frecuencia de aparición de la cadena “Cornelius Scipius”

4 En español, Escipión el Africano o Publio Cornelio Escipión Africano (236-183 a. C.).

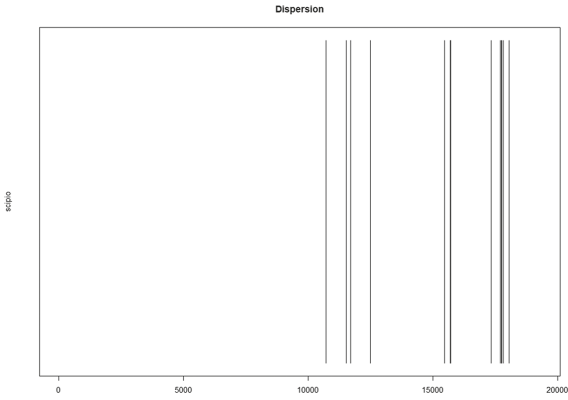


FIGURA 1.5. Frecuencia de aparición de solamente el apellido del enemigo del protagonista

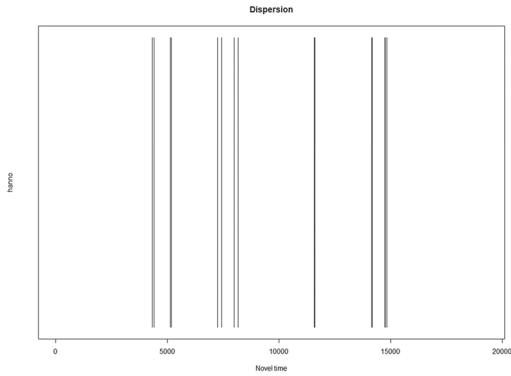


FIGURA 1.6. Resultado del análisis de aparición del nombre "Hanno the Great"

En las figuras 1.5 y 1.6 se puede observar algo interesante. Como recurso lingüístico reiterativo, el autor prefiere mencionar

por el nombre de pila a este personaje, mientras que, en otros sitios, se refiere a él por su apellido. Cornelius Scipius⁵, además de ser la respuesta a la duda, fue cónsul del Imperio romano en el año 205 a. C. A ciencia cierta, se sabe que este político fue uno de los mayores contrincantes de Hannibal durante la batalla por el Mediterráneo y quien le declaró la guerra de por vida.

Al unir las menciones que se hacen de este personaje en toda la obra, se puede hallar una fuerte presencia de este en casi todo el documento, lo que significa que es importante para el entendimiento de la historia como otro de los personajes de gran relevancia en la narración y que, según la misma historia, es quien funge de antagonista.

Pregunta 3: “¿Qué otro personaje tiene relevancia?”

La respuesta a esta pregunta es Hanno the Great.

En la historia de la antigua Roma, Hanno the Great⁶ es uno de los personajes más mencionados en relatos de estrategia militar. Fue pionero en el uso de un elefante gris como estrategia militar para transportar armas y provisiones, lo que se considera una de las primeras actividades de su tipo y una de las más exitosas de la antigüedad.

Este animal fue clave para la guerra debido a la fuerza y alta resistencia a las temperaturas del continente africano, lo que permitió abastecer a los guerreros de armas, municiones y

5 En español, Cornelio Escipión.

6 En español, Hano el Grande.

provisiones. Como se aprecia en figura 1.6, es citado en diferentes ocasiones en lo que podría referirse al inicio, nudo y desenlace.

El motivo por el que se repite en estas partes es que el elefante se usó como estrategia en varios momentos clave de la guerra, ya que, por su fuerza y gran tamaño, era útil para transportar armamento. Esto es algo nunca visto en escenarios bélicos.

Así como fue posible analizar estas preguntas sobre la vida y obra del guerrero cartaginés, también fue posible analizar otra obra literaria gracias al seguimiento del enfoque didáctico propuesto. A continuación, se describe el caso de *Romeo y Julieta*.

Caso 2. Romeo y Julieta

La obra teatral *Romeo and Juliet* (Shakespeare, 1597) sirvió como base textual para el caso de estudio que se describe en este apartado. Dicha obra se ha convertido en una de las creaciones más importantes de la literatura universal. Su influencia se puede percibir en innumerables obras posteriores, no solo literarias, sino también cinematográficas e incluso pictóricas, creadas en un soporte diferente al formato textual del original. Sin embargo, para la elaboración de este análisis, se optó por hacer una revisión discursiva más introspectiva y analizar a los dos protagonistas dentro de la propia obra.

Se sabe que la historia de amor de Romeo y Julieta es el eje central de esta obra, en torno al cual giran las preocupaciones del resto de personajes. Esto hace que nos planteáramos cuál es la importancia que cada uno de los dos protagonistas tiene en las escenas de la obra y en las conversaciones del resto de personajes. Para conocerlo, se realizó un análisis cuantitativo de la obra con R a través de su entorno de trabajo RStudio. Mediante el análisis

de las posiciones de los términos presentes en el texto, se elaboró un gráfico de dispersión de la aparición de las palabras a lo largo de la obra, en el que se muestran aquellas posiciones en las que se menciona explícitamente tanto a Romeo como a Julieta, y así conocer los momentos de mayor presencia de cada uno de ellos de cara a analizar el peso de ambos en el desarrollo de su historia.

Primeros pasos: preprocesamiento del texto

Para poder analizar el texto con R fue indispensable que este se encontrara en formato digital. Dado que la obra teatral que se ha analizado se trata de una obra de dominio público, no ha sido difícil encontrarla en un formato digital que permita manipularla cómodamente con dicho lenguaje de programación. De cara a analizar únicamente los datos que interesan, es decir, el propio contenido de la obra, se llevó a cabo un preprocesamiento en el que se aislaron los metadatos incluidos en la versión utilizada.

En este punto previo al análisis fue importante, además, tener en cuenta la estructura propia de las obras de teatro, ya que no incluyen, como sucede en los ejemplos de narrativa, el texto con la acción narrada, sino que la forman acotaciones, diálogos entre los personajes y el nombre de los personajes que intervienen en cada momento. Fue importante tenerlo presente, ya que el objetivo de este análisis fue aislar los momentos de la obra en la que se menciona a Romeo y a Julieta en las conversaciones de otros personajes, por lo que era esencial distinguir cuándo se les estaba mencionando en estas, cuándo se les nombraba dentro de las acotaciones del narrador y cuándo aparecían sus nombres por ser precisamente estos los personajes que intervienen en el diálogo.

En este momento se decidió que también era interesante para el análisis incluir aquellos momentos en los que el narrador les menciona en las acotaciones, ya que implica que dichos personajes están presentes o acaban de abandonar la escena y, por tanto, su presencia o ausencia es relevante para lo que va a suceder a continuación. Tomada esta decisión, y gracias al formato de la versión utilizada para el análisis, fue sencillo diferenciar cuándo hablaban los propios personajes y cuándo se les mencionaba en las acotaciones o en las conversaciones: en el primero de los casos sus nombres aparecen escritos por completo en mayúsculas y en el segundo aparecen únicamente con la inicial en mayúscula. Considerando que R diferencia entre mayúsculas y minúsculas al analizar texto, solo hubo que indicar que la función utilizada incluyese aquellas ocasiones en las que se mencionaba a los protagonistas únicamente con la inicial en mayúscula.

Elaboración de los gráficos de dispersión y análisis de las posiciones en el discurso

De cara a analizar las posiciones en el texto de las presencias de Romeo y Julieta en las escenas y poder hacer una comparativa de los momentos en que sus nombres son mencionados en las conversaciones, se realizaron dos gráficos en R con la función *plot*, tal y como se muestra en las figuras 1.7 y 1.8.

Sin embargo, para facilitar el análisis de la ocurrencia de ambos nombres, R permite fusionar los dos gráficos utilizando la función *par* y mostrar los datos en uno único. Así, gracias al uso de colores mediante el argumento *colors* (en este caso, azul para Romeo y rojo para Julieta) y de texturas con el argumento *lty* (utilizando una línea continua para Romeo y una punteada

para Julieta), distinguimos las presencias de cada personaje, tal y como se muestra en la figura 1.9.

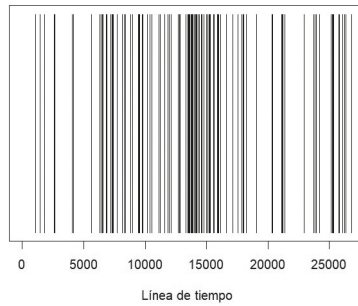


FIGURA 1.7. Gráfico de dispersión con las menciones de "Romeo"

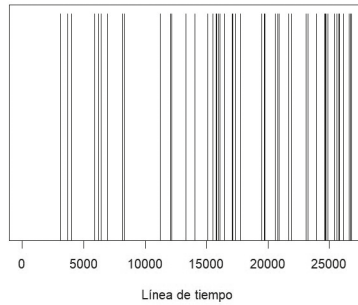


FIGURA 1.8. Gráfico de dispersión con las menciones de "Julieta"

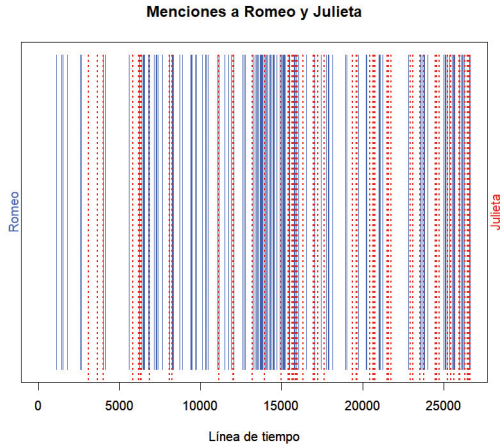


FIGURA 1.9. Gráfico de dispersión unificado con las menciones de “Romeo” y “Julietta”

Lo que estos gráficos permiten ver, en primer lugar, es que las menciones a Romeo (en azul) son mucho más numerosas que la de Julieta (en rojo) y que es precisamente este personaje el primero y el último a quien se menciona en la obra. En la primera escena del primer acto se conoce parte de la familia Montesco, a la que pertenece Romeo, por lo que solo se presencian algunas menciones a él. Es en la segunda escena en la que se menciona a los Capuleto, familia de Julieta, aunque sin menciones directas a ella por el momento, quien no aparece hasta la siguiente escena, tal y como se advierte en el gráfico con la presencia de las primeras líneas rojas. Es en la quinta y última escena del primer acto en la que se halla el primer diálogo entre los protagonistas, pero sin menciones directas a sus nombres, ya que ambos los desconocen.

El segundo acto (que comienza en torno al punto 7500 de la línea de tiempo en la figura 1.9) trae uno de los primeros momentos de mayor mención a Romeo, coincidiendo con una de las escenas por excelencia de la obra: la declaración de amor de este en el balcón de Julieta. Como se puede comprobar en la obra, es la joven quien responde a las palabras de su enamorado mencionándolo en repetidas ocasiones, lo que incrementa su presencia en el gráfico. A su vez, a pesar de aparecer en varias escenas, las menciones a Julieta no son frecuentes, lo que se comprueba con la poca presencia de líneas rojas en este punto del gráfico.

En el tercer acto (en torno al punto 12000 de la línea de tiempo) se vuelve a confirmar la mayor presencia de Romeo en las conversaciones, bien sea por estar presente en escena, como en la pelea con la que comienza este acto y en la que Romeo acaba matando a Teobaldo, miembro de la familia Capuleto, bien porque se le menciona en la conversación aunque esté ausente, como en la siguiente escena, en la que cuentan lo sucedido a Julieta, quien le nombra en repetidas ocasiones, lo que intensifica la presencia del color azul en el gráfico. A partir de este punto es donde más se entremezclan ambos nombres, por ejemplo, en la escena en la que Fray Lorenzo y el ama de Julieta planifican una cita para los dos amados.

Durante el cuarto acto (en torno al punto 18000 de la línea de tiempo) se encuentran algunos momentos con menor mención a los protagonistas seguidos de partes en las que contamos con mayor presencia de cada uno de los colores por separado. En estas escenas se puede presenciar la planificación de la boda entre Julieta y Paris, a quien la familia Capuleto acepta como su esposo, y que sirve para que Julieta confiese al fraile que oficiará en la boda

su amor por Romeo, lo que le da más presencia a él en la conversación. A continuación, es Julieta quien tiene un mayor protagonismo en el gráfico coincidiendo con el final del cuarto acto, cuando la joven ha tomado la pócima adormecedora que le hace parecer muerta, lo que centra la conversación en ella.

Para concluir, en el quinto y último acto (en torno al punto 21 000 de la línea de tiempo) se sabe que las presencias de ambos protagonistas se intercalan. Por un lado, las menciones a Julieta hacen referencia a su supuesto fallecimiento a través de los lamentos de su amado y otros personajes. Seguidamente, los nombres de ambos se combinan en los diálogos, ya que se está alcanzando el momento álgido de la historia: Romeo, creyendo muerta a Julieta, toma la misma pócima, resultando muerto y Julieta, al despertar y encontrar muerto a Romeo, decide quitarse la vida apuñalándose para poner punto final a esta trágica historia de amor.

Conclusiones

En el presente artículo se ha descrito un enfoque didáctico para apoyar el aprendizaje de análisis de textos por humanistas, que contribuye en el estado del arte de las estrategias didácticas que pueden facilitar el aprendizaje de metodologías y herramientas de análisis de datos por humanistas digitales, asunto de preocupación internacional, dado el carácter de no-programadores de este conglomerado de profesionales.

El enfoque didáctico fue evaluado a través de un conjunto de casos de estudio, de los cuales se describieron dos en este manuscrito y cuyo análisis permitió concluir que el enfoque didáctico

utilizado plantea un conjunto de pasos claros que guiaron a los alumnos hacia el logro de interpretaciones adecuadas de la literatura, favoreciendo la reflexión y mejora del propio trabajo y motivando a los estudiantes a continuar con el aprendizaje de las herramientas de análisis de datos.

Es importante señalar que, como se menciona en el texto, se utilizó R como lenguaje de soporte para el estudio debido a su capacidad de análisis estadístico; sin embargo, no es la única herramienta disponible para este propósito. Próximas investigaciones en esta área se orientarán al uso, no de una herramienta como fue el caso de R, sino de un conjunto de herramientas que pueden facilitar el aprendizaje de los alumnos en diferentes áreas del análisis de texto. Por otra parte, la inclusión de datos textuales provenientes de fuentes dinámicas como redes sociales es de interés de los autores.

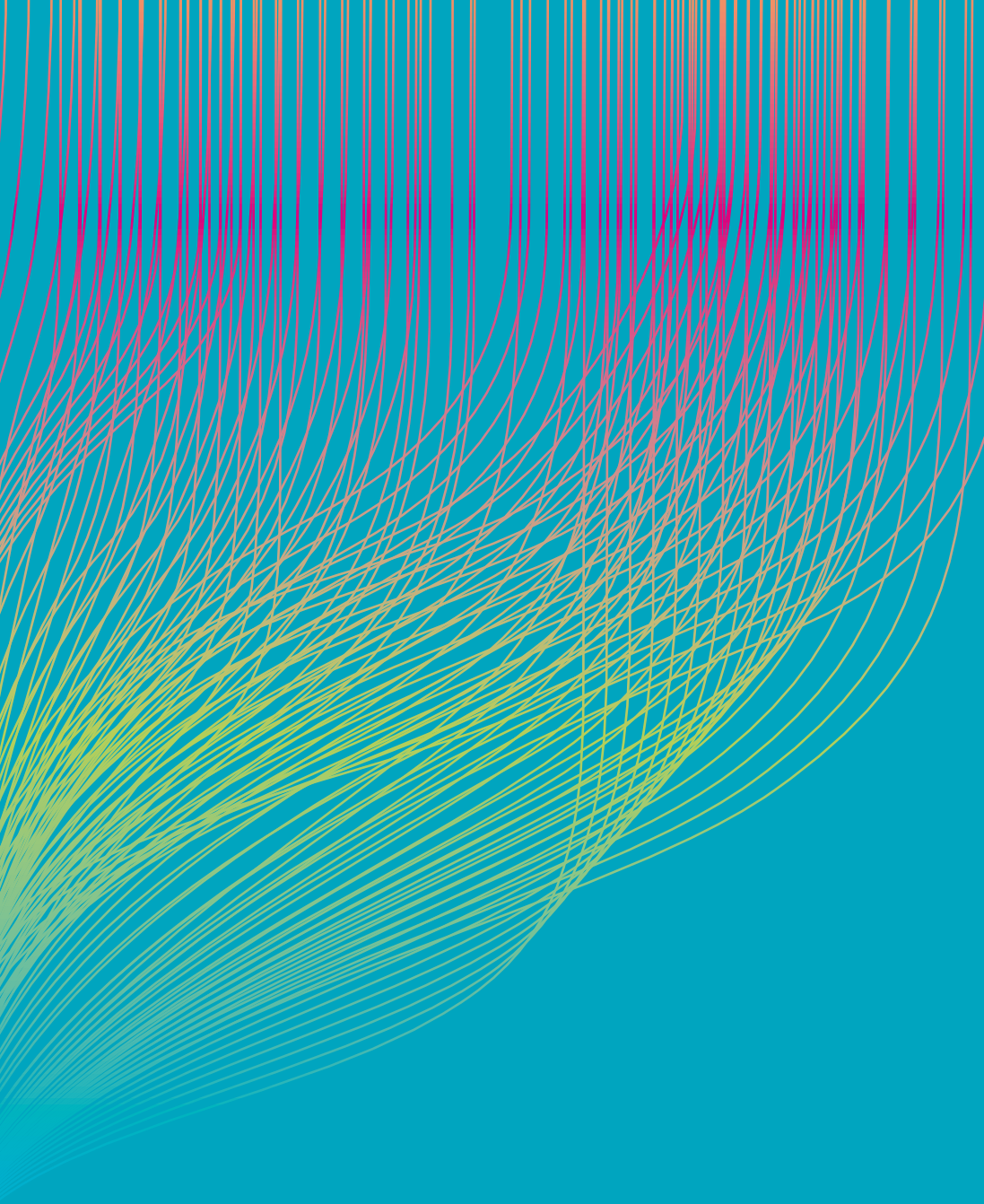
Asimismo, la llegada de la inteligencia artificial generativa (IAG) plantea un nuevo escenario que promete enriquecer los procesos de análisis textual de manera significativa. Futuras investigaciones podrían centrarse en el uso de herramientas como ChatGPT, Gemini y otras IAG en el análisis textual.



Referencias

- Abbot, J. (1901). *Hannibal the Carthaginian*. Makers of History Series. <https://www.gutenberg.org/files/27551/27551-h/27551-h.htm>
- Baldiris, S. (2022). *Herramientas para el almacenamiento, la extracción y la visualización de datos en humanidades* [Material de clase]. Universidad Internacional de La Rioja.
- Da, N. Z. (2019, 27 de marzo). The digital humanities debacle: Computational methods repeatedly come up short. *The Chronicle of Higher Education*. <https://www.chronicle.com/article/the-digital-humanities-debacle/>
- Dae-HeeKim, C. (2019). Teaching R to undergraduate business students. *Business Education Innovation Journal*, 11(1), 37-43.
- Fries, L., Son, J. Y., Givvin, K. B. y Stigler, J. W. (2021). Practicing connections: A framework to guide instructional design for developing understanding in complex domains. *Educational Psychology Review*, 33(2), 739-762. <https://doi.org/10.1007/s10648-020-09561-x>
- Hackler, R. M. y Kirsten, G. (2016). Distant reading, computational criticism, and social critique: An interview with Franco Moretti. *Genealogy+Critique*, 2(1), 1-17. <https://doi.org/10.16995/lefou.22>
- Hirsch, B. (ed.). (2012). *Digital humanities pedagogy: Practices, principles and politics*. Open Book Publishers.
- Mahony, S. y Pierazzo, E. (2012). Teaching skills or teaching methodology? En B. Hirsch (ed.), *Digital humanities pedagogy: Practices, principles and politics* (pp. 236-246). Open Book Publishers.
- Meyer, A., Rose, D. H. y Gordon, D. (2014). *Universal design for learning: Theory and practice*. CAST Professional Publishing.

- Moretti, F. (2013). *Distant reading*. Verso.
- Peng, R. D. y Matsui, E. (2017). *The art of data science*. Lean Publishing.
- Rodríguez-Ortega, N. (2018). Cinco ejes para pensar las humanidades digitales como proyecto de un nuevo humanismo digital. *Artnodes*, 22, 1-6. <https://doi.org/10.7238/a.voi22.3263>
- RStudio Team. (s. f.). *RStudio: Integrated development environment for R*. <https://rstudio.com/>
- Schreibman, S., Siemens, R. y Unsworth, J. (eds.). (2016). *A new companion to digital humanities*. Wiley-Blackwell.
- Shakespeare, W. (1597). *Romeo and Juliet*. <https://www.gutenberg.org/cache/epub/1513/pg1513.txt>
- Sinclair, S. y Rockwell, G. (2012). Teaching computer-assisted text analysis: Approaches to learning new methodologies. En B. Hirsch (ed.), *Digital humanities pedagogy: Practices, principles and politics* (pp. 236-246). Open Book Publishers.
- Son, J. Y., Blake, A. B., Fries, L. y Stigler, J. W. (2021). Modeling first: Applying learning science to the teaching of introductory statistics. *Journal of Statistics and Data Science Education*, 29(1), 4-21. <https://doi.org/10.1080/10691898.2020.1844106>
- The R Foundation. (s. f.). *R: A language and environment for statistical computing*. <https://www.r-project.org/>
- Wiedemann, G. y Niekler, A. (2017). Hands-on: A five day text mining course for humanists and social scientists. En R. En P. Bockwinkel, T. Declerck, S. Kübler y H. Zinsmeister (eds.), *Workshop on Teaching NLP for Digital Humanities (Teach4DH 2017)* (pp. 57-66). CEUR-ws. <https://www.researchgate.net/publication/319703314>



CAPÍTULO 2

***Ser vendida a él, así
comenzó nuestra historia
de amor: un estudio
discursivo asistido
por corpus de una
tendencia problemática
en los fanfictions
románticos en Wattpad***

Para citar este capítulo:

<http://doi.org/10.51573/Andes.9786287903081.9786287903098.9786287903111.2>

Resumen

Este estudio analiza la representación del amor en *fanfictions* románticos de Wattpad con la etiqueta “vendida”, en los que la trata de personas se usa como catalizador de la trama. Mediante el análisis del lenguaje en corpus, se examina un corpus especializado en comparación con otro de referencia (*fanfictions* románticos sin esta temática). Los resultados muestran que estos relatos combinan violencia (narcotráfico, secuestro) con romance, presentando a los captores como intereses amorosos y redimiendo su violencia. La protagonista es objetivada y retratada como propiedad y cuidadora emocional del captor. El análisis revela una prosodia semántica que normaliza las relaciones abusivas y está dirigida principalmente a adolescentes, lo que perpetúa ideologías patriarcales y distorsiona la percepción del amor. Se critica la falta de moderación en Wattpad y se insta a regular este contenido para proteger a los lectores jóvenes.

PALABRAS CLAVE: lingüística de corpus, Wattpad, *fanfiction*, análisis cuantitativo.



Introducción

f Desde 1930, los *fanfictions* han sido medios populares de expresión y de práctica de escritura que en su época eran publicados en revistas (Dawson, 2017). Sin embargo, hoy en día los *fanfictions* son publicados en plataformas digitales como Wattpad. Actualmente, se han propuesto iniciativas de aprendizaje de la segunda lengua basados en ejercicios de lectoescritura con los *fanfictions* (Black, 2006), por el hecho de que son escritos en su mayoría *amateur* y de cualquier tema de interés. Los *fanfictions*, como su nombre lo indica, son ficciones realizadas por fans. Son un producto literario hecho por un fan a partir de un producto de la industria cultural (García, 2009). En este sentido, hay historias que expanden la obra original como otras que se centran en el romance. En un inicio una trama romántica suena inofensiva, el problema que se plantea en este texto es el asociado a las situaciones que se desarrollan en un tipo específico de *fanfictions* románticos: los

fanfictions que utilizan la trata de personas como un catalizador para la historia de amor.

A primera vista parece inverosímil que se asocien estos dos eventos, pues concebir una historia de amor en un entorno tan intrínsecamente violento, como lo es la trata de personas, es un trabajo casi contradictorio; más en historias en las que en su mayoría las autoras son de países de América Latina. Sin embargo, en la plataforma Wattpad, bajo la etiqueta de “vendida”, se encuentran más de 40 000 *fanfictions* románticos con esta trama (Wattpad, 2023a). En este sentido, se plantea una investigación descriptiva de la prosodia semántica¹ (Sinclair, 2004) que se encuentra en un corpus especializado de estos *fanfictions*, enmarcado en un análisis discursivo asistido por corpus. Para esto, se propone la siguiente pregunta de investigación: ¿cómo se está representando semánticamente el amor en los *fanfictions* bajo la etiqueta “vendida” en la plataforma Wattpad?

Teniendo presente la pregunta de investigación, se busca alcanzar el siguiente objetivo: identificar la prosodia semántica del amor en el corpus. Para esto es necesario cumplir otros subobjetivos, como recolectar un corpus suficiente de *fanfictions*, categorizar las diferentes instancias de palabras claves y colocaciones que rodean al amor en los *fanfictions* y reflexionar sobre la manera en la que estas estadísticas conforman un imaginario semántico en el discurso de estos *fanfictions*.

1 La prosodia semántica se entiende como la valoración que el investigador hace de los términos estudiados, teniendo en cuenta el contexto de uso en un corpus. Dependiendo de la corriente de investigación, esta valoración puede ser binaria (positiva o negativa) o puede variar según la propuesta del propio investigador.

Sobre el amor romántico y los *fanfictions*

El *fanfiction* es una práctica de escritura *amateur* que implica escribir sobre un producto de la industria cultural (García, 2009), pero esta escritura también forma una comunidad (Black, 2006). Los *fanfictions* se volvieron populares aproximadamente en la década del 2000, lo más probable debido al aumento en el acceso, además de la disponibilidad de Internet, que permitió que muchas personas pudieran escribir y acceder a distintas plataformas. Algunos géneros populares de *fanfiction* son la fantasía, el romance y la ciencia ficción. El *fanfiction* romántico constituye una gran parte del *fanfiction*, con muchas historias agregadas todos los días y muchos lectores que las consultan. Según Scolari *et al.* (2018), una considerable proporción de los lectores de *fanfiction* en la plataforma Wattpad son adolescentes que potencialmente podrían ser susceptibles a las narrativas empleadas en este tipo de historias, las cuales podrían llegar a naturalizar las situaciones expuestas en los *fanfictions*.

Un tema preocupante de este subgénero es el romance *captor-captive* (captor-cautiva), que es un área poco estudiada. Varios autores han estudiado estas narrativas de cautiverio, pero nunca lo llaman por lo que es *la trata de personas*. La trata de personas se clasifica como un crimen contra la humanidad que constituye la violación de varios derechos humanos y se caracteriza por el traslado al interior o exterior del país con fines de explotación (Organización Internacional para las Migraciones [OIM], 2023; Vargas *et al.*, 2016). En este sentido, el acto de ser vendido o secuestrado por un hombre para hacer la limpieza o ser su amante

no es un romance captor-captivo o un dispositivo de trama para encontrar el amor, es un crimen.

El romance captor-captivo ha sido estudiado por Kaler (1999), Derounian-Stodola (2001) y Graham *et al.*, (1995). En la revisión se encontró que la mayoría de las tramas de estas novelas se desarrollan con el secuestro de uno de los protagonistas, no con su venta. Además, Kaler (1999) expresa que el trabajo de la mujer es ser una heroína emocional para el hombre compañero que necesita ser salvado de sus emociones negativas. Otros investigadores han examinado estas narrativas, aunque utilizando un corpus del siglo XIX, y han concluido que representan la esencia de la esclavitud de varios grupos marginalizados (Derounian-Stodola, 2001). Así, se ha puesto en evidencia que este tipo de romance ha sido objeto de estudio en diferentes investigaciones, aunque no en narrativas como el *fanfiction*, que se orientan a audiencias más jóvenes y tienen un alcance masivo. En este sentido, se hace pertinente un análisis desde la lingüística de corpus para analizar una gran cantidad de estos y compararlos con un grupo de control de otros *fanfictions*.

Por otro lado, en el romance *captor-captive* (captor-captivo), a menudo se muestran signos del síndrome de Estocolmo, que se caracteriza por el sentimiento de “amor” por las personas que abusan de la persona que padece este síndrome (Graham *et al.*, 1995). Este amor a menudo incluye minimizar el peligro en el que se encuentra la víctima. Incluso algunos autores proponen que el amor que sienten las mujeres por los hombres proviene de un síndrome de Estocolmo social (Graham *et al.*, 1994). Como tal, se podría inferir que las mujeres en estas narrativas no sienten precisamente amor, sino que es un caso de síndrome de Estocolmo, y no debería ser diferente en los *fanfictions* analizados en este estudio.

Con respecto al delito de la trata de personas, este se utiliza como un catalizador de la historia de amor en estas narrativas. El estudio de Meertens y la Oficina de las Naciones Unidas contra la Droga y el Delito [UNODC] (2009) estipula que la trata de personas tiene tres elementos constitutivos: actividad, propósito e interés. La actividad constituye cualquier acción cómplice o material para cometer el crimen. Los medios son la forma en que la persona es capturada y trasladada, ya sea por la fuerza, la amenaza, el engaño o el pago (Meertens y UNODC, 2009). El interés habla del objetivo que es la explotación de una persona que se produce a través de la instrumentalización, la objetivación y el uso de la dignidad (Meertens y UNODC, 2009). En este sentido, secuestrar o vender a alguien para que haga tareas domésticas o trabajos sexuales es, de hecho, trata de personas, lo que pone a estos “dispositivos de trama” románticos en la categoría de *delito*.

Junto a este acto criminal se le suma la objetivación de las mujeres, la cual comienza desde la racionalización de su cuerpo en los valores de la sociedad hegemónica y patriarcal. Entonces, al usar el cuerpo de la mujer, se lleva a cabo un proceso de naturalización de estas ideas como biológicas, en lugar de lo que es, una ideología patriarcal. Así, la mujer es vista como un objeto de deseo para los hombres, como una creadora de maldad o como una madre (Estramiana y Ruiz, 2006). Por tanto, además de examinar el papel que desempeña el amor en estos *fanfiction*, se observará la representación de la mujer y su posible objetivación en la narrativa. Los estudios sobre relaciones captor-cautiva señalan que el personaje femenino tiende a ser instrumentalizado como heroína emocional al servicio del interés amoroso masculino, configurándose así una objetivación de la mujer en tanto figura cuidadora.

La lingüística de corpus

En cuanto a la metodología que se puede utilizar en este tipo de estudios, hay una gran variedad, dependiendo de las necesidades de cada investigador. En primer lugar, se debe cuestionar si se va a tener un enfoque cuantitativo, cualitativo o mixto, pues cada uno tiene sus bases teóricas y métodos. Sin embargo, dado el largo volumen de corpus que se tiene, en este caso se optó por una metodología de corte mixto teniendo en cuenta la pregunta y objetivos investigativos. Para este efecto y la rama epistemológica en la que se ubica este estudio, se prefiere utilizar la lingüística de corpus para un primer análisis semántico-interpretativo del texto de los *fanfictions*. Esto se debe a las herramientas que provee la lingüística de corpus para hacer un análisis de un corpus extenso con un enfoque crítico-discursivo.

La lingüística de corpus tiene dos ubicaciones en la teoría lingüística: una es una disciplina dentro de esta ciencia como lo puede ser la sociolingüística o la etnolingüística; otra es como una metodología de cualquier tipo de enfoque lingüístico. Para este estudio, la lingüística de corpus será utilizada como una metodología. En este sentido, la lingüística de corpus es un enfoque en los estudios del lenguaje que ofrece una serie de métodos y técnicas para un análisis que antes no se creía posible gracias a distintos avances tecnológicos que facilitan el estudio de una cantidad de datos masivos de la lengua, como los son miles y millones de palabras (McEnery y Hardie, 2011).

En la lingüística de corpus no hay una sola manera de analizar la lengua ni una ruta específica que se debería seguir al ejecutar las diversas técnicas y métodos a su disposición. En este sentido,

esta metodología tiene una flexibilidad respecto de otras herramientas tradicionalmente cuantitativas para abordar muchas cantidades de corpus que requieren igual de un análisis cualitativo del contexto, algo que no pasa con el uso de modelos del lenguaje enfocados en la modelación de tópicos o el análisis de sentimientos. Si bien se basa en distintas estadísticas, es decisión del investigador cómo y cuáles usar. Tal vez lo más importante en este camino como lingüistas es observar cuidadosamente las concordancias que podemos observar en nuestro corpus.

La metodología *Corpus Assisted Discourse Studies* (CADS) es muy utilizada en la lingüística de corpus y en los estudios críticos del discurso, particularmente por el enfoque mixto que le provee al investigador. Teniendo esto en cuenta, esta metodología es útil para abordar grandes cantidades de texto y evitar que como investigadores caigan en el denominado *cherry picking* (o espiguelo). El *cherry picking* es una mala práctica en la investigación que se deriva de un corpus estudiado solo para escoger los datos que les sirven a los investigadores para probar su hipótesis. Entonces, al mezclar este enfoque estadístico se pretende primero iniciar la fase descriptiva de una investigación crítico-discursiva desde el *software* y las diferentes medidas estadísticas, para luego continuar con un análisis de las medidas en contexto y luego así iniciar la fase crítica después de un buen entendimiento semántico-descriptivo y una contextualización pragmática del discurso y los enunciados (Pardo, 2013).

En este caso en específico, nuestro universo son los *fanfictions* de Wattpad. Sin embargo, es enteramente imposible analizar de manera cualitativa siquiera una muestra representativa de esta gran cantidad de *fanfictions*. En consecuencia, se puede intentar

delimitar más el universo a los *fanfictions* en Wattpad que estén bajo la etiqueta “vendida” (que aun así siguen siendo muchos). No obstante, aquí se encuentra un pequeño problema: si el corpus solo se analizará de manera cualitativa, tomaría mucho tiempo hacerlo y sería necesario contar con un gran grupo de colaboradores para llevarlo a cabo. Por el contrario, si se abordaran cuantitativamente estos *fanfictions*, haría falta una interpretación de resultados de corte semántico que difícilmente se puede conseguir con enfoques puramente cuantitativos. Por lo anterior, se hace necesario un enfoque mixto. En efecto, para este tipo de metodologías CADs se suele recurrir a dos tipos de corpus: un corpus de referencia y otro corpus de foco.

En primer lugar, esto se hace basándose en una metodología similar a las de los diseños experimentales (Hernández *et al.*, 2018). En este sentido, el grupo de control es el corpus de referencia que no es afectado por el fenómeno lingüístico que se busca revisar. Esto deja así que el corpus de foco sea el grupo que está siendo afectado por el fenómeno específico, el cual, en este caso, son *fanfictions* cuya trama central es la trata de personas. Por consiguiente, el corpus de referencia puede perfectamente ser un corpus masivo como los TenTen (Jakubíček *et al.*, 2013) o el Corpus de Referencia del Español Actual [CREA] (Real Academia Española [RAE], 2023), pero en esta investigación se optó por elegir un corpus especializado para el de foco y el de referencia.

Además de tener que decidir una metodología, al trabajar con la lingüística de corpus se debe tener cuidado a la hora de escoger el tipo de corpus, el tamaño y específicamente si va a ser una metodología de corpus de referencia o si solo se va a tener un corpus focal (Culpeper, 2009). Esto es particularmente importante porque

a pesar de tener un enfoque basado en la estadística, la teoría de la información y una percepción “objetiva de la realidad”, es importante como investigadores tomar la decisión consciente de crear una metodología que no solo deje ver el fenómeno que estamos poniendo en evidencia, sino que activamente evitemos el *cherry picking* o el mal uso del corpora y las estadísticas que se recomiendan desde la lingüística de corpus (Duff *et al.*, 2001). Esta discusión no solo aplica para los lingüistas, sino para todas las personas que hacen investigación desde cualquier campo de conocimiento que trabajen con la lengua.

En consecuencia, como estudiosos del lenguaje debemos asumir la responsabilidad de trabajar con un producto que viene de humanos y reconocer los posibles sesgos que hayan en estas producciones, para así lograr hacer un análisis pertinente con el corpus ideal para nuestra investigación. Como resultado, se trata específicamente de asumir la idea del relativismo lingüístico como parte importante al hablar de producciones lingüísticas y al usar estas producciones para someterlas bajo distintos procesos estadísticos.

El relativismo lingüístico es la tesis que propone que la lengua afecta en mayor o menor medida en la percepción del mundo que se tiene e incluso la mirada del determinismo lingüístico plantea que el lenguaje determina el pensamiento (Cervantes, 2022). Por lo tanto, en una lectura superficial se implica que el corpus que recogemos influencia cómo otras personas piensan. Sapir y Whorf (Sapir, 1929) no fueron los únicos en pensar en esta relación desde el determinismo lingüístico, pues Van-Dijk (2016) también planteó una relación intrínseca entre el discurso, la cognición y la sociedad. Aunque no es obvia, ya se ha analizado la importancia del

corpus que se escoge para trabajar con herramientas estadísticas, pues un corpus debe tener criterios de selección para asegurarse de que se está investigando en el lugar correcto (Biber, 1990).

Por ejemplo, el no pensar en el corpus que elegimos y simplemente escoger un corpus con base en la cantidad de datos lleva a fallas intrínsecas en la construcción de modelos generativos del lenguaje (el típico caso de modelos misóginos, transfóbicos y racistas); para un analista puede incluso llevar a que las estadísticas no ayuden a develar lo que el corpus en verdad tiene por dentro, sino que oculte la ideología detrás del discurso analizado.

Como vimos, la elección del corpus es un gran factor a la hora de aplicar estas metodologías. Teniendo esto en cuenta, se decide tener los siguientes criterios de selección de corpus para este trabajo. Para el corpus foco: deben ser *fanfictions* de la plataforma Wattpad con trama romántica, deben haberse actualizado por última vez del 2020 al 2022 y deben estar bajo la etiqueta “vendida”. Por otro lado, para el corpus de referencia se tienen estos criterios: deben ser *fanfictions* de la plataforma Wattpad, deben haberse actualizado por última vez del 2020 al 2022 y deben ser *fanfictions* con trama romántica. En este sentido, se logra una estabilidad en términos de tiempo de los *fanfictions* recolectados, lo que asegura que sean muestras actuales de los *fanfictions*. Además, se está buscando encontrar las diferencias entre las representaciones del amor en estas historias bajo la etiqueta “vendida”. No obstante, para concluir específicamente qué tienen de distinto con los otros *fanfictions* románticos tradicionales es esencial que el único elemento que las diferencie sea la trata de personas.

Ambos corpora son de más de 500 000 tokens cada uno y se categorizan como corpora especializados. Se elige esto, ya que

para un enfoque de corte de CADs es importante asegurar que los corpora que se utilicen partan de los objetivos de investigación, teniendo en cuenta el *software* a utilizar.

La comparación se realiza mediante un análisis de relevancia de palabras clave en los dos corpus, descrito por Baker *et al.* (2008) como el “tema” del corpus objetivo después de eliminar las palabras conocidas como palabras vacías o *stopwords*. Las palabras clave se identifican con la ayuda del *software* de análisis de corpus AntConc (Anthony, 2022) y sus líneas de concordancia se examinan y agrupan según sus aspectos temáticos. La medida utilizada es el *mutual information* (MI) para las palabras clave. Brezina (2018) describe el MI como una medida utilizada para la comparación de textos, que se enfoca en las apariciones de palabras infrecuentes y específicas al corpus *target*. Después de este examen de resultados cuantitativos, se realiza una consulta de las concordancias utilizando la misma herramienta y un análisis cualitativo de los rasgos semánticos de las palabras clave.

La prosodia semántica se define como la función que analiza la ocurrencia de una palabra dentro de un corpus (por ejemplo, “violencia”, “amor”, “pánico”, “prevención”, “felicidad”). Dicha función dependerá del contexto de los escritos, pues estos se identifican en virtud de las palabras que rodean al lexema objeto de análisis (Sinclair, 2004). Cuando se identifican rasgos semánticos particularmente violentos en una palabra clave que no debería contenerlos, se comparan con los usos semánticos del corpus de referencia en las líneas de concordancia para ver si la prosodia semántica de hecho representa situaciones de violencia en contextos románticos y, por supuesto, si hay una tendencia a hacerlo en este tipo de *fanfictions*.

Resultados

En el análisis se hará énfasis en las primeras 10 palabras claves que se obtienen de AntConc (Anthony, 2022) comparando qué palabras clave aparecen en el corpus objetivo (vendida) con respecto a las palabras clave del corpus de referencia. Luego, se hace una comparación de los contextos de aparición de las respectivas palabras clave en la corpora. Para esto, tenemos las palabras claves del corpus “vendida” en la tabla 2.1.

En la tabla 2.1 se puede observar cómo, similar a la tabla 2.2, solo aparecen nombres de personajes. Lo que se resalta es que en la tabla 2.2 aparecen mayoritariamente los nombres de mujeres, mientras que la tabla 2.1 aparecen nombres del hombre que hace de captor y de interés romántico. Además, a simple vista se observa el recurso narrativo de algunos *fanfictions* como lo es el *tn* (abreviatura de la expresión “tu nombre”), que es un lugar para que las personas que leen el *fanfiction* puedan insertarse con su nombre propio.

TABLA 2.1. Palabras clave “vendida”

TYPE	FREQ_TAR	FREQ_REF
ian	1799	0
jack	1569	100
joe	959	0
me	10287	7516

Continúa...

Continúa...

TYPE	FREQ_TAR	FREQ_REF
samuel	703	0
ellie	691	0
jake	1104	147
ravn	570	0
tn	560	0
alfa	386	0

TABLA 2.2. Palabras clave referencia

TYPE	FREQ_TAR	FREQ_REF
ella	4382	916
jade	1458	0
perrie	984	0
su	7183	2863
jesy	552	0
lisa	508	1
qué	1739	504
leigh	389	0
labios	966	172
hinata	374	0

En el conglomerado de palabras clave de la tabla 2.2 se puede evidenciar que hay tres palabras que no son nombres: *ella*, *su* y *labios*. Una de estas palabras es un pronombre, otra un posesivo y,

por último, el nominal, *labios*. Algo curioso, pues en las primeras 100 palabras del corpus “vendida” no aparecen nominales referentes a partes del cuerpo; lo único cercano a esa categoría semántica son los nominales *sonriendo* (25)² y *lágrimas* (94). No se ahonda mucho en los morfemas léxicos como pronombres o posesivos, pues en este análisis no tienen mayor relevancia que la gramatical en el corpus.

Quizás analizar nombres propios no llame mucho la atención al analista del discurso, y tiene razón, pues lo verdaderamente importante son las palabras que están cerca de estos nombres. Al siguiente proceso de ver las ventanas de contexto, donde aparecen estas palabras, se le llama las colocaciones o *keyword in context* (KWIC). Narrativamente, se sabe que el interés romántico del corpus “vendida” es el mismo captor de *tn* o la persona protagonista. Usualmente, en estos *fanfictions*, la estructura narrativa es la siguiente:

- A. La familia de *tn* es abusiva y está en problemas financieros, así que recurre a vender a *tn*, aún siendo adolescente, al mafioso del *fanfiction*.
- B. *tn* es comprada por el mafioso del *fanfiction* y se enamora del mafioso, se casan y tienen hijos (sin ningún orden necesario de eventos en esta segunda parte).

2 El corpus se puede consultar aquí: https://github.com/AnitaBermeoo/Vendida_Corpus.git; las palabras clave, aquí: https://github.com/AnitaBermeoo/Vendida_Palabras_Clave.git. El número corresponde al orden que ocupa la palabra dentro de las primeras 100 palabras clave.

En este tipo de narraciones de *fanfiction*, desde la premisa misma se puede inferir que se aborda la temática de un crimen, específicamente la trata de personas. Lo particular de este análisis es la redención que se hace del interés romántico, pues el criminal (que suele pertenecer a una organización de narcotráfico) se convierte en el salvador de la protagonista (normalmente, cuando compra a la protagonista, la salva de un entorno en el que no se siente bien), creando una disonancia cognitiva de “amor” hacia una persona violenta. Este fenómeno se ilustra de manera contundente en los siguientes ejemplos:

- A. [...] si es la voz de Ian, me hago el muerto o la dormida. mierda no me da tiempo a reaccionar, **Ian** entra cerrando la puerta tras de sí, y yo me quedo mirándolo, un poco asustada, lleva todo el día. (...)
- B. Como se te mejoran — dijo con una sonrisa, — cuídate pequeña — me dijo y me dio un beso en la frente. — ian tengo aquí tu chaqueta antes quería buscarte para devolvértela .— le dije, — tranquila quedatela de momento, ya me la (...)Le da dos balazos, le da uno en la pierna derecha y otro en el hombro, —Jack no—lagrimas empiezan a salir por mi cara—Jack!, Agarro una piedra y golpeo al oficial,se desmaya,corro dirección hacia Jack y le estaba saliendo mucha sangre, —Jack, porfavor no me dejes— (...)Hacia el. —Te amo.— me dijo Samuel en un susurro, para después besar mi mejilla. —Yo a ti.— dije girándome, para besar ligeramente sus labios. Samuel sonrió, y me apegó más a su pecho. Estuvimos así varios minutos, hasta que Samuel se quedó dormido, igual que yo. Desperté en la (...)

Como se puede observar en las líneas de concordancia, hay una mención de casos violentos usualmente desde la narración del protagonista *tn*, en la que expresa sentir miedo, angustia hacia el interés romántico en el caso del ejemplo A y también hacia la posibilidad de que su interés romántico muera (por un enfrentamiento violento con las autoridades) en el ejemplo C. Sin embargo, en los ejemplos B y D se habla de un momento romántico con los respectivos intereses donde se besan. Esto hace difícil categorizar la polaridad semántica de cada una de estas palabras, pues sus contextos de aparición son opuestos. Entre el amor y la violencia surge la incógnita de si es una historia de amor con violencia o si es una narrativa violenta con una subtrama amorosa.

Si los nombres propios hablan de la perspectiva de *tn* hacia sus intereses románticos (mismos captores), es lógico que la palabra *tn* venga de la narración desde el punto de vista del interés romántico. La narración de *tn* es muy empática con sus captores, viendo en ellos el ser humano que son, sin importar los crímenes que cometan hacia otras personas o hacia *tn* misma. Sin embargo, la narración de los captores al hablar de *tn* es muy diciente, tratándola como una propiedad y buscando en ella un alivio emocional. Esto se puede identificar en estas concordancias:

- A. la vuelta pero se detuvo, Se acercaba a mi lentamente , me puse super nervioso y por instinto cerré los ojos hasta que siento UN BESO !!! *tn* me beso !!! O por dios díganme que no estoy soñando , aunque fue un beso en la mejilla me sentía super feliz, sentía que me (...)
- B. Jane había dicho ¿*tn* me olvido? No lo se. Durante todo el vuelo intente leer y distraerme pero me fue imposible, la idea

de que tn me haya olvidado me preocupaba. Por fin habíamos llegado, bajamos del avión e inevitablemente los recuerdos vinieron a mi mente. Uno de mis empleados (...) camine en dirección a tn quien estaba parada observando los letreros de las distintas películas. A lo lejos un hombre no dejaba de observar a tn, eso me puso celoso pero debía controlarme ya que no quiero discutir y arruinar este día. No le di más importancia hasta que ese (...)

- C. desayuno, tendré oportunidad de hablar con ella, solo espero tener valor. _____♡☆♡_____ Soy un estúpido, debí hablar con ella cuando tuve oportunidad. Hace dos semanas que tn no ha venido a mi habitación ¿por que? ¿Hice algo mal? Bueno, aparte de todas las idioteces que cometí ¿que hago para llamar su (...)

Se evidencia un claro sesgo desde la narración del hombre captor, en el que creen que son poseedores de tn como una pieza material, que no puede ser observada por nadie más que el interés romántico. Además, desarrollan un apego y temor al pensar que tn ya no siente lo mismo o que incluso los habrá olvidado, porque tn está en una posición en la que se debe enamorar de ellos. Si no, ¿qué le pasaría al no cumplir con los caprichos de una persona violenta, capaz de comprarla y capaz de matar a otras personas?

Por otro lado, hay algo muy diferente que pasa en el corpus de referencia, pues no se encuentran instancias de acciones violentas. Evidentemente, se encuentran momentos de angustia o preocupación, pero no son dadas por eventos criminales, sino por actos que pasan en la trama. Incluso con personajes que vienen de un

anime como *Naruto* (*shonen*)³, no traen un elemento violento a la historia. Lo anterior se puede observar en los siguientes ejemplos:

A. un gran anhelo por ella. Mi risa se detuvo cuando recordé la razón por la que ahora tengo un antojo tan poco saludable por Jade. Jade y yo tenemos algún tipo de ... conexión, no precisamente, pero pudimos sentir cómo nos sentíamos, como feliz, nerviosa o preocupada. No se necesitan palabras, (...)hermosa por dentro y por fuera, que era una de las mejores personas que tenía. Todavía me dolía la dura crítica, pero podría soportarlo mientras Jade no piense en mí así. Es ella lo que importa. Como pasó el tiempo, pensé que era inmune a todo el odio ... pero estaba (...)misma edad era mejor que él le causaba impotencia a la par de respeto hacia esa persona. No le molestaba hacerse cercano a alguien como Hinata, era recatada, fuerte e inteligente, muy diferente a...—¡Maldición, dobe! ¿Piensas quedarte viendo?—La vena de su cien comenzaba a hincharse de molestia cuando (...)Claro que lo es! Es mi oka-san, después de todo—respondió Sasuke con toda la seguridad que su voz podía expresar. —Vaya, me alagas Hinata-chan —la mujer comenzó a reír sutilmente. Era agradable tener allí a Hinata, necesitaba una presencia femenina en la casa aparte de ella. Era (...)

3 *Shonen*: es un género de anime dirigido a hombres adolescentes. Este es el género más popular a nivel mundial. Aquí tenemos a piezas clásicas como *Dragon Ball* y ahora *Naruto* (Vidal, 2010).

En estos ejemplos, se puede observar que hay dos situaciones de emociones altas en D y B. Sin embargo, en B se infiere que la situación de tensión es producida por una crítica que se le hace al narrador, pero el interés romántico (*jade*) es un lugar seguro para quien narra. Por otro lado, en D se puede observar que Sasuke está ensimismado viendo a su interés amoroso (*Hinata*), y alguien le recalcó con molestia que se quedara viendo a Hinata. Lo anterior es sorprendente, porque en la trama del anime *Naruto*, Sasuke es un ninja que constantemente se debate entre la oscuridad y la luz en un camino de venganza en el que hace acciones violentas y criminales (incluso para los estándares de la serie), pero en el *fanfiction* no aparecen indicios de violencia más allá de desacuerdos propios de una amistad.

Discusión

En esta sección se habla de qué implican estos resultados desde un punto de vista discursivo. Es claro que el elemento de la violencia prevalece en los *fanfictions* del corpus “vendida”. No obstante, es importante resaltar la conexión que se establece entre la trata de personas y el narcotráfico o la mafia. Como se puede observar en los resultados, las palabras clave y las líneas de concordancia muestran que en el corpus “vendida” se tienden a mezclar los temas violentos del narcotráfico con los temas románticos. Esto crea un ambiente violento en la narrativa, no solo para las personas enemigas del narco sino también para tn. Algo que no se mencionó en los resultados fueron palabras clave como *maté* (83) o la prevalencia de sustantivos asociados a lugares como *habitación* (39), *cama* (46)

y *piscina* (61), a diferencia del corpus de referencia cuya única locación en las palabras clave fue *escuela* (63).

Lo anterior parece inocente, a excepción de *maté*, pero deja de serlo cuando se considera que estos son los lugares donde *tn* está siendo retenida en contra de su voluntad. A pesar de que las protagonistas usualmente son adolescentes, solo en el corpus de referencia la palabra *escuela* resalta, pues es un lugar que debería ser común para cualquier adolescente. Esto nos deja ver una tendencia en estos *fanfictions* a instrumentalizar la violencia como un dispositivo narrativo, incluso la violencia causada por el narcotráfico. Esto se ve mucho en las palabras claves de los intereses románticos, porque en su mayoría las concordancias son de narraciones en primera persona de *tn* donde esta violencia es vista como algo que es un hilo conductor para llegar a ser feliz junto a su interés romántico.

En esta sección de la discusión se quiere hacer un pequeño análisis con respecto a las posibles ideologías y las respectivas estrategias discursivas que se pueden identificar previamente. En primer lugar, tenemos a la posible objetivación que sufre la figura femenina en el *fanfiction*: no solo es vendida por su familia a un narco, sino que también su falta de personalidad e individualidad, combinada con el tratamiento que recibe, es un indicador de la posición que ocupa en la narrativa como en la sociedad. En este sentido, la objetivación de *tn* o de los diferentes intereses románticos cumple con dos maneras de hacerlo para una mujer, desde la posición de la mujer como objeto de deseo y como madre (Estramiana y Ruiz, 2006). *tn* es un cuerpo que sirve para ser intercambiado y vendido como un bien más al ser parte de una sociedad capitalista y patriarcal, pero también es un cuerpo que está para

cuidar emocionalmente a los hombres dentro de la narrativa; como se espera de una mujer muchas veces en la sociedad.

Por otro lado, está el posible ocultamiento del conflicto con el narcotráfico en América Latina. Aunque el narcotráfico es la violencia central en la narrativa de tn, esta se agrava con un delito aún más grave: la trata de personas. En estas historias, quien vende a tn es su familia, mientras que quien termina “amándola” y ofreciéndole una supuesta vida de ensueño es el mismo capo de la mafia que la compró, a menudo con intenciones violentas. Esta vida de ensueño se representa frecuentemente en contextos lujosos, con palabras como *piscina* (61), *habitación* (39) y *cama* (46) que aparecen en el corpus “vendida”. Estos términos, en conjunto, reflejan un imaginario en el que el entorno material se convierte en un intento de redención para personajes violentos. Sin embargo, estas narrativas omiten la realidad de que estos “lugares de lujo” son, en esencia, escenarios de confinamiento y abuso. Así, estas representaciones no solo romantizan el narcotráfico, sino que redimen a personajes que, en contextos reales, encarnan conflictos profundos en los países de América Latina.

Los términos como *piscina* (61), *habitación* (39) y *cama* (46), presentes en el corpus “vendida”, reflejan un énfasis en locaciones asociadas al lujo o al espacio privado. Aunque su relevancia en las narrativas podría estar relacionada con la construcción de un imaginario de redención para los captores, este análisis exploratorio no pretende establecer causalidades. Más bien, los datos sirven como punto de partida para futuras investigaciones que podrían profundizar en cómo estas locaciones contribuyen al discurso romántico y violento en este tipo de *fanfictions*. Invitar a quien lee a insertarse en una situación en la que es víctima de la

trata de personas, obligada a convivir con capos del narcotráfico y enamorarse de su captor, plantea serios riesgos cognitivos y socio-culturales, especialmente para las lectoras adolescentes que constituyen la audiencia principal de estas historias en Wattpad.

Wattpad lo utilizan en su mayoría niñas adolescentes (Scolari *et al.*, 2018) que están expuestas a este tipo de contenido “romántico” e imaginan que son ellas quienes sufren toda esta violencia, bajo la luz de que en realidad es una gran historia de amor. Estas mismas adolescentes pueden estarse imaginando que es ideal salir con un criminal como son todos sus intereses románticos de este corpus. El interés romántico está relacionado con *tn*, pero también está relacionado con verbos como golpear y matar, y demás acciones violentas. No siempre son estas acciones dirigidas a *tn*, aunque sí a los enemigos del interés romántico y se suele retratar esta violencia de manera romántica o interesante. Lo anterior, se hace bajo con la narrativa de que “él no es así conmigo” o “lo hace para protegerme”, haciendo mucho daño en las mentes de las lectoras. Por lo que ellas pueden estar pensando que esta violencia es algo romántico o muestra de un carácter deseable, a pesar de que en América Latina muchas mujeres han sido víctimas de formas de violencia similares.

Conclusiones

Desde hace mucho tiempo estos *fanfics* sobre ser vendidas están en auge. Hay obras en Wattpad que datan del 2014 y esto solo hace que este fenómeno sea aún más preocupante. Las narrativas que se ven en estas historias son aquellas del amor retratado entre

captor-cautiva, con un elemento violento acrecentado por su conexión a las mafias o el narcotráfico. En estas narrativas se ve la figura de la mujer como un objeto, más aún en las que venden a la protagonista, utilizándola como un bien, poniéndola al mismo nivel de una caja de cigarrillos.

Las mujeres somos seres humanos, no objetos. Sin embargo, en estas narrativas y en la vida real esta parece ser una conclusión a la que es difícil llegar. La diferencia entre estas narrativas y la vida real es que el final feliz del que hablan viene cuando las mujeres en esta condición se separan y logran escapar de su captor, no cuando se casan con él. La idea de que la persona con la que serás feliz es la persona que te ve como un objeto al que puede comprar es peligrosa para las personas en Wattpad, no obstante, hay más de 40 000 historias así. La trata de personas no es una narrativa romántica, es un crimen de lesa humanidad y que Wattpad siga permitiendo la publicación de estas historias sin una revisión ni un aviso de contenido violento es dañino para nosotras como mujeres y como niñas.

Por otro lado, una recomendación grande hacia plataformas de *fanfictions* es restringir el contenido al que se pueden exponer las personas desde un trabajo de moderación riguroso, pues se puede crear una cuenta en Wattpad desde los 13 años (Wattpad, 2023b). Es claro que la actual guía de contenido de Wattpad no es suficiente para detener a las personas más vulnerables de leer este tipo de narrativas. Porque aquí las culpables no son las otras niñas que escriben estas historias, es la plataforma y productos culturales como las narconovelas que permiten esta exposición y reproducción de este contenido. Este análisis pone en evidencia cómo estas narrativas no solo reflejan, sino que también perpetúan

ideologías dañinas en una audiencia vulnerable. Urge una mayor regulación en plataformas digitales y un enfoque crítico en la educación mediática para mitigar estos impactos. Finalmente, al ser este un estudio exploratorio y descriptivo de representaciones de las relaciones en los *fanfiction* de Wattpad, puede ser útil para una comprensión más profunda de las implicaciones del amor a nivel sociocognitivo a través de los estudios críticos del discurso, haciendo énfasis en próximas investigaciones en la identificación de estrategias discursivas para lograr develar las ideologías que se plantean en estos *fanfictions*.



Anexos en línea

CORPUS: https://github.com/AnitaBermeoo/Vendida_Corpus

PALABRAS CLAVE: <https://github.com/AnitaBermeoo/>

Vendida_Palabras_Clave

Referencias

- Anthony, L. (2022). *AntConc* (versión 4.1.2) [software de computador]. Waseda University. <https://www.laurenceanthony.net/software>
- Baker, P., Gabrielatos, C., Khosravini, M., Krzyżanowski, M., McEnery, T. y Wodak, R. (2008). A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press. *Discourse & Society*, 19(3), 273-306.
- Biber, D. (1990). Methodological issues regarding corpus-based analyses of linguistic variation. *Literary and Linguistic Computing*, 5(4), 257-269.
- Black, R. W. (2006). Language, culture, and identity in online fanfiction. *E-Learning and Digital Media*, 3(2), 170-184.
- Brezina, V. (2018). *Statistics in corpus linguistics: A practical guide*. Cambridge University Press.
- Cervantes, C. V. (2022, mayo). Relativismo lingüístico. *Centro Virtual Cervantes*. https://cvc.cervantes.es/ensenanza/biblioteca_ele/diccio_ele/diccionario/relativismolinguiistico.htm
- Culpeper, J. (2009). Keyness: Words, parts-of-speech and semantic categories in the character-talk of Shakespeare's Romeo and Juliet. *International Journal of Corpus Linguistics*, 14(1), 29-59.
- Dawson, A. (2017). A short history of fanfiction. *YA Hotline*, (104).
- Derounian-Stodola, K. Z. (2001). Captivity and the literary imagination. *The Cambridge Companion to Nineteenth-Century American Women's Writing*, 105-121.

- Duff, P. A., Biber, D. y Conrad, S. (2001). Corpus-based research in TESOL: Quantitative corpus-based research: Much more than bean counting. *TESOL Quarterly*, 35(2), 331-336.
- Estramiana, J. L. Á. y Ruiz, B. F. (2006). Representaciones sociales de la mujer. *Athenea Digital. Revista de Pensamiento e Investigación Social*, (9), 65-77.
- FIVE. (1994). Women's femininity, love of men, and heterosexuality. *Loving to Survive: Sexual Terror, Men's Violence, and Women's Lives*, 3, 183.
- García, A. E. M. (2009). Introducción al mundo de las sagas. *Puertas a la Lectura*, (6), 1-243.
- Graham, D. L. et al. (1995). A scale for identifying "Stockholm Syndrome" reactions in young dating women: Factor structure, reliability, and validity. *Violence and Victims*, 10(1), 3.
- Hernández Sampieri, R., Fernández Collado, C. y Baptista Lucio, P. (2018). *Metodología de la investigación* (vol. 4, pp. 310-386). McGraw-Hill Interamericana.
- Jakubiček, M., Kilgarriff, A., Kovář, V., Rychlý, P. y Suchomel, V. (2013, julio). The TenTen corpus family. En *7th International Corpus Linguistics Conference CL* (pp. 125-127).
- Jensen, K. E. (2014). Linguistics in the digital humanities: (Computational) corpus linguistics. *MedieKultur: Journal of Media and Communication Research*, 30(57), 20. <https://doi.org/10.7146/mediekultur.v30i57.15968>
- Kaler, A. K. (1999). Conventions of captivity in romance novels. *Romantic Conventions*, 86-99.
- Kennedy, G. (2014). *An introduction to corpus linguistics*. Routledge.
- McEnery, T. y Hardie, A. (2011). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.

- Meertens, D. y Oficina de las Naciones Unidas contra la Droga y el Delito (UNODC). (2009). *Estudio nacional exploratorio descriptivo sobre el fenómeno de trata de personas en Colombia*. Universidad Nacional de Colombia.
- Organización Internacional para las Migraciones. (2023, 1.º de noviembre). Lucha contra la trata de personas. <https://www.iom.int/es/lucha-contra-la-trata-de-personas>
- Pardo A., N. (2013). *Cómo hacer análisis crítico del discurso. Una perspectiva latinoamericana* (2.ª ed.). Universidad Nacional de Colombia. Instituto de Estudios en Comunicación y Cultura (IECO).
- Real Academia Española. (2023). *Corpus de referencia del español actual* (CREA) [Banco de datos]. <http://www.rae.es>
- Sapir, E. (1929). The status of linguistics as a science. *Language*, 207-214.
- Scolari, C. A. (2018). *Teens, media and collaborative cultures: Exploiting teens' transmedia skills in the classroom*.
- Sinclair, J. M. (2004). *Trust the text: Language, corpus, and discourse*. Routledge.
- Van-Dijk, T. (2016). Análisis crítico del discurso. *Revista Austral de Ciencias Sociales*, 30, 203-222. <https://doi.org/10.4206/rev.austral.cienc.soc.2016.N.30-10>
- Vargas Trujillo, E., Flores, C. E., Mendoza Simonds, L. M. y cols. (2016). *Trata de personas en Colombia: Una aproximación a la magnitud y comprensión del problema*. Universidad de los Andes.
- Vidal Pérez, L. A. (2010). *El anime como elemento de transculturación: Caso: Naruto* [Tesis de licenciatura]. Repositorio institucional. https://ddd.uab.cat/pub/tesis/2010/100074/anime_a2010.pdf
- Wattpad. (2023a, 30 de octubre). Búsqueda: Vendida. <https://www.wattpad.com/search/vendida>

- Wattpad. (2023b, 30 de octubre). Guía para padres y familiares.
<https://support.wattpad.com/hc/es/articles/204744684-Gu%C3%ADa-para-padres-y-familiares#:~:text=Pol%C3%ADtica%20de%20edad%20en%20Wattpad,crear%20una%20cuenta%20en%20Wattpad>
- Zottola, A. (2020). Corpus linguistics and digital humanities. Intersecting paths. A case study from Twitter. *América Crítica*, 4(2), 131-141. <https://doi.org/10.13125/americacritica/4521>