

REVISTA
DE INVESTIGACIÓN
LINGÜÍSTICA

Volumen 28 2025 Murcia (España) eISSN: 1989-4554 ISSN: 1139-1146

Departamento de Lengua Española y Lingüística General
FACULTAD DE LETRAS

REVISTA DE INVESTIGACIÓN LINGÜÍSTICA (RIL)

Dirección

David Prieto García-Seco
(Universidad de Murcia)

Secretaría

Elvira Manero Richard
(Universidad de Murcia)

Consejo Editorial

Mercedes Abad Merino (Univ. de Murcia)	Xavier Laborda Gil (Univ. de Barcelona)
Beatriz Gallardo Paúls (Univ. de Valencia)	Mariano Quirós García (CSIC, Madrid)
Cecilio Garriga Escribano (Univ. Autónoma de Barcelona)	Carmen Sánchez Manzanares (Univ. de Murcia)

Consejo Científico

Pedro Álvarez de Miranda (Univ. Autónoma de Madrid y Real Academia Española ~ España)	Michael Metzeltin (Univ. de Viena ~ Austria)
Valerio Báez San José (Univ. Carlos III ~ España)	† Emilio Montero Cartelle (Univ. de Santiago de Compostela ~ España)
José Manuel Blecua Perdices (Univ. Autónoma de Barcelona y Real Academia Española ~ España)	Antonio Narbona Jiménez (Univ. de Sevilla ~ España)
Ignacio Bosque Muñoz (Univ. Complutense de Madrid y Real Academia Española ~ España)	Álvaro S. Octavio de Toledo y Huerta (CSIC ~ España)
Georg Bossong (Univ. de Zürich ~ Suiza)	Bernard Pottier (Instituto de Francia ~ Francia)
María Luisa Calero Vaquera (Univ. de Córdoba ~ España)	François Rastier (Centre national de la recherche scientifique ~ Francia)
Dolores Corbella (Univ. de La Laguna ~ España)	Emilio Ridruejo Alonso (Univ. de Valladolid ~ España)
Teresa Espar (Univ. de Venezuela ~ Venezuela)	Javier Rodríguez Molina (Univ. Complutense de Madrid ~ España)
Inés Fernández Ordóñez (Univ. Autónoma de Madrid y Real Academia Española ~ España)	M. ^a Nieves Sánchez González de Herrero (Univ. de Salamanca ~ España)
Juan Gutiérrez Cuadrado (Univ. Carlos III ~ España)	Ramón Trujillo Carreño (Univ. de La Laguna ~ España)
Covadonga López Alonso (Univ. Complutense de Madrid ~ España)	Hernán Urrutia Cárdenas (Univ. del País Vasco ~ España)
Ángel López García-Molins (Univ. de Valencia ~ España)	Agustín Vera Luján (Univ. Nacional de Educación a Distancia ~ España)
† Dietter Messner (Univ. de Salzburg ~ Austria)	† Gerd Wotjak (Univ. de Leipzig ~ Alemania)

Asesores y revisores de inglés

Moisés Almela Sánchez (Univ. de Murcia)	Teresa Marqués Aguado (Univ. de Murcia)
---	---

REVISTA DE INVESTIGACIÓN LINGÜÍSTICA (RIL)

La *Revista de Investigación Lingüística* es una publicación periódica científica dedicada al estudio de la lengua española y la lingüística general en todas sus variedades, desde cualquier modelo teórico o planteamiento metodológico oportunos para el enfoque que sugiera el autor. Los trabajos pueden adoptar una perspectiva tanto sincrónica como diacrónica. Además de acoger artículos de corte filológico tradicional, la revista pretende actuar como medio de difusión de los últimos enfoques teóricos y metodológicos desarrollados en los estudios de lengua española y lingüística general.

La línea editorial de *RIL* contempla muy diversos ámbitos de estudio: análisis del discurso, historia de la lengua, historia de la lingüística, lexicografía, lexicología, lingüística general, pragmática, semántica, sintaxis, sociolingüística, terminología, variedades del español. De acuerdo con los informes confidenciales de evaluadores externos, la revista decide sobre la publicación de los artículos recibidos, que deben ser originales inéditos.

Desde el año 2004 la *Revista de Investigación Lingüística* tiene una periodicidad anual. En la actualidad, cada número consta de una sección general, en la que se publican artículos sobre lengua española y lingüística general, y una sección dedicada a la recensión de libros. Además, la revista puede publicar monografías, cuyos artículos abordan un tema específico bajo la coordinación de uno o varios especialistas.

La *Revista de Investigación Lingüística* se publica desde 2020 exclusivamente en edición electrónica mediante el sistema OJS, disponible en la dirección <https://revistas.um.es/ril>, donde se ofrece el contenido de todos sus números en formato digital.

Dirección científica

Revista de Investigación Lingüística

Departamento de Lengua Española y Lingüística General

Facultad de Letras

Universidad de Murcia

C/ Santo Cristo, 1

30001 ~ Murcia

Dirección administrativa

Servicio de Publicaciones

Universidad de Murcia

Edificio Pleiades

Campus de Espinardo

30071 ~ Murcia

Indexación, bases de datos y catálogos

La *Revista de Investigación Lingüística* cuenta con el Sello FECYT desde 2021 (renovado en la convocatoria de 2025 para los próximos dos años), está indexada en ESCI (Clarivate) e incluida en el Catálogo Latindex 2.0 (36 de 38 características cumplidas). En Dialnet Métricas (2024) *RIL* se posiciona en el C1 de Filologías (48 de 312 revistas) y en el C2 de Lingüística (27 de 70 revistas). Según MIAR, en 2021 *RIL* tenía un ICDS de 10; en la actualidad presenta la siguiente difusión: $c1+m6+e3+x6$.

CARHUS Plus+ ~ Sistema de clasificación de revistas científicas de los ámbitos de Ciencias Sociales y Humanidades	Latindex ~ Sistema regional de información en línea para revistas científicas de América Latina, El Caribe, España y Portugal. Universidad Autónoma de México
CIRC ~ Clasificación Integrada de Revistas Científicas (de Ciencias Sociales y Humanas)	LB (Linguistic Bibliography) ~ Catálogo en línea que abarca las diferentes disciplinas lingüísticas
Dialnet ~ Portal de difusión de la producción científica hispana. Universidad de La Rioja	LLBA (Linguistics & Language Behavior Abstracts) ~ Base de datos de revistas lingüísticas
DOAJ ~ Directory of Open Access Journals	MIAR (Matriz de Información para el Análisis de Revistas) ~ Base de datos que reúne información clave para la identificación y el análisis de revistas
Dulcinea ~ Proyecto coordinado por el Ministerio de Educación y Ciencia para identificar y analizar las políticas editoriales de las revistas científicas españolas	OCLC WORLDCAT ~ Catálogo en línea que facilita el acceso a material bibliográfico
ERIH Plus ~ Índice europeo de referencia para las disciplinas humanísticas y sociales	PIO (Periodicals Index Online) ~ Base de datos internacional de revistas de Artes, Humanidades y Ciencias Sociales
ESCI (Emerging Sources Citation Index) ~ Producto de la Web of Science (WoS) editado por Thomson Reuters	REDIB ~ Red Iberoamericana de Innovación y Conocimiento Científico
Fuente Académica Plus ~ Base de datos bibliográfica	REGESTA IMPERII ~ Base de datos bibliográfica
Google Scholar ~ Buscador de Google especializado en documentos académicos con recuento de citas	SUDOC ~ Catálogo colectivo de referencias bibliográficas realizado por las bibliotecas y centros de documentación de educación superior e investigación franceses
IBZ On line ~ Bibliografía internacional de publicaciones periódicas de Humanidades y Ciencias Sociales	ULRICH'S ~ Directorio de publicaciones periódicas
IDR (Índice Dialnet de Revistas) ~ Recurso que informa sobre el impacto científico de una revista, su evolución y su posición respecto del resto de las revistas de la especialidad. Universidad de La Rioja	ZDB/EZB ~ Catálogo colectivo de revistas electrónicas
ÍnDICES-CSIC ~ Recurso bibliográfico multidisciplinar que recopila y difunde principalmente artículos de investigación publicados en revistas científicas españolas	

Derechos de autor

Las obras que se publican en la *Revista de Investigación Lingüística* están sujetas a los siguientes términos:

1. El Servicio de Publicaciones de la Universidad de Murcia (la editorial) conserva los derechos patrimoniales (copyright) de las obras publicadas y favorece y permite la reutilización de las mismas bajo la licencia de uso indicada en el punto 2.
2. Las obras se publican en la edición electrónica de la revista bajo una licencia Creative Commons Atribución/Reconocimiento-NoComercial-SinDerivados 4.0 Internacional. Consulte la versión informativa y el texto legal de la licencia.



eISSN: 1989-4554

ISSN: 1139-1146

Depósito Legal: MU-646-1988

Dirección web *RIL*: <https://doi.org/10.6018/ril>

Archivo: <https://revistas.um.es/ril/issue/archive>

Envíos: <https://revistas.um.es/ril/about/submissions>



ÍNDICE

Artículos

ALFANO, IOLANDA — Lenguaje, género y visibilidad profesional: <i>nomina agentis</i> en el ámbito médico en España e Italia	15
ÁLVAREZ MENÉNDEZ, ALFREDO IGNACIO — En la frontera de las interjecciones <i>impropias</i> : frases interjectivas, sintagmas intensivos e interjecciones formularias	39
BARGALLÓ ESCRIVÁ, MARÍA — El debate sobre la enseñanza de la gramática castellana en Chile en la segunda mitad del siglo XIX y principios del XX	67
BASTARDÍN CANDÓN, TERESA Y MARGARITA FERNÁNDEZ GONZÁLEZ — Avances en el estudio histórico del léxico andaluz: registros notariales y geografía lingüística	83
BEJARANO BEJARANO, DANIEL EDUARDO — Panorama de las fórmulas de tratamiento pronominales en el español de la Orinoquía colombiana	103
CABANES PÉREZ, SANDRA — La posición del gesto en una estructura jerárquica de la conversación: una propuesta multimodal desde el modelo Val.Es.Co.	129
CARRIAZO RUIZ, JOSÉ RAMÓN — Enigmas etimológicos en la nomenclatura de amarres y nudos marinos: <i>amante</i> , <i>ballestrinque</i> , <i>cote</i> y <i>cornamusa</i>	153
DOMÍNGUEZ VÁZQUEZ, MARÍA JOSÉ — Diseño y metodología de un etiquetador semántico-ontológico multilingüe: ESMAS-ES ⁺	175
GÓMEZ DÍAZ, SARA — La terminología como interdisciplina y transdisciplina: conexiones y aplicaciones	193
KORNFELD, LAURA MALENA — Seudocoordinación repetitiva: el caso de <i>caminan</i> y <i>caminan</i>	213
MERINO GONZÁLEZ, ALICIA — La narración escrita en menores con trastorno del espectro autista: una aproximación a los personajes principales y secundarios desde los principios explicativos	233

MONTORO DEL ARCO, ESTEBAN T. — <i>La vía compilatoria de la reflexión gramatical en los siglos XIX y XX: el género hispánico de los entretenimientos</i>	267
QUEROL-BATALLER, MARÍA — De la expresión del desacuerdo a la generación del conflicto: análisis de conversaciones conflictivas entre parejas	297
RODRÍGUEZ-GASCÓN, SARA Y DIEGO RODRÍGUEZ GASCÓN — El bilingüismo y el núcleo caudado: el control en el uso de la lengua	319
TELLEZ-PEREZ, CLARA — ¿Tiempo o evidencia? Valores del pretérito perfecto compuesto en el español de España y Ecuador	341

Reseñas

ALBALADEJO GUARINO, MANUEL — Herminia Provencio Garrigós (2024): <i>Ruta lingüística por la ciudad de Murcia y mucho más...</i> , con la colaboración de Miguel Ángel Puche Lorenzo, Mercedes Abad Merino y Esther Vivancos Mulero, Murcia, Diego Marín Librero-Editor, 227 pp.	365
JACINTO GARCÍA, EDUARDO JOSÉ — Sergio Rodríguez Tapia (2024): <i>Gestión terminológica, corpus especializados y extracción automática de terminología en español</i> , Editorial Comares, Granada, 184 pp.	371
SANFILIPPO, VINCENZO — Javier de Santiago-Guervós (2024): <i>Discurso y persuasión</i> , Madrid, Arco/Libros, 296 pp.	377
TEJERO GARCÍA, ELENA MARÍA — Antoni Nomdedeu-Rull y Sven Tarp (2024): <i>Introducción a la lexicografía en español. Funciones y aplicaciones</i> , Londres/Nueva York, Routledge, 256 pp.	383
Normas para autores	387

Sergio Rodríguez Tapia (2024): *Gestión terminológica, corpus especializados y extracción automática de terminología en español*, Granada, Editorial Comares, 184 pp.



Gestión terminológica,
corpus especializados
y extracción automática
de terminología
en español



Sergio Rodríguez-Tapia

EDITORIAL COMARES



Interlingua

El libro que aquí reseñamos, *Gestión terminológica, corpus especializados y extracción automática de terminología en español* (2024), es la primera monografía publicada por el profesor Sergio Rodríguez Tapia, de la Universidad de Córdoba. Nos encontramos ante una obra ya madura, sistemática y con un abundante aparato bibliográfico. Destaca, además, por su triple vertiente, didáctica, académica y profesional, pues va dirigida a estudiantes e investigadores del campo de la lingüística y de la computación interesados en el ámbito de las llamadas *industrias del lenguaje*. El propósito de esta obra consiste, precisamente, en plantear una herramienta de consulta que permita salvar la brecha comunicativa entre el mundo universitario y el laboral, es decir, una herramienta que ayude a adquirir no solo conocimientos teóricos, sino también competencias profesionales aplicables fuera del aula.

De los siete capítulos en los que está dividido el libro, el primero (*La tecnología en la lingüística y la terminología en la actualidad*) trata, entre otras cuestiones, sobre los objetivos de la obra y sobre la metodología aplicada en la elaboración de un catálogo de herramientas terminológicas disponibles en español, y que los lectores podrán consultar en las últimas páginas. El autor declara que el objetivo principal del libro es crear un recurso que compile los principales avances teóricos y metodológicos relativos al empleo de aquellas tecnologías orientadas a elaborar recursos terminográficos (pp. 5-6). Como ya hemos señalado, con ello quiere no solo facilitar el acceso de los lectores a un tipo de formación especializada, sino también proporcionarles los medios necesarios para el desarrollo de las competencias técnicas y profesionales que faciliten su inserción en el mundo laboral.

Los capítulos II, III y IV están dedicados a aquellos recursos tecnológicos relacionados con la gestión terminológica. Las dos primeras herramientas que se estudian son las bases de datos y los sistemas de gestión terminológicos. A ambos recursos está dedicado el capítulo II (*La gestión terminológica y las herramientas informáticas*). En el capítulo III (*El diseño y uso de los corpus especializados*) se lleva a cabo una aproximación a la lingüística de corpus y se describen las características de los corpus textuales especializados. En el IV (*La extracción terminológica*) se tratan los extractores automáticos de terminología. En el capítulo V (*Recapitulación y reflexiones*

fnales) se lleva a cabo una síntesis de los contenidos expuestos en el libro, así como una serie de consideraciones finales del autor, mientras que en el VI (*Referencias bibliográficas*) se incluye una amplia bibliografía. Finalmente, en el capítulo VII (*Catálogo de recursos tecnológicos sobre gestión terminológica en español*) se recoge la lista de herramientas digitales *sobre* terminología disponibles en español, una sección de gran utilidad para quienes se dedican al estudio del discurso especializado y que el autor ha decidido incluir acertadamente al final de la obra para su más fácil consulta.

Como señala Rodríguez Tapia (p. 11), en esta obra se recoge un total de 64 bases de datos terminológicas, 9 corpus especializados, 19 extractores de terminología y 11 sistemas de gestión terminológica (103 recursos en total), lo que nos da una idea de la labor realizada para recopilar un número tan elevado de herramientas terminológicas. Lógicamente, no todas ellas se describen de una manera pormenorizada. Para realizar una selección más restrictiva de los recursos que se van a describir, Rodríguez Tapia tiene en cuenta una serie de criterios como, por ejemplo, que los recursos se hayan actualizado en los últimos cinco años o que hayan sido objeto de trabajos científicos.

El capítulo II del libro (pp. 13-75) está dedicado a las grandes cuestiones relacionadas con la gestión terminológica y su relación con las herramientas informáticas. Una base de datos terminológica es una de las herramientas más usuales en el campo de la gestión terminológica y su uso ha estado ligado desde siempre a la labor de la traducción especializada. Tal como señala Rodríguez Tapia (p. 15), las bases de datos terminológicas comparten muchas de sus características con los diccionarios electrónicos, como son la ausencia de limitaciones de espacio, el almacenamiento estructurado de información, su fácil actualización o los sistemas de búsqueda a partir de filtros. Incluyen información sobre los conceptos de un determinado dominio y los términos empleados para designarlos. Pueden ser monolingües o plurilingües, onomasiológicos o semasiológicos. Como ocurre con otras herramientas lexicográficas, constan de una entrada y de una ficha en la que aparecen diversos campos o categorías de datos. El acceso a las bases de datos puede llevarse a cabo a través de una interfaz gráfica de usuario o bien integrándose a través de otras aplicaciones, generalmente orientadas a la traducción. La recuperación rápida de información resulta fundamental y, por ese motivo, los filtros de búsqueda o la integración de las bases de datos terminológicas en otras aplicaciones cobran una gran importancia.

De las páginas 22 a la 38 del libro el autor hace un recorrido por ocho bases de datos terminológicas. Algunas solo tienen en cuenta un dominio concreto, como DeCS/MeSH o Dicciomed, centradas en el campo de la medicina, o DiccoAdventure, que se focaliza en el ámbito del turismo de aventura, mientras que otras abarcan varios dominios, como CercaTerm, EuroVoc, IATE, TERMDAT o TERMIUM Plus. Cada base de datos es descrita siguiendo unos criterios preestablecidos, lo que permite detectar de una manera sistemática los puntos fuertes y débiles de cada recurso. Así, Rodríguez Tapia tiene en cuenta cuestiones como la variedad de campos o de categorías que se han considerado, los tipos de búsqueda por medio de filtros, la presencia de imágenes y de referencias cruzadas, la validación de los términos, la accesibilidad a las fuentes o la existencia de contextos o ejemplos de uso. Cada una de las descripciones viene

acompañada, además, de una o dos ilustraciones en las que puede verse la interfaz del recurso analizado, lo que resulta de una gran utilidad para el lector.

Tras el análisis de estas bases de datos, Rodríguez Tapia dedica las siguientes páginas a explicar cómo se lleva a cabo el proceso de gestión terminológica y, para ello, toma como punto de partida la propuesta de fases y criterios establecidos para la elaboración de diccionarios terminológicos en TERMCAT y, más concretamente, en *El diccionari terminològic* (p. 39). El proceso comienza por la delimitación de los objetivos y de los destinatarios de la base de datos que se pretende crear, lo que pasa por definir la actividad profesional en la que se utilizará generalmente el recurso y por analizar las necesidades prototípicas de los usuarios a los que va dirigido, una idea sobre la que cada vez se pone un mayor énfasis, como demuestra la aparición de la teoría de las funciones lexicográficas surgida en el seno del Centro de Lexicografía de la Universidad de Aarhus (Dinamarca). Una vez fijados los objetivos y el tipo de destinatario, se deben establecer los distintos tipos de información que se van a incluir en las fichas terminológicas, información que abarca la propia selección de las unidades terminológicas, su definición, los equivalentes, las variantes denominativas del concepto, las restricciones sintácticas, los ejemplos de uso o contextos, el dominio en el que se inserta, las ilustraciones, las restricciones de uso por normalización o estandarización, las fuentes de la información, las indicaciones de la fiabilidad de esas fuentes, etc. El autor (pp. 42-56) dedica una serie de apartados a cada una de estas informaciones, aportando, además, bibliografía que puede ser de interés al lector que desee profundizar en ellas.

El siguiente bloque incluido en el capítulo II se ocupa de los sistemas de gestión terminológica y de las ontologías. Mientras que las bases de datos terminológicas son productos ya concluidos y preparados para su consulta, los sistemas de gestión terminológica son herramientas pensadas para aquellos que tienen que elaborar su propia base de datos terminológica. Esos sistemas de gestión incluyen programas como gestores de corpus, extractores automáticos de términos, software para el diseño de entradas y de fichas terminológicas, etc. (p. 59). En esta obra se analizan concretamente tres sistemas de gestión de corpus: MultiTerm, TermStar NXT y WebTerm. Cada uno de ellos es descrito teniendo en cuenta sus características técnicas, así como las posibilidades de búsqueda y de importación y exportación de archivos.

Por último, pasan a estudiarse las ontologías o recursos de representación de conocimiento (KRR o *knowledge representation resources*) (pp. 65-73). Las ontologías son programas o sistemas de almacenamiento y recuperación de información que pueden incluir, a su vez, bases de datos terminológicas, sistemas conceptuales o modelos de redes semánticas como *WordNet*. Para su diseño juegan un papel muy importante la ingeniería informática, la inteligencia artificial y, especialmente, la ingeniería de conocimiento. Su objetivo es, como explica Rodríguez Tapia (p. 67), la representación explícita y formal de una conceptualización compartida, es decir, de las relaciones que se establecen entre conceptos, relaciones que reflejan no el conocimiento particular de un individuo, sino el compartido y consensuado por un grupo. Las ontologías se pueden dividir según su amplitud de campo (ontologías de nivel superior, como *WordNet* o *EuroWordNet*, y ontologías de nivel intermedio, como *BabelNet* o *EcoLexicon*, entre otras). También se pueden clasificar en cuanto al grado de formalización. Así, podemos encontrar nomenclaturas, glosarios, tesauros y taxonomías. Rodríguez Tapia (pp. 72-75) concluye esta

sección enumerando también algunas herramientas para el diseño de ontologías, como NeOn o Protégé, así como los componentes que participan en ellas y sus posibles aplicaciones, como aquellas que están relacionadas con la web semántica, el procesamiento del lenguaje natural, la recuperación de información y, cómo no, la gestión terminológica (p. 74).

El capítulo III de la obra (pp. 77-126) está dedicado al diseño y uso de corpus especializados, una fase que resulta fundamental en el proceso de gestión terminológica. Aunque la lingüística de corpus ha sido tratada ya en un número significativo de estudios, Rodríguez Tapia logra hacer una síntesis de los aspectos más relevantes de esta metodología, sin dejar fuera ninguna cuestión que sea de interés para el terminólogo. En primer lugar, se ocupa de los objetivos de la lingüística de corpus, su naturaleza y las ventajas instrumentales que posee (pp. 78).

En cuanto a las ventajas que trajo la lingüística de corpus al campo de la terminología, aquí se destaca el hecho de que se trata de una metodología que permite trabajar con los términos *in vivo*, aporta una perspectiva empírica y descriptiva, y permite extraer datos de conocimiento sin necesidad de consultar a expertos, así como también detectar neologismos. El autor (pp. 84-89) lleva a cabo una extensa descripción de los diferentes tipos de corpus existentes, en la que parte de clasificaciones ya realizadas por diversos autores. Para ello, realiza una clasificación multidimensional de los corpus. Aporta, además, información práctica sobre los criterios que se han de seguir a la hora de compilar un corpus especializado, tratando cuestiones como el análisis de las necesidades que implica la investigación, la fiabilidad de las fuentes o el criterio de actualización. Dentro del apartado dedicado a los criterios cualitativos de representatividad, Rodríguez Tapia (pp. 100-103) se centra en los criterios de representatividad de un corpus especializado, y lo hace teniendo en consideración, sobre todo, los grados de especialización de los textos. Las necesidades de investigación y de los objetivos que estos conllevan, determinan qué tipo de textos se van a escoger durante el proceso de compilación de un corpus: existe una variación horizontal, que marca la temática de los textos, y una variación vertical, que es la que obedece propiamente al enfoque más o menos especializado de esos textos, determinado por los participantes que actúan en el discurso y por otras variables de carácter contextual. Rodríguez Tapia (pp. 103-113) se ocupa también de otras cuestiones relativas a las dimensiones del corpus, su codificación y etiquetado, mencionando los diferentes sistemas de anotación, lo que resulta de gran utilidad para quienes se acercan a estas cuestiones por primera vez.

Concluye este importante capítulo abordando la recuperación y el análisis de la información, así como de las funciones de los gestores de corpus. Existen corpus especializados que cuentan con su propio sistema de consulta a través de una interfaz que facilita al usuario la extracción de datos y otros tipos de corpus que utilizan gestores de corpus externos e independientes como Sketch Engine o AntConc. Antes de evaluar los corpus especializados del español, Rodríguez Tapia señala la importancia de que el corpus cuente con una buena arquitectura, esto es, con una organización y división internas atendiendo a sus distintas categorías y subcategorías o ejes (p. 115). Además, señala las funcionalidades más importantes que pueden ofrecer los gestores de corpus, como la búsqueda de concordancias, el listado de frecuencias, la búsqueda de agrupaciones o coocurrencias, la búsqueda de enigramas o expresiones pluriverbales, la búsqueda de colocaciones y de palabras clave, así como la posibilidad de filtrar los textos por categorías

y de extraer tablas estadísticas. Los corpus de especialidad analizados son el *Corpus de las Sexualidades de México* (CSMX), el *Corpus Tècnic*, y el *DIACOM-es*. Este último, dirigido por De Beni, Hourani Martín y Sartor, es un corpus especializado diacrónico del ámbito del comercio que integra textos desde 1850 y, teniendo en cuenta las funcionalidades antes mencionadas, es el corpus mejor valorado por Rodríguez Tapia.

El capítulo IV está dedicado a la extracción de términos (pp. 127-153), una de las facetas más interesantes de la gestión terminológica, ya que se trata, en última instancia, del procedimiento que permite configurar listas de candidatos a términos a partir de corpus informatizados. El autor recalca la expresión *candidato*, ya que no hay extractor automático que logre una eficacia del 100 % a la hora de detectar términos. Mientras que las ventajas de los extractores son indiscutibles, en términos de sistematicidad y eficacia, Rodríguez Tapia también advierte de sus desventajas, por ejemplo, en lo que se refiere al silencio (fracaso en el reconocimiento de términos) y al ruido (detección de falsos términos). El autor distingue tres métodos de extracción: los sistemas basados en conocimiento lingüístico (se analizan patrones sintácticos, morfológicos o información semántica para determinar los posibles términos); los sistemas basados en métodos estadísticos (fundamentalmente la frecuencia de aparición en textos especializados en relación con textos de la lengua común), y sistemas híbridos, que suelen ser los que mejores resultados ofrecen (p. 138). Las siguientes páginas el autor las dedica a definir los criterios que determinan la extracción automática de términos. En primer lugar, nos encontramos con el criterio de unicidad, que consiste en establecer los límites de un término cuando este se corresponde con una unidad pluriverbal; en segundo lugar, está el criterio de la terminologización, que constituye un verdadero reto para los sistemas de extracción, ya que no solo existen las unidades crípticas, es decir, las que solo se hallan en discursos especializados, sino también las unidades terminologizadas o délficas, que son aquellas palabras usadas en el discurso común que se vuelven términos en el contexto de un discurso especializado. Como ya se ha tratado, existen métodos de extracción basados en criterios lingüísticos y otros basados en criterios estadísticos. Rodríguez Tapia (pp. 142-144) describe ambos criterios con bastante profundidad, de manera que incluso los lectores que no estén acostumbrados a tratar con algoritmos pueden llegar a entender su funcionamiento. Concluye el capítulo IV mostrando dos ejemplos de sistemas de extracción terminológica: «OneClick Terms» de Sketch Engine y TermoStat Web 3.0.

Como ya adelantamos al principio de esta reseña, el capítulo V está dedicado a las conclusiones. En ellas, Rodríguez Tapia (pp. 157-158) lamenta la escasa proliferación de corpus especializados en español, en contraste con los gestores de corpus, que no especifican muchas veces el origen de sus fuentes. El autor destaca también los avances logrados en los sistemas de extracción terminológica, aunque estos sigan sin ofrecer resultados completamente «limpios».

El capítulo VI (pp. 161-174) consiste en una exhaustiva lista de referencias bibliográficas, que se prolonga a lo largo de catorce páginas a doble columna, en la que se recoge la bibliografía más actual sobre la temática tratada en la obra y en la que hay cabida para los estudios provenientes de todas las escuelas y corrientes vinculadas a la terminología desde finales del siglo XX.

El capítulo VII, que es el último de la obra (pp. 175-184), lo forma el catálogo de todos los recursos tecnológicos sobre gestión terminológica en español recogidos en este libro, diez pági-

nas en las que estos recursos se organizan por tipologías (bases de datos terminológicas, sistemas de gestión terminológica, corpus especializados y extractores terminológicos) y por orden alfabético, siendo el nombre del recurso el que funciona como lema. Cada ficha incluye información sobre el autor, el año de la última actualización y la dirección web en la que se localiza.

Concluimos esta reseña destacando la calidad de esta obra. Rodríguez Tapia no evita la complejidad de ciertos temas, pero los afronta explicándolos con claridad; se preocupa de señalar los puntos fuertes y débiles de cada recurso, advirtiendo al lector de cuáles son las herramientas más apropiadas para alcanzar determinados objetivos y cuáles no. También se aprecia un cuidado especial por cuestiones formales como, por ejemplo, las imágenes que acompañan a cada capítulo. Por último, queremos señalar la abundante bibliografía, completamente actualizada, recogida sin ningún tipo de sesgo, y que se refleja, además, en multitud de referencias dentro de cada capítulo, así como la estructura informativa del libro, perfectamente organizada, de manera que al lector le resulta sumamente fácil encontrar la información que requiere en cada momento.

EDUARDO JOSÉ JACINTO GARCÍA
Universidad de Córdoba

RECIBIDA: 28 de marzo de 2025
ACEPTADA: 2 de mayo de 2025