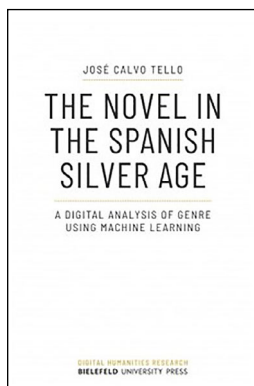


Calvo Tello, José. *The Novel in the Spanish Silver Age. A Digital Analysis of Genre using Machine Learning*

Bielefeld: Bielefeld University Press, 2021, 468 pp.
ISBN: 978-3-8394-5925-6

Dolores Romero López



Autoría:
Dolores Romero López
Universidad Complutense de Madrid, España
dromero@filol.ucm.es
<https://orcid.org/0000-0002-1230-5392>

Citación:
ROMERO LÓPEZ, Dolores. «Calvo Tello, José. *The Novel in the Spanish Silver Age. A Digital Analysis of Genre using Machine Learning*». *Anales de Literatura Española*, n.º 37, 2022, pp. 259-263. <https://doi.org/10.14198/ALEUA.2022.37.15>

© 2022 Dolores Romero López

Este trabajo está sujeto a una licencia de Reconocimiento 4.0 Internacional de Creative Commons (CC BY 4.0).



Este libro es otra Historia. Su planteamiento, desarrollo y conclusiones aportan un giro definitivo hacia la Historia Digital de la Literatura. Estamos acostumbrados a trazar la línea diacrónica de los estudios sobre la Edad de Plata siguiendo las aportaciones de Ernesto Giménez Caballero (1946), Antonio Ubieto Arteta, José María Jover Zamora y Juan Reglá (1963); Miguel Martínez Cuadrado (1973), José Carlos Mainer (1975, 1980, 2010), Pedro Laín Entralgo (1993), Hernán Urrutia Cárdenas (1999-2000), Ángela Ena Bordonada (2013), Dolores Romero López (2014), Dolores Thion (2021), etc. ¿Y ahora qué? En pleno *boom* de las llamadas Humanidades Digitales, el lector se adentra en el contenido de este libro como *Alicia en el país de las maravillas*: novelas de carne y hueso se prestan a la ciencia de la interpretación computacional. El siglo XXI apuesta por la digitalización masiva de contenidos y el desarrollo de aplicaciones y herramientas digitales que analizan el legado cultural digital para dar respuesta a nuevas preguntas. José Calvo Tello es un español que defendió su tesis doctoral en la Universidad de Würzburg (Alemania) en 2020 y ahora

es investigador y bibliotecario en la Biblioteca de la Universidad de Göttingen. *The Novel in the Spanish Silver Age. A Digital Analysis of Genre using Machine Learning* es la publicación en papel de su tesis de doctorado. Todos los datos en los que se basa su investigación están disponibles en abierto en los repositorios GitHub y Zenodo.

Su Corpus of Novels of the Spanish Silver Age (CoNNSA) contiene 358 novelas escritas desde 1880 hasta 1939 por autores como L. Alas, *Clarín*, F. Ayala, 'Azorín', los Baroja, Blasco Ibáñez, C. de Burgos, A. Carranque, R. Chacel, L. Coloma, J. Costa, A. Espina, A. Hoyos y Vinent, W. Fernández Flórez, R. Gómez de la Serna, G. Miró, Á. Ganivet, B. Jarnés, E. Pardo Bazán, J. M^a de Pereda, R. Pérez de Ayala, R. J. Sender, F. Trigo, M. de Unamuno, J. Valera, R. del Valle-Inclán... Sus novelas fueron rescatadas desde formatos digitales –procedentes de Biblioteca Digital Hispánica, ePubLibre, Archive.org, Cervantes Virtual, Wikisource, Gutenberg Project y Google Books– y transformadas a XML-TEI para etiquetarlas con metadatos específicos para esta investigación. La nómina de autores y obras puede resultar heterogénea si partimos de *a priori* historicistas, pero su heterogeneidad resulta un acierto a la hora de plantear como objetivo principal de su corpus el análisis del género 'novela' compaginando métodos tradicionales de lexicografía y teoría de los géneros con el aprendizaje automático y el procesamiento del lenguaje natural. Ningún estudio crítico ha tenido como objeto en su investigación un corpus tan amplio, tan variado, y tan extenso cronológicamente. Frente a nominalistas y postmodernos que han venido haciendo sus relecturas basándose en la fluctuación de las palabras, este libro ofrece cifras y análisis cuantitativo de datos para dar respuesta a preguntas complejas.

Partiendo de la idea de que «genres are a contradictory cultural phenomenon» (21), José Calvo Tello acota las fronteras del género novela teniendo en cuenta criterios estilísticos (novela modernista, novela realista, novela naturalista...), su grado de ficcionalización (novela histórica, novela de aventuras...) y sus temas (novela cómica, novela erótica, novela social...). ¿Cómo acotar un 'género literario' con ciencia computacional?, ¿qué datos y metadatos distinguen la novela de aventuras de la novela histórica?, ¿puede un texto pertenecer a varios géneros teniendo en cuenta el etiquetado de los textos?, ¿hay más géneros literarios aparte de los canónicos?, ¿en qué momento las novelas se hacen más cortas y por qué? A todas estas preguntas trata de dar respuesta en el segundo y tercer apartado de libro. Destacable es su rigor metodológico y la cantidad de datos que analiza, lo que suscita una gran fuente de creatividad en las hipótesis de partida. José Calvo Tello piensa con voz propia a partir del capítulo tercero, titulado «Data: Texts and Metadata», en el que tras exponer los

criterios de selección para sus corpus ofrece los primeros resultados en gráficas que representan los subgéneros de la novela durante sesenta años. Llama la atención el resultado de sus análisis de narraciones escritas por mujeres, pues solo ha conseguido analizar obras de siete escritoras –E. Pardo Bazán, C. de Burgos, R. de Castro, R. Chacel, C. Espina, E. Mulder y P. de Sinués– y concluye que «the data shows that it is harder for women to get into the canon, but once they do, they are treated similar to men» (125). La década en la que más producción narrativa hay es en los años 1920, tanto para escritores como para escritoras. Fundamental es el epígrafe en el que expone los criterios que se han seguido en su «Typology of Metadata», donde justifica la selección de metadatos de autor, texto, género (subtítulo, subgénero...), lugar (espacio, ciudad, país, continente...), tiempo (periodo, año...), acción, protagonista (sexo, nombre, profesión, clase social, edad...) y otros (narrador, acción, tipo de final...). Una de las conclusiones más significativas de su análisis tiene que ver con los rasgos autobiográficos en las novelas, que son más frecuentes hacia 1900 y van decreciendo conforme avanza el siglo; con ello se demuestra mediante análisis computacional el tránsito de la novela realista a la novela lírica. En cuanto a los espacios narrativos, Madrid es el lugar donde se desarrolla la acción en la mayor parte de estas novelas. Otra característica propia de la novela moderna es el desarrollo de los conflictos en época contemporánea; así, un 95% de las novelas tienen lugar entre 1850 y 1939.

El capítulo cinco, titulado «Analysis of Subgenre Labels», ofrece datos significativos relacionados con las primeras ediciones de obras en las que priman subgéneros como ‘memorias’, ‘historia’, ‘episodio nacional’ y ‘costumbrista’ sobre aquellos subgéneros relacionados con ‘humor’, ‘leyenda’ o ‘regionalista’. Estas etiquetas se corresponden con los nombres facilitados en títulos, subtítulos de los libros y catálogos de los editores o librerías. Lo más sorprendente es su estudio sobre el género denominado ‘ficción literaria’ que aparece al realizar cotejos de rasgos lingüísticos comparables con otros subgéneros como ‘memorias’, ‘diálogos’, ‘modernistas’, ‘naturalistas’ o ‘aventuras’. La ‘literariedad’ de los textos es una frontera en el corpus CoNNSA que no se sobrepasa, pero se pone en evidencia que esta característica es común tanto en la literatura denominada ‘seria’ o ‘canónica’ como en aquella que se considera ‘para-literatura’, ‘literatura popular’ o ‘mala literatura’. «In my opinion – argumenta el autor–, digital Literary Studies is increasingly defining a space of work in which ‘literary fiction’ is only a part, completing it with more popular genres» (268). El análisis computacional de textos literarios puede redefinir los géneros y subgéneros literarios pues «[p]erhaps subgenres would be better described in groups of subgenres or macro-subgenres: genre-literature (war, adventure, erotic

novel, etc.) against a group of literary fiction subgenres (educational, realist, naturalist novel, etc.) (268).

Del capítulo sexto, interesa el apartado dedicado a la identificación de subgéneros escondidos en el que, a través de dendrogramas se presentan dos grupos de novelas homogéneos: el «cluster 217» o «rural-intimist novels» y el «cluster 50» o «mono-dialogue novels». A este último conjunto de obras el autor propone denominarlas 'bucolic' recogiendo una tradición europea que combina la narrativa con otros géneros (drama, poesía o pastorales). Este subgénero estaría representado por las novelas de Miró, Valle-Inclán, Azorín y Jarnés en las que la descripción de la naturaleza se convierte en el eje narrativo de su ambiente rural. Por muy computacional que sea esta clasificación, término y autores traen a la memoria las investigaciones de Darío Villanueva (1983) y Ricardo Gullón (1984) sobre la novela lírica, que a su vez recogen la tesis expuesta por Ralph Freedman en 1966 y permiten equiparar las técnicas utilizadas por estos novelistas españoles con las que aportan en sus novelas escritores como, Marcel Proust, Herman Hesse, André Gide o Virginia Woolf.

En el capítulo séptimo se incide en el análisis de los subgéneros narrativos con metodología computacional y José Calvo Tello presenta los límites de su investigación al concluir que en algunos casos «the algorithm was not able to find the necessary features to distinguish the categories, or, in other terms, was not able to find the necessary and sufficient conditions» (246). Parece conveniente reconocer que en este tipo de análisis algunos subgéneros están mejor clasificados que otros porque tienen rasgos distintivos más marcados; por ejemplo, es útil demostrar con datos que las novelas con asuntos costumbristas o naturalistas están mejor representadas entre 1890 y 1900, mientras que los argumentos más poéticos van creciendo progresivamente y culminan hacia 1920.

El acierto de esta tesis es que en su punto de partida es capaz de superar el conocimiento analógico e historicista para proponer ideas fruto del análisis computacional: 1) los géneros son categorías que utilizan varios agentes (autores, casas editoriales, bibliotecas, lectores e investigadores) y no solo dependen de las características intrínsecas de los textos; 2) hay novelas que pueden pertenecer más a un subgénero que a otro, atendiendo a características lingüísticas o literarias dominantes; 3) los resultados de la clasificación son raramente perfectos, tanto en la clasificación previa como en la que ofrece la teoría computacional; 4) los subgéneros no son conceptos aislados y suelen ser difíciles de distinguir; 5) las obras principales muestran características más canónicas en contraste con otras obras que podrían considerarse periféricas; 6) Generalmente la centralidad de las categorías en los grafos se correlaciona con

los resultados de clasificación y el estatus canónico de los textos en la historia de la literatura. En definitiva, los resultados de esta tesis se adaptan a un corpus compuesto por obras de autores en su mayoría canónicos. Los subgéneros más centrales son los coinciden con los estilos canonizados (realismo, naturalismo, modernismo). Los subgéneros periféricos coinciden con los temas más populares (aventuras, erotismo, guerra...).

De gran utilidad es el epígrafe final de este volumen en el que se describen veintisiete subgéneros mostrando su clasificación, su canonicidad, sus aspectos semánticos distintivos, sus aspectos textuales y literarios, el número de textos en el corpus, su representación y su similitud con otros subgéneros en términos de características y etiquetados. Siguiendo estos parámetros se describen la novela realista, la novela social, la ficción literaria, la novela educativa, la novela filosófica, la novela naturalista, novela histórica, las memorias, la novela bucólica, la novela de aventuras, la novela biográfica, la novela modernista, la novela autobiográfica, la novela cómica, la novela psicológica, la novela espiritual, la novela fantástica, la novela de guerra, la novela dialógica, la novela mono-dialógica, la novela costumbrista, la *novela*, la novela erótica, el episodio nacional, la novela poética y la greguería.

The Novel in the Spanish Silver Age. A Digital Analysis of Genre using Machine Learning es una investigación abierta cuyos resultados deben crecer en una doble vertiente: el análisis computacional de más novelas procedentes de la denominada literatura periférica, popular o de masas; y, su comparación con otras novelas internacionales para cotejar si los resultados vertidos en esta investigación tienen parangón en la llamada literatura global publicada entre 1880 y 1939. Queda mucho por construir aún, pero la piedra angular ya está puesta.