

1. Introducción

Este libro se sitúa en el marco de las humanidades digitales, un área de investigación (y también de enseñanza e incluso creación) en la que convergen las humanidades y las tecnologías de la información. Las humanidades digitales abarcan desde el diseño y almacenamiento de materiales en soportes físico y digital hasta el análisis de datos culturales a gran escala. En este libro nos ocupamos, a través de las humanidades digitales, de una disciplina humanística concreta: la literatura. Y lo hacemos con propósitos analíticos. De forma más precisa, estudiamos la producción narrativa de Benito Pérez Galdós, de la que ofreceremos un análisis crítico utilizando un enfoque novedoso en el marco de los estudios literarios: la estilística de corpus. Este enfoque combina metodologías propias de la lingüística de corpus con motivaciones teóricas y marcos de estudio tanto de la crítica literaria como de la estilística tradicional para ofrecer valoraciones críticas del texto literario. Aunque consolidada en el mundo anglosajón, se trata de una disciplina todavía incipiente en el ámbito hispánico —con notables excepciones, naturalmente, como demuestran los trabajos de Parodi (2010), Rojo (2021), Cantos-Gómez y Howe (2022) o Ruano San Segundo (2024), entre otros. Podría decirse que este libro pretende así contribuir a demostrar el potencial de los enfoques de corpus en el análisis de textos literarios en nuestra lengua.

Antes de explicar cómo planteamos nuestro análisis de la obra de Galdós mediante un enfoque de estilística de corpus, conviene detenerse brevemente en el carácter heterogéneo de las humanidades digitales. En el caso de la literatura, por ejemplo, las humanidades digitales se relacionan tanto con la producción como con la recepción del texto literario. En lo que a la producción se refiere, las humanidades digitales han alterado, por así decirlo, la concepción del propio texto literario. Así, el texto literario tradicional entendido como un tipo de información claramente estructurada o secuenciada ha dado

paso, con el uso de las nuevas tecnologías de la información, a una nueva realidad inconcebible hace no demasiado tiempo: el texto literario como un tipo de información no estructurada (o débilmente estructurada) capaz de ser procesada con programas y consumida atendiendo a parámetros distintos a los habituales. Dicho de otro modo, el texto literario ha dejado de asociarse indisolublemente a una edición impresa de este. Es así como ha surgido, por ejemplo, el concepto de literatura electrónica, un nuevo paradigma en el que la creación artística ha de entenderse desde la lógica del contexto digital, que no es traducible a otros soportes sin perder su esencia (Navas Ocaña y Romero López 2013). Aunque en el texto literario electrónico se engloban elementos literarios tradicionales, también se aprovechan las capacidades y los contextos informáticos, lo que hace difícil (y estéril) su análisis con parámetros de la crítica literaria tradicional o teorías de la literatura. No es nuestra intención profundizar en esta dimensión de las humanidades digitales y la literatura, pues poco tiene que ver con la de Galdós. Sin embargo, consideramos útil mencionarla, dado que nuevas realidades como la poesía hipertextual o la ficción interactiva en la que el lector participa en la configuración de una obra virtual han pasado a formar parte del paisaje literario asociado a las humanidades digitales.

En este libro –insistimos– no nos interesa esa dimensión de las humanidades digitales y la literatura, sino la relación entre ambas en lo que a la recepción del texto literario tradicional se refiere. Por ello, cabe destacar que la relación de la literatura con las humanidades digitales es, como las propias humanidades digitales, muy heterogénea (Moretti 2018), y abarca desde la digitalización de textos hasta el análisis crítico de estos. Aunque no se trata aquí de detallar aspectos que abordaremos con detenimiento en el capítulo 2 de este libro, podemos decir que las humanidades digitales han permitido desarrollar nuevas formas de acercarse al texto literario. El almacenamiento de obras y su difusión es tal vez el mejor ejemplo de ello. Efectivamente, la digitalización y creación de archivos electrónicos, su organización en repositorios digitales o el etiquetado de los textos han facilitado no solo una mayor accesibilidad a los textos literarios en general, sino también que el lector pueda interactuar con ellos y con otros aspectos con los que guardan relación a través de, por ejemplo, la utilización de hipervínculos. Por su parte, en lo que al análisis crítico de textos literarios se refiere, el desarrollo de las nuevas tecnologías ha permitido implementar enfoques de tipo cuantitativo, como la estilometría, el análisis estadístico de textos (más conocido por sus siglas en inglés PCA: Principal Component Analysis), el TF-IDF (Term Frequency-Inverse Document Frequency)

y, de forma más reciente, los métodos de clasificación de la IA para identificar autores o clasificar textos según sus estilos –SVM (Support Vector Machine) o los clasificadores Naive Bayes, entre otros–. Es cierto que estos enfoques han sido siempre recibidos por un sector de la crítica literaria con cierta renuencia (véase apartado 2.1.2). Desde luego hay motivaciones legítimas en algunas de las reticencias de ese sector de la crítica literaria tradicional, pues en muchos de estos análisis el componente cuantitativo desplaza la visión crítica del texto. Sin embargo, con el paso del tiempo, y gracias sobre todo al desarrollo de nuevos programas para el procesamiento de textos en el campo de la lingüística, la valoración de textos literarios utilizando recursos computacionales se ha ido refinando. Es así como han surgido nuevas vías de acercamiento, cada vez más aceptadas y con una dimensión cualitativa capaz de dotar al enfoque de un aparato crítico lo suficientemente sólido como para dar validez a hallazgos de tipo artístico.

En este libro utilizamos un enfoque de estilística de corpus para llevar a cabo nuestro estudio, una disciplina relativamente reciente y que, como decíamos al inicio de estas páginas de introducción, se sitúa entre la lingüística de corpus, por un lado, y la crítica literaria y la estilística tradicional, por otro. Al contrario que otras disciplinas cuantitativas como la estilometría o el análisis estadístico de textos que mencionábamos más arriba, la estilística de corpus se caracteriza por una mayor presencia de la dimensión cualitativa en sus análisis, situándola al mismo nivel que la cuantitativa. De hecho, los presupuestos teóricos sobre los que se cimientan los análisis de la estilística de corpus son similares a los de los estudios de estilística tradicional (véase apartado 3.1.1). Sin embargo, al contrario de lo que ocurre con un enfoque de estilística tradicional, la estilística de corpus es capaz de manejar cantidades de texto inasumibles para enfoques basados en una lectura atenta de los textos. Por esta razón, la estilística de corpus ofrece claras ventajas sobre disciplinas como la antedicha estilística tradicional o la crítica literaria. En este libro esperamos ser capaces de demostrar cómo este enfoque permite realizar análisis sistemáticos y sólidos sobre la producción narrativa de Benito Pérez Galdós. Como se verá, gracias a un enfoque de corpus es posible detectar patrones que tradicionalmente han pasado desapercibidos para la crítica literaria tradicional pero que constituyen hábitos estilísticos que contribuyen a definir el estilo del autor canario. De esta manera confiamos en ser capaces de demostrar cómo las humanidades digitales se relacionan con la dimensión exegética de los textos y cómo el análisis de estos puede beneficiarse del uso de las nuevas tecnologías.

Precisamente sobre cómo las tecnologías se relacionan con la valoración de los textos literarios en el marco de las humanidades digitales es importante detenerse en el espíritu conciliador del libro. Aunque debería resultar obvio para cualquier lector, a lo largo del libro no posicionaremos en ningún caso la estilística de corpus como una opción de análisis más eficaz que otras de corte más tradicional, sino como un enfoque que puede complementarse con otras disciplinas y del que el análisis de textos literarios se puede beneficiar en gran medida. El espíritu de este libro, como referimos, es conciliador. De hecho, somos conscientes de que, precisamente por su naturaleza todavía incipiente en el ámbito hispánico, al acercarse a la estilística de corpus uno ha de tener en cuenta esa falta de antecedentes teóricos propia de su estatus novel. A pesar de ello, sin embargo, estamos convencidos del potencial analítico de la estilística de corpus y de que, en ocasiones, estos enfoques pueden resultar el modo más eficiente de investigar aspectos que, bien por su numerosidad bien por su dispersión textual en la obra de un autor, han pasado desapercibidos para la crítica tradicional. A lo largo del libro esperamos ser capaces de demostrar cómo tecnología y análisis literario pueden conjugarse para detectar matices estilísticamente relevantes en la obra de Galdós¹ que no podrían haberse analizado de otra manera y extraer así conclusiones que nos ayuden a comprender mejor la figura del autor canario como novelista.

En cuanto a sus contenidos, el libro se estructura en torno a bloques temáticos bien diferenciados, organizados en diferentes capítulos que van desde lo general a lo particular. En primer lugar, en el capítulo 2 se ofrece una contextualización sobre las humanidades digitales y su relación con la literatura, prestando especial atención a cuestiones que tienen que ver con la comentada relación de las humanidades digitales con la valoración de textos literarios mediante el uso de las nuevas tecnologías de la información. Para ello ofreceremos una definición del campo que nos sirva de punto de partida y que ayude a que el lector entienda el enfoque de corpus que empleamos para analizar el estilo de Pérez Galdós. También nos detendremos en la distinción entre los conceptos de humanista digital y humanista tradicional y explicaremos cómo se han configurado las humanidades digitales como área desde la que estu-

¹ Como se explica con más detenimiento en el capítulo 3 (apartado 3.4.1) nuestro corpus de estudio de las novelas de Galdós (89 novelas; ca. 6,4 millones de palabras) está formado por textos descargados en su totalidad del repositorio digital de la Biblioteca Virtual Miguel de Cervantes (s.f.). La calidad de la versión digitalizada de los textos de este repositorio, fundamental en los estudios de corpus, está fuera de toda duda, pues la Biblioteca cuenta con un consejo científico que avala el rigor de los materiales alojados en el repositorio (véase apartado 2.2.1).

diar el texto literario, ocupándonos de dos de sus pilares fundamentales en el análisis crítico de textos literarios: los datos y su procesamiento. Todo ello servirá para definir, con precisión, el marco de estudio de este libro.

El capítulo 3, por su parte, se adentra en la estilística de corpus como enfoque empleado en este libro para llevar a cabo nuestro análisis. Debido a su carácter novel, dedicaremos el espacio del capítulo a acomodarla en el campo de los estudios literarios, poniendo de manifiesto cómo se relaciona con (y en qué se diferencia de) la lingüística de corpus y la estilística tradicional. Ofreceremos asimismo un contexto en el marco de los estudios hispánicos, donde a pesar de ser una disciplina aún por explotar, existen trabajos seminales en los que se emplean estos enfoques, así como precedentes teóricos de otros campos relacionados con el análisis cuantitativo de la literatura, como la estilometría o el análisis estadístico de textos. En este capítulo explicaremos también cuestiones relacionadas con la compilación de un corpus e introduciremos los que utilizamos en nuestro análisis.

Tanto el capítulo 2 como el capítulo 3 son el contexto necesario para abordar el análisis que se lleva a cabo en los capítulos 4, 5, 6, 7, 8 y 9, donde se investiga la producción narrativa de Galdós². Estos capítulos constituyen la parte verdaderamente novedosa de este libro, pues constituyen un análisis de las novelas del autor canario desde un prisma inédito hasta ahora. En el capítulo 4, en primer lugar, se presenta un análisis de palabras clave (*keywords*) en *La Fontana de Oro*, primera novela de Pérez Galdós. En concreto, se compara la primera novela del autor con el resto de su producción con el fin de identificar hábitos estilísticos del Galdós más joven. Este análisis no solo permitirá identificar aquellas palabras estadísticamente más significativas en *La Fontana de Oro* que en el resto de las novelas de Galdós y calibrar su representatividad, sino analizar de forma sistemática el uso que el autor canario hace de ellas y examinar su valor desde un punto de vista literario. El análisis se centrará en los verbos de comunicación (*decir, contestar, preguntar, etc.*), un elemento claramente protagonista en *La Fontana de Oro*. Como se explicará, además de para dar voz a sus personajes, Galdós utiliza estos verbos con unos fines estilísticos concretos. El análisis abordará aspectos de tres dimensiones bien diferenciadas: la psicológica, la física y, por supuesto, la discursiva. El análisis

² El contenido de los capítulos 4, 5 y 6 ha sido expuesto en distintos congresos científicos (Nieto Caballero 2016a, 2016b, 2017a, 2017b, 2017c). Más tarde, este contenido se publicó en distintos artículos (Nieto Caballero, 2019b, 2020a, 2020b). Aunque no se trata de una reproducción exacta de tales publicaciones, parte del contenido y los ejemplos que se analizan en los capítulos 4, 5 y 6 han sido publicados en estos artículos.

sis de estas tres dimensiones servirá para arrojar luz sobre la configuración de determinados aspectos de la historia, como la caracterización de los personajes o la creación del universo ficticio que el autor canario nos plantea.

En el capítulo 5, en segundo lugar, llevamos a cabo un análisis de los adverbios en *-mente* a lo largo de toda la producción narrativa de Galdós (89 novelas; ca. 6,4 millones de palabras). Mediante un análisis de concordancias, mostraremos el uso que Galdós hace de este elemento como unidad estilísticamente significativa en el marco de la proyección del discurso de los personajes que pueblan sus historias. Como se verá, la frecuencia con la que Galdós recurre a este elemento no solo dobla a la de los novelistas decimonónicos con los que compartió plana, sino que además recurre a patrones similares a lo largo de sus novelas para configurar distintos aspectos, tanto formales como funcionales. El estudio se centra en tres patrones relacionados con la voz, el plano mental y el lenguaje gestual de los personajes, que servirán para mostrar hábitos relacionados con la caracterización que sugieren estrategias en el diseño de la construcción de sus personajes en el plano textual que hasta ahora no han sido exploradas en el marco de las técnicas de representación del discurso del autor canario.

En el capítulo 6, en tercer lugar, analizamos el valor del espacio como uno de los ejes sobre los que se vertebra la creación de los universos ficticios galdosianos. Para ello realizamos un análisis de *clusters* tanto en el corpus de estudio de novelas de Galdós como en un corpus de referencia de novela española decimonónica, que nos servirá para identificar patrones tanto formales como funcionales típicamente galdosianos que han pasado desapercibidos para la crítica tradicional. De manera más concreta, nos centraremos en segmentos de texto que contienen referencias espaciales que, además de encontrarse entre las construcciones más utilizadas por Galdós en sus novelas, son empleadas con un valor similar a lo largo de toda su producción. Los ejemplos localizados cumplen distintas funciones –que abarcan desde la identificación de personajes hasta su uso como referencia para la finalización de escenas en momentos de despedida–, cuyo análisis nos servirá para arrojar luz sobre los hábitos estilísticos de Galdós en su tratamiento del espacio, uno de los aspectos por los que goza de mayor reconocimiento.

En el capítulo 7, en cuarto lugar, presentamos un estudio de la representación de la oralidad en la producción narrativa de Galdós. Para realizar el análisis, recurrimos, como en el capítulo anterior, a la identificación de *clusters* en el corpus de estudio. Sin embargo, en este capítulo nos centramos en aquellas secuencias de palabras utilizadas únicamente en construcciones de

estilo directo, de tal suerte que la masa de texto que analicemos se corresponda exclusivamente con la representación directa del discurso de los personajes. Este matiz nos permitirá explorar fórmulas empleadas por Galdós de manera recurrente para construir el habla de los personajes que pueblan sus universos ficticios, lo que servirá de complemento a los conocidos recursos de caracterización mencionados habitualmente por la crítica, como las muletillas de los personajes —«en toda la extensión de la palabra», de doña Lupe, en *Fortunata y Jacinta*, por ejemplo—. Al contrario que estos recursos tan llamativos, que sirven para individualizar el habla de un personaje concreto en el marco de la novela en la que aparece, los *clusters* de los que nos ocupamos en este capítulo actúan, como veremos, como una suerte de base textual desde la que Galdós construye el habla de sus personajes en general y desde la que, como sugerimos, dota de realismo y verosimilitud a la caracterización de los personajes, uno de los recursos por los que el autor canario goza de mayor reconocimiento.

En el capítulo 8, en quinto lugar, también recurrimos a un análisis de *clusters* del corpus de estudio para investigar el lenguaje corporal en las novelas del autor. El propósito de este capítulo es identificar y valorar el uso construcciones repetidas de las que Galdós se sirve para configurar hábitos gestuales con los que dotar de autenticidad y realismo a sus personajes. Para ello nos centraremos en las partes del cuerpo empleadas de forma más frecuente en sus novelas (fundamentalmente las manos, los ojos y la cabeza, pero también otras partes del cuerpo como la espalda y los hombros). Como en el caso del capítulo 7, nuestro interés no es tanto explicar idiosincrasias del lenguaje corporal de personajes concretos como ser capaces de demostrar la regularidad en el uso de construcciones a lo largo de sus novelas relacionadas con el uso de lenguaje corporal, en las que se apoya para definir a los personajes en el plano textual. Como se podrá comprobar, existen patrones estilísticamente relevantes que desempeñan funciones literarias concretas a lo largo de sus novelas, como la conceptualización de emociones mediante el lenguaje gestual que acompaña al discurso de los personajes.

En el capítulo 9, por último, analizamos el estilo de Galdós desde una perspectiva distinta. En concreto, utilizamos la estilística de corpus como enfoque desde el que investigar la influencia que Charles Dickens ejerció en el estilo del autor canario. El análisis se centra en las suspensiones, que pueden definirse como una interrupción del habla del personaje por parte del narrador. Una suspensión puede tener varias funciones, como organizar el discurso, ofrecer información relevante sobre la caracterización de los personajes o

crear efectos como la sensación de simultaneidad entre las palabras del personaje y la acción (si la hubiera) que se describe en la propia suspensión. El uso de suspensiones ha sido identificado como un hábito literario característico de Dickens que, gracias a los avances en los estudios de corpus, ha podido ser analizado de forma sistemática a lo largo de toda su producción. En este capítulo investigamos el uso que Galdós hace de este elemento en *Doña Perfecta*. Para ello utilizamos un sistema de anotación, que nos ha permitido identificar las suspensiones en el texto galdosiano de forma automática. Como se podrá comprobar, el uso que Galdós hace de las suspensiones es similar al de Dickens, tanto desde un punto de vista formal como funcional. La sistematicidad con la que el autor canario recurre a hábitos textuales que podríamos calificar como dickensianos, así como su utilización para configurar efectos que también caracterizan el uso que Dickens hace de este elemento, servirán para reforzar la consabida influencia que el autor victoriano ejerció sobre Galdós desde una perspectiva inédita en la exégesis del autor canario.

El libro finaliza con un capítulo de conclusiones sobre los hallazgos más importantes realizados en nuestro análisis, relacionándolos con las metodologías y los enfoques que los han hecho posibles. También nos referiremos al potencial de los estudios de estilística de corpus en el ámbito de la literatura española, así como a las posibilidades de análisis que estos abren y el futuro que tiene por delante la estilística de corpus en el mundo hispánico como enfoque de análisis crítico. Como esperamos ser capaces de demostrar tanto desde la teoría como la praxis, este enfoque atesora un potencial en la dimensión exegética de los autores de nuestra lengua que, impulsado por la constante evolución a la que estamos asistiendo en los últimos tiempos en el marco de las tecnologías de la información en general, merece la pena explorar³.

En suma, en *Humanidades digitales y análisis estilístico: la lengua de Benito Pérez Galdós* nos ocupamos tanto de analizar la figura de Galdós como de ofrecer una explicación razonada (y respaldada por el análisis que realizamos de la obra del novelista canario) del potencial de los estudios de corpus. Esta doble vertiente por la que discurre el libro le confiere, por así decirlo, una utilidad también doble. De un lado, los galdosistas que se acerquen al

³ Huelga decir, antes de cerrar de este capítulo de introducción, que los análisis que se presentan en los capítulos 4 a 9 que acabamos de detallar no son sino casos de estudio concretos desarrollados con herramientas específicas. Desde una perspectiva general, por lo tanto, estos análisis pretenden ser un botón de muestra del potencial de los enfoques de estilística de corpus en la interpretación de textos literarios en español, pero en ningún caso agotan las posibilidades de análisis ni los planteamientos metodológicos que las herramientas de corpus ponen a disposición del investigador.

libro encontrarán un análisis novedoso, que, construido sobre la teoría de la crítica literaria y la estilística tradicional, ofrece nuevos hallazgos sobre el estilo del autor canario. De otro lado, los humanistas digitales que se acerquen al libro encontrarán una dimensión de las humanidades digitales poco frecuente en el ámbito literario: el análisis exegético de literatura tradicional. Confiamos en ser capaces de conjugar ambos aspectos para que cualquier lector pueda aprovechar plenamente sus contenidos sin importar su área de especialización —ya sean los estudios galdosianos o las humanidades digitales—. Si lo conseguimos, habremos sido capaces de dar cumplimiento a otros de los objetivos que nos hemos marcado: lograr que la estilística de corpus sea un punto de encuentro para las nuevas tecnologías de la información y el análisis crítico de textos literarios, algo habitual en el ámbito internacional pero que, desafortunadamente, todavía constituye una vía por explorar en el ámbito hispánico.