

ÍNDICE

1. INTRODUCCIÓN	11
1.1. Contexto académico.....	12
1.2. Estructura del libro.....	14
1.3. La aportación de este libro.....	17
2. LA LINGÜÍSTICA DE CORPUS	19
2.1. ¿Qué es la lingüística de corpus?	19
2.1.1. ¿Permite un enfoque de corpus investigar cualquier hipótesis lingüística?	23
2.2. Dos dimensiones bien definidas.....	24
2.3. ¿Disciplina de estudio o metodología?	26
2.4. Enfoques basados en corpus y enfoques guiados por corpus	29
3. EL CORPUS	33
3.1. Sobre el concepto de corpus	33
3.2. Principios básicos para la elaboración de un corpus.....	36
3.2.1. Representatividad.....	37
3.2.1.1. Tamaño	40
3.2.1.2. Especificidad.....	43
3.2.1.3. Equilibrio.....	44
3.2.2. Consideraciones finales.....	45
3.3. Anotación de un corpus.....	46
3.3.1. Qué es la anotación de un corpus.....	46
3.3.2. Cómo se lleva a cabo la anotación	47
3.3.2.1. Fases.....	48
3.3.3. Tipos de la anotación.....	51
3.3.4. Críticas a la anotación	54
3.4. Cuestiones legales.....	56
3.4.1. Derechos de autor.....	56
4. HERRAMIENTAS Y OPCIONES DE ANÁLISIS	59
4.1. Herramientas	60
4.1.1. Herramientas integradas en sitios web.....	60

4.1.2. <i>Software</i> independientes.....	65
4.1.3. <i>Software</i> de anotación	68
4.2. Opciones de análisis.....	71
4.2.1. Concordancias	71
4.2.2. Listas de palabras	73
4.2.3. Palabras clave.....	75
4.3. Una nota de precaución.....	78
5. LA LINGÜÍSTICA DE CORPUS Y EL ESTUDIO DE LA LITERATURA.....	81
5.1. Marco teórico	81
5.1.1. La dimensión exegetica de la estilística de corpus.....	83
5.1.2. Conceptualización de la estilística de corpus y antecedentes en el ámbito hispánico.....	83
5.2. Análisis de intertextualidad entre Charles Dickens y Benito Pérez Galdós..	87
5.2.1. Ecos dickensianos en la obra de Galdós	90
5.2.2. Metodología de estudio	93
5.2.3. Resultados	97
5.2.4. Análisis.....	99
5.2.4.1. Forma y función de las suspensiones	99
5.2.4.2. Proposición proyectada: primera parte.....	102
5.2.4.3. Proposición proyectada: segunda parte.....	104
5.3. Conclusiones.....	108
6. LA LINGÜÍSTICA DE CORPUS Y LOS ESTUDIOS DE TRADUCCIÓN.....	111
6.1. Marco teórico	111
6.1.1. Enfoques lingüísticos y enfoques culturales en los estudios de traducción	112
6.1.2. Traducción y estudios de corpus	114
6.1.2.1. Traducción y estudios de corpus en el ámbito literario	115
6.2. Un caso práctico: Galdós como traductor de Charles Dickens.....	115
6.2.1. Galdós y su única traducción: <i>Las aventuras de Pickwick</i>	116
6.3. Los verbos de habla: valor literario en la obra de Dickens y traducción de Galdós.....	118
6.3.1. Localización de ejemplos en <i>Pickwick Papers</i> , de Charles Dickens ..	121
6.3.2. Localización de ejemplos en <i>Las aventuras de Pickwick</i> , de Galdós .	124
6.3.3. Análisis de ejemplos.....	127
6.3.3.1. Omisión del verbo de habla	128
6.3.3.2. Utilización sistemática del verbo “decir”	131
6.3.3.3. Traducciones libres	134
6.3.3.4. Errores de traducción.....	137

6.4. Los adverbios en <i>-mente</i> : valor literario en la obra de Dickens y traducción de Galdós	139
6.4.1. Localización de ejemplos en <i>Pickwick Papers</i>	140
6.4.2. Localización de ejemplos en <i>Las aventuras de Pickwick</i> , de Galdós	145
6.4.3. Análisis de ejemplos.....	146
6.5. Consideraciones finales	149
7. LA LINGÜÍSTICA DE CORPUS Y LA ENSEÑANZA Y APRENDIZAJE DE LENGUAS	151
7.1. Marco teórico	151
7.1.1. Tipos de aplicaciones de enfoques de corpus en la enseñanza de lenguas	152
7.2. Incorporar la enseñanza de lengua en el aula de literatura en el ámbito universitario: claves del estilo directo en torno a la literatura inglesa victoriana.....	155
7.2.1. El estilo directo como estructura susceptible de análisis	158
7.2.2. Metodología y resultados	163
7.2.2.1. ¿Qué ocurre en versiones adaptadas de los textos?.....	170
7.2.3. Consideraciones finales.....	171
8. CONSIDERACIONES FINALES	173
8.1. Cuál es la principal contribución de este libro.....	173
8.1.1. La lingüística de corpus y la estilística	174
8.1.2. La lingüística de corpus y la traducción literaria	175
8.1.3. La lingüística de corpus y la enseñanza de lenguas	176
8.2. Sugerencias y desafíos	177
8.3. A modo de cierre	178
REFERENCIAS BIBLIOGRÁFICAS.....	181

INTRODUCCIÓN

Este libro se presenta ante el lector con la intención de ilustrar distintas dimensiones del componente anglosajón presente en el ámbito hispánico mediante distintos casos de estudio con un denominador común: la metodología de lingüística de corpus empleada para llevar a cabo los análisis. El libro discurre por una doble vertiente. Por un lado, la intención del libro es presentar, de una forma ordenada y lo más didáctica posible, una fundamentación teórica de los principios básicos de la lingüística de corpus, deteniéndose en los aspectos más importantes de esta disciplina. Por otro lado, se ofrecen tres análisis relacionados con distintas áreas de estudio de la lengua y la literatura en las que las metodologías de lingüística de corpus pueden aplicarse. En concreto, nos ocupamos de la relación que puede establecerse entre la lingüística de corpus y (i) la estilística (capítulo 5), (ii) la traducción literaria (capítulo 6) y (iii) la enseñanza de lenguas extranjeras (capítulo 7), en un decidido intento de demostrar la aplicabilidad de las metodologías de corpus en áreas de estudio que van más allá de la lingüística. El denominador común de los tres casos de estudio que planteamos es, como ya se ha dicho, el componente anglosajón en el ámbito hispánico, pues en cada caso se abordan aspectos de la lengua y la literatura inglesas que son analizados en el marco de un contexto de nuestra cultura claramente definido (véase apartado 1.2).

Sobre la motivación académica que subyace a la publicación de este libro, consideramos que un trabajo como el que aquí se presenta se justifica por la falta de obras de temática similar escritas en español —sobre todo en comparación con el ámbito internacional, especialmente en el mundo anglosajón (Biber *et al.* 1998; Tognini-Bonelli 2001; McEnery y Wilson 2001; Meyer 2002; Sinclair 2004; Baker *et al.* 2006; O’Keeffe y McCarthy 2010; Gries *et al.* 2010; McEnery y Hardie 2012; Cheng 2012; Biber y Reppen 2015; McEnery y Brezina 2022, entre otros)—. Como recientemente apuntaba Rojo, “no existen introducciones a esta corriente [la lingüística de corpus] escritas

en español que utilicen corpus de español y muestren cómo se pueden tratar e intentar resolver con datos de corpus problemas de lingüística española” (Rojo 2021: x). Esta situación no solo justifica, sino que parece demandar un trabajo como el que aquí presentamos. Esto no quiere decir, naturalmente, que no existan trabajos en el ámbito hispánico que se hayan ocupado de la lingüística de corpus o de su aplicabilidad. De hecho, en los últimos quince años se han publicado tres trabajos de referencia que no podemos dejar de mencionar en estas páginas de introducción, como son el libro de Parodi (2010), el del propio Rojo (2021) y más recientemente el de Parodi, Cantos-Gómez y Howe (2022). En ningún caso debe interpretarse la antedicha necesidad del presente libro como un menoscabo a los tres trabajos que acabamos de citar. Más bien al contrario, pues, como se podrá comprobar en la lectura del libro, hacemos referencia a ellos de forma habitual. La pregunta que cabe hacerse, por tanto, es por qué es necesario otro libro más sobre lingüística de corpus. Pues bien, como señalan McEnery y Brezina (2022: 254), precisamente en su monografía sobre lingüística de corpus recientemente publicada, la lingüística de corpus no es estática, pues su rigor epistemológico se basa precisamente en un constante estado de revisión. Tal vez por ello Rojo (2021) y Parodi *et al.* (2022) en estos dos últimos años han publicado sendas monografías, quizá animados porque la “lingüística hispánica llegó con cierto retraso a la lingüística de corpus” (Rojo 2021: ix) y existe una carencia patente de trabajos de referencia en este sentido (Rojo 2021: x; Parodi: Cantos-Gómez y Howe 2022: 2). Desde luego, esa es una de las motivaciones principales de este libro, en el que aspiramos, no solo a ofrecer una introducción de tipo teórico a los estudios de lingüística de corpus, sino a ilustrar su relevancia y utilidad en el ámbito hispánico.

1.1. Contexto académico

Los contenidos de este libro se sitúan, de forma más amplia, en el marco de las Humanidades Digitales. Las Humanidades Digitales pueden definirse como un “campo interdisciplinario que busca entender el impacto y la relación de las tecnologías de cómputo en el quehacer de los investigadores en las Humanidades” (Galina Russell *et al.* 2018). Ese es precisamente el lugar que ocupa este libro y por ello consideramos necesario detenernos brevemente en las Humanidades Digitales como marco contextual en el que se acomoda este

trabajo. Para algunos académicos, la eclosión de las Humanidades Digitales ha supuesto una revolución parangonable a la de la imprenta de Gutenberg (Trigo 2017: 25), al menos en lo que a la ruptura con el *modus operandi* como investigadores, académicos y docentes se refiere. Spence conceptualiza esta revolución como un cambio de paradigma que implica necesariamente rupturas “técnicas, metodológicas e intelectuales profundas en nuestro modo de investigar que apenas hemos empezado a identificar mucho menos resolver, y que se manifiestan de manera más aguda en la interfaz entre el mundo académico y un público más amplio” (Spence 2014: 118). Estas rupturas son de tal envergadura y han provocado tal seísmo en el estudio de las ciencias humanas que resulta complicado encontrar un consenso sobre los pilares epistemológicos de las Humanidades Digitales, así como sobre cómo estas se relacionan con otras áreas de conocimiento. Esa es tal vez la razón por la que el acercamiento a las Humanidades Digitales suele enfocarse con visión de futuro. La razón para que esto sea así parece tener que ver con esa tendencia a converger cada vez mayor entre las ciencias humanas y las tecnologías de la información, que nos empujan a creer que, con el paso del tiempo, nos resultará más sencillo explicar una realidad que a día de hoy no goza de gran tradición. Sin embargo, las Humanidades Digitales ya son una realidad plenamente consolidada que se manifiesta de muy diversas formas y que dibujan un panorama con reconocimiento a nivel institucional (proyectos investigación, puestos de trabajo, titulaciones oficiales, asociaciones, etc.) (Spence 2018). Así pues, y aunque su alcance pueda ser objeto de debate, es un hecho inopinable que las Humanidades Digitales han pasado de ser una actividad con un enfoque relativamente limitado (a veces simplemente metodológico; es decir, de aplicar tecnología al estudio de una disciplina) a comprender un campo más amplio que estudia varios espacios convergencia entre las tecnologías de la información y el humanismo. Sin duda, el espectro que ocupan las Humanidades Digitales es muy amplio, entre los que encontramos campos tan variados como

el análisis geoespacial en la historia, la aplicación de metodología para Big Data en estudios literarios, las visualizaciones en la presentación de interpretación filológica, la representación de espacios perdidos con tecnologías 3D, el activismo digital, el empleo de los medios interactivos para estudiar el patrimonio cultural o la tecnología como objeto de estudio (Spence 2018).

Este libro es un ejemplo más de ello. En él nos ocupamos de cómo las metodologías de lingüística de corpus pueden aplicarse a distintos ámbitos que

trascienden la propia lingüística, como son (i) la estilística (capítulo 5), (ii) la traducción literaria (capítulo 6) o (iii) la enseñanza de lenguas extranjeras (capítulo 7).

Precisamente sobre cómo las tecnologías se relacionan con las Humanidades, es importante detenerse en el espíritu conciliador del libro. Aunque debería resultar obvio para cualquier lector, no posicionaremos en ningún caso la lingüística 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 la lengua desde distintos ángulos se puede beneficiar en gran medida. A lo largo del libro esperamos ser capaces de demostrar cómo la tecnología y el análisis de la lengua pueden conjugarse para llevar a cabo estudios sólidos desde un punto de vista académico y capaces de explorar hipótesis en el campo de las Humanidades.

Además de este contexto de Humanidades Digitales, si tuviéramos que precisar las áreas de conocimiento en las que se sitúa el libro, podríamos decir que, por la naturaleza de los estudios que se presentan en el bloque práctico (capítulos 5, 6 y 7), nos encontramos a caballo entre la lengua y la literatura españolas, por un lado, y la filología inglesa, por otro. Efectivamente, aunque este libro gira en gran medida en torno a la fundamentación teórica de la lingüística de corpus, que podríamos circunscribir al área de conocimiento de la lingüística, los contenidos abarcan parcelas de diversas áreas de conocimiento. Y aunque es un libro que en gran medida puede circunscribirse al ámbito hispánico, no cabe duda de que en los casos prácticos la filología inglesa es un pilar fundamental. De hecho, en los capítulos 5, 6 y 7 encontramos un hilo conductor en la figura de Charles Dickens, como se detalla en el apartado 1.2. Lejos de ser una coincidencia, este aspecto sirve para ilustrar uno de los pilares en torno a los que se construye este trabajo: la presencia destacada de elementos del acervo popular inglés en el ámbito hispánico.

1.2. Estructura del libro

El libro se divide en dos grandes bloques. El primero comprende los capítulos 2, 3 y 4, de corte eminentemente teórico y que van desde lo general a lo particular. El capítulo 2 constituye un acercamiento a la lingüística de corpus como disciplina de estudio. Además de proponer una definición basada en los componentes que la caracterizan, también se abordan cuestiones que sirven

para entender mejor tanto su desarrollo como su consolidación. A lo largo del capítulo ofrecemos un breve repaso histórico de las dimensiones de la lingüística que sirvieron de germen para el surgimiento de la lingüística de corpus, así como algunas actitudes más reticentes a este tipo de enfoques históricamente —personificadas habitualmente en la figura de Chomsky, como veremos—. Nos ocupamos asimismo de cuestiones como la institucionalización de la lingüística de corpus en el ámbito hispánico e intentamos arrojar luz sobre algunos aspectos que, todavía hoy, resultan conflictivos a la hora de explicar la lingüística de corpus, como la cuestión de si es una metodología al servicio de la lingüística o si, por el contrario, comporta un enfoque con motivaciones teóricas propias.

En el capítulo 3, nos ocupamos del corpus como rasgo definitorio de la lingüística de corpus. De forma más concreta, nos detenemos en los principios fundamentales que deben guiar la compilación de un corpus, que abarcan desde el formato electrónico de los textos hasta las cuestiones de representatividad (tamaño y especificidad de los textos, equilibrio en la distribución de estos, etc.), fundamentales en la compilación de una muestra de estudio sólida. También ofrecemos una explicación sobre aspectos relacionados con su procesamiento (como la anotación) y su uso y difusión (cuestiones legales).

En el capítulo 4, abordamos los recursos empleados para el procesamiento de los corpus y hacemos una distinción entre las herramientas propiamente dichas y las opciones de análisis. Por un lado, presentamos al lector los distintos tipos de herramientas que se suelen utilizar para procesar los corpus. Nos ocupamos tanto de los *software* integrados en algunos sitios web (como la Biblioteca Virtual Miguel de Cervantes o los de la Real Academia de Española) como de los independientes (como WordSmith Tools o AntConc). También nos detendremos en los *software* de anotación. Por otro lado, en la explicación de las distintas opciones de análisis detallaremos cuáles son los procedimientos más habituales a la hora de procesar corpus con herramientas independientes. Nos centraremos en las concordancias, listas de palabras, las palabras clave, que suelen ser los tres pilares sobre los que se diseñan las herramientas más empleadas en los estudios de lingüística de corpus

En el segundo bloque, que comprende los capítulos 5, 6 y 7, nos movemos de la teoría a la práctica, presentamos tres casos de estudios que servirán para ilustrar distintas dimensiones del componente anglosajón presente en la lengua y la literatura españolas. En el capítulo 5, por un lado, analizamos la influencia de Charles Dickens en el estilo de Benito Pérez Galdós a través de

una comparación textual de la ópera prima de ambos autores: *La Fontana de Oro* y *Pickwick Papers*. Ambas novelas se analizarán a partir de una anotación de estas que permite aislar segmentos de texto concretos para posteriormente analizarlos de forma metódica. En concreto, gracias a la anotación de ambas novelas, en este estudio calibramos la presencia de suspensiones¹ en *La Fontana de Oro* y *Pickwick Papers*, y analizamos tanto su forma y sus funciones como su relación con las construcciones de estilo que las rodean, en un decidido intento de demostrar que la influencia que Charles Dickens ejerció sobre Benito Pérez Galdós es más profunda de lo apuntado hasta ahora.

En el capítulo 6, nos detenemos en la traducción de Benito Pérez Galdós del *Pickwick* de Dickens, que comparamos con la traducción de José María Valverde. Para ello, nos centraremos en dos elementos estilísticamente significativos en la producción literaria de Dickens en general y en *Pickwick Papers* en particular que pueden ser identificados y analizados de forma sistemática tanto en el texto original como en las traducciones de este: los verbos de habla y los adverbios en *-mente*. Para localizar los verbos y adverbios en la novela inglesa recurriremos a un *software* de concordancias, mientras que para examinar la traducción de Pérez Galdós de estos elementos utilizaremos un corpus alineado con la novela inglesa y las traducciones de Pérez Galdós y de Valverde. Con este análisis pretendemos demostrar cómo la lingüística de corpus puede aplicarse a los estudios de traducción para la identificación de patrones textuales estilísticamente relevantes y el análisis de estos en el proceso de traducción.

En el capítulo 7, por último, planteamos una propuesta para integrar el estudio de la lengua en la enseñanza de la literatura en el marco del Espacio Europeo de Educación Superior. De forma más precisa, nos centramos en la lengua y la literatura inglesas y su enseñanza como lengua extranjera en el sistema educativo español. Para ello, nos ocupamos de las estrategias de representación del discurso en general y del estilo directo en particular. En concreto, exponemos un caso de estudio a través de la figura de Charles Dickens y de su obra *A Tale of Two Cities*. Además de mostrar las implicaciones del estilo directo como estrategia comunicativa, mediante este análisis esperamos ser capaces de demostrar cómo la enseñanza del estilo directo puede ser aprovechada para lograr un mayor grado de concreción y aislar aspectos

¹ Una suspensión es una “protracted interruption by the narrator of a character’s speech” (Lambert 1981: 6). Para una información más detallada, véase el apartado 5.2.

susceptibles de análisis por su relevancia estilística y lingüística que sirvan para ilustrar claves literarias del período que se esté enseñando, al tiempo que se pueda utilizar como recurso para incorporar la enseñanza de aspectos propios de la lengua. Combinar ambas dimensiones puede resultar beneficioso, no solo para explicar claves literarias de la época, sino para compensar la falta de atención que la enseñanza de estrategias de representación del discurso en general ha recibido tradicionalmente en el currículum de la enseñanza secundaria y bachillerato en España.

1.3. La aportación de este libro

Los capítulos del segundo bloque del libro constituyen la verdadera aportación de este trabajo. Desde distintas perspectivas, los estudios de caso que se presentan en los capítulos 5, 6 y 7 sirven para medir distintos componentes de la dimensión anglosajona en el ámbito hispánico inéditos hasta la fecha. En el caso del estudio de intertextualidad presentado en el capítulo 5, por un lado, descubrimos un aspecto inédito en la exégesis de Galdós que tiene un claro eco en la narrativa de Dickens. Es cierto que existen numerosos estudios sobre la influencia que Dickens ejerció en el estilo de Pérez Galdós, pero no hay análisis sistemáticos del estilo de ambos autores que permitan explorar de manera metódica hábitos lingüísticos que sí pueden explorarse gracias a una metodología de corpus como la que aquí planteamos. Por otro lado, en el estudio sobre la traducción de Pérez Galdós del *Pickwick Papers* de Dickens llevado a cabo en el capítulo 6 advertimos cómo, a pesar de la clara influencia del autor victoriano en el estilo de Galdós demostrada en el capítulo anterior, en su traducción de la novela de Dickens deja de trasladar al español con la fidelidad necesaria elementos idiosincrásicos del estilo de Dickens. La traducción de Galdós de los verbos de habla y los adverbios en *-mente* de los que nos ocupamos a lo largo del capítulo, y que en el caso de Dickens comportan un hábil recurso de caracterización de los personajes que pueblan su universo ficticio, se materializan en una clara mella en el resultado final de su traducción, sobre todo cuando la comparamos con la traducción de José María Valverde. Además de mostrar las implicaciones de la traducción de Galdós, este estudio nos permitirá observar con precisión cómo, a pesar de la clara influencia que Dickens ejerció en el novelista canario mostrada en el capítulo 5, este no supo identificar ese componente del estilo dickensiano en su traducción del

Pickwick Papers. Finalmente, en el estudio de corpus sobre la enseñanza de la literatura en el ámbito del Espacio Europeo de Educación de Superior presentado en el capítulo 7, veremos cómo en la enseñanza de la literatura inglesa es posible localizar claves relacionadas con la lengua en general que pueden ser parte del aprendizaje del inglés como lengua extranjera. A través del ejemplo de Dickens mostraremos cómo la estrategia de estilo directo puede ser analizada —y enseñada— gracias a su uso en las novelas del autor victoriano como un recurso en el marco de las estrategias de representación del discurso desde el que poder profundizar en la enseñanza de distintos elementos del léxico, como los verbos de habla. A través del ejemplo concreto de la novela *A Tale of Two Cities*, demostraremos cómo la enseñanza del inglés como lengua extranjera en España podría utilizar textos literarios ingleses como fuente desde la que abordar dimensiones del léxico que tradicionalmente han estado desatendidas desde un punto de vista pedagógico.

Por último, antes de cerrar estas páginas de introducción conviene destacar que los distintos ámbitos de los que nos ocupamos a lo largo del libro no son, en ningún modo, compartimentos estancos ni pretenden ofrecer una foto fija de la lingüística de corpus como disciplina de estudio o de su aplicabilidad. Tanto en el bloque teórico como en el práctico ofrecemos una serie de contenidos que están sujetos, como no podía ser de otra manera, a ser revisados y ajustados. El constante —y rápido— desarrollo de la tecnología en general y de los bancos de datos y las herramientas para procesar textos en los estudios de lingüística de corpus en particular, por ejemplo, lleva aparejado una necesidad casi constante de revisión del estado de la cuestión. Esa es precisamente una de las razones por las que se siguen sucediendo nuevas publicaciones sobre la lingüística de corpus, como ya apuntábamos más arriba. Por su parte, la aplicación de las metodologías de corpus a distintos campos —en nuestro caso, la estilística, la traducción y la enseñanza de lenguas extranjeras— deberá ajustarse en cada caso a las circunstancias y las necesidades correspondientes. Por eso nos gustaría finalizar estas páginas de introducción incidiendo en que el carácter de este libro no es prescriptivo. Por el contrario, este libro pretende ser una herramienta de consulta y, por qué no, de ayuda, con la que el lector pueda explorar, cuestionar y desarrollar aspectos de la lengua desde una perspectiva innovadora pero eficiente en el estudio de la lengua española: los enfoques de corpus.