

# ELUA



## Los géneros digitales en la interfaz del trabajo de corpus y el análisis del discurso especializado

Coordinado por  
Chelo Vargas-Sierra y  
José A. Sánchez Fajardo

E  
38 L

ELUA Estudios de Lingüística

A

Universidad de Alicante

2022



Universitat d'Alacant  
Universidad de Alicante

**ESTUDIOS DE LINGÜÍSTICA**  
UNIVERSIDAD DE ALICANTE  
(ELUA)

2022

NÚMERO 38

MONOGRAFÍA

**LOS GÉNEROS DIGITALES EN LA INTERFAZ  
DEL TRABAJO DE CORPUS Y EL ANÁLISIS DEL  
DISCURSO ESPECIALIZADO**

COORDINADO POR CHELO VARGAS-SIERRA Y JOSÉ A. SÁNCHEZ FAJARDO

REVISTA DE LENGUA ESPAÑOLA  
Y LINGÜÍSTICA GENERAL



Universitat d'Alacant  
Universidad de Alicante

<https://revistaelua.ua.es>

#### DIRECTORA

María Belén Alvarado Ortega

#### SECRETARIAS

Ruth María Lavale Ortiz

María del Carmen Méndez Santos

#### CONSEJO DE REDACCIÓN

Elisa Barrajón López (U. Alicante)  
Cristina Cacciari (U. Modena and Reggio Emilia)  
Carmen Cazorla Vivas (U. Complutense de Madrid)  
Dmitrij O. Dobrovol'skij (U. Stockholm)  
Jorge Fernández Jaén (U. Alicante)  
Catalina Fuentes Rodríguez (U. Sevilla)  
María del Mar Galindo Merino (U. Alicante)  
M.<sup>a</sup> Ángeles García Aranda (U. Complutense de Madrid)  
Elena Hoicka (U. Bristol)  
Juan Luis Jiménez Ruiz (U. Alicante)  
Johannes Kabatek (U. Zürich)  
Esther Linares Bernabéu (U. Valencia)  
Óscar Loureda Lamas (U. Heidelberg)  
Carmen Marimón Llorca (U. Alicante)  
José Joaquín Martínez Egido (U. Alicante)  
María Antonia Martínez Linares (U. Alicante)

Francisco Matte Bon (U. Studi Internazionali di Roma)  
María Águeda Moreno Moreno (U. Jaén)  
Elena de Miguel Aparicio (U. Autónoma Madrid)  
Inés Olza Moreno (U. Navarra)  
Manuel Padilla Cruz (U. Sevilla)  
Xose A. Padilla García (U. Alicante)  
Susana Pastor Cesteros (U. Alicante)  
Inmaculada Penadés Martínez (U. Alcalá)  
Herminia Provencio Garrigós (U. Murcia)  
Alberto Rodríguez Lifante (U. Alicante)  
Susana Rodríguez Rosique (U. Alicante)  
Leonor Ruiz Gurillo (U. Alicante)  
Ventura Salazar García (U. Jaén)  
Isabel Santamaría Pérez (U. Alicante)  
Augusto Soares da Silva (U. Católica Portuguesa)  
Larissa Timofeeva Timofeev (U. Alicante)  
Alexandre Veiga Rodríguez (U. Santiago de Compostela)

#### CONSEJO ASESOR

Dolores Azorín Fernández (U. Alicante)  
Ignacio Bosque Muñoz (U. Complutense Madrid y RAE)  
Antonio Briz Gómez (U. Valencia)  
José Luis Cifuentes Honrubia (U. Alicante)  
Delia Chiaro (U. Bolonia)  
Javier de Santiago Guervós (U. Salamanca)  
Nicole Delbecque (U. Católica Lovaina)  
Milagros Fernández Pérez (U. Santiago de Compostela)  
Francisco Gimeno Menéndez (U. Alicante)  
Salvador Gutiérrez Ordóñez (U. León y RAE)  
Ángel López García (U. Valencia)  
Humberto López Morales (RAE)

María Antonia Martín Zorraquino (U. Zaragoza)  
Elena Méndez García de Paredes (U. Sevilla)  
Ana Isabel Navarro Carrasco (U. Alicante)  
Neal R. Norrick (U. Saarland)  
Lourdes Ortega (U. Georgetown)  
María Elena Placencia (U. Londres)  
Emilio Ridruejo Alonso (U. Valladolid)  
Agustín Vera Luján (UNED)  
Juan Andrés Villena Ponsoda (U. Málaga)  
Gerd Wotjak (U. Leipzig)  
Jef Verschueren (U. Amberes)

ISSN: 2171-6692

Diseño de la cubierta: Xose. A. Padilla García

Publicada gracias a las ayudas para la publicación de revistas científicas de la Facultad de Filosofía y Letras y del Vicerrectorado de Investigación de la Universidad de Alicante.

Composición: Compobell, S.L.

Normas para envío, evaluación y edición de contribuciones:

<https://revistaelua.ua.es/>

## ÍNDICE

Introducción.....	7
<i>Chelo Vargas-Sierra y José A. Sánchez Fajardo</i>	
<b>ARTÍCULOS</b>	
BLOQUE I. <i>ANÁLISIS DEL DISCURSO ESPECIALIZADO</i>	
La resolución administrativa publicada en portales web institucionales: Un análisis basado en corpus.....	17
<i>Sara Pistola y Iria da Cunha</i>	
Usos y funciones de la metáfora conceptual en discurso pedagógico sobre meditación	47
<i>Antonio-José Silvestre-López</i>	
¿De qué hablamos cuando divulgamos sobre lingüística? Análisis de un corpus de textos divulgativos y aplicaciones al estudio terminológico de la semántica léxica ...	73
<i>Marta Sánchez-Saus Laserna y Vanesa Álvarez Torres</i>	
Las emociones en los géneros textuales especializados: Poder, manipulación y persuasión desde la AST, o <i>Teoría del Espectro Afectivo</i> .....	99
<i>María Ángeles Orts Llopis</i>	
Adaptación de las escalas de calidad de la traducción automática a los nuevos géneros digitales: el caso del contenido generado por el usuario .....	117
<i>Miguel A. Candel-Mora</i>	
Análisis de la comunicación experto-lego por medio del consentimiento informado: legibilidad y relación médico-paciente.....	137
<i>Sofía Zakhir Puig</i>	
¿Quién es quién? Análisis de corpus de los procesos de desfocalización del agente en español y en francés en dos géneros digitales .....	165
<i>Émeline Pierre</i>	
Los marcadores del discurso en la narrativa financiera: análisis de las cartas a los accionistas.....	187
<i>Ana García Toro</i>	

## BLOQUE II. CORPUS Y SUS APLICACIONES

El corpus Iberia como fuente para el estudio del léxico especializado.....	217
<i>Luis Pablo Nuñez</i>	
Estudio basado en corpus de 4-gramas en el artículo científico.....	241
<i>Eva Lucía Jiménez-Navarro</i>	
Verbos de corte en la construcción media: examinando la co-especificación composicional de las medias pacientes y agentivo-instrumentales .....	263
<i>Macarena Palma Gutiérrez</i>	
Extracción automática de unidades terminológicas guiada por Framenet: una aplica- ción al corpus electrónico CORD-19 .....	281
<i>Mario Crespo Miguel</i>	
La empatía en la experiencia turística online. Adjetivos y patrones de uso más fre- cuentes del lenguaje promocional en inglés y en español .....	301
<i>Eva María Llorens Simón</i>	

## RESEÑAS

Hernández, E. y P. Martín Butragueño (eds.) (2020). <i>Las palabras como unidades lingüísticas</i> . Madrid/ Ciudad de México: csic. 467 Páginas. ....	327
<i>Raquel Rodríguez Fernández</i>	
Albadejo García, M. D. (2020). <i>Aprovechamiento didáctico de textos narrativos en el aula de español como lengua extranjera</i> . Editorial Universidad de Jaén. ....	333
<i>Neus Gimeno Gimeno</i>	
Rojo, Guillermo (2021): <i>Introducción a la lingüística de corpus</i> . Londres: Routledge. 380 páginas. DOI: <a href="https://doi.org/10.4324/9781003119760">https://doi.org/10.4324/9781003119760</a> .....	339
<i>María Eugenia Conde Noguerol</i>	
Jorge Fernández Jaén (2019). <i>El abecé de la lingüística cognitiva</i> . Arco Libros. 96 páginas. ISBN: 978-84-7635-996-9. ....	343
<i>Irene Arance Rusillo</i>	
La investigación en marcha. Reseña a Santos Gargallo, Isabel y Pastor Cesteros, Susana (dirs.) (2022). <i>Metodología de la investigación en la enseñanza-aprendizaje del español como segunda lengua (L2)/ lengua extranjera (LE)</i> . Arco Libros, 391 pp., ISBN 978-84-71353-864-8.....	347
<i>Miguel Ángel Mora Sánchez</i>	
Galindo Merino, M. M. y M. C. Méndez (eds.) (2022). <i>La lingüística del amor: de la pasión a la palabra</i> . Pie de página editorial, 322 páginas.....	353
<i>Anastasia Khaylina</i>	

## INDEX

Introduction.....	7
<i>Chelo Vargas-Sierra &amp; José A. Sánchez Fajardo</i>	
<b>PAPERS</b>	
BLOQUE I. <i>ANALYSIS OF SPECIALISED DISCOURSE</i>	
The administrative resolution published on institutional web portals: a corpus-based analysis.....	17
<i>Sara Pistola &amp; Iria da Cunha</i>	
Uses and functions of conceptual metaphor in pedagogical meditation discourse.....	47
<i>Antonio-José Silvestre-López</i>	
What do we talk about when we communicate about linguistics as a science? Analysis of a corpus of science communication texts and applications to the terminological study of lexical semantics .....	73
<i>Marta Sánchez-Saus Laserna &amp; Vanesa Álvarez Torres</i>	
Emotions in specialised genres: Power, manipulation and persuasion from the <i>Affect Spectrum Theory</i> .....	99
<i>María Ángeles Orts Llopis</i>	
Fine-tuning machine translation quality-rating scales for new digital genres: The case of user-generated content.....	117
<i>Miguel A. Candel-Mora</i>	
Analysis of expert-patient communication through informed consent: legibility and the doctor-patient relationship.....	137
<i>Sofía Zakhir Puig</i>	
Who is who? A corpus-based study of agent-defocussing mechanisms in Spanish and French in two digital genres.....	165
<i>Émeline Pierre</i>	
Discourse markers in the financial narrative: analysis of letters to shareholders .....	187
<i>Ana García Toro</i>	

BLOQUE II. CORPUS AND ITS APPLICATIONS

Iberia corpus as a source for the study of the spanish technical vocabulary.....	217
<i>Luis Pablo Núñez</i>	
A corpus-based study of 4-grams in the research article genre.....	241
<i>Eva Lucía Jiménez-Navarro</i>	
Verbs of cutting in the middle construction: Examining the compositional cospecification of patientive vs. Agent-instrument middles .....	263
<i>Macarena Palma Gutiérrez</i>	
Automatic framenet-guided extraction of terminology: an application to the cord-19 electronic corpus.....	281
<i>Mario Crespo Miguel</i>	
On empathy in online tourism experience. High-frequency adjective patterns in english and spanish promotional language .....	301
<i>Eva María Llorens Simón</i>	
<b>REVIEWS</b>	
Hernández, E. y P. Martín Butragueño (eds.) (2020). <i>Las palabras como unidades lingüísticas</i> . Madrid/ Ciudad de México: csic. 467 Páginas. ....	327
<i>Raquel Rodríguez Fernández</i>	
Albadejo García, M. D. (2020). <i>Aprovechamiento didáctico de textos narrativos en el aula de español como lengua extranjera</i> . Editorial Universidad de Jaén. ....	333
<i>Neus Gimeno Gimeno</i>	
Rojo, Guillermo (2021): <i>Introducción a la lingüística de corpus</i> . Londres: Routledge. 380 páginas. DOI: <a href="https://doi.org/10.4324/9781003119760">https://doi.org/10.4324/9781003119760</a> .....	339
<i>María Eugenia Conde Noguerol</i>	
Jorge Fernández Jaén (2019). <i>El abecé de la lingüística cognitiva</i> . Arco Libros. 96 páginas. ISBN: 978-84-7635-996-9. ....	343
<i>Irene Arance Rusillo</i>	
La investigación en marcha. Reseña a Santos Gargallo, Isabel y Pastor Cesteros, Susana (dirs.) (2022). <i>Metodología de la investigación en la enseñanza-aprendi- zaje del español como segunda lengua (L2)/ lengua extranjera (LE)</i> . Arco Libros, 391 pp., ISBN 978-84-71353-864-8 .....	347
<i>Miguel Ángel Mora Sánchez</i>	
Galindo Merino, M. M. y M. C. Méndez (eds.) (2022). <i>La lingüística del amor: de la pasión a la palabra</i> . Pie de página editorial, 322 páginas.....	353
<i>Anastasia Khaylina</i>	

## Introducción

CHELO VARGAS-SIERRA  
Universidad de Alicante, España  
[chelo.vargas@ua.es](mailto:chelo.vargas@ua.es)  
<https://orcid.org/0000-0002-4026-4372>

JOSÉ A. SÁNCHEZ FAJARDO  
Universidad de Alicante, España  
[jasanchez@ua.es](mailto:jasanchez@ua.es)  
<https://orcid.org/0000-0002-6136-0582>

La comunicación digital ha experimentado un crecimiento exponencial en los últimos años gracias, entre otras cosas, al acceso a Internet, al uso de dispositivos electrónicos y, por supuesto, a plataformas digitales que se han convertido en herramientas de mediación entre los usuarios y los profesionales. Los medios tecnológicos actuales se han adaptado a un nuevo marco de análisis en el que la comunicación y, por consiguiente, el discurso, están mediados por ordenador. En este contexto de comunicación digitalizada —en particular en los ámbitos académicos, científicos y técnicos— se abren paso nuevos productos discursivos de naturaleza digital, dando lugar a nuevos géneros digitales (Belda-Medina y Casañ-Pitarch 2020), géneros que constituyen una nueva línea de investigación lingüística (Luzón y Pérez-Llantada 2019). Asimismo, cabe destacar que la investigación sobre el lenguaje o el discurso especializado ha puesto de relieve la funcionalidad y relevancia de la lingüística del corpus, puesto que el trabajo con corpus aporta una metodología sólida que se nutre de la observación y la experimentación sistemática, así como de la verificación de los resultados. Afortunadamente, algunas barreras tecnológicas, como fueron en su momento la velocidad y la capacidad de los ordenadores, están más que superadas en la actualidad y resulta bastante sencillo obtener, compilar y procesar textos digitales, aspecto que ha sido determinante para popularizar el trabajo computacional en el ámbito, por ejemplo, del análisis del discurso. El propio Sinclair se refiere a la lingüística del corpus y al análisis del discurso como “twin

---

**Para citar este artículo:** Vargas-Sierra, C. y Sánchez Fajardo, J. A. (2022). Introducción. *ELUA*, (38), págs. 7-12. <https://revistaelua.es/article/view/23109>

© 2022 Chelo Vargas-Sierra, José A. Sánchez Fajardo



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



pillars of language research” (Sinclair 2004: 11) y es por ello que un gran número de publicaciones y eventos académicos se han dedicado al estrecho vínculo existente entre estas dos disciplinas en el contexto actual (Taylor y Marchi 2018).

El análisis del discurso y el trabajo con corpus, por otra parte, han sufrido cambios metodológicos en lo que se refiere a la investigación lingüística. Esto se ha debido, más concretamente, al uso de nuevos géneros textuales y discursivos en la comunicación interpersonal (Paltridge 2006; Flowerdew 2012; Friginal y Hardy 2020). La digitalización de estrategias discursivas repercute en el modo en que adquieren nuevas formas o funciones los paradigmas léxicos, pragmáticos, semánticos y sintácticos. Así, es importante que los corpora, como herramienta metodológica de naturaleza empírica y, como ya hemos apuntado, de gran importancia en los estudios lingüísticos, se adapten a los nuevos géneros digitales.

Este volumen tiene como objetivo principal la recopilación de trabajos innovadores, tanto teóricos como aplicados, relacionados con el discurso especializado, los géneros textuales o el trabajo con corpus en la era digital. El análisis del discurso especializado o de géneros textuales que presentamos en este volumen no se limita a un aspecto lingüístico o a la micro o macroestructura de un género concreto, sino que recoge diferentes niveles del lenguaje especializado, poniendo especial énfasis en cómo estos niveles reflejan los contextos situacionales, las funciones lingüísticas, o las variaciones específicas que se derivan del uso en diferentes ámbitos o situaciones comunicativas. Para facilitar la comprensión de los contenidos y los resultados que aquí se muestran, los artículos, un total de trece, se encuentran agrupados en dos grandes bloques temáticos de acuerdo con su grado de especialización. El bloque I recoge artículos que incluimos bajo la etiqueta de “Análisis del discurso especializado” y el bloque II está dedicado a “Corpus y sus aplicaciones”. Cabe señalar que esta distribución no limita los contenidos de los artículos recopilados en cada uno de los bloques, ya que todos trabajan la lingüística de corpus, pero desde diferentes perspectivas.

El bloque I es el más numeroso, con un total de ocho trabajos. Siguiendo la línea temática del bloque sobre el discurso especializado o digitalizado, “La resolución administrativa publicada en portales web institucionales: Un análisis basado en corpus”, de Sara Pistola Grille e Iria da Cunha Fanego, realiza un estudio pormenorizado del lenguaje de la Administración electrónica, sobre todo las resoluciones publicadas en portales institucionales. Con un corpus formado por 20 textos, este estudio parte de un análisis textual y léxico para poder determinar los contenidos prototípicos de este género textual. Asimismo, las autoras aportan un conjunto de unidades fraseológicas que, además de arrojar luz sobre los rasgos digitales de este tipo de género discursivo, facilitan la redacción de los contenidos para “promover el uso de un lenguaje más sencillo por parte de las instituciones públicas”.

El artículo titulado “Uses and functions of conceptual metaphor in pedagogical meditation discourse”, firmado por Antonio-José Silvestre-López, parte del concepto de la metáfora conceptual para entender cómo se usan en manifestaciones discursivas del dominio fuente del espacio. El corpus utilizado en este estudio está formado por charlas introductorias sobre meditación, y la identificación de las metáforas se ha realizado a través de *Deliberate Metaphor Identification Procedure* (DMIP). Una de las contribuciones más relevantes del estudio es la de establecer una asociación entre el uso de las metáforas y la función explicativa en este tipo de textos. Las metáforas, además de facilitar la comprensión del mensaje mediante la reconceptualización, aportan coherencia y cohesión textual según las intenciones comuni-

cativas del propio hablante. El estudio, como podrá constatar el lector, desvela una novedosa técnica metodológica donde intervienen procesos inductivos y de codificación.

El tercer artículo del bloque I, bajo el título “¿De qué hablamos cuando divulgamos sobre lingüística? Análisis de un corpus de textos divulgativos y aplicaciones al estudio terminológico de la semántica léxica”, por Marta Sánchez-Saus Laserna y Vanesa Álvarez Torres, se centra un corpus compuesto por textos periodísticos divulgativos escritos por lingüistas publicados en diarios o revistas digitales. El estudio, como destacan las autoras, permite determinar qué aspectos del trabajo y la investigación lingüística están trasladando los lingüistas a la sociedad en la prensa en línea y, por otra parte, establecer el estado terminológico de las unidades léxicas especializadas del ámbito de la semántica léxica que aparecen en contextos divulgativos.

Siguiendo con el ámbito de los géneros especializados, “Emotions in specialised genres: power, manipulation and persuasion from the *affect spectrum theory*”, firmado por María Ángeles Orts Llopis, establece una conexión entre el constructivismo social y la *basic emotion theory*, que constituye, en esencia, una de las aportaciones más importantes de este estudio. Basado en la premisa de *affect spectrum theory*, el trabajo se centra en el papel constructivo de las emociones en la elaboración de textos especializados; en particular, en los textos jurídicos. Los resultados demuestran que los géneros textuales reflejan, como mecanismos intrínsecos de comunidades de habla, las tensiones o dinámicas de las propias comunidades, donde las emociones, desde una perspectiva biosocial, constituyen la raíz de las manifestaciones sociales, y por ende, de la propia comunicación interpersonal.

El artículo titulado “Fine-tuning machine translation quality-rating scales for new digital genres: The case of user-generated content”, por Miguel A. Candel-Mora, redefine la conocida premisa de que “cada género requiere escalas de valoración de la calidad específicas”, a través de patrones de error y características textuales de reseñas de usuarios. Dichas reseñas conforman el corpus usado por el autor para el análisis de los resultados, el cual permitirá mejorar exponencialmente los sistemas de traducción automática. En la presentación y análisis de resultados, el autor realiza una breve descripción de los factores que influyen en la calidad de la traducción automática, como son la subjetividad y la carencia de estructura específica en este género digital.

Por su parte, la autora Sofía Zakhir Puig, con el trabajo titulado “Análisis de la comunicación experto-lego por medio del consentimiento informado: legibilidad y relación médico-paciente”, toma como punto de partida el lenguaje especializado en el ámbito médico-jurídico. El estudio se propone como objetivo analizar el nivel de legibilidad y de experiencia de encuestados sobre el consentimiento informado, que constituye un género que pocas veces se lee antes de cualquier tratamiento médico. Basada en varias sesiones de entrevistas con 40 sujetos, la investigación refleja que la problemática de la poca claridad de este tipo de texto viene dada por el hecho de que la terminología usada carece de explicación o algún tipo de adaptación para usuarios legos. Asimismo, el porcentaje de entrevistados que consideran que la comunicación se ha realizado eficazmente es muy bajo, aspecto que explica por qué este tipo de investigación es importante para fomentar la adaptación de géneros textuales altamente especializados.

El siguiente artículo, titulado “¿Quién es quién? Análisis de los procesos de desfocalización del agente en español y en francés en dos géneros digitales”, por Emeline Pierre, se propone analizar tres estrategias de desfocalización del agente en español y en francés (la

pasiva perifrástica, la construcción con *se* – ambas disponibles en español y en francés – y el pronombre impersonal *on* en francés). El estudio, que se basa en corpus digitales, logra demostrar la influencia de variables como el género y la lengua en el lenguaje producido por los participantes en los foros digitales. Para ello, se ha tenido en cuenta la imagen de los interlocutores que participan en el foro digital, puesto que dicha interacción se basa en preguntas y respuestas. Asimismo, con el uso de la enciclopedia en línea se ha tenido en cuenta la intencionalidad de la objetividad, lo que permite contar con un estudio más amplio del proceso de desfocalización en textos discursivos de diferente categoría.

Finalmente, para cerrar este primer bloque del volumen sobre el discurso especializado, contamos con un artículo sobre marcadores pragmáticos en el discurso titulado “Los marcadores del discurso en la narrativa financiera: ¿Indicadores de ganancias y pérdidas?”. Con este sugerente título, la autora, Ana García Toro, combina dos procedimientos metodológicos perfectamente complementarios: un estudio cuantitativo de la frecuencia de los marcadores del discurso (MD) y otro estudio contrastivo entre dos corpus formados, por un lado, por cartas de accionistas que han obtenido ganancias y, por otro, por cartas de accionistas que han tenido pérdidas. Esta distribución cuantitativa de los MD y la tipología textual permitirá establecer si existe una relación entre los MD y el texto, así como corroborar si este tipo de herramienta de la lingüística computacional podría arrojar más luz al análisis de la argumentación en los discursos financieros.

El segundo bloque del presente volumen, titulado “Corpus y sus aplicaciones”, no se aleja demasiado de los contenidos del bloque anterior, puesto que varios de los trabajos sobre discursos especializados y digitales parten de la lingüística de corpus como herramienta metodológica de recopilación y análisis de datos. El bloque cuenta con cinco trabajos que aportan un amplio espectro sobre las aplicaciones de corpus en ámbitos de estudio diferentes.

El primer trabajo, titulado “El corpus Iberia como fuente para el estudio del léxico especializado”, por Luis Pablo Núñez, estudia el alcance del corpus sincrónico Iberia en la investigación del léxico especializado. Para constatar el alcance y las limitaciones de dicho corpus, el autor utiliza un número cerrado de términos extraídos de los ámbitos de la astronomía y la medicina. Dichos términos constituyen el *tertium comparationis* entre Iberia y los corpus de la Real Academia Española y del IULA.

El segundo artículo del bloque, titulado “A corpus-based study of 4-grams in the research article genre”, firmado por Eva Lucía Jiménez-Navarro, se centra en un apartado importante de los estudios de corpus: el estudio de la fraseología en textos discursivos especializados. En este trabajo en particular, la autora parte de grupos de palabras compuestas por cuatro gramas (4-gramas) en el lenguaje científico para determinar la secuencia de palabras y su función dentro del texto. El trabajo de investigación, como es de esperar, se basa en la recopilación, en forma de corpus, de textos científicos, que luego fueron procesados por el software Sketch Engine para obtener el listado de 4-gramas. Uno de los resultados más relevantes del trabajo ha sido identificar que el tipo de estructura más frecuente es el sintagma nominal, cuya función giraba alrededor del tipo de investigación y de la autoría del texto.

Por su parte, Macarena Palma Gutiérrez, con su contribución titulada “Verbs of Cutting in the Middle Construction: Examining the Compositional Cospecification of Patientive vs. Agent-Instrument middles”, examina dos tipos de medias (*middles* en inglés): los ergativos (p. ej. ‘*The cake cuts well*’) y los agentivo-instrumentales (p. ej. ‘*The saw cuts like a*

*dream*’). Para desarrollar los modelos colostruccionales de unidades complejas, la autora recopila los ejemplos del corpus; en particular, los que siguen la colocación de verbo + adverbio, así como la clasificación de Levin (1993), que hace una distribución de los verbos de “cortar” en inglés. Los resultados demuestran que los efectos prototípicos de la construcción de medias refuerzan los diferentes procesos de la cospecificación composicional que siguen los dos tipos de medias, tanto los ergativos como los agentivo-instrumentales.

El siguiente trabajo, titulado “Extracción automática de unidades terminológicas guiada por Framenet: una aplicación al corpus electrónico CORD-19”, firmado por Mario Crespo Miguel, se basa en la importancia de los estudios terminológicos en el desarrollo de herramientas tecnológicas que permitan una mejor comprensión de cómo se estructuran los discursos especializados. Este marco, como bien indica el autor, se basa “en el hecho de que ciertas palabras evocan determinadas situaciones en las que tienen lugar determinados participantes”. El estudio cuantitativo que aquí se presenta muestra los marcos de FrameNet que podrían tener un valor representativo en lo que se refiere a la COVID-19 y, de esta manera, determinar qué unidades léxicas funcionan como unidades terminológicas del corpus. Como es de esperar, el trabajo propone una metodología de investigación que, además de explicitar cómo se pueden extraer unidades terminológicas, permite que los datos puedan utilizarse para estructurar el nuevo sistema de unidades dentro del discurso especializado al que pertenece.

Como colofón de este segundo bloque, la autora Eva M. Llorens Simón, con el artículo titulado “On empathy in online tourism experience. Highest-frequency adjective patterns in English and Spanish promotional language”, propone un análisis de patrones adjetivales recurrentes en textos digitales sobre promoción turística en inglés y en español. De esta manera, y partiendo de los ejemplos extraídos de corpus en ambas lenguas, la autora establece segmentos prototípicos para conocer su potencial como transmisores de empatía. Estos segmentos permiten, a su vez, determinar la relevancia emocional de los adjetivos más habituales, lo cual podría tener un impacto significativo en la propia capacidad de predicción de la carga emocional del propio texto. En concreto, el estudio demuestra la existencia de puntos en común y divergentes entre los patrones adjetivales, así como los efectos emocionales que surgen entre los hablantes de cada idioma.

En definitiva, con todas estas contribuciones pretendemos iniciar una reflexión profunda sobre aspectos que son fundamentales en el campo de los nuevos géneros digitales, el análisis del discurso especializado y el trabajo con corpus; pensamos que todas las temáticas que abordan cada uno de los artículos incluidos en los dos grandes bloques donde los hemos insertado resultan de gran interés en una monografía de este tipo y, además, pretenden servir de apoyo a futuras investigaciones sobre cualquiera de los temas abordados por los autores.

## AGRADECIMIENTOS

Este número especial ha contado con la financiación de la Conselleria de Innovación, Universidades, Ciencia, y Sociedad Digital de la Generalitat Valenciana dentro del Programa para la promoción de la investigación científica, el desarrollo tecnológico y la innovación en la Comunitat Valenciana, con referencia del proyecto AORG/2021/118.

## REFERENCIAS BIBLIOGRÁFICAS

- Belda-Medina, J. y R. Casañ-Pitarch. (Eds.). (2020). *Análisis del discurso en la era digital. Una recopilación de casos de estudio*. Granada: Comares.
- Flowerdew L. (2012). How is Corpus Linguistics Related to Discourse Analysis? En *Corpora and Language Education. Research and Practice in Applied Linguistics*. London: Palgrave Macmillan.
- Friginal, E. y J. A. Hardy. (Eds.) (2020). *The Routledge Handbook of Corpus Approaches to Discourse Analysis*. Taylor & Francis. <https://doi.org/10.4324/9780429259982>
- Luzón, M.J. y C. Pérez-Llantada. (Eds.) (2019). *Science Communication on the Internet. Old genres meet new genres*. John Benjamins Publishing Company.
- Charlotte, T., y A. Marchi. (2018). *Corpus Approaches to Discourse: A Critical Review* (1st ed.). Routledge. <https://doi.org/10.4324/9781315179346>
- Paltridge, B. (2006). *Discourse Analysis*. London: Continuum.
- Sinclair, J. (2004). *Trust the text: language, corpus and discourse*. London/New York: Routledge.

# ARTÍCULOS



## Bloque I

### *Análisis del discurso especializado*





## La resolución administrativa publicada en portales web institucionales: Un análisis basado en corpus

The administrative resolution published on institutional web portals:  
A corpus-based analysis

SARA PISTOLA

Universidad Nacional de Educación a Distancia (UNED)  
spistola@flog.uned.es  
<https://orcid.org/0000-0002-6011-8639>

IRIA DA CUNHA

Universidad Nacional de Educación a Distancia (UNED)  
iriad@flog.uned.es  
<https://orcid.org/0000-0002-7707-1574>

### Resumen

La Administración pública española se comunica con la ciudadanía a través de una serie de textos especializados. Hasta hace unos años, estas comunicaciones se producían en formato papel. Sin embargo, con la aparición de la denominada *Administración electrónica*, internet se ha convertido en el canal de comunicación principal entre la Administración y la ciudadanía. Esto hace que los textos administrativos preexistentes se hayan tenido que adaptar a las posibilidades técnicas que ofrece la red, lo que los convierte en géneros digitales. Este trabajo se centra en el análisis de un género textual del

### Abstract

The Spanish public administration communicates with its citizens through a series of specialized texts. Until a few years ago, these communications were produced in paper format. However, with the emergence of the so-called e-administration, the internet has become the main channel of communication between the administration and its citizens. This means that pre-existing administrative texts have had to be adapted using the available technical possibilities, becoming digital genres. This paper focuses on the analysis of a textual genre of the administration that is currently disseminated

---

**Para citar este artículo:** Pistola, S. y da Cunha Fanego, I. (2022). La resolución administrativa publicada en portales web institucionales: Un análisis basado en corpus. *ELUA*, (38), 17-45. <https://doi.org/10.14198/ELUA.22442>

Recibido: 08/04/2022, Aceptado: 23/05/2022

© 2022 Sara Pistola, Iria da Cunha



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

ámbito de la Administración que en la actualidad se difunde casi exclusivamente a través de medios digitales, la resolución publicada en portales institucionales. Para ello, se toman como objeto de estudio las resoluciones publicadas en la página web del Boletín Oficial del Ayuntamiento de Madrid (BOAM), y se lleva a cabo un análisis lingüístico a nivel textual y léxico de un corpus formado por 20 textos. Asimismo, se realiza un análisis cualitativo de los rasgos digitales de este género textual. Como principal resultado de este estudio se obtiene una estructura modelo que contiene los apartados, los títulos y los contenidos prototípicos del género textual objeto de estudio. Además, se incluye una selección de unidades fraseológicas que se pueden utilizar para redactar esos contenidos. Estas unidades se adaptan a las recomendaciones de lenguaje claro más frecuentes para el español, con las que se pretende promover el uso de un lenguaje más sencillo por parte de las instituciones públicas. En cuanto al análisis de los rasgos digitales, se observan algunas diferencias en cuanto a la publicación de este tipo de textos en formato papel y a través de internet, como la posibilidad de ampliar la información a través de enlaces, o el uso de Códigos Seguros de Verificación (CSV).

**PALABRAS CLAVE:** Administración electrónica; análisis lingüístico; géneros digitales; lingüística de corpus; resolución administrativa.

## 1. INTRODUCCIÓN

La Administración pública se comunica continuamente con la ciudadanía en el marco de situaciones diversas, como la publicación de una convocatoria, la gestión de un contrato, el uso de un servicio, o la consulta de un portal web institucional por parte de un ciudadano. En este contexto de comunicación, se emplean una serie de textos prototípicos del ámbito especializado de la Administración, que son “el principal medio de comunicación entre ciudadanos y Administración” (MAP 2003: 13). Diversos autores hacen referencia a estos textos, los enumeran y clasifican a lo largo de los años. La lista más reciente de géneros textuales del ámbito de la Administración, clasificada en función de su emisor y su receptor, se recoge en Pistola y Viñuales-Ferreiro (2021). En este estudio, se citan 20 géneros redactados por la Administración y dirigidos a la ciudadanía, como el acta, el acuerdo, el certificado, el contrato, el convenio, la licencia y el reglamento.

El personal de la Administración remite algunos de esos géneros directamente a los interesados. A modo de ejemplo, cuando una persona debe aportar información o documen-

almost exclusively through digital media: the resolution published on institutional web portals. For this purpose, the resolutions published on the website of the Boletín Oficial del Ayuntamiento de Madrid (BOAM) are taken as the subject of study for this research. We carry out a linguistic analysis (on a textual and lexical level) of a corpus consisting of 20 texts. Likewise, a qualitative analysis of the digital features of this textual genre is performed. The main result of this study is a model structure containing the prototypical sections, titles and contents of the textual genre under analysis. In addition, a selection of phraseological units that can be used to write these contents is included. These units are adapted to plain language recommendations, which advocate for the use of simpler language by public institutions. Regarding the analysis of the digital features, some differences are observed between the publication of this type of text in paper format and on the internet. For example, with online texts it is possible to expand the information through web links or to use Secure Verification Codes (CSV).

**KEYWORDS:** E-administration; linguistic analysis; digital genres; corpus linguistics; administrative resolution.

tación en el contexto de un procedimiento administrativo, el responsable de la Administración le hace llegar un requerimiento. Tradicionalmente, estas comunicaciones se remitían a la ciudadanía mediante correo postal. Sin embargo, con la aparición de la Sociedad de la Información, el desarrollo de las Tecnologías de la Información y la Comunicación (TIC), y la democratización del acceso a internet, la Administración pública ha tenido que adaptarse e iniciar la transición de una Administración en papel a una Administración electrónica (Araguàs 2012). De esta forma, tal y como señala Cotino (2017: 398-399): “En virtud de la legislación administrativa reciente (Leyes 39 y 40/2015), el canal de relación de la ciudadanía con la Administración casi por defecto es el electrónico. De igual modo, el soporte básico de la información es el electrónico”.

Asimismo, existen algunos géneros redactados por la Administración y dirigidos a la ciudadanía que no se remiten a un ciudadano en particular, sino que se deben poner a disposición de la ciudadanía en general. Por ejemplo, un anuncio o un reglamento. El canal mediante el cual se comunica este tipo de géneros también ha cambiado mucho durante los últimos años. En la actualidad, estos géneros se ponen en conocimiento de la ciudadanía mayoritariamente a través de los portales web de las instituciones públicas. Un ejemplo de estos portales institucionales o sedes electrónicas son los Boletines Oficiales. El Boletín Oficial del Estado (BOE) es quizás el más conocido. En él se publican leyes, disposiciones y actos de diversos tipos. Dispone de una versión en papel y de una versión electrónica. Esta última está disponible desde el 1 de enero de 2009 en <https://www.boe.es/>.

Además, otros organismos públicos, tanto de carácter autonómico como local, disponen de sus propias sedes electrónicas y Boletines Oficiales. Por ejemplo, en el ámbito local, se encuentra el Boletín Oficial del Ayuntamiento de Madrid (BOAM). En él se publican disposiciones y actos de los organismos municipales, así como otro tipo de anuncios e informaciones cuando se considera que deben ponerse en conocimiento de la ciudadanía. Actualmente, el BOAM se publica únicamente en formato digital en la web <https://sede.madrid.es/>.

Uno de los géneros administrativos más importantes que se publican en los portales institucionales en el ámbito local es la resolución. Este género adquiere especial relevancia porque es el que se utiliza para informar a los ciudadanos sobre cuestiones de su día a día, como convocatorias de empleo, ayudas sociales, decisiones sobre la gestión del tráfico y muchas otras. Asimismo, este tipo de resoluciones puede dar inicio a un procedimiento (p. ej. convocar unas ayudas para el transporte), finalizarlo (p. ej. conceder o denegar las ayudas), funcionar como actos de trámite (p. ej. solventar errores en la publicación de la convocatoria), y para plasmar otro tipo de atribuciones regulatorias y de policía conferidas a las Administraciones locales (p. ej. abrir el paso de ciertos vehículos autorizados al centro de la ciudad). Por lo tanto, mediante este género textual la ciudadanía puede conocer cuáles son sus derechos y sus deberes, y cómo hacerlos valer ante las instituciones.

En el marco de este proceso de adaptación y modernización de la Administración pública española también destaca la creciente preocupación por el uso de un lenguaje más sencillo y comprensible. De hecho, existe una corriente internacional para promover el uso de un lenguaje claro por parte de las instituciones públicas con la que se pretende “disminuir las barreras de acceso que tiene el ciudadano a la información institucional” (Pedraza 2019: 107).

En este contexto, en este estudio se toman como caso de estudio las resoluciones publicadas en el BOAM y se plantean tres objetivos:

1. Determinar la estructura y contenidos prototípicos de la resolución que se publica en portales institucionales o sedes electrónicas (análisis del nivel textual), a partir de un análisis de corpus.
2. Extraer fraseología que se pueda utilizar para redactar los contenidos de este género textual y adaptarla a lenguaje claro (análisis del nivel léxico).
3. Analizar los rasgos digitales de este género textual cuando se publica en portales institucionales o sedes electrónicas (análisis de la vertiente digital).

Para ello, se propone una metodología basada en el análisis de un corpus formado por 20 resoluciones. Estas resoluciones se han extraído del portal web del BOAM, que genera una “gran oferta y disponibilidad textual” (Vargas-Sierra 2005: 1) para la compilación de corpus en el ámbito de la Administración. Como resultado, se obtiene una ficha con la estructura, los contenidos y la fraseología prototípicos del género textual *resolución publicada en portales institucionales*. Esta ficha se podría emplear como plantilla para crear recursos de escritura que ayudasen al personal público a redactar de forma fácil, clara y sencilla este tipo de textos dirigidos a la ciudadanía. También se podría utilizar para el diseño de un redactor asistido de textos administrativos electrónicos en lenguaje claro. Además, el análisis de los rasgos digitales permite determinar el grado de digitalización de estos textos.

En el apartado 2 se ofrece un estado de la cuestión que relaciona los conceptos de Administración electrónica, géneros digitales y lenguaje claro. En el apartado 3 se detalla la metodología. En el apartado 4 se exponen los resultados del análisis textual (apartado 4.1), de la extracción y adaptación de la fraseología (apartado 4.2), y del análisis de los rasgos digitales (apartado 4.3). Finalmente, en el apartado 5 se incluyen las principales conclusiones y las líneas de trabajo futuro.

## 2. ESTADO DE LA CUESTIÓN

En los últimos años, la Administración pública ha tenido que evolucionar para adaptarse a los cambios de la sociedad. En este sentido, Sánchez Acevedo (2015: 231) señala que “las nuevas tecnologías ofrecen oportunidades para el intercambio de información, la participación del público y la colaboración entre ciudadanos”. Asimismo, en relación con los cambios en el proceso de comunicación entre Administración y ciudadanía, Pedraza (2019: 279) establece que:

El usuario de la Administración pública es de carácter colectivo, heterogéneo y, desde el punto de vista comunicativo, es solo un destinatario, mas no un interlocutor. Sin embargo, la irrupción de las TIC ha propiciado un cambio de visión donde el ciudadano adquiere un rol activo en el circuito de la comunicación. Esto ha acarreado un cambio significativo en la forma como las instituciones comunican, dada la exigencia de comunicar y no solo de informar.

En este contexto de cambio, en las últimas décadas han surgido dos corrientes de estudio que hacen frente a las necesidades de la sociedad actual como sociedad de la información y que tienen implicaciones en los procesos de comunicación entre la Administración y la ciudadanía: el lenguaje claro y la Administración electrónica. Estas dos corrientes son una muestra del proceso de modernización que está viviendo la Administración española.

Por un lado, el término *lenguaje claro* procede del inglés *plain language*. El movimiento *Plain Language* surge en Reino Unido en 1979 para promover el uso de un lenguaje más sencillo en las instituciones públicas. La International Plain Language Federation (en línea) define *lenguaje claro* de la siguiente forma: “A communication is in plain language if its wording, structure, and design are so clear that the intended readers can easily find what they need, understand what they find, and use that information”. En el ámbito jurídico, encontramos una definición reciente de lenguaje claro en Meza et al. (2020: 22):

We understand plain writing in Law as the process of constructing legal discourses aimed at promoting the intelligibility of the various legal institutions or institutions of Law, such as sentences, regulations, contracts, and so on, in order to offer or strengthen security in the legal field.

En el marco de esta corriente, algunos estudios ofrecen recomendaciones concretas para la redacción en lenguaje claro de textos del ámbito jurídico-administrativo (Carretero 2019; Montolío y Tascón 2020; Jiménez y Sancho 2021). En concreto, en Da Cunha y Escobar (2021) se recopilan las recomendaciones sobre lenguaje claro en español más frecuentes y se clasifican en función del nivel de la lengua al que se refieren. Algunas de estas recomendaciones son: el uso de oraciones cortas (nivel discursivo), el uso de la voz activa en vez de la voz pasiva (nivel morfosintáctico) y el uso de palabras comunes en vez de términos, en caso de ser posible (nivel léxico).

La adaptación a lenguaje claro del discurso jurídico-administrativo también contribuye a adecuar los textos al medio digital. Tal y como señala García-Cabrera (2018), en la web se tiende a leer más lentamente, y a comprender y a recordar menos que mediante la lectura en papel. Por lo tanto, es necesario adaptar la forma de escribir al medio digital. Para ello, esta autora recomienda evitar información redundante, usar oraciones cortas y simples, utilizar palabras lo más comunes y sencillas posible, eliminar las palabras oscuras, y usar verbos en voz activa, entre otras cuestiones. Estos consejos para escribir en la web coinciden con algunas de las recomendaciones de lenguaje claro más frecuentes para el español.

Por otro lado, el concepto de *Administración electrónica* hace referencia al uso de medios digitales por parte de las Administraciones públicas. Araguàs (2012: 129) ofrece la siguiente definición:

Concebimos la Administración electrónica como un nuevo modelo de gestión de la actividad administrativa, fundamentado en el uso intensivo de las tecnologías de la comunicación y la información, al que van ligadas consecuencias tanto de carácter interno, relacionadas con los procesos administrativos y con la organización, estructura y personal de la propia Administración; como consecuencias de carácter externo, esencialmente, modificando las formas y canales de comunicación entre la Administración y los ciudadanos, así como cambiando las formas en las que se venían prestando los servicios públicos tradicionales y prestando servicios públicos nuevos.

La mayoría de las investigaciones sobre Administración electrónica se hacen desde una perspectiva jurídica. Por consiguiente, se centran en la regulación de esta nueva realidad, y los derechos de los ciudadanos y las obligaciones de la Administración que se derivan de ella. Por ejemplo, Sánchez Acevedo (2015: 232) habla sobre el derecho a la buena Administración electrónica y lo relaciona con el principio de transparencia:

El derecho a la buena eAdministración, desde la óptica de la transparencia, impone la necesidad contar con instrumentos regulados y tecnológicamente desarrollados que permitan que las administraciones publiquen la información y que los interesados tengan acceso a la misma. [...] La información sujeta a las obligaciones de transparencia debe estar publicada en las correspondientes sedes electrónicas o páginas web de las administraciones, de una manera clara, estructurada y entendible para los interesados y, preferiblemente, en formatos reutilizables.

En este mismo sentido, Cotino (2017: 399) señala que el nexo entre la Administración electrónica, la transparencia y la información pública “es amplio e ineludible”. Asimismo, Toscano (2019) establece una conexión entre la Administración electrónica y la simplificación administrativa, que hace referencia a los trámites.

Desde la perspectiva lingüística, se han llevado a cabo diversos estudios en los que se emplea el análisis de corpus de textos jurídico-administrativos. Entre otros, se pueden citar Castellón (2001), Montolío (2011), Cervera y Torres (2016), Da Cunha y Montané (2020), y Prodigioso Volcán (2020, 2021). Al respecto del empleo de la lingüística de corpus para el estudio del discurso jurídico-administrativo, Lorente (2016: 178-179) afirma que:

La construcción de un corpus con textos especializados no difiere en los aspectos principales de la compilación de un corpus de lengua general, pero presenta elementos específicos que conviene tener en cuenta. Suelen ser corpus más reducidos en los que la cantidad de información no es tan determinante para el grado de representatividad lingüística, pero en los que una buena estructuración interna puede garantizar que la descripción del discurso especializado sea más exhaustiva. Los corpus especializados se organizan por dominios y subdominios temáticos, y la selección de sus documentos o fragmentos de documentos atiende a criterios de tipología textual y de variación de géneros discursivos.

También se han elaborado desde el punto de vista lingüístico algunos estudios relacionados con la vertiente digital de la Administración. Por ejemplo, Pedraza (2019) promueve la democratización del discurso administrativo a través de una propuesta lexicográfica digital panhispánica. Asimismo, Prodigioso Volcán (2020, 2021) elabora dos informes sobre la claridad con la que la Administración se dirige a la población en general y a los públicos vulnerables en particular. Para ello, tienen en cuenta dimensiones relacionadas con la lingüística (lenguaje claro) y la tecnología (usabilidad, accesibilidad y diseño de los portales institucionales donde se vuelcan los textos).

Sin embargo, se han encontrado pocos estudios en los que se profundice sobre el estudio de los géneros del ámbito de la Administración como géneros digitales, a pesar de que la regulación actual obliga a las Administraciones públicas a emplear medios electrónicos en sus relaciones con la ciudadanía (Toscano 2019). Esto puede deberse a que los géneros administrativos que se vuelcan en la Administración electrónica no son realmente géneros digitales, sino que lo que se produce es un trasvase de los géneros por escrito al medio digital. Es decir, los géneros preexistentes en papel se adaptan a la comunicación electrónica. En palabras de Toscano (2019: 351), “es, digámoslo así, el mismo derecho administrativo con otros medios”.

A esta distinción entre géneros puramente digitales y géneros preexistentes que se adaptan al medio digital hace referencia Fernández Sánchez (2003: 40) en su estudio sobre los géneros digitales del ámbito bancario. En concreto, señala que:

Los géneros digitales, también denominados *cybergenres* (Shepherd & Watters, 1999), *web genres* (Crowston & Williams, 1997), o *Internet genres*, podrían definirse como los equivalentes virtuales de documentos tradicionales, que se han adaptado a las condiciones del medio digital y, a su vez, han aprovechado las posibilidades técnicas que ofrece la red (Fortanet et al., 1999; Schmid-Isler, 2000; Peticca, 2003). La interactividad, los enlaces, o la libertad a la hora de estructurar el contenido de los documentos electrónicos están provocando la transformación de géneros tradicionales en subgéneros análogos (*extant subgenres*), esto es, géneros digitales basados en otros ya existentes en medios como el papel o el vídeo, o en subgéneros totalmente nuevos (*novel subgenres*), que han surgido gracias a las posibilidades técnicas que ofrece la *World Wide Web* y que carecen de réplicas en los medios tradicionales (Sheperd & Watters, 1998).

Más prolífico es el estudio de los géneros digitales de otros ámbitos especializados como el periodismo y el turismo. En estos sectores, surgen nuevos géneros digitales y los géneros preexistentes sufren cambios significativos al trasladarse al canal digital. A modo de ejemplo, Pérez-Llantada (2021) analiza un género digital emergente en el ámbito científico, el proyecto de ciencia orientado al micromecenazgo en España. Por su parte, Fernández-Vallejo (2017) analiza un género preexistente en el ámbito de la comunicación (la disculpa corporativa) que se adapta a la comunicación digital (a través de Twitter). Asimismo, González García (2013) profundiza en la caracterización de los géneros textuales del ámbito del turismo en internet.

### 3. METODOLOGÍA

La metodología de esta investigación se basa en la empleada por Da Cunha y Montané (2020) y Da Cunha et al. (2021) para el análisis lingüístico de géneros textuales de ámbitos especializados, en concreto, de la Administración y el turismo. A esta metodología se le introducen algunas modificaciones. Finalmente, incluye seis etapas, que se detallan a continuación.

En la etapa 1 se compila el corpus, que está formado por 20 textos correspondientes al género textual *resolución publicada en portales institucionales*. Para determinar el número de textos que conforman el corpus se siguen las investigaciones de Da Cunha y Montané (2020) y Da Cunha et al. (2021). En ellas, se lleva a cabo un análisis estadístico en el que se concluye que un corpus de 20 textos resulta representativo para analizar géneros de ámbitos especializados como el de la Administración. El corpus incluye un total de 29.463 palabras. El texto más corto tiene 497 palabras y el texto más largo, 6.505.

Los textos se seleccionan de la página web del BOAM. Para ello, se introduce la palabra *resolución* en el buscador y se realiza una selección aleatoria de los resultados que cumplen los siguientes requisitos:

- Se corresponden con el género *resolución publicada en portales institucionales*.
- Tienen como emisor al personal del Ayuntamiento de Madrid y como receptor a la ciudadanía. No se tienen en cuenta los documentos internos ni los redactados por la ciudadanía y dirigidos al Ayuntamiento.
- Se publican entre el 1 de enero de 2017 y el 31 de diciembre de 2021. Las resoluciones anteriores al 2017 se descartan para garantizar que los textos sean actuales y cumplan con la legislación vigente. En concreto, se garantiza que los documentos cumplan



con la Ley 39/2015, de 1 de octubre, del Procedimiento Administrativo Común de las Administraciones Públicas, y la Ley 40/2015, de 1 de octubre, de Régimen Jurídico del Sector Público. Ambas normas constituyen la regulación básica del procedimiento administrativo y del régimen jurídico del sector público en España.

- Tienen una extensión de 1-20 páginas.
- Pertenecen a áreas del Ayuntamiento diversas. Se plantea un umbral mínimo de cinco áreas u organismos, y un umbral máximo de cinco textos por área. Es decir, los 20 textos compilados para cada género textual deben proceder de, al menos, cinco áreas u organismos del Ayuntamiento de Madrid. Además, no puede haber más de cinco textos pertenecientes a un mismo área u organismo.
- Son textos reales que se han hecho llegar a la ciudadanía. Como se trata de documentos públicos, no es necesario anonimizarlos.
- Están constituidos mayoritariamente por texto escrito. También pueden contener tablas y gráficos, pero estos no pueden ser el único contenido del documento.

En la etapa 2 se selecciona el marco teórico de la investigación. De esta forma, se parte de las consideraciones de Swales (1990) y Van Dijk (1977, 1989) sobre los géneros textuales. Para Swales (1990), los géneros textuales son estructuras convencionalizadas a través de las cuales se organizan los intercambios comunicativos de una determinada comunidad discursiva, como puede ser la Administración pública. Por su parte, Van Dijk (1977, 1989) afirma que los géneros textuales siguen un patrón claramente codificado y ampliamente aceptado, que él identifica con los conceptos de *macroestructura* y *superestructura*.

En esta investigación se emplea el concepto de superestructura de Van Dijk (1977, 1989), que se corresponde con el esquema organizativo de un texto. En general, la superestructura se muestra mediante distintos apartados, que pueden incluir un título y diferentes contenidos. En este contexto, la fraseología prototípica de los textos especializados puede ayudar al usuario a expresar un contenido concreto. Asimismo, se utiliza el concepto de *moves* de Swales (1990), que representan los segmentos de texto que cumplen una función comunicativa concreta (Swales 1990; Upton y Cohen 2009) y que suelen insertarse en alguno de los apartados del texto. En este trabajo, los *moves* se equiparan a los contenidos.

Teniendo esto en cuenta, se seleccionan los rasgos lingüísticos que se considerarán para el análisis del corpus. Son: los apartados, los títulos y los contenidos para el nivel textual, y la fraseología para el nivel léxico. Los apartados se utilizan para estructurar el texto y pueden ser claramente identificados ya sea porque tienen un espacio en blanco antes y después, o porque el tema sobre el que tratan es diferente. Los títulos pueden aparecer al inicio de los apartados y se suelen distinguir por la aparición de espacios en blanco antes y después, un cambio en los elementos tipográficos, o el uso de mayúsculas y negritas. Los contenidos (*moves*) se identifican porque forman parte de los apartados del texto, cumplen una función comunicativa concreta y funcionan como unidades semánticas. La fraseología se emplea para expresar los contenidos.

En la etapa 3 se anota el corpus. Para ello, se transforman los textos de la resolución a texto plano y se importan los archivos en la herramienta ATLAS.ti, un programa para el análisis cualitativo de datos que se utiliza frecuentemente para la anotación de corpus. Mediante esta herramienta, un anotador con formación en lingüística y en derecho administrativo anota los rasgos correspondientes con el nivel textual (es decir, los apartados, los

títulos y los contenidos), así como la fraseología prototípica que se utiliza en el texto para expresar los contenidos.<sup>1</sup> En la figura 1 se muestra la anotación de uno de los textos del corpus mediante la herramienta ATLAS.ti.

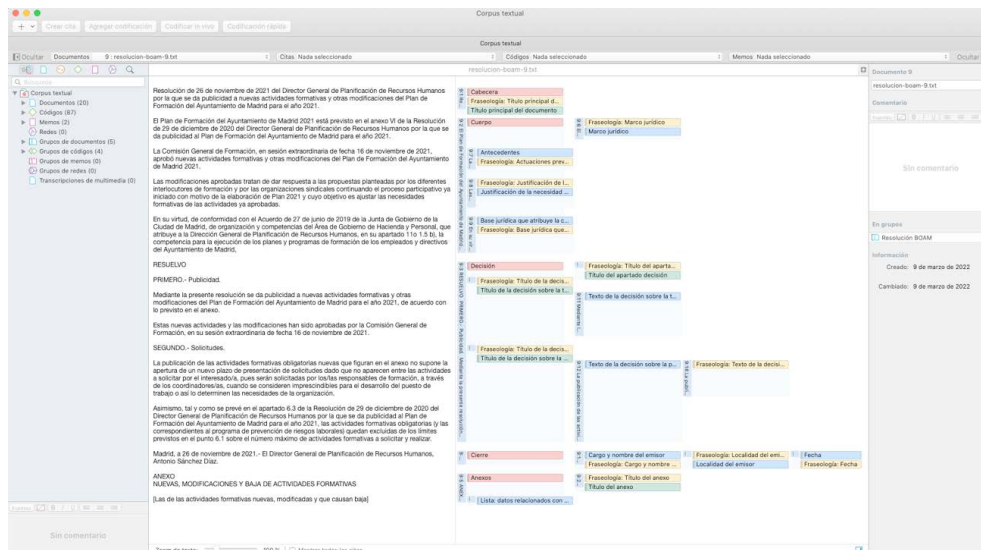


Figura 1. Anotación del corpus con ATLAS.ti.

En la etapa 4 se determinan los rasgos textuales prototípicos del género *resolución publicada en portales institucionales*. Para ello, se cuantifican los textos en los que aparecen los distintos apartados, títulos y contenidos identificados en la etapa anterior. Se establece un umbral de frecuencia del 50 % para determinar la relevancia de un rasgo, siguiendo la metodología validada en Da Cunha y Montané (2020) y Da Cunha et al. (2021). Es decir, si un apartado, título o contenido aparece en un 50 % o más de los textos del corpus, se considera que dicho rasgo es prototípico de la resolución. Adicionalmente, en nuestro trabajo, los apartados, títulos y contenidos que aparecen en entre un 40 % y un 50 % de los textos se marcan como opcionales mediante las siguientes marcas: *[apartado opcional]*, *[título opcional]* y *[contenido opcional]*.

En la etapa 5 se selecciona la fraseología más representativa para expresar cada contenido y se adapta a lenguaje claro. La selección se lleva a cabo de forma cualitativa. Se extraen tanto oraciones completas como fragmentos de oraciones, que pueden ser invariables o incluir algún tipo de información variable. En este segundo caso, la información variable se marca entre corchetes. Por ejemplo, la unidad detectada en el corpus *Gerencia del Organismo Autónomo Madrid Salud* se recoge como *[área/órgano/institución]*. Asimismo, si se observan cuestiones de formato o aspectos lingüísticos llamativos, se aporta una indicación entre paréntesis al final de la frase. A modo de ejemplo, se detecta un contenido que

<sup>1</sup> La selección de apartados, títulos, contenidos y frases del corpus fue validada por otras dos investigadoras del proyecto de investigación en el que se enmarca este trabajo, una especialista en lingüística y otra en derecho administrativo.

siempre comienza por un verbo en infinitivo, por lo que en la fraseología correspondiente se incluye la siguiente indicación: *(siempre empieza por un verbo en infinitivo)*. Para que la fraseología resulte representativa, se descartan las frases que hacen referencia a temáticas muy específicas, y que son difícilmente extrapolables a resoluciones sobre otras temáticas o redactadas por Administraciones públicas diferentes.

En cuanto a la adaptación a lenguaje claro de la fraseología seleccionada del corpus, se tienen en cuenta las recomendaciones que se recogen en Da Cunha y Escobar (2021). Esta investigación establece las recomendaciones más frecuentes de lenguaje claro en español a nivel discursivo (p. ej. *se recomienda redactar oraciones cortas y se recomienda redactar párrafos breves*), morfosintáctico (p. ej. *se recomienda utilizar la voz activa en vez de la voz pasiva y se recomienda evitar los verbos en gerundio*), y léxico (p. ej. *se recomienda eliminar las palabras innecesarias y se recomienda evitar los arcaísmos*). La adaptación de la fraseología a lenguaje claro se lleva a cabo de forma manual.

Por último, en la etapa 6 se determinan los rasgos digitales del género textual objeto de estudio. Para ello, se describen, por un lado, la página web mediante la cual la ciudadanía accede a este tipo de texto y, por otro lado, el texto en sí mismo, tanto en su versión HTML como en el formato PDF descargable.

Para ilustrar esta metodología se ha diseñado la figura 2. En ella, se resumen las seis fases explicadas en este apartado.

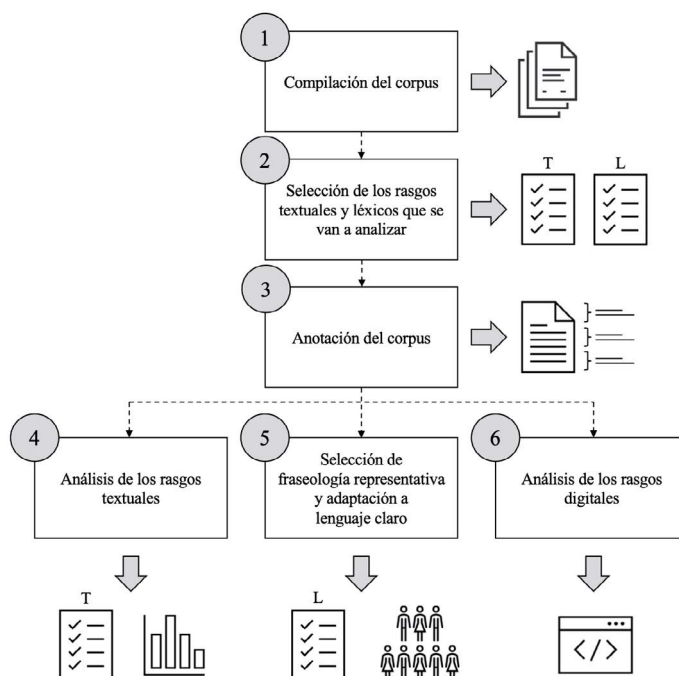


Figura 2. Fases de la metodología.

#### 4. RESULTADOS: ANÁLISIS DEL GÉNERO *RESOLUCIÓN PUBLICADA EN PORTALES INSTITUCIONALES*

En este apartado se incluye una propuesta de estructura modelo del género *resolución publicada en portales institucionales*, con la indicación de sus apartados, títulos y contenidos prototípicos (apartado 5.1). Asimismo, se muestra la adaptación a lenguaje claro de una selección de fraseología que puede ser utilizada para expresar los contenidos de este género textual (apartado 5.2). Finalmente, se ofrece el resultado del análisis del portal web en el que se insertan las resoluciones y de los rasgos digitales que se observan en los textos (apartado 5.3).

##### 4.1. Rasgos textuales: estructura y contenidos

Tal y como se explica en el apartado 4, se consideran prototípicos del género *resolución publicada en portales institucionales* los apartados, títulos y contenidos que aparecen en al menos un 50 % de los textos del corpus. Asimismo, se incluyen como opcionales los apartados, títulos y contenidos que aparecen en entre un 40 % y un 50 % de los textos. Teniendo esto en cuenta, en la tabla 1 se muestra la ficha con la estructura modelo propuesta.

Apartados	%	Contenidos	%
Cabecera	100	Título principal del documento	100
Cuerpo	100	Marco jurídico	90
		Antecedentes	65
		Objetivo de la resolución	55
		Base jurídica que atribuye la competencia para dictar la resolución	100
Decisión	100	Título del apartado	100
		Título de la decisión sobre la temática de la resolución (puede haber varias)	100
		Texto de la decisión sobre la temática de la resolución (puede haber varias)	100
		Título de la decisión sobre la publicación de la resolución en el portal institucional/Boletín Oficial/sede electrónica [título opcional]	40
		Texto de la decisión sobre la publicación de la resolución en el portal institucional/Boletín Oficial/sede electrónica [contenido opcional]	45
		Título de la decisión sobre el acceso a información relacionada con la resolución [título opcional]	40
		Texto de la decisión sobre el acceso a información relacionada con la resolución [contenido opcional]	40
Información sobre recursos	80	Fórmula de introducción de los recursos posibles	75
		Recursos que cabe interponer	75
		Información sobre la posibilidad de interponer otros recursos	75

Cierre	100	Localidad del emisor	95
		Fecha	95
		Cargo y nombre del emisor	100
Anexos [apartado opcional]	40	Título del anexo	40
		Posibles contenidos de los anexos: tablas/listas con datos relacionados con la resolución, formularios, regulación de aspectos concretos de la resolución, etc.	-

Tabla 1. Estructura modelo de la resolución publicada en portales institucionales.

Como se puede observar, cuatro de los cinco apartados que superan el umbral del 50 % lo hacen con una frecuencia del 100 %. Es decir, estos cuatro apartados aparecen en todos los textos del corpus. Además, el apartado *Información sobre recursos* aparece en un 80 % de los textos. Estos resultados van en la línea de lo expuesto por autores como De Miguel (2000: en línea), quien señala que: “el texto administrativo, según se ha señalado repetidamente en la bibliografía, se caracteriza por la rigidez de su estructura, un esquema invariable establecido de antemano para cada modalidad”. Asimismo, se incluye un apartado opcional, el correspondiente a los *Anexos*.

En cuanto a los títulos, tres de los seis apartados detectados comienzan por un título. Además, dentro del apartado *Decisión* se puede observar que el texto de cada una de las decisiones (sobre la temática, sobre la publicación y sobre el acceso a información relacionada con la resolución) va precedido de un título con una indicación numérica. Esta forma de marcar los diferentes apartados y contenidos del texto promueve el orden interno, que es una de las cuestiones por las que se caracteriza el lenguaje administrativo modélico según Castellón (2000, 2001). También facilita la lectura de los textos publicados en la web (García-Cabrera 2018). En el ejemplo (1) se muestra un fragmento original extraído del corpus en el que se observa la distribución de los contenidos del apartado *Decisión*:

(1) RESUELVO

PRIMERO.- Acumular los expedientes individuales de solicitud de ayudas asistenciales de la convocatoria de 2021 dada su identidad sustancial.

SEGUNDO.- Denegar las solicitudes de ayudas asistenciales convocatoria 2021 relacionadas en el expediente, al personal municipal, jubilado y pensionista del Ayuntamiento de Madrid, por no cumplir los requisitos establecidos.

TERCERO.- Por el órgano competente se procederá a la publicación de la presente resolución en el Boletín Oficial del Ayuntamiento de Madrid, en la intranet/extranet municipal ayre (<https://ayre.madrid.es>) y en la web <https://jubilacion.madrid.es>.

CUARTO.- Ordenar la inserción de la información individual de cada solicitante, relativa a las solicitudes de ayudas asistenciales denegadas, en la aplicación de acción social con fecha de la publicación de la presente resolución en el Boletín Oficial del Ayuntamiento de Madrid.

QUINTO.- La relación de ayudas asistenciales denegadas al personal pensionista, que no haya realizado su solicitud de forma electrónica, se publicará como anexo de la presente resolución en el Boletín Oficial del Ayuntamiento de Madrid.

En relación con los contenidos, se observan algunos muy frecuentes. En el apartado *Cuerpo*, los más frecuentes son *Marco jurídico* y *Base jurídica que atribuye la competencia para dictar la resolución*. El motivo de que estos contenidos aparezcan en un gran número de textos del corpus puede encontrarse en el preámbulo de la Ley 39/2015. En él, se dice que uno de los instrumentos mediante los cuales se protegen los derechos de la ciudadanía frente a la actuación de la Administración es el procedimiento administrativo, “que es la expresión clara de que la Administración pública actúa con sometimiento pleno a la Ley y al Derecho” (preámbulo I, L 39/2015, de 1 de octubre). De esto se deduce que las decisiones que toma la Administración y que plasma en las resoluciones deben estar fundamentadas en normas jurídicas previas. En el ejemplo (2) se recoge un fragmento del corpus relativo al *Marco jurídico*, mientras que en el ejemplo (3) se muestra un fragmento del contenido *Base jurídica que atribuye la competencia para dictar la resolución*:

- (2) Ordenanza 10/2021, de 13 de septiembre, por la que se modifica la Ordenanza de Movilidad sostenible de 5 de octubre de 2018 (en adelante, OMS), fue aprobada por el Pleno del Ayuntamiento de Madrid el 13 de septiembre y publicada en el Boletín Oficial de la Comunidad de Madrid el 21 de septiembre, regula, entre otras cuestiones, las nuevas zonas de bajas emisiones de la ciudad y, en particular, la Zona de Bajas Emisiones de Especial Protección (ZBEDEP) Distrito Centro, que introduce algunas modificaciones respecto al régimen anterior de la Zona de Bajas Emisiones “Madrid Central”. El artículo 23 sigue siendo el que define los motivos de su creación, delimita su ámbito territorial y establece los criterios de acceso y funcionamiento y estos últimos son desarrollados en su apartado 3 en función de la clasificación de los vehículos por su potencial contaminante y la función y necesidad de acceso de dichos vehículos a la zona.
- (3) En su virtud, de conformidad con lo dispuesto en el apartado 13o.1.3. c) y d) y 13.2.1 del Acuerdo de 27 de junio de 2019 de la Junta de Gobierno de la Ciudad de Madrid de organización y competencias del Área de Gobierno de Hacienda y Personal, modificado, en cuanto a las competencias de esta Dirección General, por el Acuerdo de 30 de enero de 2020 de la Junta de Gobierno,

En el apartado *Decisión*, los contenidos más frecuentes tienen que ver con el título y el texto de la decisión sobre la temática de la resolución. Este tipo de decisiones son las que deciden sobre el fondo del asunto, es decir, contienen el propósito general de la resolución, así como algunas particularidades relacionadas con esta. Por eso, aparecen en todas las resoluciones. Pueden hacer referencia a distintas temáticas, por ejemplo: convocar puestos de trabajo, adjudicar ayudas sociales, prohibir la circulación por una calle de la ciudad, conceder un permiso, subsanar errores, etc. En el ejemplo (4) se observa un fragmento del corpus correspondiente a los contenidos *Título de la decisión sobre la temática de la resolución* (indicación numérica) y *Texto de la decisión sobre la temática de la resolución*. La temática de esta decisión es la convocatoria de puestos de trabajo:

- (4) PRIMERO.- Convocar los puestos de trabajo que se relacionan en el Anexo I.

El resto de las decisiones sobre *publicación de la resolución en el portal institucional/Boletín Oficial/sede electrónica y acceso a información relacionada con la resolución* son de carácter formal. Por lo tanto, son independientes de la temática de la resolución y no deciden

sobre cuestiones relacionadas con el fondo del asunto. Los títulos y textos relacionados con estas decisiones se incluyen en la estructura modelo de la tabla 1 con la marca de opcionales. En el ejemplo (5) se muestra un fragmento del corpus correspondiente a los contenidos *Título de la decisión sobre la publicación de la resolución en el portal institucional/Boletín Oficial/sede electrónica* y *Texto de la decisión sobre la publicación de la resolución en el portal institucional/Boletín Oficial/sede electrónica*. Entretanto, en el ejemplo (6) se recoge un fragmento relativo a los contenidos *Título de la decisión sobre el acceso a información relacionada con la resolución* y *Texto de la decisión sobre el acceso a información relacionada con la resolución*:

- (5) TERCERO. Publicar en el Boletín Oficial del Ayuntamiento de Madrid la presente resolución.
- (6) TERCERO.- Ordenar la inserción de la información individual relativa a las solicitudes de ayudas a la promoción interna del personal, concedidas, denegadas y desistidas solicitadas del año 2021 al personal de Madrid Salud en la intranet municipal “Ayre”, en la ruta “Mi área personal/Mis ayudas”.

En el apartado *Información sobre recursos*, los tres contenidos prototípicos aparecen con la misma frecuencia en el corpus. Los recursos administrativos constituyen otro de los mecanismos mediante los cuales se protegen los derechos de la ciudadanía frente a la actuación de la Administración, junto al control realizado por jueces y tribunales (preámbulo I, L 39/2015, de 1 de octubre). Con carácter general, los interesados pueden interponer recursos administrativos contra las resoluciones que publican las Administraciones públicas. No obstante, se establecen excepciones. A modo de ejemplo, la Ley 39/2015 dice que contra las resoluciones de acumulación de los expedientes “no procederá recurso alguno” (artículo 57, L 39/2015, de 1 de octubre). En el ejemplo (7) se observa un fragmento original del corpus anotado como *Información sobre recursos* en el que se incluyen los tres contenidos prototípicos de este apartado, es decir, *Fórmula de introducción de los recursos posibles*, *Recursos que cabe interponer* e *Información sobre la posibilidad de interponer otros recursos*:

- (7) Contra la expresada resolución, que es definitiva en vía administrativa, podrá interponerse cualquiera de los recursos que se indican a continuación:

I.- Recurso ordinario de reposición con carácter potestativo, ante el mismo órgano que adoptó la resolución, en el término de un mes al ser este un acto expreso, contado a partir del día siguiente al de recibir esta notificación (artículo 124 de la Ley 39/2015, de 1 de octubre, del Procedimiento Administrativo Común de las Administraciones Públicas).

Transcurrido un mes desde la interposición del recurso sin que se dicte y notifique su resolución, se entenderá desestimado por silencio administrativo, pudiendo en este caso interponer contra la desestimación presunta recurso contencioso-administrativo dentro del plazo de seis meses (artículo 46 de la Ley 29/1998, de 13 de julio, Reguladora de la Jurisdicción Contencioso- Administrativa).

II.- Recurso contencioso-administrativo ante el Juzgado de lo Contencioso-Administrativo, en cuya circunscripción tenga su domicilio el demandante o se halle la sede del órgano autor

del acto impugnado, a elección de aquel, en el plazo de dos meses contados desde el día siguiente al de la notificación del acto que ponga fin a la vía administrativa (artículos 8.1, 14.1.2a, 46 de la Ley 29/1998).

Todo ello sin perjuicio de que pueda ejercitar, en su caso, cualquier otro que estime procedente.

En cuanto al apartado *Cierre*, incluye prototípicamente los siguientes contenidos: *Localidad del emisor*, *Fecha*, y *Cargo y nombre del emisor*. Estos contenidos aparecen en el corpus con una frecuencia muy alta. Es necesario precisar que en el caso de las resoluciones publicadas en portales institucionales no se requiere firma electrónica. Tal y como indica el artículo 26 de la Ley 39/2015, para que los documentos públicos administrativos emitidos a través de medios electrónicos sean considerados válidos deben incorporar las firmas electrónicas que correspondan según la normativa aplicable. No obstante, el mismo artículo en su apartado tercero indica que: “No requerirán de firma electrónica los documentos electrónicos emitidos por las Administraciones Públicas que se publiquen con carácter meramente informativo, así como aquellos que no formen parte de un expediente administrativo. En todo caso, será necesario identificar el origen de estos documentos” (artículo 26.3, L 39/2015, de 1 de octubre). En consecuencia, las resoluciones que se publican para conocimiento general de la ciudadanía deben indicar el cargo del emisor, en el que se incluye la información sobre el área, organismo o institución que dicta la resolución. Por ejemplo: *Director General de la Policía Municipal*. En el ejemplo (8) se recoge un fragmento correspondiente al apartado *Cierre* con sus tres contenidos prototípicos:

- (8) Madrid, a 18 de noviembre de 2021.- El Gerente del Organismo Autónomo Madrid Salud, Antonio Prieto Fernández.

El último apartado que se incluye en la estructura modelo de la tabla 1 es *Anexos*. Este apartado se recoge con la marca [*apartado opcional*], ya que se encuentra en el umbral de entre el 40 % y el 50 % de los textos del corpus. Dentro de *Anexos*, se incluyen dos contenidos. El primero de ellos, *Título del anexo*, supera el umbral del 40 % de frecuencia en el corpus. Sin embargo, los posibles contenidos correspondientes a cada anexo no superan el umbral del 40 % de aparición. En el corpus, se anotaron los siguientes contenidos en relación con los anexos: *formulario de solicitud* (15 %), *lista: datos relacionados con la resolución* (10 %), *regulación de aspectos concretos de la resolución* (5 %) y *tabla: datos relacionados con la resolución* (30 %). En cuanto a esto, se toma la decisión de incluir en la estructura una breve descripción de los posibles contenidos de los anexos que se detectan en la anotación del corpus, que es la que se puede ver en la tabla 1.

## 4.2. Rasgos léxicos: fraseología prototípica

En este trabajo también se recopilan unidades fraseológicas del corpus que se utilizan para expresar los contenidos detectados en los diferentes apartados de la resolución publicada en portales institucionales. Tal y como se explica en el apartado 4, se recogen tanto oraciones completas como fragmentos de oraciones. En el ejemplo (9) se muestra una oración completa seleccionada para expresar el contenido *Texto de la decisión sobre la publicación de*



la resolución en el portal institucional/Boletín Oficial/sede electrónica. En el ejemplo (10) se ofrece un fragmento de oración relacionada con el mismo contenido:

- (9) Publicar en el Boletín Oficial del Ayuntamiento de Madrid la presente resolución.
- (10) y se publicará en el Boletín Oficial de la Comunidad de Madrid y del Ayuntamiento de Madrid y electrónicamente en el portal web municipal.

Asimismo, se detectan frases invariables y otras que incluyen algún tipo de información variable. En este segundo caso, la información variable se marca entre corchetes. En el ejemplo (11) se muestra una frase sin información variable, mientras que en el ejemplo (12) se recoge una frase con información variable entre corchetes. Ambos ejemplos se utilizan para expresar el contenido *Antecedentes*:

- (11) una vez acreditada la observancia del procedimiento debido, así como el cumplimiento por parte de los candidatos elegidos de los requisitos y especificaciones exigidos en la convocatoria
- (12) Concluidos los procedimientos previstos y a la vista de la propuesta de resolución del Concurso realizada por la Comisión de Valoración nombrada al efecto conforme a la base [nº] de las Bases Generales de convocatoria de [temática de la convocatoria], constatándose que se han llevado a efecto los procesos previstos en las bases y disposiciones de aplicación supletoria,

Además, si se observan cuestiones de formato o aspectos lingüísticos llamativos, se aporta una indicación entre paréntesis al final de la frase, como se puede ver en el ejemplo (13). Este ejemplo se asocia al contenido *Recursos que cabe interponer*:

- (13) Recurso potestativo de reposición, en el Registro [tipo de registro], ante el mismo órgano que dictó la resolución. (si se pueden interponer varios recursos, se recomienda indicarlos en una lista)

Finalmente, se seleccionan 46 unidades fraseológicas representativas para expresar los contenidos del género objeto de estudio. Estas unidades originales extraídas del corpus se adaptan a las recomendaciones más frecuentes de lenguaje claro (Da Cunha y Escobar 2021).

Las recomendaciones de lenguaje claro a nivel discursivo que se emplean para la adaptación de la fraseología son: redactar oraciones cortas, redactar párrafos breves, utilizar conectores discursivos para enlazar y ordenar las ideas e incluir un solo tema en cada párrafo. En el ejemplo (14) se muestra una unidad fraseológica original extraída del corpus con cuestiones que no siguen las recomendaciones de lenguaje claro a nivel discursivo:

- (14) Recurso contencioso-administrativo, ante los Juzgados de lo Contencioso-Administrativo de conformidad con lo dispuesto en el artículo 8.1 de la Ley 29/1998, de 13 de julio, Reguladora de la Jurisdicción Contencioso-Administrativa, siendo el plazo para su interposición el de dos meses contados desde el día siguiente a la fecha de su publicación en el Boletín Oficial de la Comunidad de Madrid, de conformidad con lo dispuesto en el artículo 46.1 de la Ley 29/1998, de 13 de julio.

El fragmento del ejemplo anterior constituye un párrafo formado por una única oración de 71 palabras. Para adaptarlo a lenguaje claro, se segmenta en tres oraciones más bre-

ves, independientes y con un sentido completo. Estas tres oraciones se muestran en los ejemplos (15), (16) y (17):

- (15) Recurso contencioso-administrativo, ante los Juzgados de lo Contencioso-Administrativo. (si se pueden interponer varios recursos, se recomienda indicarlos en una lista)
- (16) El plazo para interponer el recurso contencioso-administrativo es de dos meses desde el día siguiente a la fecha en que se publica la resolución en el Boletín Oficial del/de la [entidad territorial].
- (17) Los artículos 8 y 46 de la Ley 29/1998, de 13 de julio, Reguladora de la Jurisdicción Contencioso-Administrativa regulan los requisitos y plazos del recurso contencioso-administrativo.

Además, se detectan en el corpus fragmentos de oraciones que se utilizan para expresar un contenido concreto, pero que forman parte de oraciones más largas en las que también se hace referencia a otros contenidos. Estos fragmentos se modifican para convertirlos, cuando sea posible, en oraciones completas. En el ejemplo (18) se observa un fragmento de oración del corpus y en el ejemplo (19) la oración completa que se crea con base en la anterior:

- (18) y se publicará en el Boletín Oficial de la Comunidad de Madrid y del Ayuntamiento de Madrid y electrónicamente en el portal web municipal.
- (19) Esta resolución se publicará en el Boletín Oficial del/de la [entidad territorial] y en el portal web [estatal/autonómico/municipal/etc.].

Las recomendaciones de lenguaje claro a nivel morfosintáctico que se emplean para la adaptación de la fraseología son: utilizar la voz activa en vez de la voz pasiva, utilizar la estructura *sujeto + verbo + complemento*, evitar el abuso de oraciones subordinadas, evitar los verbos en gerundio, formular las ideas en positivo en vez de en negativo y mencionar el sujeto de las acciones. En el ejemplo (20) se muestra una unidad fraseológica original extraída del corpus:

- (20) Transcurrido un mes desde la interposición del recurso sin que se dicte y notifique su resolución, se entenderá desestimado por silencio administrativo, pudiendo en este caso interponer contra la desestimación presunta recurso contencioso-administrativo dentro del plazo de seis meses.

En el ejemplo (20), se detecta que en la oración no se menciona el sujeto. Además, se utiliza un verbo en gerundio (*pudiendo*). Finalmente, en los ejemplos (21) y (22) se muestra el resultado de la adaptación a lenguaje claro teniendo en cuenta, entre otras, estas recomendaciones morfosintácticas:

- (21) Si la persona interesada interpone el recurso y al cabo de un mes no se ha dictado y notificado la resolución, se entenderá que se desestima por silencio administrativo.
- (22) Si el órgano administrativo desestima un recurso por silencio administrativo, la persona interesada puede interponer un recurso contencioso-administrativo dentro del plazo de seis meses.

En cuanto a las recomendaciones de lenguaje claro a nivel léxico, en este trabajo se tienen en cuenta las siguientes: utilizar palabras comunes en vez de términos en caso de ser posible, explicar el significado de los términos si es necesario utilizarlos, utilizar palabras precisas en vez de palabras ambiguas, eliminar palabras innecesarias, evitar los arcaísmos, y utilizar palabras concretas en vez de abstractas. En el ejemplo (23) se incluye una frase original extraída del corpus en la que se incluyen expresiones que podrían ser consideradas arcaicas y que tienen una variante más sencilla y más actual (*de conformidad con y referente a*) y palabras innecesarias (*lo dispuesto en*):

- (23) De conformidad con lo dispuesto en el artículo 4.4 de las bases generales para 2021, sobre la forma en la que debe difundirse y publicarse la información referente a las Ayudas de Acción Social,

Las expresiones se cambian por otras más sencillas (*según y sobre*, respectivamente). Además, se eliminan las palabras innecesarias. Estas modificaciones, entre otras, dan lugar a la frase adaptada a lenguaje claro del ejemplo (24):

- (24) Según el artículo [nº] del/de la [normativa aplicable: ley/reglamento/acuerdo/etc.], sobre la forma en la que debe difundirse y publicarse la información sobre [información relacionada con la resolución],

El resultado final de este análisis es una lista de fraseología adaptada a lenguaje claro que se puede utilizar para expresar los contenidos prototípicos de la resolución publicada en portales institucionales. En el anexo I se puede ver una tabla con la estructura modelo y la fraseología en lenguaje claro de la resolución.

### 4.3. Rasgos digitales: particularidades de la difusión por medios electrónicos

Las resoluciones administrativas que deben ponerse en conocimiento de un colectivo o de la ciudadanía en general se publicaban tradicionalmente a través de Boletines Oficiales y tabloneros de anuncios en papel. En la actualidad, este tipo de textos se ha trasladado al medio digital gracias a los Boletines Oficiales en formato electrónico, los portales web institucionales y las sedes electrónicas. De esta forma, tal y como indica la Ley 39/2015: “La publicación de los diarios o boletines oficiales en las sedes electrónicas de la Administración, Órgano, Organismo público o Entidad competente tendrá, en las condiciones y con las garantías que cada Administración Pública determine, los mismos efectos que los atribuidos a su edición impresa” (artículo 131, L 39/2015, de 1 de octubre). En el caso del Ayuntamiento de Madrid, estas resoluciones se publican en el BOAM, que en la actualidad está disponible únicamente en formato digital.

Para acceder al BOAM, el usuario debe entrar en la página web <https://sede.madrid.es/>, acceder a la pestaña *Publicaciones Oficiales* y, dentro de esta, a *Boletín Oficial del Ayuntamiento*. Una vez hecho eso, se muestra un buscador que permite localizar documentos por fecha, y por palabras en el texto y en el título. Asimismo, ofrece unos filtros avanzados para ajustar los resultados de la búsqueda en función del número de boletín, la fecha de publicación, el tipo de contenido y el número asignado al documento. Para compilar el corpus de este trabajo se empleó este buscador y se introdujeron los siguientes criterios de búsqueda:

- Mostrar los documentos que contengan la palabra “resolución” en el título.
- Limitar los resultados a los boletines publicados entre el 1 de enero de 2017 y el 31 de diciembre de 2021.

En la figura 3 se muestra una captura de pantalla del buscador del BOAM. Se han introducido los criterios de búsqueda que se mencionan arriba.

The screenshot shows the search interface of the Boletín Oficial del Ayuntamiento de Madrid (BOAM). The search criteria are: "Buscar en texto" (Search in text), "Buscar en título" (Search in title) with the keyword "resolución", and "Ir al BOAM de fecha (dó/m/a/a/a)" (Go to BOAM by date) with the date range "01/01/2017" to "31/12/2021". The interface also shows a calendar for April 2022 and a "Buscar BOAM" button.

Figura 3. Buscador del BOAM.

Cuando se pincha en un documento publicado en el BOAM, aparece el texto en HTML. Además, se ofrece la versión descargable en PDF tanto del documento como del BOAM completo correspondiente a ese día. Los anexos también se muestran en formato descargable en PDF.

Tras el análisis del corpus de resoluciones publicadas en el BOAM se observan similitudes y diferencias entre la versión en HTML y la versión en PDF. En cuanto a las similitudes, se determina que el texto de ambas versiones es idéntico. Asimismo, en ambas versiones se incluye la posibilidad de abrir una URL si se pincha sobre ella, lo cual es un rasgo propio de los textos digitales. Todas las URL que se incorporan en los textos del corpus enlazan con portales web del propio Ayuntamiento de Madrid. Por lo tanto, se excluyen los enlaces directos a webs no institucionales o de instituciones públicas distintas a la que dicta la resolución. No siempre se incorpora esta posibilidad de hacer click sobre las URL del texto, aun cuando hacen referencia a portales del Ayuntamiento de Madrid. De esto se deduce que queda a discreción del personal público incluir o no este tipo de elementos. En el ejemplo (25) se incluye un fragmento original del corpus en el que se incluye la URL a la intranet municipal *Ayre*, mientras que en el ejemplo (26) se hace referencia a dicha intranet sin incorporar la URL.

- (25) TERCERO.- Por el órgano competente se procederá a la publicación de la presente resolución en el Boletín Oficial del Ayuntamiento de Madrid, en la intranet/extranet municipal ayre (<https://ayre.madrid.es>) y en la web <https://jubilacion.madrid.es>.
- (26) SEGUNDO.- Hacer pública la adopción de la Resolución del Gerente del Organismo Autónomo de Madrid Salud, por la que se procede a la concesión, denegación y desistimiento de ayudas a la promoción interna del personal, del año 2021, en el Boletín Oficial del Ayuntamiento de Madrid, en la intranet municipal “Ayre/Área de actividades/Salud/Personas/Acción Social”.

En relación con las diferencias, la más destacable es que la versión en PDF incorpora lo que se denomina *Código Seguro de Verificación* (CSV). Este código se incluye en los documentos electrónicos generados por el Ayuntamiento de Madrid y permite comprobar la autenticidad del documento. El motivo por el cual este código no se incluye en la versión HTML es que se presume la autenticidad del texto por estar integrado en el portal institucional. En cambio, la versión descargable en PDF permite al usuario imprimir el documento o guardarlo en local, por lo que se desvincula de la institución de origen y requiere de una verificación de su autenticidad.

En los textos analizados, este código se muestra a través de tres elementos: una URL, un código alfanumérico y un código QR. Estos elementos se encuentran en el margen inferior derecho de todas las páginas del documento. Para verificar una resolución del Ayuntamiento de Madrid con CSV, es necesario acceder al Servicio de Verificación de Documentos Electrónicos ([https://csv.madrid.es/VECSV\\_WBCONSULTA/VerificarCove.do](https://csv.madrid.es/VECSV_WBCONSULTA/VerificarCove.do)). El interesado puede acceder fácilmente al servicio en línea a través de la URL del documento o del código QR. La URL será más útil si el fichero se encuentra guardado en local, mientras que el código QR se utilizará en el caso de que el PDF se imprima. Una vez se accede al sistema, es necesario introducir el código alfanumérico que aparece en el documento. Como resultado, se obtiene una copia electrónica del texto que se puede cotejar con la que ya tenía previamente el usuario. Esto permite verificar la coincidencia de los datos.

Este tipo de códigos de verificación también se incluye en otros documentos oficiales. A modo de ejemplo, en el BOE se utiliza el denominado *Código de Verificación Electrónica* (CVE). Tiene un funcionamiento muy similar al CSV que emplea el Ayuntamiento de Madrid. Se muestra en los documentos en papel y en los descargables en PDF en el margen inferior derecho de todas las páginas mediante un código alfanumérico y una URL. En este caso, no se incluye el código QR. Para contrastar la autenticidad de un documento del BOE, hay que acceder a [https://www.boe.es/diario\\_boe/](https://www.boe.es/diario_boe/). Una vez dentro, en el apartado *Verificar documentos mediante CVE* se inserta el código alfanumérico correspondiente y se obtiene una copia electrónica del texto.

## 5. CONCLUSIONES

Como se ha visto, en los últimos años se ha transformado la manera en la que se comunican la Administración y la ciudadanía, debido a la aparición de la Administración electrónica y de los portales web institucionales. Asimismo, se está promoviendo el uso de un lenguaje claro en este ámbito.

Nuestro trabajo se centra en el estudio de un género textual del ámbito especializado de la Administración, la resolución, que en la actualidad se publica principalmente a través de

portales web institucionales. De esta forma, los objetivos generales que se han planteado son: determinar la estructura y contenidos prototípicos de este género digital, extraer fraseología que se pueda utilizar para expresar esos contenidos y adaptarla a lenguaje claro, y analizar los rasgos digitales de la resolución cuando se publica por medios electrónicos. Para la consecución de estos objetivos, se toma como caso de estudio el Ayuntamiento de Madrid y se recopila un corpus de 20 textos extraídos del portal web de su Boletín Oficial.

En cuanto al primer objetivo, tras la anotación y el análisis del corpus se obtiene una estructura modelo de la resolución publicada en portales institucionales, con indicación de los apartados, títulos y contenidos prototípicos. En trabajos previos de Da Cunha y Montané (2020) y Da Cunha et al. (2021) también se obtienen estructuras modelo de géneros textuales especializados. En el primer caso, se establece la estructura prototípica de cinco géneros del ámbito de la Administración redactados por la ciudadanía y dirigidos a la Administración. Por lo tanto, en nuestro estudio se invierte el emisor y el receptor de los textos objeto de estudio. En el segundo caso, se determinan los apartados, títulos y contenidos de tres géneros del ámbito del turismo. En ninguno de estos estudios se hace referencia a la digitalización de los textos, a pesar de que todos ellos se pueden remitir al destinatario a través de medios electrónicos y de que algunos, como la entrada en blog de viajero, surge en el entorno de internet.

En relación con el segundo objetivo, se seleccionan una serie de unidades fraseológicas que se pueden usar para expresar los contenidos que se detectan en el análisis textual. Esa fraseología original extraída del corpus se adapta a las recomendaciones más frecuentes de lenguaje claro en español (Da Cunha y Escobar 2021). En relación con el lenguaje claro, se publican multitud de estudios en los últimos años de tipo descriptivo y en los que se ofrecen recomendaciones para hacer que los textos jurídico-administrativos sean más fáciles de entender para la ciudadanía (Carretero 2019; Montolío y Tascón 2020; Jiménez y Sancho 2021). Menos frecuentes son las investigaciones en las que se ofrecen adaptaciones a lenguaje claro de textos especializados dirigidos a la población general. En el ámbito jurídico, Carretero (2022) ofrece una versión alternativa de las actas de información de derechos a los detenidos. Por su parte, en el ámbito puramente administrativo, destaca la adaptación de la multa de tráfico llevada a cabo en el Ayuntamiento de Madrid<sup>2</sup>. También es relevante la tesis doctoral de Bayés (2021), en la que ofrece varias versiones clarificadas de un documento administrativo. En cada una de estas versiones aplica una recomendación de lenguaje claro y evalúa empíricamente su impacto en la percepción de claridad de estudiantes universitarios.

En relación con el primer y segundo objetivos, cabe destacar el uso que se ha hecho de las posibilidades que ofrece la informática en el estudio de la lingüística (González Fernández 2017), concretamente, para la compilación y el análisis del corpus. Primero, el corpus se ha obtenido en línea y de forma sencilla gracias a la publicación de los textos en abierto. Segundo, se ha hecho uso de una herramienta tecnológica para anotar el corpus, ATLAS.ti.

Los resultados relativos al tercer objetivo muestran que el género *resolución publicada en portales institucionales* es un género digital en la medida en que se trata de un texto preexistente que se ha adaptado a las condiciones del medio digital y se ha aprovechado

---

2 Se puede consultar más información en la web del Ayuntamiento: <https://www.madrid.es/portales/munimadrid/es/Inicio/Cultura-ocio-y-deporte/Actividades-y-eventos/Comunicacion-Clara/?vgnextfmt=default&vgnextoid=a01f1905bacde510VgnVCM1000001d4a900aRCRD&vgnnextchannel=6381f073808fe410VgnVCM2000000c205a0aRCRD&idCapitulo=10559046> (23-05-2022).

de las posibilidades técnicas que ofrece la red. Sin embargo, se determina que los rasgos digitales introducidos en estos textos no suponen cambios sustanciales con respecto a las versiones tradicionales publicadas en papel. En este sentido, se reafirma la idea de que la Administración electrónica supone en muchos casos un mero cambio en el canal de comunicación (Toscano 2019).

Como propuesta de mejora de cara a la digitalización de los textos, sugerimos ampliar el uso de las URL, que permiten al lector obtener más información sobre cuestiones que se citan en la resolución. A modo de ejemplo, podría enlazarse la normativa aplicable, de forma que el usuario pueda acceder a ella fácilmente si está leyendo el texto desde un dispositivo electrónico. También se podrían incorporar recursos gráficos que simplifiquen la información y que contribuyan a que el texto sea más claro (Prodigioso Volcán 2021). Además, destacamos la necesidad de utilizar un lenguaje claro en los textos administrativos publicados en la web, ya que el uso de este tipo de lenguaje contribuye a que la lectura por medios digitales sea más ágil y sencilla (García-Cabrera 2018). Asimismo, tal y como se indica en el informe de Prodigioso Volcán (2020), el uso de un lenguaje claro contribuye a que se obtengan traducciones automáticas más adecuadas. En el contexto administrativo, esto podría ser útil para las personas extranjeras, que obtendrán mejores resultados si traducen a su lengua materna los textos que reciben de la Administración a través de un traductor automático en internet. En ese mismo informe se indica que el lenguaje claro también contribuye al mejor funcionamiento de sistemas de procesamiento de lenguaje, como los lectores de pantalla. Por lo tanto, las personas con discapacidad visual que tengan que hacer uso de este tipo de lectores automáticos también se verían beneficiadas.

Actualmente, los resultados de esta investigación se están utilizando en un proyecto de investigación relacionado con el uso de la tecnología en la escritura. En el marco de ese proyecto se desarrolla en la actualidad arText, una herramienta de ayuda a la redacción automática de textos en español que ayuda a redactar géneros textuales de ámbitos especializados y textos administrativos en lenguaje claro. El sistema arText está disponible en línea de forma gratuita en el siguiente enlace: <http://sistema-artext.com/>. Los resultados de esta investigación se incorporarán al sistema arText con el objetivo de aportar al usuario información sobre la estructura, los contenidos y la fraseología del género *resolución publicada en portales institucionales*.

## FUENTES DE FINANCIACIÓN DEL TRABAJO

Este trabajo se deriva de un contrato de formación predoctoral FPI, financiado por el Ministerio de Ciencia e Innovación en la convocatoria 2019 del Subprograma Estatal de Formación del Programa Estatal de Promoción del Talento y su Empleabilidad en I+D+i, en el marco del Plan Estatal de Investigación Científica y Técnica y de Innovación 2017-2020 (referencia predoctoral PRE2019-088419). El contrato fue concedido a Sara Pistola Grille, que lleva a cabo su tesis doctoral como investigadora en formación en la Escuela Internacional de Doctorado de la Universidad Nacional de Educación a Distancia (UNED), “Programa de Doctorado en Filología. Estudios Lingüísticos y Literarios: teoría y aplicaciones”, Departamento de Filologías Extranjeras y sus Lingüísticas, Facultad de Filología, UNED. El contrato está asociado al proyecto de investigación titulado “Tecnologías de la Información y la Comunicación para la e-Administración: hacia la mejora de la comuni-

cación entre Administración y ciudadanía a través del lenguaje claro (TIC- eADMIN)”, financiado por el Ministerio de Ciencia, Innovación y Universidades en la convocatoria 2018 de Proyectos I+D del Subprograma Estatal de Generación de Conocimiento (referencia PGC2018-099694-A-I00), y desarrollado en el Departamento de Filologías Extranjeras y sus Lingüísticas de la Facultad de Filología de la UNED, en el marco del grupo de investigación ACTUALing, con la colaboración del grupo IULATERM (IULA-UPF). La investigadora principal del proyecto es Iria da Cunha. En el marco de este proyecto se ha firmado un “Protocolo de colaboración en investigación científica en materia de comunicación clara” con el Ayuntamiento de Madrid, a quien agradecemos su colaboración. También agradecemos a Susana Viñuales su ayuda en la revisión de los resultados de la presente investigación.

## REFERENCIAS BIBLIOGRÁFICAS

- Araguàs, I. (2012). La Administración electrónica en España: de la «administración en papel» a la «e-administración». *Revista chilena de derecho y ciencia política*, 3(2), 109-139. <https://portalrevistas.uct.cl/index.php/RDCP/article/view/425>
- Bayés, M. (2021). Análisis del impacto de una selección de (meta)indicaciones de redacción clara en la percepción de claridad de un documento administrativo: estudio de caso [Tesis doctoral, Universitat de Barcelona]. <https://www.tdx.cat/handle/10803/672521#page=1> (23-05-2022)
- Carretero, C. (2019). *Comunicación para juristas*. Tirant lo Blanch.
- Carretero, C. (2022). Personas detenidas y la comprensión de sus derechos con base en la Directiva 2012/13/UE. *Revista General de Derecho Procesal*, 56, 1-51. <http://hdl.handle.net/11531/65255>
- Castellón, H. (2000). *Los textos administrativos*. Arco/Libros, S.L.
- Castellón, H. (2001). *El lenguaje administrativo: formas y uso*. Editorial La Vela.
- Cervera, Á. y J. Torres. (2016). Estrategias lingüísticas en el interrogatorio judicial español: una aproximación pragmalingüística. *ELUA*, 30, 61-78. <https://doi.org/10.14198/ELUA2016.30.03>
- Cotino, L. (2017). La regulación del uso de medios electrónicos en la difusión activa de información pública y el ejercicio del derecho de acceso. En I. Martín (Dir.), *La reforma de la Administración electrónica: Una oportunidad para la innovación desde el Derecho* (pp. 398-433). Instituto Nacional de Administración Pública (INAP).
- Da Cunha, I. y M. A. Montané. (2020). A corpus-based analysis of textual genres in the administration domain. *Discourse Studies*, 22(1), 3-31. <https://doi.org/10.1177/1461445619887538> (23-05-2022)
- Da Cunha, I. y M. A. Escobar. (2021). Recomendaciones sobre lenguaje claro en español en el ámbito jurídico-administrativo: análisis y clasificación. *Pragmalingüística*, 29, 129-148. <https://doi.org/10.25267/Pragmalinguistica.2021.i29.07> (23-05-2022)
- Da Cunha, I., M. A. Montané, B. Fisas y M. Á. Escobar. (2021). Comparing textual genres in Spanish: the case of the tourism domain. *Ibérica*, 42, 163-190. <https://revistaiberica.org/index.php/iberica/article/view/603/193> (23-05-2022)
- De Miguel, E. (2000). El texto jurídico-administrativo: análisis de una orden ministerial. *Círculo de lingüística aplicada a la comunicación*, 4. <https://webs.ucm.es/info/circulo/no4/demiguel.htm> (23-05-2022)
- Fernández Sánchez, C. (2003). Análisis de géneros digitales y contenido visual en la web de bancos tradicionales y nuevos bancos virtuales. *Ibérica*, 6, 39-63. <https://www.redalyc.org/articulo.oa?id=287026395011>
- Fernández-Vallejo, A. M. (2017). La disculpa corporativa a través de Twitter en Iberia y Banco Santander. *ELUA*, 31, 151-170. <https://doi.org/10.14198/ELUA2017.31.08>
- García-Cabrera, L. (2018). Metodología remediavagos para escribir textos eficientes para la web: 20 buenas prácticas. En R. Marfil, S. Osuna y P. González (Coords.). *Innovación y esfuerzo investigador en la educación mediática contemporánea* (pp. 25-38). Ediciones Egregius.



- González Fernández, A. (2017). Estudio de neologismos a través de big data en un corpus textual extraído de Twitter. *ELUA*, 31, 171-186. <https://doi.org/10.14198/ELUA2017.31.09>
- González García, V. (2013). El discurso del turismo en Internet: hacia una caracterización de sus géneros. En J. Sanmartín (Ed.), *Discurso turístico e Internet* (pp. 13-49). Iberoamericana Vervuert.
- Jiménez, R. y I. Sancho. (2021). Resoluciones judiciales correctas, claras y precisas. *InDret*, 4, 432-452. <https://indret.com/resoluciones-judiciales-correctas-claras-y-precisas/>
- Lorente, M. (2016). La lingüística de corpus para el estudio del discurso especializado de los ámbitos jurídico y administrativo. En M. Eurrutia (coord.), *El lenguaje jurídico y administrativo en el ámbito de la extranjería. Estudio multilingüe e implicaciones culturales* (pp. 167-193). Peter Lang.
- Meza, P., F. González-Catalán, C. López-Ferrero e I. Gutiérrez. (2020). Plain writing in the legal field: an approach from the discourse of specialists. *Discourse Studies*, 22(3), 356-383. <https://doi.org/10.1177/1461445620906027> (23-05-2022)
- Ministerio para las Administraciones Públicas (MAP). (2003). *Manual de documentos administrativos* (3ª ed.). Editorial Tecnos.
- Montolio, E. (dir.). (2011). *Estudio de campo: lenguaje escrito. Comisión para la modernización del lenguaje jurídico*. Ministerio de Justicia. <https://lenguajeadministrativo.com/wp-content/uploads/2015/10/CMLJ-Lenguaje-escrito.pdf> (23-05-2022)
- Montolio, E. y M. Tascón. (2020). *El derecho a entender: la comunicación clara, la mejor defensa de la ciudadanía*. Prodigioso Volcán y Catarata.
- Pedraza, M. B. (2019). *Democratización de la lengua de la administración pública: retos de una propuesta lexicográfica digital panhispánica*. [Tesis doctoral, Universitat Autònoma de Barcelona (España)]. Tesis Doctorals en Xarxa. <https://www.tdx.cat/handle/10803/667363#page=1> (23-05-2022)
- Pérez-Llantada, C. (2021). Grammar features and discourse style in digital genres: The case of science-focused crowdfunding projects. *Revista Signos. Estudios de Lingüística*, 54(105), 73-96. <http://www.revistasignos.cl/index.php/signos/article/view/363/267> (23-05-2022)
- Pistola, S. y S. Viñuales-Ferreiro. (2021). Una clasificación actualizada de los géneros textuales de la Administración pública española. *Revista de Lengua i Dret*, 75, 181-203. <https://doi.org/10.2436/rld.i75.2021.3587>
- Prodigioso Volcán. (2020). *¿Habla claro la Administración pública?* Prodigioso Volcán. <https://comunicacionclara.com/lenguaje-claro/prodigioso-volcan-habla-claro-la-administracion.pdf> (23-05-2022)
- Prodigioso Volcán. (2021). *¿Habla claro la Administración a los públicos vulnerables?* Prodigioso Volcán. <https://comunicacionclara.com/habla-claro-la-administracion/> (23-05-2022)
- Sánchez Acevedo, M. E. (2015). *Derecho a la buena Administración electrónica* [Tesis doctoral, Universidad de Valencia]. <http://hdl.handle.net/10550/50882> (23-05-2022)
- Swales, J. M. (1990). *Genre analysis: English in academic and research settings* (1ª ed.). Cambridge University Press.
- Toscano, F. (2019). Análisis de las transformaciones actuales del derecho administrativo en España. *Revista Digital De Derecho Administrativo*, 22, 337-363. <https://doi.org/10.18601/21452946.n22.13>
- Upton, T. A. y M. A. Cohen. (2009). An approach to corpus-based discourse analysis: The move analysis as example. *Discourse Studies*, 11(5), 585-605. <https://doi.org/10.1177/1461445609341006>
- Van Dijk, T. A. (1977). *Text and context: Explorations in the semantics and pragmatics of discourse* (1ª ed.). Longman.
- Van Dijk, T. A. (1989). *La ciencia del texto: un enfoque interdisciplinario* (1ª ed.). Paidós Comunicación.
- Vargas-Sierra, C. (2005). Diseño de un corpus especializado con fines terminográficos: el corpus de la piedra natural. *Debate terminológico*, 2, 1-20. <http://hdl.handle.net/10045/9426>

## LEGISLACIÓN

España. Ley 39/2015, de 1 de octubre, del Procedimiento Administrativo Común de las Administraciones Públicas [Internet]. BOE, de 2 de octubre de 2015, núm. 236, pp. 89343 a 89410. [Última consulta: 23/05/2022]. Disponible en: <https://www.boe.es/eli/es/l/2015/10/01/39/con>

España. Ley 40/2015, de 1 de octubre, de Régimen Jurídico del Sector Público [Internet]. BOE, de 2 de octubre de 2015, núm. 236, pp. 89411 a 89530. [Última consulta: 23/05/2022]. Disponible en: <https://www.boe.es/eli/es/l/2015/10/01/40/con>

## ANEXO I

Fraseología en lenguaje claro para expresar los contenidos prototípicos de la resolución publicada en portales institucionales

Apartados	Contenidos	Fraseología
Cabecera	Título principal del documento	Resolución de [día] de [mes] de [año] del [cargo del emisor] del [área/órgano/institución] por la que se [resuelve/convoca/establece/aprueba/da publicidad/concede/deniega/etc.] el/la [objeto de la resolución].
Cuerpo	Marco jurídico	Según los artículos [nº] y siguientes del/de la [normativa aplicable: ley/reglamento/etc.], de [fecha], en adelante, [sigla por la que se conoce la normativa aplicable],
		El/La [normativa aplicable: ley/reglamento/etc.] (en adelante, [sigla por la que se conoce la normativa aplicable]) señala en su artículo [nº] que “[cita textual]”.
		El/La [artículo o parte de la norma] de la [normativa aplicable: ley/reglamento/etc.], que entró en vigor el [fecha], establece
		La normativa vigente sobre [temática de la norma] establece
		El [órgano colegiado], en sesión ordinaria de [fecha], aprobó el [reglamento/acuerdo/etc.] por el que
		La convocatoria establece en el apartado [nº] que: “[cita textual]”.
	Antecedentes	Se ha concluido el procedimiento y la Comisión de Valoración ha dictado la propuesta de resolución del concurso.
		La Comisión de Valoración se nombró conforme a la base [nº] de las Bases Generales de la convocatoria de [temática de la convocatoria].
		Se confirma que se han llevado a cabo los procesos que se prevén en las bases y en las disposiciones supletorias.
		Se ha acreditado que el procedimiento se ha llevado a cabo adecuadamente.
	Se ha comprobado que los candidatos que se han elegido cumplen los requisitos que exige la convocatoria.	
	Se han obtenido los datos para formular la propuesta provisional sobre las entidades que se han admitido y excluido de la convocatoria pública para solicitar el/la [objeto de la solicitud].	
	Se han revisado las solicitudes de [objeto de la solicitud] para [colectivo de solicitantes] relacionadas con este expediente.	

Apartados	Contenidos	Fraseología
		Se ha comprobado que los solicitantes no cumplen los requisitos para que se les conceda el/la [objeto de la solicitud].
		Las solicitudes se han denegado por los motivos que se indican en esta resolución o porque se ha superado el gasto máximo anual que se fijó para el año [año].
	Objetivo de la resolución	Por lo tanto, se considera necesario dictar esta resolución para [propósito de la resolución].
	Base jurídica que atribuye la competencia para dictar la resolución	Según todo lo anterior, y conforme a las competencias que se atribuyen al [cargo del emisor] en el apartado [nº] del/de la [base jurídica en la que se atribuye la competencia: ley/reglamento/acuerdo/estatutos/etc.]
		De acuerdo con las facultades que me otorga el/la [base jurídica en la que se atribuye la competencia: ley/reglamento/acuerdo/estatutos/etc.],
		De acuerdo con las facultades que se recogen en el artículo [nº] del/de la [base jurídica en la que se atribuye la competencia: ley/reglamento/acuerdo/estatutos/etc.], que se modifica en el/la [base jurídica en la que se atribuye la competencia: ley/reglamento/acuerdo/estatutos/etc.],
		Por tanto, y según las competencias que se otorgan en el apartado [nº] del/de la [base jurídica en la que se atribuye la competencia: ley/reglamento/acuerdo/estatutos/etc.], así como en el/la [base jurídica en la que se atribuye la competencia: ley/reglamento/acuerdo/estatutos/etc.],
		Según la base [nº] de la convocatoria, como órgano instructor,
Decisión	Título del apartado	RESUELVO
	Título de la decisión sobre la temática de la resolución (puede haber varias)	[PRIMERO/SEGUNDO/TERCERO/etc.]. [PRIMERO/SEGUNDO/TERCERO/etc.]. [Breve indicación del tema de la decisión].
	Texto de la decisión sobre la temática de la resolución (puede haber varias)	[Adjudicar/Convocar/Acordar/Aprobar/Publicar/Conceder/Acumular/Subsanar/etc.] [texto de la decisión] (siempre empieza por un verbo en infinitivo)

Apartados	Contenidos	Fraseología
	Título de la decisión sobre la publicación de la resolución en el portal institucional/ Boletín Oficial/sede electrónica [título opcional]	[SEGUNDO/TERCERO/CUARTO/etc.]. [SEGUNDO/TERCERO/CUARTO/etc.]. [Breve indicación del tema de la decisión].
	Texto de la decisión sobre la publicación de la resolución en el portal institucional/ Boletín Oficial/sede electrónica [contenido opcional]	Publicar esta resolución en el Boletín Oficial del/de la [entidad territorial]. Esta resolución se publicará en el Boletín Oficial del/de la [entidad territorial] y en el portal web [estatal/autonómico/municipal/etc.]. El órgano competente publicará esta resolución en el Boletín Oficial del/de la [entidad territorial], en la intranet/extranet [estatal/autonómica/municipal/etc.] correspondiente ([URL]) y en la web ([URL]). Publicar esta resolución, por la que se [propósito de la resolución].
	Título de la decisión sobre el acceso a información relacionada con la resolución [título opcional]	[SEGUNDO/TERCERO/CUARTO/etc.]. [SEGUNDO/TERCERO/CUARTO/etc.]. [Breve indicación del tema de la decisión].
	Texto de la decisión sobre el acceso a información relacionada con la resolución [contenido opcional]	El/La [información relacionada con la resolución] se publicará como anexo de esta resolución en el Boletín Oficial del/de la [entidad territorial]. El expediente original de esta resolución se guarda en la sede del/de la [área/órgano/institución] ([dirección postal]). Las personas interesadas podrán consultar la información sobre [información relacionada con la resolución] en la [aplicación informática/intranet/sede electrónica/etc.], en la sección “[indicación de la ruta para acceder a la información en la web]”, desde el día que se publique esta resolución en el Boletín Oficial del/de la [entidad territorial].

Apartados	Contenidos	Fraseología
		<p>Insertar la información sobre [información relacionada con la resolución] en la [aplicación informática/intranet/sede electrónica/etc.], en la ruta “[indicación de la ruta para acceder a la información en la web]” el mismo día que se publique esta resolución en el Boletín Oficial del/de la [entidad territorial].</p> <p>Según el artículo [nº] del/de la [normativa aplicable: ley/reglamento/acuerdo/etc.], sobre la forma en la que debe difundirse y publicarse la información sobre [información relacionada con la resolución],</p>
	Fórmula de introducción de los recursos posibles	<p>Contra la resolución, que pone fin a la vía administrativa, las personas interesadas podrán interponer:</p> <p>Contra esta resolución, que pone fin a la vía administrativa, podrá interponerse cualquiera de los recursos que se indican a continuación:</p>
	Recursos que cabe interponer	<p>Recurso potestativo de reposición, ante el [cargo del emisor] en el plazo de un mes desde el día siguiente a la fecha en la que se publica esta resolución. (si se pueden interponer varios recursos, se recomienda indicarlos en una lista)</p> <p>Recurso potestativo de reposición, en el Registro [tipo de registro], ante el mismo órgano que dictó la resolución. (si se pueden interponer varios recursos, se recomienda indicarlos en una lista)</p> <p>El recurso potestativo de reposición debe interponerse en el plazo de un mes a partir del día siguiente a la fecha en que se publica la resolución en el/la [portal institucional/Boletín Oficial/sede electrónica].</p> <p>Los artículos 30, 123 y 124 de la Ley 39/2015, de 1 de octubre, del Procedimiento Administrativo Común de las Administraciones Públicas regulan los plazos administrativos y el recurso potestativo de reposición.</p> <p>Si la persona interesada interpone el recurso y al cabo de un mes no se ha dictado y notificado la resolución, se entenderá que se desestima por silencio administrativo.</p>
	Recursos que cabe interponer	<p>Si el órgano administrativo desestima un recurso por silencio administrativo, la persona interesada puede interponer un recurso contencioso-administrativo dentro del plazo de seis meses.</p> <p>Recurso contencioso-administrativo, ante los Juzgados de lo Contencioso-Administrativo. (si se pueden interponer varios recursos, se recomienda indicarlos en una lista)</p>
		<p>Directamente recurso contencioso-administrativo, ante el órgano jurisdiccional competente, en el plazo de dos meses desde el día siguiente a la fecha en que se publica esta resolución. (si se pueden interponer varios recursos, se recomienda indicarlos en una lista)</p> <p>El plazo para interponer el recurso contencioso-administrativo es de dos meses desde el día siguiente a la fecha en que se publica la resolución en el Boletín Oficial del/de la [entidad territorial].</p> <p>Los artículos 8 y 46 de la Ley 29/1998, de 13 de julio, Reguladora de la Jurisdicción Contencioso-Administrativa regulan los requisitos y plazos del recurso contencioso-administrativo.</p>

Apartados	Contenidos	Fraseología
		El personal laboral podrá interponer demanda contra esta resolución ante la vía judicial laboral. (si se pueden interponer varios recursos, se recomienda indicarlos en una lista)
		La posibilidad para el personal laboral de acudir a la vía judicial laboral se regula en el artículo 69 de la Ley 36/2011, de 10 de octubre, reguladora de la jurisdicción social.
		Las demandas del personal laboral contra esta resolución se interpondrán en la forma y plazos que se establecen en la Ley 36/2011, de 10 de octubre, reguladora de la jurisdicción social y en el Estatuto de los Trabajadores.
	Información sobre la posibilidad de interponer otros recursos	Las personas interesadas podrán interponer cualquier otro recurso que sea procedente para el caso. La posibilidad de interponer otros recursos además de los que se indican en la resolución se establece en el artículo 40.2 de la Ley 39/2015, de 1 de octubre, del Procedimiento Administrativo Común de las Administraciones Públicas.
Cierre	Localidad del emisor	[Localidad],
	Fecha	a [día] de [mes] de [año].-
	Cargo y nombre del emisor	El/La [cargo del emisor], [nombre completo].
Anexos [apartado opcional]	Título del anexo	ANEXO [nº]
Anexos [apartado opcional]	Posibles contenidos de los anexos: tablas/listas con datos relacionados con la resolución, formularios, regulación de aspectos concretos de la resolución, etc.	ANEXO [nº]. [Título del anexo]



## Uses and functions of conceptual metaphor in pedagogical meditation discourse

### Usos y funciones de la metáfora conceptual en discurso pedagógico sobre meditación

ANTONIO-JOSÉ SILVESTRE-LÓPEZ

Universitat Jaume I, España

[asilvest@uji.es](mailto:asilvest@uji.es)

<https://orcid.org/0000-0002-4006-1364>

#### Abstract

In mindfulness and meditation instructional contexts, conceptual metaphor is a powerful facilitator of intersubjective communication about inner events and first-person experiences. Previous studies in meditation discourse have described space as a productive source domain that is used by meditation teachers to characterise metaphorically inner events and abstract concepts related to the meditative practice (Silvestre-López 2019, 2020). This study examines the discourse manifestations of the space source domain used to metaphorically characterise the most representative metaphorical topics (targets) in a corpus of introductory talks about meditation addressed to an audience of non-expert meditators. The analysis is sensitive to the communicative dimension of the use of metaphors in that it distinguishes between non-deliberate and deliberate uses (Steen 2015) and considers the functions they fulfil in discourse (Semino 2008; Goatly 2011). The corpus was analysed

#### Resumen

La metáfora conceptual es un potente facilitador de la comunicación de eventos internos y experiencias en primera persona en contextos de instrucción de mindfulness y meditación. Estudios previos sobre el discurso de la meditación han descrito la noción del espacio como un dominio productivo en el discurso de maestros de meditación utilizado como fuente para caracterizar metafóricamente eventos y conceptos abstractos relacionados con la práctica meditativa (Silvestre-López 2019, 2020). Este artículo explora las manifestaciones discursivas del dominio fuente del espacio utilizadas para caracterizar metafóricamente temas metafóricos (dominios meta) representativos en un corpus de charlas introductorias sobre meditación dirigidas a una audiencia de no expertos. El análisis es sensible a la dimensión comunicativa del uso de las metáforas y para ello distingue entre usos no deliberados y deliberados (Steen 2015), y considera también sus funciones principales en el discurso (Semino 2008; Goatly 2011). El corpus fue analizado cualitativamente

---

**To cite this article:** Silvestre-López, A. J. (2022). Uses and functions of conceptual metaphor in pedagogical meditation discourse. *ELUA*, (38), 47-72. <https://doi.org/10.14198/ELUA.22349>

Recibido: 02/04/2022, Aceptado: 13/05/2022

© 2022 Antonio-José Silvestre-López



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



qualitatively following a bottom-up approach. Metaphor identification was carried out using DMIP (Reijnierse et al. 2018). Target and source domain coding was done with the help of dictionary information and the domains already identified in a metaphor compendium created in previous meditation discourse analyses (see Coll-Florit & Climent 2019; Silvestre-López & Navarro 2017; Silvestre-López 2020). This approach allowed different layers of domain specificity to be unveiled, with remarkable differences found between non-deliberate and deliberate metaphor use. Among them, differences in degree of source domain granularity (Langacker 2008) ranging from vaguer notions like UNDEFINED SPACE OF SPACIOUSNESS to conceptually richer and finer-grained scenarios like the HOME AND SHELTER, OCEAN, LANDSCAPE, OR SKY models. The speakers were found to use these source domains consistently with explanatory purposes (explanatory function) to help their audience reconceptualise some of the most frequent target topics in their talks (e.g., THE PRESENT MOMENT, THE MEDITATOR, THE MIND, AWARENESS, OR MEDITATION) from more mindfully-oriented perspectives, hence allowing to become aware of and get rid of potential misconceptions. Deliberate uses proved particularly useful in promoting understanding through reconceptualisation, which was in turn facilitated by the vivid imagery evoked through detailed elaborations of source-domain scenery. Besides ideational functions, deliberate metaphors were also found to facilitate positive attitudinal changes towards the practice, as well as to provide for argument structuring and create textual cohesion. Overall, non-deliberate and deliberate uses were found to combine in the speakers' production to aid them in skilfully achieving their own communicative intentions, which underscores the need for further studies that explore potential ways of applying similar findings to real meditation practice, and in instructional settings.

**KEYWORDS:** Meditation discourse; pedagogical talks; conceptual metaphor; deliberate metaphor, metaphor functions.

siguiendo un enfoque inductivo. La identificación de metáforas se llevó a cabo aplicando DMIP (Reijnierse et al. 2018). La codificación de los dominios meta y fuente se realizó con la ayuda de la información del diccionario y de los dominios ya identificados en un compendio de metáforas elaborado en anteriores estudios sobre el discurso de la meditación (véase Coll-Florit & Climent 2019; Silvestre-López & Navarro 2017; Silvestre-López 2020). El enfoque aquí adoptado ha permitido desvelar diferentes grados de especificidad en los dominios metafóricos, con diferencias relevantes entre metáforas deliberadas y no deliberadas. Entre ellas, los resultados revelan diferentes grados de granularidad de los dominios fuente (Langacker 2008) desde nociones menos definidas como ESPACIO NO DEFINIDO O ESPACIOSIDAD hasta escenarios conceptualmente más ricos como los modelos CASA Y REFUGIO, OCÉANO, PAISAJE O CIELO. El estudio ha permitido desvelar cómo los hablantes utilizan estos dominios fuente de forma coherente con fines explicativos (función explicativa) para ayudar a su audiencia a reconceptualizar algunos de los temas (dominios meta) más frecuentes en sus charlas (por ejemplo, EL MOMENTO PRESENTE, EL MEDITADOR, LA MENTE, LA CONSCIENCIA O LA MEDITACIÓN), aportando perspectivas más cercanas a la filosofía del mindfulness, facilitando de este modo que la audiencia pueda tomar de conciencia de posibles ideas erróneas sobre la práctica. En concreto, los usos metafóricos deliberados se han revelado especialmente útiles para promover la comprensión a través de la reconceptualización, facilitada ésta por las imágenes vívidas evocadas a del dominio fuente. Además de las funciones ideacionales, las metáforas deliberadas también son utilizadas para facilitar cambios de actitud positivos hacia la práctica, así como proporcionar estructura argumental y facilitar la cohesión textual. En general, el estudio ilustra cómo los usos deliberados y no deliberados se combinan en la producción de los hablantes para alcanzar sus propias intenciones comunicativas, subrayando la necesidad de llevar más estudios que exploren posibles formas de aplicar resultados similares a la práctica real de la meditación y su instrucción.

**PALABRAS CLAVE:** Discurso de la meditación; charlas pedagógicas; metáfora conceptual; metáfora deliberada; funciones de la metáfora.

## 1. INTRODUCTION

In Conceptual Metaphor Theory (CMT, Lakoff & Johnson 1980; Lakoff 1993) conceptual metaphors are conceived of as cognitive mechanisms that allow us to conceptualise, reason and communicate about relatively abstract areas of experience (the target domains, henceforth TDs) in terms of other experiences that are simpler, more concrete, or more familiar to us (the source domains, henceforth SDs). For example, the metaphor *THOUGHTS ARE OBJECTS*,<sup>1</sup> allows us to structure part of our knowledge about *THOUGHTS* (the TD) according to part of our knowledge about *OBJECTS* (the SD) in the physical space so that we can understand and produce expressions like “the idea never *crossed* my mind” or “let your thoughts *pass by*, do not *cling* to them”.

Conceptual metaphor facilitates the structuring of concepts and communication about abstract experiences related to inner-life events like thoughts, feelings, and emotions (Kövecses 2020). Due to its power to establish connections between perceptual, conceptual, and emotional experience and to verbalise such relationships through metaphorical expressions, conceptual metaphor is a powerful catalyst for intersubjective communication about first-person experiences such as those dealt with in mindfulness or meditation practice settings. Metaphorical language is a productive resource in instructional settings where meditation instructors and novel meditators are expected to share their inner experiences wholeheartedly (Silvestre-López 2016). Among the systematic conceptual metaphors about meditation unveiled in recent research, the metaphor *MEDITATION IS A JOURNEY* has been found to pervade a wide range of lay, spiritual and religious meditation practice contexts (Silvestre-López & Navarro 2017; Silvestre-López 2019). This metaphor, which inherits its structure from the *PROCESSES ARE PATHS* primary metaphor (Grady 1997), allows one to conceptualise and talk about the *process* of meditation as a *journey*. As the meditation journey is expected to include several stages or places that the meditator is to experience and (re)visit, the notion of *space* is thus also expected to be an important metaphorical source in meditation discourse. Several studies have described space as an important area of experience recruited in figurative uses of language in the context of contemplative practices (Silvestre-López & Navarro 2017; Silvestre-López 2019), but the particular discourse manifestations of *SPACE* as a SD for metaphorical expressions in the context of meditation have not been explored yet. This study aims to explore the metaphorical elaborations of the *SPACE* SD in a corpus of videos of pedagogical talks about meditation.

The analysis of conceptual metaphors conducted in this study is framed in CMT and its extended version, the Deliberate Metaphor Theory (DMT, Steen 2015, 2017; Reijnierse 2017; Reijnierse *et al.* 2018), which makes a distinction between the levels of *language*, *thought*, and *communication*. From this perspective, conceptual metaphors can be manifested in *language* (metaphorical linguistic expressions) mainly as *direct*, *indirect*, or *implicit* metaphorical linguistic expressions (see Steen *et al.* 2010 for a detailed account).

---

<sup>1</sup> In this paper, formulae like *THOUGHTS ARE OBJECTS* (TD IS SD) are used to represent conceptual metaphors, that is, the associations between two conceptual domains at the level of thought. When discourse topics coincide with metaphorical conceptual domains (TD or SD), these are also capitalised (e.g. *MEDITATION*, *THE MEDITATOR*). The linguistic manifestations of these associations (linguistic metaphorical expressions) are signalled in italics in the examples.

In direct metaphors, there is a direct connection between the TD and the SD which is signalled in language, as in *A IS B* formulae, or as juxtaposition (e.g., *the ocean does not try to stop its waves, so you don't need to try to stop your thoughts*), whereas this connection is not made explicit in indirect metaphors. Finally, implicit metaphors are usually realised through cohesive devices like demonstratives or personal pronouns that refer to either indirect or direct metaphors in the co-text.

At the level of *cognition* or *thought*, metaphorical expressions can be classified as *conventional* or *novel* depending on their degree of entrenchment. Conventional metaphors are highly entrenched in the conceptual, cultural, and linguistic systems of a linguistic community; they are thus very frequent in everyday language and speakers seldom realise that they are using metaphorical language. *The idea never crossed my mind* is an example of a well-entrenched, conventional metaphor in the English language. Novel metaphors, in contrast, are creative or unconventional ways of developing the metaphorical association at the conceptual level.<sup>2</sup> *Let your thoughts soar through the sky of your mind* could be regarded as a novel metaphorical expression (see Vargas-Sierra & Moreno-Sandoval 2021: 44-45, for a detailed description of the conventional-novel dichotomy).

At the level of *communication*, conceptual metaphors may be manifested as *non-deliberate* or *deliberate* uses. Deliberate metaphors, in particular, are ostensive uses of metaphor *as* metaphor in communication that are usually associated with particular communicative intentions on the part of the speaker (Steen 2015, 2017). According to DMT, in deliberate metaphors “the source domain plays a role in the representation of the referential meaning of the utterance” (Reijnierse et al. 2018: 134), and when metaphors are used deliberately, their structure often “signals that the addressee has to move away their attention momentarily from the target domain of the utterance or even phrase to the source domain that is evoked by the metaphor-related expression” (Steen 2015: 68). Attention to the SD is thus an essential feature of deliberate metaphors and an important factor distinguishing it from non-deliberate metaphor:<sup>3</sup> deliberate metaphors invite the addressee to reconsider the TD from the perspective profiled by the SD and to interpret the message according to this reconsideration, but this does not happen in non-deliberate metaphors (Reijnierse et al. 2018; Steen 2015, 2018). This distinction becomes particularly apparent in cases of prototypical non-deliberate and deliberate metaphor. Non-deliberate metaphor accounts for the bulk of metaphor realisations in discourse, and they are prototypically manifested through *indirect* and *conventional* metaphor like those described above. In contrast, in cases of prototypical deliberate metaphor, the invitation to reconsider the TD according to the SD perspective in deliberate metaphors is often manifested through *novel direct metaphor*,<sup>4</sup> *metaphor repetition*, or *metaphorical extendedness*, among other types (see

2 In metaphor identification procedures like the Vrije Universiteit Metaphor Identification Procedure, MIPVU (Steen et al. 2010), and the Deliberate Metaphor Identification Procedure, DMIP (Reijnierse et al. 2018), a metaphorical expression is regarded as novel if the contextual meaning of the expression in discourse contrasts with its basic meaning, but it is not registered as an independent dictionary entry (see Steen et al., 2010, particularly chapter 3, for a detailed description of this procedure in MIPVU, and Reijnierse et al. 2018, 2020 for how novel and conventional metaphors may be identified as deliberate metaphors at the level of communication in DMIP).

3 This is in fact the main criterion on which DMIP (Reijnierse et al. 2018) pivots to determine the potential deliberate status of metaphorical expressions in discourse.

4 In DMT, novelty and directness are two metaphor traits (at the level of cognition and language respectively) marking deliberateness (Steen 2015, 2017). Thus, both direct and novel metaphors are regarded as cases of deliberate metaphor by definition (Reijnierse et al. 2018: 135).

Reijnierse, 2017 for a detailed description of potential deliberate metaphor realisations in discourse). For example, the creative component of *novel* metaphorical expressions like “soar through” and “the sky of your mind” in *let your thoughts soar through the sky of your mind* is a particularly relevant element in determining deliberateness, as it triggers attention towards the SD of the metaphor, so that the TD is reconsidered according to the SD aspects highlighted in the expression (see Reijnierse *et al.* 2018: 133, 135-137). This example of novel metaphor is also a prototypical case of a deliberate metaphor. Besides novelty, moreover, the metaphorical expression “the sky of your mind” in the example also facilitates, through direct metaphorical language (*B OF A* is a direct metaphor flag), the activation of *SKY* as a key element in the referential meaning for the representation of *thoughts* and the *mind* as *clouds* and the *sky*. Direct language thus works in tandem with novelty to make this example a prototypical case of deliberate metaphor.

Due to the overt communicative intention with which deliberate metaphors are used by speakers in discourse, they are useful tools in meditation pedagogical contexts. Not only in terms of the discourse and communicative functions they fulfil (Silvestre-López 2019), but they have also been found to influence metacognition and affective factors in short guided meditation exercises (Silvestre-López *et al.* 2021), suggesting that the effects of considering the TD from the SD perspective go well-beyond discourse and communication. In light of these considerations, it seems sensible to include the deliberate/non-deliberate metaphor contrast in any analysis focussing on meditation discourse. This study surveys the different discourse manifestations of *SPACE* as a SD to metaphorically characterise the most representative TDs in a corpus of introductory talks about meditation by well-known media speakers and meditation experts. Due to the pedagogical nature of the talks, some communicative value was assumed to hold for at least some of the metaphors used by the speakers. The analysis is sensitive to the communicative dimension of metaphor use in that it distinguishes between non-deliberate and deliberate realisations of metaphor in discourse (Steen 2015, 2017; Reijnierse 2017) and considers the main functions with which deliberate metaphors are used in the talks.

## 2. METHODOLOGY

The corpus is composed of a selection of eight introductory talks about meditation recorded on video and transcribed for this analysis.<sup>5</sup> The talks are original oral deliveries by four world-renowned meditation experts and teachers in the media (two talks per meditation speaker) that are currently highly influential and widely followed online by western audiences. All speakers are fluent English users. All talks were delivered in a masterclass (monologue) format, and in all of them the speakers directly address the audience. In two of the videos (Kabat-Zinn 2020, March 30, March 31), the masterclass about meditation is followed by a question-answer session, but only the monologue part was selected and transcribed for analysis in order to adhere to a consistent discursive format throughout. Altogether, the corpus amounts to 30,321 words.

---

<sup>5</sup> Full video references are provided in the bibliography list.

Besides the criterion with respect to the speakers' international profile, the talks were selected on the basis of the topic covered and their intended audience, namely, video talks introducing the notion of meditation and explaining the basic principles to an audience of non-expert meditators (adults interested in meditation but with little or no meditative experience at all). A careful talk selection process was followed to guarantee that all talks addressed only the early stages of meditation, and that the conception of meditation was in line with meditative practices designed to develop self-awareness and to provide basic training in attention regulation. Specifically, all talks address the elementary stages of meditation for the uninitiated and are related to the so-called mindfulness regulation practices (focussed-attention in combination with open-monitoring practices) (Dahl *et al.* 2015). Attention regulation practices enable the development of sustained attention and provide the meditator with a clearer view of reality. These practices are therefore necessary for the meditator to develop a solid foundation for any further meditative experience, both in spiritual and secular contexts. In religious and spiritual contexts, these practices also constitute the first stage in the development of the spiritual path, as they quieten the mind and thereby allow access to other areas of inner experience (areas that are more detached from the mundane and closer to divinity). In secular contexts of meditation instruction, these practices underpin a considerable part of today's mindfulness and compassion programmes (Cebolla *et al.* 2014; Matko & Sedlmeier 2019). All practices discussed in the talks, therefore, can be equated with those used for the development of mindfulness and are the prelude to more advanced practices such as those related to the field of loving-kindness, compassion, or ego-deconstructive practices (Dahl *et al.* 2015; García-Campayo 2020).

To ensure the correct analysis of all types of uses (especially deliberate uses related to topic incongruencies, cases of topic-driven metaphor, extended metaphor, or metaphoric echoes within and across talks), and also to adhere to the metaphor identification procedure used in this study (see below), the corpus was analysed following a bottom-up qualitative approach. The analysis was aided by the qualitative analysis software ATLAS.ti and it was conducted in three stages: metaphor *identification*, metaphor *codification*, and metaphor *reconstruction*.

The first stage in the analysis involved the *identification* of conceptual metaphors in the corpus was made through their expression in discourse by applying DMIP (Reijnierse *et al.* 2018) to the whole corpus. DMIP is currently the standard method used widely by the scientific community to identify deliberate metaphors in discourse. DMIP requires a bottom-up approach like the one adopted in this analysis to identify all potential cases of deliberate metaphor. Firstly, and as proposed by DMIP (first step), the talk transcripts were read to guarantee the correct understanding of each textual unit as a whole. DMIP draws on MIPVU as a second step to identify metaphor-related words (MRWs), or metaphor vehicles. The transcripts were thus analysed following MIPVU as part of the metaphor identification stage. Basically, MIPVU consists in identifying linguistic metaphors (indirect, direct, or implicit) by reading the whole text, establishing the contextual meaning of each unit (the meaning with which each unit is used in the text), using dictionary entry information to determine whether the unit has a more basic meaning (one that is more concrete, bodily-related, precise, or historically older), and deciding whether the contextual meaning contrasts with the basic meaning but can be understood in comparison with it. If this is the case, then the unit is marked as a MRW (indirect/implicit metaphor). MIPVU

also permits cases of direct metaphor and metaphor signals (e.g., *A IS B* or *B OF A* metaphor, metaphorical similes, counterfactual reality expressions, etc.) to be identified and considered as part of metaphorical realisations in discourse. Thus, if the lexical unit is signalled in discourse, or if it is part of a direct metaphor, it is also marked as an MRW (direct metaphor). Otherwise, the unit is discarded for metaphorical analysis (not metaphorical).<sup>6</sup> The third step of DMIP involves deciding on the deliberate or non-deliberate status of each MRW identified in the previous step. If the SD of the MRW is part of the referential meaning of the utterance, then this condition is made explicit and the deliberate status of the MRW is marked as such. Otherwise, the MRW is considered non-deliberate.<sup>7</sup>

Once the status of all MRWs had been identified, the second analytical stage involved the *codification* of the metaphorical TDs and SDs. All MRWs were marked with ATLAS.ti quotations. A compositional coding strategy was adopted (Kimmel 2012) in which TDs and SDs were coded separately. This allows multiple TD-SD combinations to be identified and retrieved (in the metaphor reconstruction stage) as well as their classification in terms of frequency. TD and SD coding was done with the help of dictionary information and the domains already identified in a metaphor compendium (see Coll-Florit & Climent 2019) elaborated in previous studies of meditation discourse (Silvestre-López & Navarro, 2017; Silvestre-López, 2020). The specificity of TD and SD codes was reviewed and, when necessary, the codes were minimally restructured so that their formulation would adhere as closely as possible to metaphor use in the corpus. Following the compositional coding strategy, deliberate metaphor uses were also assigned a deliberate metaphor code. Deliberate metaphor functions were also coded separately. They were selected and assigned to deliberate metaphor uses from a pre-defined list of potential ideational, interpersonal, and textual functions that had previously been uploaded into the ATLAS.ti code list. The list of functions is based on previous literature (Charteris-Black 2004; Semino 2008; Steen *et al.* 2010; Goatly 2011; Kimmel 2012; Steen 2015; Reijniere 2017; Reijniere *et al.* 2018). Each metaphorical expression (or ATLAS.ti quotation) was assigned only one TD code, one SD code, and, if deliberate, one deliberate code. However, as the same expression may fulfil a variety of functions simultaneously, each expression could be assigned several function codes.

The last stage of the analytical process involved the *reconstruction* of conceptual metaphors, which entailed retrieving the correspondences between each TD and SD to reconstruct relevant TD-SD pairings (i.e., conceptual metaphors) and their discourse realisations in the corpus. This process also involved the distribution of non-deliberate and deliberate uses of each TD-SD pairing as well as the functions with which they were used. This process was aided by the ATLAS.ti query tool and its main results are shown in the tables in section 3.

6 For a detailed description of the steps and intricacies of MIPVU, see Steen *et al.* (2010: 25-42).

7 For a detailed and fully-illustrated account of DMIP steps and decisions in all stages of the procedure, see Reijniere *et al.* (2018: 135-141).

### 3. RESULTS AND DISCUSSION

#### 3.1. Main TD-SD Correspondences and Deliberate Metaphor Functions

TD \ SD	SPACE: UNDEFINED SPACE/ SPACIOUSNESS	SAFE STRUCTURES: HOME AND SHELTER	HERE (a nearby place)	NATURAL SPACE: MASS OF WATER	OPPRESSIVE/ CONFINEMENT STRUCTURES	AWAY (a distant place)	NATURAL SPACE: FIELDS AND LANDSCAPES	NATURAL SPACE: SKY	TOTAL INSTANCES (per TD)
THE PRESENT MOMENT IS...	51	2	0	2	2	0	1	0	58
THE MEDITATOR IS...	16	21	0	17	0	0	1	1	56
ATTENTION IS...	0	8	44	0	0	0	0	0	52
THE MIND IS...	22	2	0	3	0	0	0	2	29
AWARENESS IS...	14	3	1	4	0	0	5	1	28
THOUGHT IS...	3	0	0	7	13	5	0	0	28
CONSCIOUSNESS IS...	24	0	0	0	0	0	0	0	24
MEDITATION IS...	7	13	0	0	0	0	2	0	22
DISTRACTION IS...	0	0	0	0	1	9	0	0	10
THE FUTURE IS...	4	0	0	0	0	0	0	0	4
EMOTION IS...	2	0	0	0	1	0	0	0	3
<b>TOTAL INSTANCES (per SD)</b>	<b>143</b>	<b>49</b>	<b>45</b>	<b>33</b>	<b>17</b>	<b>14</b>	<b>9</b>	<b>4</b>	<b>314</b>

Table 1. Space metaphors in the corpus (TD-SD combinations): General overview.

Table 1 shows all space-related metaphor occurrences in the corpus, arranged by TD-SD combinations. The first row in the table relates to the most frequent SDs in the corpus, arranged from the highest to the lowest frequency scores. The SDs show different degrees of granularity or elaboration (Langacker 2008). Domains with a low degree of elaboration include, for example, UNDEFINED SPACE/SPACIOUSNESS, HERE (a nearby place), and AWAY (a distant place); in these cases, the metaphorical characterisation entails the limited selection and elaboration of the conceptual structure of the SD to be projected onto the TD. Richer or more elaborated SDs include the HOME scenario (Musolff 2016) or nature-related spaces like the MASS OF WATER, LANDSCAPE, and SKY models. These domains provide rich conceptual structure relative to specific elements (e.g., home, owner, visitors; ocean, waves) and their relationships and roles, which is selected and projected onto the structure of the TD, so that the TD elements and relationships are understood in terms of the SD structure (see Silvestre-López 2019; Kövecses 2020)

The first column in Table 1 presents the most frequent TDs (metaphorised topics in the corpus) arranged from highest to lowest frequency. The most frequent TDs relate to highly frequent metaphorised topics in the corpus. In the table, TDs combine with SDs (the external areas of experience providing the conceptual structure with which the TD is characterised or understood) to form conceptual metaphors (i.e., each TD-SD pairing is to be conceived of as a conceptual metaphor formulation). Accordingly, the TD THE PRESENT MOMENT combines with the SD UNDEFINED SPACE/SPACIOUSNESS to form the most frequent conceptual metaphor in the corpus (THE PRESENT MOMENT IS SPACE). Since *the now* is experienced in the physical space immediately surrounding the meditator, *the here* and *the now* correlate in the meditator's experience. The 51 occasions on which THE PRESENT MOMENT IS SPACE is manifested through metaphorical expressions in the corpus suggest that the experiential correlation between *the present moment* and *the present space* is also an important metaphorical topic in the talks.

Table 1 shows 35 different TD-SD combinations (that is, 35 different space-based conceptual metaphors), which indicates that the domain of space is an important and creative source of metaphoric images in the corpus. A substantial part of the speakers' production refers to the figure of the meditator. Since the process of meditation takes place in the meditator, a good number of metaphorical realisations found in the corpus (56/314) characterise the meditator as the space where all the phenomena of meditation take place. This may be a generic metaphorical space, as in the conceptual metaphor *THE MEDITATOR IS SPACE*, but it can also be developed in richer scenarios grounded in more concrete SDs like *HOME AND SHELTER* or *MASS OF WATER*. As we shall see, when these domains become the source of metaphors like *THE MEDITATOR IS HOME* or *THE MEDITATOR IS AN OCEAN* their conceptual structure is often used to highlight or expound, predominantly with deliberate uses, particular aspects or desirable attitudes of the meditator towards the meditative practice.

The topic *MEDITATION* is metaphorised less frequently in the corpus (22 times in total), but it is certainly another salient topic addressed by all speakers in the talks. In this case, meditation is characterised not only as the space where the meditator opens up to the experience of the present moment (*MEDITATION IS SPACE*), but also as a process of *coming back home* (*MEDITATION IS COMING HOME*), which is directly related to the *inner journey* of the meditator towards their *inner home* (their inner self), and inextricably linked to metaphors like *THE MEDITATOR IS HOME*.

Similarly, the *MIND* is a recurrent topic (the fourth most metaphorised TD in the corpus, with a total of 29 instances) in the discourse of the speakers, since many of the processes involved in the early stages of meditation are related to *what to do with the mind* (one of the common preconceptions of non-meditators who want to start meditating), or rather, how to conceive of and relate to the mind. Helping novice meditators to reconceptualise the mind not as *something you have to do something with*, but as a *SPACE* where mental phenomena occur (a *SPACE* in which the meditator does not necessarily have to do something) is certainly one of the most difficult tasks of meditation instructors. This is precisely the reason why the *MIND* is recurrently metaphorised as the *SPACIOUSNESS* in which mental phenomena occur, and one of the main instances of reconceptualisation phenomena in the corpus (see metaphorical functions below).

Since the talks address the early stages of meditation (i.e., meditation practices focusing on self-awareness and attention training and regulation), other recurring topics are *AWARENESS* (also conceptualised as *SPACIOUSNESS* with relative frequency; 14 instances), *CONSCIOUSNESS* (exclusively conceptualised as *SPACIOUSNESS*; 24 instances), and the processes of *ATTENTION* and *DISTRACTION*. The metaphorical conceptualisation of the latter retrieves basic spatial image-schematic patterns such as the proximity-distance image schema (Johnson 1987; Peña-Cervel 2003) to characterise *ATTENTION* as *presence here* (*PAYING ATTENTION IS BEING HERE*) and *DISTRACTION* as *being away* (*BEING DISTRACTED IS BEING IN A DISTANT PLACE*). The *JOURNEY* metaphor mentioned in the introduction, therefore, takes on a new dimension in the attentional processes described in the corpus.

Table 1 provides an at-a-glance panoramic overview of the main TDs, SDs, and TD-SD combinations in the corpus, presenting the domains according to general metaphor frequency counts (discourse manifestations of TD-SD pairings as metaphorical expressions) regardless of their deliberate status (Steen 2015; Reijnierse *et al.* 2018). In other words, the table follows a general frequency criterion to present the information, but it does not



discriminate between non-deliberate and deliberate uses. When a conceptual metaphor (or a TD-SD pairing, as in the table) is used deliberately in discourse, the metaphorical nature of the expressions becomes a salient element in the communicative dimension, an ostensive act of meaningful communication that is usually bound to a particular communicative intention (Reijnierse 2017; Reijnierse *et al.* 2018). Considering the relevance of the communicative dimension and the overtly explanatory nature of the material in the corpus, metaphorical deliberateness is an important aspect to consider in the analysis. Table 2 shows the main TD-SD combinations arranged according to deliberate metaphor frequency counts.<sup>8</sup> The table gives 23 different possible TD-SD combinations, indicating the extensive use of deliberate metaphor in the corpus (of the 35 different conceptual metaphors in Table 1, 23 involve at least one deliberate realisation in discourse), thus confirming the importance of metaphorical deliberation in the pedagogical talks about meditation analysed.

TD \ SD	SPACE: UNDEFINED SPACE / SPACIOUSNESS	SAFE STRUCTURES: HOME AND SHELTER	HERE (a nearby place)	NATURAL SPACE: MASS OF WATER	OPPRESSIVE/ CONFINEMENT STRUCTURES	AWAY (a distant place)	NATURAL SPACE: FIELDS AND LANDSCAPES	NATURAL SPACE: SKY	TOTAL INSTANCES (per TD)
THE PRESENT MOMENT IS...	51	2	0	2	2	0	1	0	58
THE MEDITATOR IS...	16	21	0	17	0	0	1	1	56
ATTENTION IS...	0	8	44	0	0	0	0	0	52
THE MIND IS...	22	2	0	3	0	0	0	2	29
AWARENESS IS...	14	3	1	4	0	0	5	1	28
THOUGHT IS...	3	0	0	7	13	5	0	0	28
CONSCIOUSNESS IS...	24	0	0	0	0	0	0	0	24
MEDITATION IS...	7	13	0	0	0	0	2	0	22
DISTRACTION IS...	0	0	0	0	1	9	0	0	10
THE FUTURE IS...	4	0	0	0	0	0	0	0	4
EMOTION IS...	2	0	0	0	1	0	0	0	3
<b>TOTAL INSTANCES (per SD)</b>	<b>143</b>	<b>49</b>	<b>45</b>	<b>33</b>	<b>17</b>	<b>14</b>	<b>9</b>	<b>4</b>	<b>314</b>

Table 2. Space metaphors in the corpus (TD-SD combinations): Deliberate uses.

From the TD perspective, the data indicate that *THE MEDITATOR* is the most frequently used TD in deliberate metaphors (50 deliberate uses out of 56 in the corpus, see Total Instances columns in Tables 1 and 2) and keeps its overall distribution among the three most prolific SDs in deliberate metaphor uses (Table 2), namely, *HOME*, *SPACIOUSNESS*, and *MASS OF WATER*. *THE PRESENT MOMENT* is in second position, with 19 deliberate metaphor instances (out of 51) also largely drawing on the notion of *UNDEFINED SPACE*. The third most frequent TD used deliberately by the speakers is *MEDITATION*, which moves from eighth position in Table 1 to third position in Table 2, with 15 deliberate uses out of a total of 22 (about two thirds of the uses are deliberate).

Some of the most interesting changes between general (Table 1) and deliberate metaphor (Table 2) uses occur at the SD level. A comparison of Tables 1 and 2 reveals that, with 49 instances, *HOME AND SHELTER* is exclusively realised deliberately in discourse in all metaphors where it appears in the corpus (related to 6 different TDs). Additionally, and despite the

<sup>8</sup> Besides the total frequency counts, the italicised text in the second line of each cell shows the breakdown of the frequency distribution of each TD-SD combination by speaker (where B stands for Brach, F for Foster, K for Kabat-Zinn, and T for Tolle). For example, from the 21 deliberate metaphor realisations of the *THE MEDITATOR IS HOME* metaphor 20 are from Brach's production and only 1 is a metaphorical expression by Kabat-Zinn.

fact that spaciousness is a relatively abstract and undefined notion, the SD UNDEFINED SPACE/SPACIOUSNESS shows a considerable degree of deliberation; specifically, it is used deliberately by the speakers on 42 occasions (out of a total of 143 uses) as discursive manifestations of four conceptual metaphors (four different TDs in Table 2). Finally, it is also worth noting that nature-related spaces appear in the next three positions in Table 2, which indicates that - regardless of whether they are images of water, land, or air - nature experiences seem to have particular relevance in the images used by speakers; particularly, the images used to talk about THE MEDITATOR and their AWARENESS of the MIND and its contents (THOUGHT).

The frequency distribution of each TD-SD among speakers in Table 2 also provides insight into the saliency of some metaphorical domains in the discourse of some of the speakers. From the TD perspective, THE PRESENT MOMENT stands as a particularly relevant metaphorical topic in Tolle's talks, accounting for 17 out of 19 deliberate metaphor instances in the corpus, 16 of which are deliberate realisations of the THE PRESENT MOMENT IS SPACIOUSNESS metaphor. AWARENESS is also a recurrent topic in Kabat-Zinn's production, with 9 out of 12 instances distributed, in this case, across five different metaphors; this is the widest range of deliberate metaphorical images in the corpus related to a TD, and is indicative not only of the importance of the topic in the speaker's discourse, but also of his attempts to help his audience understand the notion of mindful awareness from multiple perspectives. From the SD perspective, the speakers' TD-SD frequency distribution breakdown reveals HOME AND SHELTER as a notably important SD that structures a variety of metaphors in Brach's discourse. Specifically, 4 of the 5 TD-SD pairings in which the HOME AND SHELTER model works as a SD are found in Brach's talks through metaphors like THE MEDITATOR IS HOME (20/21 instances), MEDITATION IS HOME (13/13 instances), and ATTENTION IS HOME (8/8 instances). Brach's talks are in fact entitled after the home metaphor, and this widespread use indicates that the talks are conceptually structured around it too. This and other related aspects are addressed and illustrated with examples in section 3.2.

Overall, both the changes in the TD and SD order in Table 2 with respect to Table 1 and the information in Table 2 suggest that, when talking about the most frequent domains, speakers consistently use explicit metaphorical language or salient metaphorical imagery related to some aspect of the SDs, most likely with specific communicative intentions. Table 3 ranks the main functions with which deliberate metaphors are used in the corpus.

DELIBERATE METAPHOR FUNCTION	INSTANCES	%
Explanatory	116	35.58
Evoke vivid, quasi-perceptual imagery	70	21.47
Metaphoric framing (reframing) fostering reconceptualisation	66	20.24
Compress inferences	31	9.51
Create textual cohesion or provide for textual/argument structuring	16	4.91
Create argumentative impact and grab the audience's attention	10	3.08
Highlight/hide aspects of a topic	9	2.76
Other	8	2.35

Table 3. Main deliberate metaphor functions in the corpus.

In the talks, the speakers introduce and try to explain some of the most basic notions of meditation to a non-expert audience. This is not only reflected in the diversity of conceptual metaphors used deliberately in the corpus, but also in the spread of functions with which they are used by speakers. Specifically, 35.58% of the uses fulfil an explanatory function. It is important to note that the same deliberate metaphor may be used with more than one function simultaneously,<sup>9</sup> and in several cases, the explanatory intention is also fulfilled with a metaphor that evokes vivid imagery through SD structure activation (21.47%). For example, in order to describe the impermanent nature of thought and the non-reactive observation that the meditator is expected to work on during their meditation practice, the meditation experts sometimes talk about the MIND (TD) as if it were the SKY (SD), where thoughts are to be passively observed as if they were drifting clouds. It is also common for the meditation teachers to use deliberate metaphor as a tool to shift the perspective of their audience, that is to say, to help them to reconceptualise a topic from an external perspective or point of view. For example, in order to foster acceptance and cultivate non-reactivity towards aversive inner events (e.g., negative experiences, intrusive thoughts, or annoying memories), at some points the speakers try to change the “ordinary” (that is, non-meditative) attitude of their audience (non-experts in meditation) by comparing them (MEDITATORS, TD) with an OCEAN (SD) so that, in meditation, these experiences are just superficial phenomena, i.e., waves on the surface of the ocean “that they are” (an ocean in whose depths there is a timeless state of calm). This illustrates one of the cases of the *metaphoric framing (reframing) fostering reconceptualisation* function (20.24%). The bulk of deliberate metaphor functions in the talks, (together with *highlight/hide aspects of a topic*, used in 2.76% of instances) can be classified as ideational functions, as they are used mainly to promote conceptual and attitudinal understanding towards the practice aspects targeted in the metaphors. However, these uses also merge with other textual and interpersonal dimensions, with some metaphors used as a resource to *compress inferences* (9.51% of uses), *create textual cohesion or help to structure an argument* (4.91%), and *create argumentative impact and grab the audience’s attention* (3.08%).

### 3.2. Discourse Manifestations of Metaphorical Space in the Talks

This section analyses a selection of the most representative conceptual metaphors in the talks illustrating, with anchor examples from the corpus, their main discourse realisations through non-deliberate and deliberate uses, together with the main functions with which the latter are used in discourse. Specifically, metaphors related to the notion of UNDEFINED SPACE or SPACIOUSNESS are presented first (section 3.2.1), before more concrete representations of space through less abstract scenarios like HOME AND SHELTER (section 3.2.2) and the nature-related spatial domains MASS OF WATER, FIELDS AND LANDSCAPES, and SKY are explored (section 3.2.3).

#### 3.2.1. Spaciousness Metaphors

The interconnection between physical and mental space is a key idea in the field of meditation, as *the now* is experienced in the immediate present space – referred to as the *space of presence* or *space of consciousness* by many meditators (Silvestre-López, 2016). In this

<sup>9</sup> This means that the same metaphorical realisation may fulfil one or of the various functions specified in Table 3.

space, sensory perceptions of the physical and material merge with mental perceptions of the meditator's inner life. It is a generic space of fully open awareness, of the mindfulness of all that unfolds in present-moment experience. It is thus a space in which the barriers between the physical and the mental are diluted to give way to direct attention to the experience *as it is* in the present moment.

Due to its peculiarity, it is not surprising that this conception of the meditative space appears in the widest range of conceptual metaphors in the corpus. Table 4 shows the conceptual metaphors that draw on the notion of UNDEFINED SPACE/SPACIOUSNESS as an SD in the corpus. This SD is the source of 9 different conceptual metaphors (TD-SD pairings in the table) realised non-deliberately in the talks. Four of these metaphors also include deliberate uses in the discourse of the speakers, namely, THE PRESENT MOMENT (16 instances), THE MEDITATOR (14 instances), CONSCIOUSNESS (8 instances), and AWARENESS (4 instances).

TD \ SD	UNDEFINED SPACE/ SPACIOUSNESS (total counts)	NDM	DM
THE PRESENT MOMENT IS...	51	35	16
CONSCIOUSNESS IS...	24	16	8
THE MIND IS...	22	22	0
THE MEDITATOR IS...	16	2	14
AWARENESS IS...	14	10	4
MEDITATION IS...	7	7	0
THE FUTURE IS....	4	4	0
THOUGHT IS...	3	3	0
EMOTION IS...	2	2	0
<b>TOTAL INSTANCES</b>	<b>143</b>	<b>101</b>	<b>42</b>

Table 4. SPACOUSNESS conceptual metaphors and non-deliberate (NDM) /deliberate (DM) uses in the talks.

Overall, most non-deliberate metaphor realisations instantiate highly conventional entrenched (Charteris-Black, 2004) metaphorical uses in non-specialised contexts of everyday language use, most often exploiting the notion of UNDEFINED SPACE. This vaguely-defined metaphorical space is activated, on most occasions, by expressions that evoke spatial characteristics of containment or three-dimensionality with a low degree of conceptual elaboration as in, for example, “arises *in* your mind”, “comes *into* your mind” (THE MIND IS A SPACE), “lead you *into* meditation” (MEDITATION IS A SPACE), “coming *into* presence” (CONSCIOUSNESS IS SPACE<sup>10</sup>), “tap *into* [...] awareness” (AWARENESS IS SPACE), “go back *into* the present moment” (THE PRESENT MOMENT IS SPACE), and “*in* what you are” (THE MEDITATOR IS SPACE) in the examples below.

- (1) ... one thing after another *arises in* your mind, and you identify with everything that *comes into* your mind, amplified as I said by the digital stuff... (Tolle, 2014, March 2, 0:44)

<sup>10</sup> *Consciousness* and *presence* are used as synonyms by the speakers in the talks. A methodological decision was taken in the metaphor identification and codification stage such that all metaphorical expressions targeting consciousness and presence (topics, target domains) in the corpus were coded with the umbrella code “TD\_Consciousness”.

- (2) So, we spend perhaps a little bit of time getting here, and you'll notice, when I *lead* you *in* the meditation, I'll spend a longer period of time just to give you the sense of what it's like to listening [sic] to the sounds, listening to and feeling the sensation. (Brach, 2012, July 10, 7:12)
- (3) ... but even then, there was underneath that a profound interest in *coming into* presence and understanding the nature of reality. (Brach, 2012, July 10, 1:25)
- (4) It really depends on our willingness to not take personally things that aren't personal and learn to *tap* deeply *into* this profound interior resource, perhaps the most fundamental aspect of our being human: awareness itself. (Kabat-Zinn, 2020, March 31, 22:25).
- (5) ... now we *go back into* the present moment and, as you become more aware of your sense perceptions... (Tolle, 2014, March 2, 13:07)
- (6) That's just something else that is allowed, deeply allowed to *arise* and *dissolve in* what you are. (Foster, 2012, May 24, 18:57)

When the conceptual metaphors are used deliberately, their discursive realisations take on a more specific character than conventional everyday uses. This qualitative distinction between non-deliberate and deliberate uses is deployed in different ways in the corpus. In some cases, for example, deliberate uses make explicit mention of the SD, as in the deliberate discourse realisations of CONSCIOUSNESS IS SPACE and AWARENESS IS SPACE in (7) and (8), where the TD is characterised as a *space* or a *place* through direct metaphors.

- (7) ... and *the only place* that we can really discover the source of our creativity, our... see truly the nature of reality, *is when we're fully present* [...] because we know that *presence* is really the source of everything. (Brach, 2012, July 10, 14:05)
- (8) *A mindful awareness* does not oppose or interfere or judge or do anything to mess with what's right here. It just allows. *It's space*. (Brach, 2012, July 10, 21:08)

In some other cases, like in the metaphor THE PRESENT MOMENT IS SPACE, the conventional mappings of non-deliberate realisations like (5) are further expanded in deliberate uses such that *the present moment* is not only overtly addressed as a space, but some *spaciousness qualities* of that metaphorical space are also elaborated and described explicitly by the speakers. As suggested in the introduction, the importance of the domain of space in meditation discourse may be due to the widespread conceptualisation of meditation as a journey process. In this case, THE PRESENT MOMENT IS SPACE metaphor is licensed by the common-ground metaphor MEDITATION IS A JOURNEY, which is made explicit in one of Tolle's talks as follows:

- (9) But the now to most people is an abstract thing, and not interesting enough; so, let's see what that means and let's make it into meditation. So, let's call this, our *meditation* this afternoon, "*the journey into the now*." (Tolle, 2014, March 2, 3:42)

This metaphor, appearing in minute 3:42 of the talk, allows Tolle to set the ground for his audience to better follow his description of the steps of meditation and the "*journey into the now*" (i.e., the journey into the present moment); in other words, once the JOURNEY

domain is activated, it is readily available for the interpretation of subsequent discourse contents. This deliberate use allows conventional expressions like “first *step*”, “*step into*”, or “*go a little bit deeper*” that would otherwise have been classified as non-deliberate uses to be reconsidered as deliberate metaphor instances. Example (10) includes a selection of talk fragments distributed across different times to illustrate the development and elaboration of the metaphor throughout the talk (the example illustrates realisations of the metaphor that extend over 20 minutes in the talk, but they are in fact used throughout the whole talk).

- (10) [05:16] So, as we *enter the now*, we become more aware of our surroundings, it's the *first step into the now*.  
 [5:58] ...and that *is the space of now*, so, the word to use is “you acknowledge what is”, and close to acknowledging you *go a little bit deeper*...  
 [06:20] And there are some people who are not yet ready to *go any deeper* than that, and that's fine, it's enough to just *go two steps into the now*; *well, let's just call it three (people like the idea of steps), let's call it step one*: you become more aware of your surroundings; *step two*: you acknowledge, and with the acknowledgement comes a sense of the goodness of the ways in which life manifests continuously around you. And then you *go a little deeper still: the next step takes you deeper*.<sup>11</sup>  
 [15:34] Now, the next *step down into the now* is...  
 [20:23] ... and now we *can go even deeper*: sense perceptions, acknowledging, appreciating, feeling the energy field in the body, all these things *arise in the present moment*.  
 [29:12] So, when we *go to the deepest level of the present moment, when you go to the deepest level of the moment, what do you find?* And what do you *discover?* Yourself, the essence of you. So, the essence of you, which is conscious space, is inseparable from the essence of the present moment. (Tolle, 2014, March 2)

The expressions in (10) help to see Tolle's characterisation of the “*journey into the now*” as a vertical journey, that is, an exploratory process that takes the meditator deeper into the present moment until s/he can discover her/his very essence. But the italicised THE PRESENT MOMENT IS A SPACE expressions work in the talk not only as reconceptualisation boosters; they are also used recurrently by Tolle throughout the talk as discourse structuring devices to create textual cohesion and to produce a coherent line of reasoning based on an expected script of steps that need to be taken throughout the process.

Repetition is a key factor in this example. On the one hand, and in tandem with the activation of the JOURNEY metaphor, repetition allows the SD (the notion of physical *space* and the notion of a *journey* down into a physical space) of conventional expressions like “*first step*” and “*go deeper*” to be revitalised and then reconsidered as an important area of experience with which the TD is revisited or understood. In other words, repetition works here as a metaphorical extendedness mechanism that allows highly conventional and

<sup>11</sup> At the point where Tolle states “*people like the idea of steps*”, Tolle gesticulates with his hands, which reveals his awareness of this *loose use* of language, and he and his audience start laughing. This reinforces the awareness of all the parts involved in this communicative event of the fact that Tolle is using metaphorical language, and is deliberately playing on words (metaphorical words) to make clear that there are no “real steps” in meditation, but that the use of words (“*let's call it step x*” / “*people like the idea of steps*”) may help the audience to reconceptualise meditation and the present moment in terms of previous and more familiar experiences (i.e., journeys, stairs-steps, and “profound spaces”).

otherwise non-deliberate expressions to become salient elements in the meaning of the utterance, and hence be reconsidered as deliberate metaphor uses (see Reijnierse 2017, 2020 for an extended account of the role of repetition, extendedness, and co-text in deliberate metaphor identification and use). On the other hand, repeating metaphorical expressions of progress (e.g., *go, first, second*, etc.) and depth (e.g., *step down, go deeper, the deepest level*) at very distant points in the talk make these expressions work as *attractors* (Cameron & Deignan 2006). That is, their use allows the speaker to echo previous explanations and quickly compress inferences so that the audience can understand the whole talk as a full informative unit: the talk has allowed the audience to experience “the meditation journey” and to take home a structured set of steps to continue the journey on their own.

The last line in (10) explicitly associates the essence of the present moment with the essence of the meditator, and both are described as “conscious space”. This is a recurrent idea in the talks that is often also developed through the metaphorical characterisation of *THE MEDITATOR AS SPACE*. Whereas conventional and non-deliberate uses of metaphor like (6) simply refer to *UNDEFINED SPACES*, deliberate uses like (11) and (12) often portray novel conceptions of the meditator as *SPACIOUSNESS*.

(11) We’re not trying to stop sensations or emotions, we’re not trying to have perfect feelings, are not trying to have enlightened feelings, we’re not trying to get rid of feelings. The same thing... *we’re being the space* for all feelings, any feelings as they arise, stay for a while, express themselves, and pass. (Foster, 2015, January 3, 5:12).

(12) ... all sounds or smells, all tastes, all memories, all ideas about the future, *they’re all allowed, they’re all embraced, they are all held in this vastness, in this vastness that you are*. (Foster, 2015, January 3, 7:00)

This process of reification often brings forth an emphasis on certain attitudes that the meditator is expected to develop through the process of meditation, like adopting an open-hearted, non-reactive stance towards any kind of inner event that s/he experiences. The attitudes of the meditator are thus characterised as qualities of the space with which the meditator identifies. Accordingly, the meditator in (11) and (12) is pictured as a vast, open, or boundless space in which all phenomena (even affective ones) are to be “welcomed”.

### 3.2.2. Home Metaphors

“Welcoming” all inner phenomena (as if they were part of the meditator’s inner home) is a widespread metaphorical image in the talks that draws on the *HOME* scenario, one of the most productive sources of metaphorical images in the corpus. The rich structure of this domain in terms of participants and roles (home, owner-host, receiving/welcoming guests-visitors, etc.) is congruent with its exclusive realisation of *HOME AND SHELTER* through deliberate metaphors in discourse, as shown in Table 5, which ties in with the trend of greater SD detail and elaboration in deliberate uses described in studies like Silvestre-López (2019, 2020).

<b>TD \ SD</b>	<b>SAFE STRUCTURES: HOME AND SHELTER</b>	<b>NDM</b>	<b>DM</b>
THE MEDITATOR IS...	21	0	21
MEDITATION IS...	13	0	13
ATTENTION IS...	8	0	8
AWARENESS IS...	3	0	3
THE PRESENT MOMENT IS...	2	0	2
THE MIND IS...	2	0	2
<b>TOTAL INSTANCES</b>	<b>49</b>	<b>0</b>	<b>49</b>

Table 5. HOME AND SHELTER conceptual metaphors and non-deliberate (NDM)/deliberate (DM) uses in the talks.

This space-based model also draws on the JOURNEY metaphor, and it basically helps to profile THE MEDITATOR as a safe space in which to take shelter, namely, his/her own HOME. If THE MEDITATOR IS HOME (one of the journey spaces), then MEDITATION is conceptualised as the journey process in which the meditator comes back home (MEDITATION IS COMING BACK HOME).

As mentioned in previous sections, the talks address the notion of meditation from a beginner's perspective. Meditation thus involves a series of practices aimed at regulating the novel meditator's attention while also helping him/her to adopt an open, non-striving attitude towards the practice. THE MEDITATOR IS HOME and MEDITATION IS COMING BACK HOME are enriched metaphorical elaborations of a highly-entrenched metaphor related to everyday processes of awareness, attention, and distraction, namely, PAYING ATTENTION IS BEING HERE/ BEING DISTRACTED IS BEING AWAY. It is thus not surprising that notions like AWARENESS and ATTENTION also appear as significant TDs that combine with the HOME AND SHELTER SD. Example (13) illustrates how the PAYING ATTENTION IS BEING HERE metaphor is used first by the speaker to talk about the process of regaining attention ("come back"), and then how this conventional, non-deliberate metaphorical use is elaborated deliberately by drawing on the HOME SD ("coming back" becomes "coming home"), as illustrated by expressions like "coming home" or "being intimate with".

- (13) And what would it be like if instead of racing through to the end line we learn to pause and *come back*? *Come back*. Okay... Let's *come back here*... And what would it be like if we paused and *came back* and really just *stayed* with what's here, enough? [...] And this is the promise, in a way, of a path of presence, that we give ourselves this blessing of *coming home* and *being intimate with our inner* life and also *intimate with* the world around us? (Brach, July 10, 49:38)

Although the HOME scenario is used by all the speakers in the corpus, it is particularly relevant in Brach's talks (see Table 2 and its description in section 3.1.). This is a very prolific image that has already been found to permeate different social spheres of meditation practice (see Silvestre-López 2019, 2020), and which, in the case of Brach's talks, is taken from a



famous poem by the 13<sup>th</sup>-century Sufi poet Rumi. Not only does Brach mention Rumi as the inspiration for this metaphor in her talks, but she also states explicitly that the talk series is entitled “Introduction to Meditation: Do you make regular visits to yourself?” after this poem. Example (14) reproduces this acknowledgement and illustrates how the metaphors drawing on the HOME SD in her talks become a prototypical case of *topic-driven* deliberate metaphor uses (Reijnierse 2017).

- (14) And one of the lines I most love, this is “*Rumi*”, is this, he says: “*Do you make regular visits to yourself?*” That’s it. Isn’t that wonderful? *Do you make regular visits to yourself?* So, I’m re-titling these two classes with that: “*Do you make regular visits to yourself?*” And what we find is that there’s suffering when we leave home - and we all leave home. I mean, if you look at today... What was today like? I mean, how many moments did you leave your body in some way? Just lose [sic] track of your body and your senses? How many moments did you leave your heart - leave the kind of tender quality of the heart, that sensitivity? We leave home. We leave... We disconnect from our belonging to this natural world, to the Earth. We leave home. And what happens is that when we leave home regularly, we lose touch with what we most value. (Brach, July 3, 6:53)

Besides the clear topic-driven metaphor case, deliberateness is also made apparent in the example through other textual characteristics like conceptual juxtaposition (target and source concepts are simultaneously tackled throughout the fragment), elaboration (explanations of what “making regular visits” means and of the implications of “leaving home”), or extendedness (the conceptual metaphor smears over the fragment with a range of metaphorical expressions like “*leave home*” and “*leave your body*” that are repeated to put her points across). The vivid image evoked by the HOME metaphor is thus a perspective changer used with the clear intention of fostering the reconceptualisation of the processes of ATTENTION, MEDITATION, and THE MEDITATOR. Moreover, the talk title (and topic) and the repeated echoes of the metaphor throughout her two talks work as well as effective discourse structuring devices in that they provide textual cohesion and argument coherence.

Interestingly, the HOME SD does not simply remain as a reverberation of a well-established traditional model in contemplative traditions. Brach manages to recontextualise (Semino et al., 2013) it in various ways, as in (15).

- (15) Now, what’s your attitude when you’re visiting yourself? Because you might consider, if you’re visiting another person, what’s your attitude? You know, if it’s Aunt Myrtle who you’re doing out of duty, right? You’re visiting her out of duty and you’re thinking: “Oh, she’s just going to repeat the same stories and she’s kind of judgemental.” And [you] could be talking about your own mind, right? Same story: it’s judgemental. But you know you’re going to get bored or you know you’re going to get reactive. What’s your attitude in going towards that? So, we start to sense that when we visit ourselves, it might not always be comfortable what we come to. And what’s our approach to that? (Brach, July 3, 11:30)

Brach craftily uses the metaphor to address two important attitudinal components in early meditation practice by specifying particular details of a fully-fledged scenario (probably well-known to her audience), namely, the non-judgemental and non-reactive attitudes of the practice (Kabat-Zinn, 2003). Moreover, comparing the meditator’s judgemental mind to aunt Myrtle’s attitudes and behaviour also allows Brach to “split roles”, so that the meditator

is no longer identified with his/her mind. This disidentification process is also an important step in early meditation practice in that it may favour the development of more desirable attitudes and perspectives towards the practice.

THE MEDITATOR is the prototypical HOME in most examples, but some parts of the meditator are also often profiled as salient instances as, for example, *the body*, *the heart* (as in example 14), or *the breath* (15). In the corpus, metonymies typically combine with HOME and JOURNEY metaphors to address the process of attention, licensing the metonymically motivated metaphors like OBJECTS OF ATTENTION ARE PARTS OF THE MEDITATOR'S HOME, vehicled, for example, with novel metaphorical expressions like “home base” in (15).

- (15) And what we can do [...] is pick a *home base* - and there are many *home bases* that you can choose from. Some of you [...] will choose to be with *the breath* - that you'll let *the breath* be right at the centre, kind of in the foreground, and you'll let... When *the mind travels*, you'll go: “Oh, *come back, come back*, and relax”, and you'll *find your way back* to the breath. *The breath is a valuable home base*, especially if you want to refine the attention and get it very, very quiet and precise [...] and just know that whenever *the mind travels* and you want to have a *place to come home* to, you can feel your body right here (Brach, July 3, 37:27)

Finally, meditation can also be seen as a safe space, a place to find refuge and relief when in troubled times. This facet is highlighted not only in the MEDITATION and MEDITATOR TDs, but through the metaphors AWARENESS IS HOME and THE PRESENT MOMENT IS HOME, as in (16), (17) and (18).

- (16) This is, *in some sense, a new place to take refuge, a new place to reside* that's always been here [...]. The capacity we've had all along of which requires a certain kind of cultivation to learn how to *inhabit our own awareness*. (Kabat-Zinn, March 30, 8:40)
- (17) So, seeing if right *in this moment* there isn't a certain quality of *refuge*, of actually being *out of the storm, out of the ferocity of the winds and the turbulence of what we're experiencing...* (Kabat-Zinn, March 30, 14:00)
- (18) So, just this moment, just this breath coming in, just this breath leaving the body, just this *resting in this timeless moment we call now* [...] and *taking up residency here*, as if your very life depended on it. (Kabat-Zinn, March 31, 13:36)

Overall, HOME AND SHELTER metaphors are realised in the talks as rich deliberate elaborations that promote the reconceptualisation of meditation as a journey to revitalise positive back-home aspects and thus establish a positive affective connection with meditative practice. In other words, this reconceptualisation process has the function of highlighting desirable aspects of meditation practice towards present-moment experience (like warmth and openness), as well as evoking the protection sensation we feel when we return home after having been away on a long trip or for a long period of time.

### 3.2.3. *Natural Space Metaphors*

The generic notion of space is also represented metaphorically in the talks through more elaborated nature-related scenes. When metaphorical SDs belonging to the realm of nature

are used deliberately,<sup>12</sup> the emphasis is placed on the *lack of agency* with which natural phenomena take place. In other words, metaphorically characterising TDs like those shown in Table 6 allows the speakers to reconceptualise them as the spaces in which natural and weather phenomena simply happen, without any human agent causing or controlling the process.

SD \ TD	NATURAL SPACE: MASS OF WATER	NDM	DM
THE MEDITATOR IS...	17	4	13
THOUGHT IS...	7	1	6
AWARENESS IS...	4	3	1
THE MIND IS...	3	0	3
THE PRESENT MOMENT IS...	2	2	0
<b>TOTAL INSTANCES</b>	<b>33</b>	<b>10</b>	<b>23</b>

SD \ TD	NATURAL SPACE: FIELDS AND LANDSCAPES	NDM	DM
AWARENESS IS...	5	3	2
MEDITATION IS...	2	0	2
THE PRESENT MOMENT IS...	1	1	0
THE MEDITATOR IS...	1	0	1
<b>TOTAL INSTANCES</b>	<b>9</b>	<b>4</b>	<b>5</b>

SD \ TD	NATURAL SPACE: SKY	NDM	DM
THE MIND IS...	2	0	2
AWARENESS IS...	1	0	1
THE MEDITATOR IS...	1	0	1
<b>TOTAL INSTANCES</b>	<b>4</b>	<b>0</b>	<b>4</b>

Table 6. MASS OF WATER, FIELDS AND LANDSCAPES, and SKY conceptual metaphors and non-deliberate (NDM)/deliberate (DM) uses in the talks.

This is indeed the common ground upon which metaphors like THE MEDITATOR IS A LANDSCAPE (19), THE MEDITATOR IS THE SKY (20), and THE MEDITATOR IS THE OCEAN (20 and 21) are built. Such images, moreover, allow speakers to highlight the *ephemeral* nature of all types of present-moment experience manifestations. This is illustrated, for example, in (21) where objectively fleeting experiences like sounds and smells, along with more abstract inner events like emotions, thoughts, and feelings are all described as entities that *come* and *go*, *arise* and *dissolve*, *appear* and *disappear* in the *everlasting* space in which they are allowed to exist.

(19) So, I often *think of us, of our inner life, as a kind of wilderness*, and there's all parts of our being, *there's inner weather*, there's these vulnerable parts of us... (Brach, 2012, July 10, 24:25)

(20) Many people think that meditation is about *stopping thought*, but *the sky doesn't need to stop the clouds, the ocean doesn't need to stop its waves. We're not trying to stop thoughts*. (Foster, 2015, January 3, 4:45)

(21) Now I'd invite you right now, again just as an experiment, without any expectation, just to begin to notice what's appearing in present experience, that "*the waves*" – *as I call them* – of

<sup>12</sup> Natural-space-related non-deliberate uses belong to conventional expressions widely used in everyday language such as "the field of awareness" (AWARENESS IS A FIELD-LANDSCAPE). Due to space constraints, and because deliberate uses are more representative of the "natural spaces" SD category, the discussion here deals mainly with deliberate uses.

*present experience, all the waves that are appearing in the ocean that you are. So, thoughts, sensations, feelings, sounds, smells: these are all just waves in the ocean that you are. So just begin to notice these waves appearing and disappearing, notice thoughts coming and going. [...] Just notice thoughts arising and dissolving. [...] And what you begin to notice is that thoughts come and go; is that it seems to be the nature of thought; it seems to be the nature of these waves, is [sic] that they appear and disappear. And what you are, somehow, doesn't appear and disappear. What you are, you could say, is the space in which waves, in which thoughts are allowed to appear and disappear. It's like what you are is the ocean in which all these waves are allowed to come and go. Not because you are allowing them: they're naturally allowed. They're naturally allowed to arise and dissolve in you. (Foster, 2012, May 24, 2:50-5:04)*

The same holds for AWARENESS – AWARENESS IS THE SKY (22), AWARENESS IS THE OCEAN, 23) – and MIND – THE MIND IS THE SKY (24), THE MIND IS THE OCEAN (25).

- (22) So, we can simply rest in awareness and let whatever unfolds unfold, *as if our awareness... sometimes the images used is "like the sky"*. (Kabat-Zinn, 2020, March 31, 20:24)
- (23) And yet if we *drop down below the surface*, there's *gentle undulations and stillness, and that's as true for our awareness, if not much more true than it is for the ocean*. And so, even in the midst of the full catastrophe of the human condition, and we certainly are, we can, in taking our seats, really take a stand in our relationship to whatever's going on, and the *turbulence* and the challenges of it at all, and find a kind of domain of strength and clarity and stability and calm [sic]. (Kabat-Zinn, 2020, March 30, 3:35)
- (24) And I think I might have mentioned this yesterday: that *storms can arise and weather patterns in the space of the sky, of the atmosphere, turbulence, calmness, light, darkness, rain, snow... and in the mind*, also, again depending on conditions in the world [...]. For now, simply let it all play out *in the space of the mind, like weather patterns*, and resting in awareness that is capable of holding it all (Kabat-Zinn, 2020, March 31, 20:35)
- (25) That's just *the nature of the mind: to wave; just like it's in the nature of the ocean to wave, depending on the atmospheric conditions and storms; and so, the mind can be very turbulent at the surface but just like the ocean: drop down twenty or thirty feet, and even in tempest or hurricane there's simply gentle undulations [...]*. The more we work with whatever *turbulence and resistance, or impatience, or whatever it is, boredom, that arises, those are all just waves on the surface of the ocean of the mind...* (Kabat-Zinn, 2020, March 30, 19:55-21:47)

Natural-space metaphors like the ones described above are powerful inner-event reificators in that the vivid, quasi-perceptual SD imagery evoked (e.g., the sky and the clouds, or the ocean and the waves) allows emotional and mental events to be reconsidered as if they were impersonal phenomena whose existence does not depend on the meditator. In the examples involving the SKY SD, moreover, the physical distance between the clouds and the observer also helps to highlight a convenient distance between the thinker and the thoughts which – in tandem with non-agentive relationship described above – promotes the representation of the meditator as a passive, non-reactive observer (instead of an active agent involved in the process of *generating* and *clinging to* thoughts).

By recalling relatively pleasant experiences, the images used in examples like (20), (21) and (22) facilitate the transfer of this feeling of tranquillity to the practice of meditation. This helps to reconceptualise mental contents as entities beyond the meditator's control, but most importantly, it also helps to stimulate a positive attitude towards the practice, since mental phenomena are not always pleasant experiences. The affectively positive and negative valence of mental events is illustrated by comparing the relative calm induced in (20), (21), and (22) and the images of turbulent atmospheric phenomena in (19), (23), (24), and (25). Finally, it is worth noting that, in the talks, these kinds of negative images are usually evoked by the speakers to refer to ordinary, non-meditative states of consciousness. A clear example is the *THOUGHT IS A MASS OF WATER* metaphor.

- (26) Some people are so out of touch with their surroundings... they are never where they are, they are *in the stream of their thinking*, with only enough attention not to bump into things. And a lot of that [is] negative too, critical, critical, complaining, having an issue with this or that, thinking about how unpleasant was yesterday, or 10 years ago, and thinking what might go wrong tomorrow, or tonight, in all kinds of [sic], it takes you to that. That consciousness is continuously *being absorbed by thinking* and people are *in the grip of the stream of thinking*, it *pulls you along* amplified these days by the digital gadgets. [...] and *you're drowning in stuff*, and I'm not meaning physical stuff, some people also *drown in physical stuff*, but *you're drowning in mental stuff*, one thought after another, one text message, after another one... (Tolle, 2014, March 2, 9:27-11:57)

This example illustrates very vividly how the metaphorical nature (initially concealed, due to the entrenchment of this expression in everyday language use) of the highly-conventional expression *stream of thought* becomes more apparent and explicitly metaphorical through the repeated use of *MASS OF WATER*-related expressions like *drown*, *absorb*, *pulls along*, or *stream* in the production of the meditation teacher. The image of thought as a *stream* in which the thinker is *absorbed*, *pulled along*, or even *drowned* becomes a stronger source in the speaker's mental and discursive representation of thought. It thus becomes part of the referential meaning of the utterance and takes part in its purported representation in the receiver's mind, which evidences, in turn, and as pointed to in the DMT framework (Reijnierse *et al.* 2018) the progressively more deliberate nature of the metaphor.

#### 4. CONCLUSION

The bottom-up qualitative analysis approach to metaphor analysis adopted in this study has made it possible to reveal the ways the joint use of deliberate and non-deliberate metaphor can foster understanding, facilitate communication and allow speakers to achieve particular communicative ends in their talks. *UNDEFINED SPACE-SPACIOUSNESS* METAPHORS are mainly used in the corpus to highlight attitudes of the acceptance of present-moment experience, particularly in relation to TDs like *THE PRESENT MOMENT*, *THE MEDITATOR*, *THE MIND*, *AWARENESS* and *CONSCIOUSNESS*. These TDs are characterised metaphorically as spaces in which all present-experience phenomena are allowed to arise and take place. The results and discussion section has also illustrated how the *UNDEFINED SPACE/SPACIOUSNESS* SD can also be elaborated through more detailed scenery related to the domain of nature. In these cases, the SDs *MASS OF WATER*, *LANDSCAPE*, and *SKY* are used in the talks to also emphasise

the need of the mediator to *become the space* in which all physical and mental perceptions are allowed. The peculiarity of nature-related conceptual metaphors lies in their particular emphasis on non-volitional aspects. That is to say, portraying THE MEDITATOR, AWARENESS, THE MIND, OR THE PRESENT MOMENT as natural spaces helps to remove the agency component from the meditation scene. This hiding/highlighting capacity of metaphors is skillfully exploited in the talks to favour the understanding (in a non-expert audience) of the need to adopt an open and non-reactive attitude towards all present experience phenomena. The conceptual metaphors drawing on the HOME AND SHELTER SD also highlight an open or *welcoming* attitude to present-moment experience. However, rather than downplaying agency, the cosiness and familiarity of home-related experiences is exploited to evoke warm feelings and a friendly attitude towards all kinds of mental and perceptual phenomena. Finally, another important aspect emphasised in HOME metaphors is the central role of the meditator, who is portrayed as both the departure - and endpoint of the journey metaphor.

Since its inception, DMT has had a variety of both supporters and detractors and has generated a heated academic debate that is still active (see Steen, 2017). This paper takes the stance that deliberate uses of metaphor do have an important bearing on communication. The importance of incorporating the non-deliberate/deliberate distinction in metaphor analyses has been illustrated by placing special analytical focus, on the one hand, on domain granularity differences and, on the other, on how deliberate uses are realised discursively with a variety of functions.

The unveiling of SD granularity differences would not have been possible without the methodological approach adopted here, that is, a qualitative bottom-up analysis involving a flexible approach to TD and SD identification and coding. The findings subscribe to the aforementioned tendency for non-deliberate uses to appear as less-defined domains or as vaguer elaborations in discourse, which is illustrated mainly in the analysis described in section 3.2.1. For example, some conceptual metaphors are elaborated in the corpus mainly through conventional mappings both in non-deliberate and deliberate discourse realisations. In such cases (e.g., THE PRESENT MOMENT IS SPACE metaphor), non-deliberate realisations instantiate, on the whole, broad elaborations of conventional expressions of the TD-SD mapping (see example 5), whereas deliberate uses exploit more dimensions of the mapping, thereby describing more qualities of the TD in terms of salient SD features (see example 10). Examples like these illustrate how non-deliberate and deliberate discourse realisations of the same metaphor can both draw on conventional mappings. However, the contrast between non-deliberate and deliberate uses is commonly grounded in novelty differences. Specifically, non-deliberate metaphors are usually realised through conventional metaphors and deliberate uses through more creative TD-SD elaborations like, for example, the contrasts between (6) and (11, 12) in THE MEDITATOR IS SPACE.

The strong discourse and communicative implications of deliberateness in discourse have also been made apparent by analysing the functions with which deliberate metaphors are used in the talks. Due to the instructional character of the talks, ideational functions have been identified as a pervasive category, particularly those involving the reconceptualisation of a topic according to the SD perspective (usually associated with the reproduction of vivid, quasi-perceptual physical SD imagery) in order to promote understanding or to trigger an attitudinal change towards the practice. While these sometimes overlap with the interpersonal dimension, many deliberate uses have also been found to fulfil a range

of functions traditionally belonging to the textual dimension, such as creating textual cohesion and providing for argument structuring by compressing inferences.

Overall, these findings suggest that metaphorical language is part and parcel of meditation instructional discourse. Thus, it may be desirable to undertake further research aimed at raising the awareness of instructors and meditation teachers about the importance and implications of the use of metaphor for pedagogical purposes, but also in terms of exploring the potential ways in which the deliberate use of metaphoric images may have a bearing on the understanding of the practice.

## ACKNOWLEDGEMENTS

This research was carried out within the framework of the projects GV/2019/101 (Generalitat Valenciana) and UJI-B2018-59 (Universitat Jaume I).

## REFERENCES

- Cameron, L. & A. Deignan. (2006). The emergence of metaphor in discourse, *Applied Linguistics*, 27(4), 671-690. <https://doi.org/10.1093/applin/aml032>
- Cebolla, A.J., J. García-Campayo & M. Demarzo. (2014). *Mindfulness y ciencia: De la tradición a la modernidad*. Alianza.
- Charteris-Black, J. (2004). *Corpus approaches to critical metaphor analysis*. Palgrave Macmillan.
- Coll-Florit, M. & S. Climent. (2019). A new methodology for conceptual metaphor detection and formulation in corpora. A case study on a mental health corpus. *SKY Journal of Linguistics*, 32, 43-74.
- Dahl, C.J., A. Lutz & R. J. Davidson.. (2015). Reconstructing and deconstructing the self: cognitive mechanisms in meditation practice. *Trends in Cognitive Sciences*, 19(9), 515-523. <https://doi.org/10.1016/j.tics.2015.07.001>
- García-Campayo, J. (2020). *Vacuidad y no dualidad. Meditaciones para deconstruir el “yo”*. Kairós.
- Goatly, A. (2011). *The language of metaphors* (2nd Ed.). Routledge.
- Grady, J. (1997). Theories are buildings revisited, *Cognitive Linguistics*, 8(3), 267–290.
- Johnson, M. (1987). *The body in the mind*. University of Chicago Press.
- Kabat-Zinn, J. (2003). Mindfulness-based interventions in context: Past, present, and future. *Clinical Psychology: Science and Practice*, 10, 144–156. <https://doi.org/10.1093/clipsy/bpg016>
- Kimmel, M. (2012). Optimizing the analysis of metaphor in discourse: how to make the most of qualitative software and find a good research design. *Review of Cognitive Linguistics*, 10(1), 1-48. <https://doi.org/10.1075/rcl.10.1.01kim>
- Kövecses, Z. (2020). *Extended conceptual metaphor theory*. Cambridge University Press.
- Lakoff, G. & M. Johnson. (1980). *Metaphors we live by*. University of Chicago Press.
- Lakoff, G. (1993). The contemporary theory of metaphor. In A. Ortony (Ed.), *Metaphor and thought* (2nd ed., pp. 202-251). Cambridge University Press.
- Langacker, R. (2008). *Cognitive grammar. A basic introduction*. Cambridge University Press.
- Matko, K. & P. Sedlmeier. (2019). What is meditation? Proposing an empirically derived classification system. *Frontiers in Psychology*, 10, 2276. <https://doi.org/10.3389/fpsyg.2019.02276>
- Musolff, A. (2006). Metaphor scenarios in public discourse. *Metaphor and Symbol*, 21(1), 23-38. [https://doi.org/10.1207/s15327868ms2101\\_2](https://doi.org/10.1207/s15327868ms2101_2)
- Musolff, A. (2016). *Political metaphor analysis. Discourse and scenarios*. Bloomsbury.
- Peña-Cervel, M.S. (2003). *Topology and cognition: What image-schemas reveal about the metaphorical language of emotions*. Lincom Europa.

- Reijnierse, W. G. (2017). *The value of deliberate metaphor*. LOT Dissertation Series, 469. Ph.D. thesis. Utrecht: LOT. Retrieved from [https://www.lotpublications.nl/Documents/469\\_fulltext.pdf](https://www.lotpublications.nl/Documents/469_fulltext.pdf)
- Reijnierse, W. G., C. F. Burgers, T. Krennmayr & G. J. Steen. (2018). DMIP: A method for identifying potentially deliberate metaphor in language use. *Corpus Pragmatics*, 2(2), 129-147. <https://doi.org/10.1007/s41701-017-0026-7>
- Reijnierse, W. G., C. F. Burgers, T. Krennmayr & G. J. Steen. (2020). The role of co-text in the analysis of potentially deliberate metaphor. In C. Di Biase-Dyson & M. Egg (Eds.), *Drawing attention to metaphor: Case studies across time periods, cultures and modalities* (pp. 15-38). John Benjamins. <https://doi.org/10.1075/ftl.5.02rei>
- Semino, E. (2008). *Metaphor in discourse*. Cambridge University Press.
- Semino, E., A. Deignan & J. Littlemore. (2013). Metaphor, genre, and recontextualization. *Metaphor and Symbol*, 28(1), 41-59. <https://doi.org/10.1080/10926488.2013.742842>
- Silvestre-López, A. J. (2016). The Discourse of Mindfulness: What Language Reveals about the Mindfulness Experience. In P. Ordóñez-López & N. Edo-Marzà (Eds.), *New insights into the analysis of medical discourse in professional, academic and popular settings* (pp.173-198). Multilingual Matters.
- Silvestre-López, A. J. (2019). Deliberate metaphors in Buddhist teachings about meditation. In I. Navarro (Ed.), *Current approaches to metaphor analysis in discourse* (pp. 205–234). Mouton de Gruyter.
- Silvestre-López, A. J. (2020). Conceptual Metaphor in Meditation Discourse: An Analysis of the Spiritual Perspective. *Gema Online® Journal of Language Studies*, 20(1), 35-53. <http://doi.org/10.17576/gema-2020-2001-03>
- Silvestre-López, A.J. and I. Navarro-Ferrando. (2017). Metaphors in the conceptualization of meditative practices. *Metaphor and the Social World*, 7(1), 26-46. <https://doi.org/10.1075/msw.7.1.03sil>
- Silvestre-López, A.J., D. Pinazo & A. Barrós-Lorcerales. (2021). Metaphor can influence meta-thinking and affective levels in guided meditation. *Current Psychology* <https://doi.org/10.1007/s12144-021-01655-1>
- Steen, G. J. (2015). Developing, testing, and interpreting deliberate metaphor theory. *Journal of Pragmatics*, 90, 67-72. <https://doi.org/10.1016%2Fj.pragma.2015.03.013>
- Steen, G. J. (2017). Deliberate Metaphor Theory: Basic assumptions, main tenets, urgent issues. *Intercultural Pragmatics*, 14(1), 1–24. <https://doi.org/10.1515/ip.2017.0001>
- Steen, G. J. (2018). Attention and deliberateness in metaphor processing. In V. Cuccio (Ed.), *Attention to metaphor. From neurons to representations* (pp. 89–109). John Benjamins. <https://doi.org/10.1075/milcc.7>
- Steen, G. J., Dorst, A. G., Herrmann, J. B., Kaal, A. A., Krennmayr, T. & Pasma, T. (2010). *A Method for Linguistic Metaphor Identification: From MIP to MIPVU*. John Benjamins. <https://doi.org/10.1075/celcr.14>
- Vargas-Sierra, C. & A. Moreno-Sandoval. (2021). War and health metaphors in financial discourse: The case of “Letter to Shareholders” in annual reports. In J. Mateo & F. Yus (Eds.), *Metaphor in Economics and Specialised Discourse* (pp. 41-71). Peter Lang.

## Corpus references

- Brach, T. [Tara Brach] (2012, July 3). Introduction to Meditation: Do you make regular visits to yourself? (Part 1). [Video file]. [https://www.youtube.com/watch?v=DKR5mcu2J60&ab\\_channel=TaraBrach](https://www.youtube.com/watch?v=DKR5mcu2J60&ab_channel=TaraBrach)
- Brach, T. [Tara Brach] (2012, July 10). Introduction to Meditation: Do you make regular visits to yourself? (Part 2). [Video file]. [https://www.youtube.com/watch?v=YmHTmYKr5JM&ab\\_channel=TaraBrach](https://www.youtube.com/watch?v=YmHTmYKr5JM&ab_channel=TaraBrach) (January 5, 2022)
- Foster, J. [The Deepest Acceptance-Jeff Foster] (2012, May 24). “Who meditates?” A 20-minute alternative guide to meditation. [Video file]. <https://www.youtube.com/watch?v=XbQzZg0neKE> (December 05, 2021)



- Foster, J. [The Deepest Acceptance-Jeff Foster] (2015, January 3). True meditation has no goal. [Video file]. <https://www.youtube.com/watch?v=tZQshhxqQSs> (December 5, 2021)
- Kabat-Zinn, J. [Wisdom 2.0.] (2020, March 30). Jon Kabat-Zinn Meditation and Q & A |Episode 1. [Video file]. [https://www.youtube.com/watch?v=hcKGv06U\\_KU&list=RDCMUCEh14Grr46IolhzGhZ3agNw&start\\_radio=1&t=271s&ab\\_channel=Wisdom2.0](https://www.youtube.com/watch?v=hcKGv06U_KU&list=RDCMUCEh14Grr46IolhzGhZ3agNw&start_radio=1&t=271s&ab_channel=Wisdom2.0) (December 5, 2021)
- Kabat-Zinn, J. [Wisdom 2.0.] (2020, March 31). Jon Kabat-Zinn Meditation and Q & A |Episode 2. [Video file]. [https://www.youtube.com/watch?v=ZsTqCXxemGw&ab\\_channel=Wisdom2.0](https://www.youtube.com/watch?v=ZsTqCXxemGw&ab_channel=Wisdom2.0) (December 5, 2021)
- Tolle, E. (2009). *What is meditation? Realizing the life that you are*. [DVD-Video file]. Eckhart Teachings.
- Tolle, E. [Wisdom 2.0.] (2014, March 2). Meditation: Eckhart Tolle. [Video file]. <https://youtu.be/foU1qgOdtwg> (January 6, 2022).

## ¿De qué hablamos cuando divulgamos sobre lingüística? Análisis de un corpus de textos divulgativos y aplicaciones al estudio terminológico de la semántica léxica

What do we talk about when we communicate about linguistics as a science?  
Analysis of a corpus of science communication texts and applications to the  
terminological study of lexical semantics

MARTA SÁNCHEZ-SAUS LASERNA  
Instituto Universitario de Investigación en Lingüística Aplicada, Universidad de Cádiz, España  
[marta.sanchezsaus@uca.es](mailto:marta.sanchezsaus@uca.es)  
<https://orcid.org/0000-0002-1123-2873>

VANESA ÁLVAREZ TORRES  
Instituto Universitario de Investigación en Lingüística Aplicada, Universidad de Cádiz, España  
[vanesa.alvarez@uca.es](mailto:vanesa.alvarez@uca.es)  
<https://orcid.org/0000-0003-0477-4065>

### Resumen

En este trabajo presentamos la confección y el análisis de un corpus de textos digitales conformado por columnas periodísticas de corte divulgativo, escritas por lingüistas, que tratan sobre la lingüística como ciencia y que han sido publicadas en diarios digitales o revistas digitales de naturaleza divulgativa. Lo hacemos con dos objetivos: por un lado, observar qué aspectos del

### Abstract

In this paper we describe the compilation and analysis of a corpus of digital texts conformed by science communication articles written by linguists that discuss linguistic topics and that have been published in digital newspapers or digital magazines. The aims are twofold: on the one hand, to observe what aspects of linguistic work and research linguists are conveying

---

**Para citar este artículo:** Sánchez-Saus Laserna, M. y Álvarez Torres, V. (2022). ¿De qué hablamos cuando divulgamos sobre lingüística? Análisis de un corpus de textos divulgativos y aplicaciones al estudio terminológico de la semántica léxica. *ELUA*, (38), 73-98. <https://doi.org/10.14198/ELUA.22384>

Recibido: 31/03/2022 Aceptado: 14/05/2022

© 2022 Marta Sánchez-Saus Laserna, Vanesa Álvarez Torres



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

trabajo y la investigación lingüística estamos trasladando los propios lingüistas a la sociedad en los textos divulgativos que aparecen en la prensa *online* y, por otro, analizar en este corpus la presencia de términos de la semántica léxica, lo que nos permite establecer para ellos un estatus terminológico propio de la divulgación e incluir, en el trabajo terminográfico, ejemplos de uso de estos términos en contextos divulgativos. Esto lo hacemos dentro del marco que nos proporcionan los proyectos sobre terminología de la semántica y bases de datos terminológicas que se están desarrollando en el Instituto Universitario de Investigación en Lingüística Aplicada de la Universidad de Cádiz<sup>1</sup>.

**PALABRAS CLAVE:** lingüística digital, lingüística de corpus, textos divulgativos, terminología, semántica léxica

to society in the informative texts that appear in the online press, and, on the other hand, to analyze the presence of terms related to lexical semantics in this corpus, which allows us to establish for them a terminological status specific to science communication and to include, in the terminographic work, examples of the use of these terms in science communication contexts. We do this within the framework provided by the projects on semantic terminology and terminological databases that are being developed at the Instituto Universitario de Investigación en Lingüística Aplicada of the University of Cádiz.

**KEYWORDS:** digital linguistics, corpus linguistics, science communication texts, terminology, lexical semantics

## 1. INTRODUCCIÓN

### 1.1. Comunicación mediada por ordenador y lingüística de corpus

El mundo desarrollado actual no se puede comprender sin la comunicación digital. En este contexto, tanto general como también especialmente en el académico, científico, técnico y profesional, es inevitable que surjan nuevos géneros digitales (Belda-Medina y Casañ-Pitarch 2020), que constituyen una línea de investigación lingüística emergente y novedosa (Luzón y Pérez-Llantada 2019). Las nuevas tecnologías y, sobre todo, las nuevas formas de transmisión de la información de manera electrónica han supuesto la aparición de nuevos géneros de carácter digital y la adaptación al entorno digital de géneros ya existentes, como el artículo periodístico en la prensa digital, que es el que abordamos en este trabajo. Los estudios lingüísticos no han sido ajenos a esta irrupción de lo digital y empieza a ser frecuente encontrar investigaciones que se sitúan dentro del campo de la “lingüística digital”, que Crespo Miguel (2020: 151) define así:

Cuando hablamos de lingüística digital debemos entender un conjunto de técnicas y herramientas enfocadas a la manipulación de datos lingüísticos en formato digital, lo que incluye el almacenamiento, la representación, la manipulación y la difusión de este tipo de datos. También se ocupa de cómo extraer y representar la información lingüística en formato electrónico, así como de las mejores prácticas para trabajar con esos datos, utilizando todos los beneficios de las tecnologías actuales.

<sup>1</sup> Los dos proyectos en los que se encuadra esta investigación, cuyo IP, en ambos casos, es Miguel Casas Gómez, son los siguientes: 1) “Lingüística y nuevas tecnologías de la información: la creación de un repositorio electrónico de documentación lingüística” (FEDER-UCA18-107788) de la Convocatoria 2018 de Ayudas a proyectos de I+D+i en el marco del Programa Operativo FEDER Andalucía 2014-2020, y 2) “Lingüística y Humanidades Digitales: base de datos relacional de documentación lingüística” (PY18-FR-2511) de la convocatoria 2018 de Ayudas a proyectos I+D+i (Modalidad “Frontera Consolidado”) del Plan Andaluz de Investigación, Desarrollo e Innovación.

Más concretamente, el marco de estudio de los géneros digitales está íntimamente vinculado con el empleo de la web como corpus, que se ha convertido en un tema de investigación muy fructífero en español, con distintos objetivos y desde distintas disciplinas de estudio. De Benito Moreno (2022: 481), en una interesante revisión de este campo a la que remitimos, señala que “la popularización de la comunicación digital ha conllevado el surgimiento de una nueva rama de la lingüística: el análisis de la comunicación y el discurso mediados por ordenador” y que la lingüística de corpus ha sido el marco metodológico en el que se ha desarrollado, debido a las grandes cantidades de texto que pone a nuestra disposición internet.

Los estudios que analizan las características propias del lenguaje empleado en las comunicaciones a través de internet comenzaron con el nacimiento de la propia internet. En inglés encontramos publicaciones que hacen referencia a la *computer mediated communication* (CMC) desde la década de 1980. Baron (1984) ya se centra en los posibles efectos que el empleo de ordenadores pueda tener para el lenguaje:

But there is yet a third dimension to the study of computers and language. That is the growing use of computers as conduits of natural language. Instead of speaking face-to-face with another or using traditional means of writing, we express ourselves at the computer keyboard. (Baron 1984: 119)

El acercamiento a la CMO supone, como hemos dicho, aproximarse a la lingüística de corpus. Como indican diversos autores (Parodi 2008, Zulaica Hernández 2015), la lingüística de corpus es una metodología de estudio de la lengua y el lenguaje que nos permite trabajar directamente en contextos auténticos con grandes cantidades de texto, asegurando la fiabilidad de los análisis cuantitativos y cualitativos y la comparación posterior con otros trabajos. Este enfoque es, además, de gran utilidad para el análisis de producciones orales y escritas de los hablantes de una lengua, como muestra la variedad de campos de estudio del ámbito de la lingüística que aplican los avances en la lingüística de corpus a sus investigaciones. En este trabajo vamos a enfocarnos hacia dos aplicaciones de los corpus: el análisis del discurso y el trabajo terminográfico.

Para el análisis del discurso son numerosísimos los trabajos que emplean corpus, tanto generales como especializados. Destacamos, como referencias básicas para el trabajo en español, los estudios de Mancera Rueda y Pano Alamán (2013) y (2014) y Pano Alamán y Moya Muñoz (2015), así como los de Cantamutto y Vela Delfa (2016) y Vela Delfa y Cantamutto (2015).<sup>2</sup> En nuestro caso, la perspectiva que adoptamos para el análisis de los temas tratados en el corpus es un análisis de contenido que parte de la extracción de las palabras clave o de las combinaciones léxicas significativamente más frecuentes en el corpus de estudio que en un corpus de comparación, asumiendo que existe una relación entre la frecuencia de la palabra y su relevancia en el discurso y en la construcción de representaciones (Scott y Tribble 2006; Rayson 2008; Duque 2014). En cuanto al uso de corpus para el trabajo terminográfico, como el que llevamos a cabo en la segunda parte de nuestro análisis, remitimos especialmente al trabajo de Cabré y Da Cunha (2022) –aunque nos centraremos en esto más abajo–.

El corpus que hemos confeccionado para este trabajo está formado por textos divulgativos escritos por especialistas en lingüística. Es cierto que ya existen para el español

<sup>2</sup> Llevamos a cabo una revisión más exhaustiva del uso de internet como corpus y, especialmente, de los textos extraídos de redes sociales, en Sánchez-Saus Laserna (2022).

proyectos en desarrollo muy interesantes basados en corpus que han tomado los textos periodísticos que tratan sobre la lengua como objeto de estudio, como los proyectos METAPRES y METAPRES-COLING<sup>3</sup> cuya IP es en ambos la profesora de la Universidad de Alicante Carmen Marimón. Estos proyectos están centrados en el análisis de las columnas sobre la lengua como género metalingüístico desde 1940 hasta la actualidad, aunque se encuentra aún en desarrollo<sup>4</sup>. Nuestro corpus es más modesto en tamaño, pero también es de naturaleza distinta: abarca solo textos de los últimos diez años, publicados directamente en prensa digital y escritos por lingüistas con una intención divulgativa.

## 1.2. La divulgación de la ciencia y los textos divulgativos como textos especializados

La divulgación científica tiene un papel fundamental para la mejora del conocimiento que la sociedad debe tener del mundo que le rodea. Es de gran relevancia que se conozca de primera mano la labor que realizan los investigadores, precisamente para que la sociedad valore la investigación en su justa medida. Esto, que es válido para cualquier disciplina, es especialmente importante para las humanidades y, en particular, para la lingüística, cuya imagen de cara a la sociedad está inexplicablemente alejada de la idea de investigación y se desconoce la aplicabilidad de muchas de sus disciplinas.

Sin embargo, la divulgación no es algo novedoso, pues, según indican Llácer Llorca y Ballesteros (2012: 58), se inició en 1632 con la obra *Dialogo sopra i due massimi sistemi del mondo tolemaico e copernicano* de Galileo Galilei, que, para llegar a un público más amplio, la redactó en italiano, y no en latín. Por su parte, Galán Rodríguez (2003: 138) señala que “la divulgación científica es una actividad indisoluble del proceso secular moderno que, a partir del siglo XVII, configuró en Europa un nuevo orden económico, político, científico, social y cultural”.

Aunque la divulgación de la ciencia no supla la formación realizada en las aulas por los docentes, ni convierta en especialista a un público lego en la materia, permite que “se privilegie la posibilidad de que el ciudadano pueda ubicarse racionalmente frente a la importancia de la ciencia y la tecnología en su devenir cotidiano, y que le permita recorrer y aprovechar la porción de conocimientos que le sean necesarios, útiles, o al menos interesantes” (Olmedo Estrada 2011: 139).

De este modo, difusión de la ciencia a la sociedad y los textos divulgativos están estrechamente relacionados, tanto que, como indica Galán Rodríguez (2021: 665), “el grado de acceso de la sociedad al conocimiento de una ciencia o una tecnología determina la posibilidad de producir textos divulgativos”. Esta autora emplea, en lugar de *ciencia*, un término más abarcador, *tecnociencia*, que genera dos tipos de discursos: 1) “de especialidad científica” y 2) “de divulgación” (Galán Rodríguez 2021: 657-658) y designa:

- 1) la actividad intelectual que trata de representar y explicar de forma objetiva el mundo y sus fenómenos;

<sup>3</sup> <http://metapres.es/>.

<sup>4</sup> El volumen de Marimón Llorca (2020) recoge resultados de las distintas direcciones en las que el equipo del Proyecto METAPRES ha abordado el examen del columnismo lingüístico como tradición discursiva. Por su parte, el proyecto de la Universidad de Cádiz “LinPePrensa. Ideas lingüísticas y pedagógicas en la prensa española del siglo XIX” analiza también las ideas lingüísticas en la prensa española, aunque se centra en textos del XIX. La polémica lingüística como tipología discursiva se analiza, por ejemplo, en Gaviño Rodríguez (2021).

- 2) el entramado del descubrimiento científico y sus aplicaciones sociales.

En cuanto a la divulgación de la lingüística, además de lo señalado anteriormente para la ciencia en general, si bien todos los hablantes de una comunidad no poseen un conocimiento específico de este campo de estudio, sí que cuentan con la facultad del lenguaje y un código lingüístico compartido por toda una comunidad de hablantes con el que se comunican entre sí: más que suficiente, por tanto, para que sea adecuado y pertinente transmitir a la sociedad los avances generados por la ciencia del lenguaje. Al fin y al cabo, como afirma Escandell Vidal (2011: XV):

El lenguaje está presente en todas las facetas de nuestra existencia: cada individuo es capaz de comunicarse en su lengua en multitud de situaciones y sin esfuerzo aparente. Sin embargo, es muy poco lo que el usuario común sabe de manera positiva y sistemática acerca de ese poderoso instrumento y de su propia capacidad de utilizarlo.

A partir de este razonamiento, el interés por los temas lingüísticos no solo debería surgir en lingüistas y filólogos u otros especialistas de otros ámbitos de estudio interesados en el lenguaje, sino también en los hablantes profanos en la materia. Así, si la sociedad actual apuesta por una ciencia al alcance de todos, en la que las nuevas tecnologías y el uso de nuevos medios de comunicación prestan servicio a la difusión de la misma, esto se presenta como una oportunidad para promulgar la proyección y el alcance del estudio científico del lenguaje a toda la sociedad.

Esta labor de divulgar el conocimiento científico se lleva a cabo, sobre todo, en los medios de comunicación. En ellos, divulgadores expertos en la materia transmiten conocimientos científicos al público en general de una manera más cercana y atrayente –adaptando un texto científico a un texto sobre ciencia para un público no experto en ese campo–, y, de este modo, se convierten en “traductores y propagandistas” (Galán Rodríguez y Montero Melchor 2002: 53). Sin embargo, hay que diferenciar la traducción realizada en el proceso adaptativo de los textos divulgativos de la traducción propiamente dicha, entre diferentes códigos lingüísticos, de ahí que Galán Rodríguez y Montero Melchor (2002: 59) establezcan las siguientes diferencias:

- a) mientras que en la traducción no hay ninguna referencia al texto primario (de hecho, deben ser dos textos distintos), en la divulgación se hace mención explícita a las fuentes originales porque, evidentemente, el divulgador actúa como un transmisor de la palabra de otros;
- b) el traductor es solo un intermediario sin voz propia que debe reflejar fielmente lo que se dice, pero no interpretarlo; el divulgador también es un intermediario, pero como su función es explicar un texto y adaptarlo (reformularlo), puede opinar;
- c) así como en la traducción no hay restricciones que afecten a la tipología o contenido de los textos, la divulgación está condicionada por los intereses del público.

Podemos añadir que los divulgadores, además de traductores y propagandistas, también son mediadores, en la línea que señalan Casas Gómez y Fernández Smith (2021: 724) cuando hacen referencia al género discursivo periodístico en el que se tratan temas específicos, como puede ser la ciencia.

De este modo, el redactor de textos divulgativos adecua a un público lego un contenido científico destinado en un principio a los propios científicos. Si bien los textos divulgativos

suelen estar redactados por expertos o periodistas con conocimientos específicos en la materia, ambos “se ajustan a un esquema semejante, con lo que las diferencias son de índole léxica y, en menor grado, gramaticales” (Galán Rodríguez y Montero Melchor 2002: 61). El divulgador, por tanto, se encuentra con la labor de mediar entre esos diferentes niveles de especialización, esto es, transmitir un conocimiento que cuenta con un alto grado de especialización a una sociedad que, por lo general, no cuenta con una formación tan específica en el ámbito en cuestión. Para ello, como indica Galán Rodríguez (2003: 150-152), el divulgador reformula o redefine haciendo uso de una serie de mecanismos metalingüísticos, “tipografía especial (comillas, letra itálica, paréntesis), verbos del ámbito semántico «llamar» (*llamar, llamarse, designar, nombrar, decir*, etc.), verbos que indican equivalencia o identidad (*ser, se define como, se considera como, consiste en, se trata de*), definiciones analíticas (por descomposición) y definiciones por la función”. Asimismo, el divulgador debe “reelaborar el sentido del trabajo científico original para que el texto divulgativo mantenga la coherencia semántica”, que implican reformulaciones no solo en el nivel léxico, sino también en otros niveles de análisis lingüístico, como el sintáctico.

Por tanto, si bien este tipo de textos se enmarca dentro de la comunicación especializada, caracterizada por el uso de un lenguaje especializado, debemos resaltar que el divulgador sí que hace uso de este tipo de lenguaje, pero los destinatarios del texto divulgativo, normalmente, al no contar con conocimientos específicos de la materia, carecen del lenguaje específico de ese ámbito. Esto constituye una muestra de la relación existente entre las dos dicotomías que algunos autores (Cabré 2004: 20) suelen relacionar: *lenguaje especializado/lenguaje general* y *conocimiento especializado/conocimiento general*.

### 1.2.1. Nivel de especialización de los textos de especialidad

El lenguaje especializado puede manifestarse en diferentes grados de especialización y, precisamente, el discurso divulgativo constituye, según Guantiva Acosta, Cabré y Castellà Lidon (2008: 18), el tercer nivel de especialización:

- 1) el discurso especializado (con distintos grados de especialización), dirigido a los especialistas;
- 2) el discurso de tipo didáctico, orientado al personal en formación, y
- 3) el discurso de carácter divulgativo, enfocado al público general.

En relación con los textos especializados y los textos divulgativos, Cabré (2002) establece algunas diferencias básicas entre ellos, que engloban la función, la temática, los usuarios, la situación comunicativa y el discurso y que pueden allanar el camino de nuestro estudio para comenzar con una base teórica fundamentada. No obstante, aunque es necesario determinar grados en el nivel de especialización de un texto, se ha señalado (Cabré 2002; Rodríguez Tapia 2016) que entre los distintos grados, en cuyos extremos están el texto especializado y el texto divulgativo, existe un *continuum*, en cuyo punto medio encontramos el texto semiespecializado, aunque con fronteras ciertamente difusas.

Los textos especializados pueden presentar mayor o menor grado de especialización dependiendo del conocimiento que tenga el destinatario sobre la materia. El destinatario puede tratarse de otro experto, un estudiante de universidad o una persona carente de conociemien-

tos específicos del campo en cuestión, de ahí que, como afirman Cabré y da Cunha (2022: 198), “esta diferenciación entre niveles de especialización de los textos es clave a la hora de definir los géneros textuales que contendrá el corpus”. Para estas autoras la producción científica dirigida a expertos en la materia, como los artículos científicos y las monografías científicas, serían géneros textuales de nivel de especialización alto; el material generado por docentes en el ámbito universitario (por ejemplo, apuntes de clase) se enmarcaría en el nivel intermedio; y, por último, los textos divulgativos constituirían el nivel de especialización bajo. En estos tres niveles de especialización nos encontramos con una *comunicación especializada*, esto es, “el proceso mediante el cual se imparte o comparte conocimiento especializado” (Guantiva Acosta, Cabré y Castellà Lidon 2008: 17), pero, evidentemente, como ya hemos comentado, no del mismo grado de especialización, puesto que los factores que conciernen al propio acto comunicativo varían, como pueden ser los destinatarios. Así pues, en el lenguaje especializado empleado en este tipo de comunicación, estudiado en el ámbito de la terminología, encontramos una *variación de especialidad*, que forma parte de una lingüística del hablar (Casas Gómez y Fernández Smith 2021: 715).

### 1.3. Terminología y términos propios de la divulgación

La terminología, aunque surgiera fuera de la lingüística, se ha convertido en los últimos años en una disciplina fundamental para las nuevas directrices y orientaciones de la lingüística moderna, especialmente de corte aplicado, así como para el estudio de los nuevos contenidos y perspectivas de la semántica léxica (Casas Gómez y Hummel 2017, Casas Gómez 2020a). Desde los proyectos de investigación desarrollados por el grupo de investigación “Semaínein” (HUM-147) en el Instituto Universitario de Investigación en Lingüística Aplicada de la Universidad de Cádiz se trabaja en la elaboración de un diccionario especializado de los usos terminológicos de la semántica léxica. Y es fundamental recalcar esto último: el objetivo no es llevar a cabo un glosario de tecnicismos o unidades terminológicas, sino de usos técnicos de la semántica léxica, como señala Casas Gómez (2020b: 3), quien insiste en:

the necessity and usefulness of this type of specialised dictionary, since it is not a lexicon of technicisms or terminological units, but, rather, of the technical uses of lexical semantics. For this reason, starting from explanatory sub-indices, terminological datasheets must be created, for each acception of the different terms.

Con la extracción de los términos propios de la semántica presentes en el corpus de textos divulgativos pretendemos, en este contexto de investigación, por un lado, contribuir a uno de los objetivos que se plantean Casas Gómez y Varo Varo (2022, en prensa) en relación con la fase actual de trabajo de los proyectos citados:

Rescatar lingüísticamente una disciplina como la semántica, que parecía estar olvidada en el panorama de nuestra ciencia y que, sin embargo, se está convirtiendo, cada vez más, por su carácter básico, naturaleza interdisciplinar y dimensión transversal, en una materia central en las investigaciones lingüísticas actuales.

Y, por otro, ayudar a completar la base de datos de usos terminológicos sobre la que está construido el diccionario, que es metalingüística y multilingüe y ha conllevado la digitaliza-



ción de numerosas fuentes doctrinales y lexicográficas correspondientes a la semántica léxica (Álvarez Torres y Gómez Díaz 2022). Como señalan Casas Gómez y Varo Varo (2022, en prensa) resulta fundamental, como reto futuro para alcanzar resultados más representativos, ampliar el número de contextos metalingüísticos.

El documento terminográfico básico de estos proyectos es la ficha terminológica, patentada como *DOCUTERM: Modelo de documentación de usos terminológicos* (Casas Gómez, 2020), que consta de 24 campos. En este trabajo nos proponemos ampliar las posibilidades de dos de ellos: el estatus del término y los posibles contextos metalingüísticos significativos. En lo referente al estatus de los términos, los modelos de fichas no presentan información demasiado abundante sobre ellos. Cabré y Da Cunha (2022: 201) señalan, en la descripción de los posibles campos de una ficha: “Estatus del término (normativo, normalizado, etc.)”; Casas Gómez (2020c), por su parte, aporta como posibilidades dentro de este campo “estandarizado” / “no estandarizado”, “documentado en diccionarios o textos” / “no documentado en diccionarios o textos”, “propuesto para el traductor” / “no propuesto para el traductor”. Todas estas posibilidades dependen de su presencia en el corpus que se emplee y, como hemos indicado, el corpus empleado por los proyectos del ILA consta de diccionarios terminológicos y documentos de investigación científica. La propuesta que traemos nosotras en este trabajo es la de considerar un estatus más básico que todos estos, el estatus “documentado en textos divulgativos”, que manifiesta un uso mucho más extendido.

En cuanto a los contextos metalingüísticos, como se puede observar en la revisión de los modelos de documentación terminográfica que lleva a cabo Casas Gómez (2006), el apartado de contexto aparece en los modelos de Auger y Rousseau (1977), Dubuc (1999) y Cabré (1993), así como en el de Casas (2020c). Se prevé, para el diccionario de terminología de la semántica léxica, que estos contextos sean definitorios, representativos y adecuados para determinar y completar los diferentes usos terminológicos de las entradas léxicas. Sobre esto mismo, Casas (2020b: 5) indica:

It is not a mere testimonial use in which its presence is documented in a text, but, rather, a definitory or metalinguistic context, providing information on the content or giving a linguistic explanation of its use as a terminological unit of the discipline.

Como desarrollamos en el apartado de análisis, en este trabajo aportamos, para cada uno de los términos de la semántica léxica extraídos del corpus de artículos divulgativos, tres contextos metalingüísticos significativos que sean definitorios (del tipo “este término significa X”), que se empleen para denominar ejemplos de fenómenos lingüísticos (del tipo “este fenómeno X se denomina ‘término’”) o bien que, en el contexto en que aparecen, se vinculen con otros términos o sean aclaradas cuestiones de su definición o de su uso que puedan resultar interesantes desde el punto de vista terminológico.

## 2. METODOLOGÍA

El corpus creado para este trabajo está compuesto por los textos de 226 columnas de opinión de corte divulgativo, sobre temas relacionados con la lingüística y escritas por lingüistas. Están publicadas en prensa digital o en publicaciones periódicas digitales especializadas en divulgación, de acceso libre (es decir, para cuyo acceso no es necesario estar registrado en

la web del medio que las publica) en los últimos diez años (desde 2012). No se han tomado propiamente noticias, sino columnas, es decir, escritos de opinión redactados por lingüistas especializados. Para este estudio, que en principio se limita a la terminología en español, se han tomado solo publicaciones de España y en lengua española.

El corpus empleado está formado por 229.927 palabras extraídas de las siguientes fuentes:

- columnas de Elena Álvarez Mellado en *elDiario.es* ([https://www.eldiario.es/autores/elena\\_alvarez\\_mellado/](https://www.eldiario.es/autores/elena_alvarez_mellado/), desde el 15 de enero de 2017 hasta el 7 de febrero de 2022, 53 columnas): 50.785 palabras;
- columnas de María del Carmen Horno Chéliz en *Letras Libres* (<https://letraslibres.com/author/mamen-horno/>, desde el 10 de abril de 2012 hasta el 8 de marzo de 2022, 34 columnas): 27.622 palabras;
- columnas de Lola Pons en el suplemento de ciencia de *El País*, *Verne* ([https://verne.elpais.com/autor/maria\\_dolores\\_pons\\_rodriguez/a/](https://verne.elpais.com/autor/maria_dolores_pons_rodriguez/a/), desde el 2 de julio de 2017 hasta el 20 de febrero de 2021, 71 columnas): 71.220 palabras<sup>5</sup>;
- artículos publicados en la revista *Archiletras* en su sección de firmas (<https://www.archiletras.com/firmas/>, 68 artículos, de diversos autores, desde el 17 de diciembre de 2018 hasta el 11 de febrero de 2022): 80.300 palabras.

Para la extracción de los textos se ha empleado la herramienta de Sketch Engine para la confección de corpus desde la web. En concreto, se le han proporcionado los enlaces de cada uno de los artículos de cada periódico o revista, cuyo contenido ha sido descargado y con él se ha creado el corpus de trabajo. Se han añadido como metadatos de cada artículo el nombre del autor, la fecha de publicación, el medio en que fue publicado y el enlace al artículo. De nuevo con Sketch Engine, se han extraído las palabras clave del texto, así como los denominados por esta plataforma *multi-word terms*. Estos son combinaciones léxicas frecuentes “which appear more frequently in the focus corpus than in the reference corpus and, additionally, match the typical format of terminology in the language. The format is defined in the term grammar”<sup>6</sup>. El formato para el español es nombre + adjetivo, nombre + *de* + nombre o adjetivo + nombre + *de* + nombre, es decir, las estructuras que generalmente presentan los términos complejos, tal y como señalan Estopà (1999) y Vivaldi (2001), entre otros.

Estas palabras clave y *multi-word terms* son elementos cuya frecuencia de aparición en el corpus de análisis es significativamente mayor que en el corpus de comparación, a partir de un *keyness score*. El cálculo de este *keyness score* lo lleva a cabo Sketch Engine de la siguiente manera (Kilgarriff *et al.* 2014):

Key words are words typical of a focus corpus (a corpus we are interested in) in contrast to a reference corpus (usually a general corpus in the same language as the focus corpus). The *keyness score* of a word is calculated according to the following formula:

$$\frac{fpm_{focus} + n}{fpm_{ref} + n}$$

5 Las columnas de Lola Pons han continuado publicándose en *El País* tras el cierre del suplemento *Verne* en 2021, pero sujetas a suscripción al diario, por lo que no nos ha sido posible incluirlas en el corpus.

6 <https://www.sketchengine.eu/guide/keywords-and-term-extraction/#toggle-id-2>.

where  $f_{m_{\text{focus}}}$  is the normalized (per million) frequency of the word in the focus corpus,  $f_{m_{\text{ref}}}$  is the normalized (per million) frequency of the word in the reference corpus,  $n$  is the simple Maths (smoothing) parameter ( $n = 1$  is the default value).

El corpus de comparación que empleamos en la plataforma es el *Spanish Web 2018* (esTenTen18), que cuenta con más de 20.000 millones de palabras, extraídas de la web a partir de textos tanto de español de España como americano.

Esta selección de elementos léxicos a partir de su frecuencia en el corpus nos da un listado de elementos significativamente frecuentes, que evaluamos de manera manual para poder analizar con ellos las temáticas más frecuentes en los textos divulgativos sobre lingüística. Eliminamos de la lista apellidos (como Chomsky, Moliner o Nebrija), títulos de obras (*Diccionario de Autoridades*), palabras relacionadas con el contexto inmediato en que se redactó el artículo (*pandemia, confinamiento*) y, sobre todo, los ejemplos de fenómenos de los que se habla en los artículos del corpus, que, por su excepcionalidad, su carácter neológico o su pertenencia al registro coloquial o al nivel de lengua vulgar no tienen presencia en el corpus de comparación y, por tanto, aparecen entre las primeras palabras clave del corpus (*almóndiga, malamente, kellys, güeno, verdul, portavoza, señor, piolines, hemos* y otros muchos).

Para el análisis terminológico que llevamos a cabo en este trabajo vamos a centrarnos en la lista de palabras clave (pues los términos propios de la semántica que aparecen en el corpus son simples en todos los casos). Emplearemos las combinaciones léxicas frecuentes (los *multi-word terms*) únicamente para la descripción temática del corpus de divulgación, que tratamos en el epígrafe 3.1. Para esta descripción temática no necesitamos discriminar entre términos propiamente dichos (*principio de economía lingüística*, por ejemplo) y simples combinaciones de término (sustantivo) + modificador (*hablante nativo*), pues no pretendemos hacer un análisis terminológico con ellas.

La plataforma Sketch Engine se ha empleado, también, para extraer las concordancias de los términos que nos interesan, es decir, los contextos en los que los términos se encuentran dentro del corpus, y seleccionar de entre ellos los que sean apropiados para el trabajo terminográfico. La herramienta de correspondencias de esta plataforma incluye una opción especialmente interesante cuando se quieren extraer contextos significativos de cara a una base de datos terminológica: el sistema de evaluación de oraciones llamado GDEX, “Good Dictionary Examples” (Kilgarriff *et al.* 2008)<sup>7</sup>:

It is a system for evaluation of sentences with respect to their suitability to serve as dictionary examples or good examples for teaching purposes. Sentences are evaluated with respect to their length, use of complicated vocabulary, presence of controversial topics (politics, religion...), sufficient context, references pointing outside of the sentence (e.g. pronouns), brand names and other criteria.

7 <https://www.sketchengine.eu/guide/gdex/>.

### 3. ANÁLISIS

#### 3.1. Análisis de contenido del corpus: ¿de qué se habla en los artículos de divulgación sobre lingüística?

En el primer apartado del análisis nos centramos en las palabras clave y las combinaciones léxicas frecuentes (*multi-word terms*) del corpus que nos permite extraer Sketch Engine. Estos, recordamos, son los elementos léxicos simples y complejos significativamente frecuentes en nuestro corpus en comparación con un corpus normal. Como se ha señalado más arriba, conocer sobre qué temas se habla cuando se hace divulgación de la lingüística tiene una gran importancia de cara a la visibilidad y la percepción que la sociedad tiene de la disciplina: ¿qué estamos transmitiendo a la sociedad sobre lo que se investiga en lingüística?, ¿qué temas aparecen más a menudo?

	Palabras clave	FAce	FRce	FRce	Score		Combinaciones léxicas frecuentes	FA	FRce	FRce	Score
1	<b>hablante</b>	476	1784,31	3,87	366,87	1	<b>lengua materna</b>	71	266,15	1,70	99,03
2	<b>anglicismo</b>	63	236,16	0,44	164,36	2	<b>género gramatical</b>	28	104,96	0,07	98,87
3	<b>léxico</b>	134	502,31	3,26	118,02	3	<b>lenguaje inclusivo</b>	22	82,47	0,08	77,39
4	<b>lingüista</b>	63	236,16	1,27	104,41	4	<b>lengua estándar</b>	19	71,22	0,05	68,50
5	<b>lingüístico</b>	409	1533,16	14,19	100,98	5	<b>cambio lingüístico</b>	19	71,22	0,07	67,41
6	<b>gramatical</b>	102	382,35	3,09	93,69	6	<b>norma culta</b>	17	63,73	0,03	62,79
7	<b>uve</b>	30	112,46	0,27	89,69	7	<b>fenómeno lingüístico</b>	17	63,73	0,07	60,69
8	<b>diminutivo</b>	39	146,19	0,66	88,91	8	<b>hablante nativo</b>	20	74,97	0,26	60,07
9	<b>hache</b>	33	123,70	0,41	88,30	9	<b>máxima de Grice</b>	12	44,98	0	45,88
10	<b>leísmo</b>	23	86,22	0,05	83,31	10	<b>uso lingüístico</b>	12	44,98	0,11	41,44
11	<b>monolingüe</b>	27	101,21	0,27	80,42	11	<b>sexismo lingüístico</b>	11	41,23	0,02	41,42
12	<b>RAE</b>	121	453,57	4,82	78,13	12	<b>lenguaje juvenil</b>	11	41,23	0,02	41,39
13	<b>participio</b>	29	108,71	0,41	77,97	13	<b>español europeo</b>	11	41,23	0,04	40,74
14	<b>hispanohablante</b>	46	172,43	1,24	77,46	14	<b>diccionario académico</b>	11	41,23	0,06	39,67
15	<b>espanghish</b>	21	78,72	0,04	77,02	15	<b>castellano antiguo</b>	11	41,23	0,09	38,83
16	<b>latín</b>	178	667,24	7,77	76,17	16	<b>variedad lingüística</b>	11	41,23	0,10	38,43
17	<b>presuposición</b>	24	89,97	0,24	73,57	17	<b>lengua hija</b>	10	37,49	0	38,38

	Palabras clave	FAce	FRce	FRcc	Score		Combinaciones léxicas frecuentes	FA	FRce	FRcc	Score
18	<b>multilingüismo</b>	22	82,47	0,15	72,40	18	<b>historia del español</b>	10	37,49	0,03	37,37
19	<b>diccionario</b>	268	1004,61	13,03	71,69	19	<b>lengua de signos</b>	16	59,98	0,64	37,27
20	<b>morfema</b>	25	93,71	0,33	71,09	20	<b>español americano</b>	10	37,49	0,07	36,09
21	<b>eñe</b>	22	82,47	0,19	70,35	21	<b>español de España</b>	10	37,49	0,09	35,21
22	<b>lengua</b>	1311	4914,35	71,75	67,57	22	<b>signo de puntuación</b>	13	48,73	0,44	34,51
23	<b>yeísmo</b>	18	67,47	0,03	66,41	23	<b>lengua natural</b>	10	37,49	0,12	34,35
24	<b>subjuntivo</b>	25	93,71	0,45	65,20	24	<b>economía del lenguaje</b>	9	33,74	0,03	33,82
25	<b>gramática</b>	91	341,12	4,25	65,14	25	<b>complemento directo</b>	10	37,49	0,16	33,15
26	<b>multilingüe</b>	29	108,71	0,73	63,49	26	<b>lengua del mundo</b>	9	33,74	0,11	31,39
27	<b>preposición</b>	33	123,70	0,97	63,40	27	<b>idioma nacional</b>	9	33,74	0,11	31,38
28	<b>vocativo</b>	18	67,47	0,11	61,72	28	<b>lengua romance</b>	11	41,23	0,35	31,33
29	<b>neologismo</b>	27	101,21	0,67	61,22	29	<b>lengua española</b>	38	142,44	3,58	31,31
30	<b>consonante</b>	37	138,70	1,29	61,01	30	<b>máxima cooperativa</b>	8	29,99	0	30,99
31	<b>verbo</b>	166	622,26	9,55	59,06	31	<b>economía lingüística</b>	8	29,99	0,01	30,59
32	<b>lingüística</b>	52	194,92	2,35	58,54	32	<b>lenguaje claro</b>	10	37,49	0,29	29,90
33	<b>sufijo</b>	33	123,70	1,14	58,39	33	<b>género neutro</b>	8	29,99	0,05	29,54
34	<b>pronombre</b>	36	134,95	1,33	58,33	34	<b>grupo de palabra</b>	8	29,99	0,10	28,20
35	<b>lexicográfico</b>	18	67,47	0,24	55,07	35	<b>signo de apertura</b>	7	26,24	0,02	26,59
36	<b>lexicógrafo</b>	16	59,98	0,12	54,57	36	<b>área de Broca</b>	7	26,24	0,04	26,09
37	<b>sintáctico</b>	37	138,70	1,60	53,79	37	<b>español peninsular</b>	7	26,24	0,07	25,42
38	<b>blending</b>	14	52,48	0,00	53,41	38	<b>unidad léxica</b>	7	26,24	0,11	24,57
39	<b>pronunciación</b>	48	179,93	2,43	52,71	39	<b>sistema lingüístico</b>	7	26,24	0,13	24,09
40	<b>filólogo</b>	27	101,21	1,01	50,97	40	<b>sexo débil</b>	7	26,24	0,13	24,05
41	<b>tilde</b>	32	119,95	1,38	50,88	41	<b>desdoblamiento de género</b>	6	22,49	0	23,48
42	<b>adjetivo</b>	97	363,61	6,25	50,30	42	<b>principio de economía lingüística</b>	6	22,49	0	23,47
43	<b>dialecto</b>	48	179,93	2,78	47,90	43	<b>letra de mujer</b>	6	22,49	0,01	23,37

	Palabras clave	FAce	FRce	FRcc	Score		Combinaciones léxicas frecuentes	FA	FRce	FRcc	Score
44	<b>etimológico</b>	23	86,22	0,86	46,82	44	<b>español caribeño</b>	6	22,49	0,01	23,33
45	<b>ortografía</b>	64	239,91	4,47	44,00	45	<b>enunciado performativo</b>	6	22,49	0,01	23,29
46	<b>seseo</b>	12	44,98	0,08	42,54	46	<b>lengua actual</b>	6	22,49	0,02	23,08
47	<b>prefijo</b>	32	119,95	1,84	42,53	47	<b>diversidad lingüística</b>	7	26,24	0,20	22,67
48	<b>ideófono</b>	11	41,23	0,00	42,18	48	<b>niño bilingüe</b>	6	22,49	0,04	22,62
49	<b>performativo</b>	16	59,98	0,49	41,03	49	<b>lengua vehicular</b>	7	26,24	0,21	22,44
50	<b>aranés</b>	12	44,98	0,16	39,56	50	<b>función sintáctica</b>	6	22,49	0,07	22,03

Tabla 1. Palabras clave y combinaciones léxicas frecuentes en el corpus de estudio (FAce: frecuencia absoluta en el corpus de estudio, FRce: frecuencia relativa por millón de palabras en el corpus de estudio, FRcc: frecuencia relativa por millón de palabras en el corpus de comparación).

Más allá de términos muy generales, empleados en cualquier tema que esté relacionado con la lingüística (*hablante, lingüista, lingüístico, lengua, fenómeno lingüístico*, etc.), con los niveles del lenguaje (*sintáctico*) o con las categorías gramaticales (*verbo, preposición, adjetivo*, etc.), observamos que los temas preponderantes son varios. En primer lugar, es muy frecuente que se traten cuestiones relacionadas con los diccionarios y la labor lexicográfica (*diccionario, lexicógrafo, lexicográfico*), y, por extensión, con el papel de la RAE, así como la norma culta y la lengua estándar. Este papel de la RAE y del diccionario académico está también vinculado a los neologismos, de cuya inclusión en el diccionario o existencia fuera de él se habla a menudo (especialmente cuando la RAE publica la lista de palabras y acepciones nuevas incluidas en el *DLE*). Es recurrente en el corpus, en este sentido, la idea de que las palabras existen más allá de su registro en los diccionarios, que los diccionarios no incluyen solo palabras de la norma culta y que los usos de las palabras pueden ir más allá de la información que registre de ellas el diccionario, particularmente el *DLE*:

- (1) Los diccionarios no incluyen solo las formas que consideramos prestigiosas sino también otras que están marcadas por ser restringidas, por ejemplo, porque están en desuso, se limitan a una zona hispanohablante o tienen un empleo restringido a determinados contextos muy formales o muy informales. (“*Toda la verdad sobre ‘almóndiga’*”, Lola Pons, *Verne*, 10/03/2018, [https://verne.elpais.com/verne/2018/03/06/articulo/1520325761\\_808313.html](https://verne.elpais.com/verne/2018/03/06/articulo/1520325761_808313.html)).
- (2) Muchos han impugnado la palabra *desescalada* porque no está en el *DLE* (ahora presente en la última actualización), olvidando que nuestro diccionario es un puzzle que, además de registrar palabras concretas, incluye también las piezas necesarias para construir otras nuevas. (“*Desescalando la nueva normalidad*”. Isaiás Lafuente, *Archiletras*, 02/12/2020, <https://www.archiletras.com/firma/desescalando-nueva-normalidad/>).
- (3) Nada más lejos de la realidad: lo que dota de significado a las palabras no es su etimología, ni lo que significaron en el pasado, ni siquiera su definición en un diccionario; lo que dota

de significado a una palabra es el uso que los hablantes hagan de ella. (“La falacia etimológica”, Elena Álvarez Mellado, *Archiletras*, 04/10/2019, <https://www.archiletras.com/firma/la-falacia-etimologica/>).

Por tanto, si lo lexicográfico y lo académico tienen esa presencia en el corpus, es comprensible que sean muy frecuentes términos como *lengua estándar*, *norma culta*, así como *neologismo* y, en relación con esto, *anglicismo* o *blending* como procedimientos para crear palabras nuevas. En relación, asimismo, con la norma culta y las cuestiones normativas otro tema recurrente es la ortografía, sobre la que encontramos términos como *uve*, *hache*, *ortografía*, *signo de puntuación*.

El enorme debate que suscita el lenguaje inclusivo está, como era de esperar, en el corpus: *género gramatical*, *lenguaje inclusivo*, *sexismo lingüístico*, *género neutro*, *desdoblamiento de género* muestran la relevancia de este tema en los artículos de divulgación. Las posturas son diversas y, en algunos casos, muy críticas:

- (1) Finalmente, desde el punto de vista sociolingüístico y estilístico, teniendo en cuenta que la mayoría de los registros de estas voces han partido de chicas jóvenes adictas a los clubs de fans, al ser un morfema no marcado en cuanto a género gramatical ofrece una ventaja a las hablantes sensibilizadas –y obsesionadas– con la cuestión del sexismo y del género inclusivo por mor de una mal entendida corrección política. (“Un nuevo sufijo en español: -er”, Félix Rodríguez, *Archiletras*, 23/03/2021, <https://www.archiletras.com/firma/un-nuevo-sufijo-en-espanol-er/>).
- (2) Pues bien, si los fenómenos sexistas discursivos reflejan estas formas de discriminación más allá del género gramatical, las recomendaciones no pueden ser ni la duplicación sistemática ni la feminización de cualquier apelativo, ni tampoco el uso de términos neutros. (“Una mirada otra al sexismo lingüístico”, Aguas-Vivas Catalá González y Enriqueta García Pascual, *Archiletras*, 05/03/2019, <https://www.archiletras.com/firma/sexismo-linguistico/>).

En otros, en cambio, se justifica el uso de, por ejemplo, el morfema “-e”:

- (1) La propuesta de construir un género neutro en -e soluciona muchos de los escollos que las anteriores propuestas dejaban sin resolver: fácil de pronunciar, morfológicamente claro, lingüísticamente económico, socialmente inclusivo. (“Todas, tod@s, todxs, todes: historia de la disidencia gramatical”, Elena Álvarez Mellado, *elDiario.es*, 27/06/2017, [https://www.eldiario.es/opinion/zona-critica/todas-todes-historia-disidencia-gramatical\\_129\\_3307468.html](https://www.eldiario.es/opinion/zona-critica/todas-todes-historia-disidencia-gramatical_129_3307468.html)).
- (2) Lo interesante del “todes” radica en que no solo se está proponiendo un morfema novedoso, sino que introduce un nuevo valor en oposición al masculino y femenino. Es decir, la morfología del castellano distingue masculino y femenino, y esta distinción es encarnada por diversos pares de morfemas (-o/-a, or/riz, -és/esa, etc.). La propuesta del género en -e no solo propone un nuevo morfema, sino que introduce todo un valor morfológico nuevo que sí tendría un valor neutro en oposición al masculino y femenino (“Lenguaje inclusivo: algunas claves lingüísticas”, Elena Álvarez Mellado, *elDiario.es*, 30/07/2021, [https://www.eldiario.es/opinion/zona-critica/lenguaje-inclusivo-claves-linguisticas\\_129\\_8021559.html](https://www.eldiario.es/opinion/zona-critica/lenguaje-inclusivo-claves-linguisticas_129_8021559.html)).

Las cuestiones que rodean a la lengua española son, pues, preponderantes. Se habla a menudo sobre sus variedades (*español americano*, *español de España*, *español europeo*,

*español peninsular, español caribeño*). Y no solo en su sincronía, sino también en su diacronía, lo que hace que términos como *castellano antiguo, historia del español* y *latín* se encuentren en la lista.

La diversidad lingüística, por tanto, tiene poca presencia. Como decimos, mayoritariamente se habla sobre el español, sus variedades, sus fenómenos, sus diccionarios, su academia y su historia. Solo en casos puntuales se adopta una perspectiva que recoja otras lenguas, normalmente asociada al multilingüismo y la convivencia de lenguas, especialmente en España, y en alguna ocasión puntual a cuestiones tipológicas y, también, a la importancia del reconocimiento de las lenguas maternas, especialmente cuando son minoritarias (y de ahí que aparezcan en el listado elementos como *monolingüe, multilingüe, lengua del mundo, diversidad lingüística* o *lengua materna*):

- (1) Cada cierto tiempo suelen aparecer noticias en diversos medios de comunicación y voces en redes sociales que se dedican a exponer los peligros y las desventajas de la diversidad lingüística, es decir, del hecho de que en el mundo coexistan lenguas diferentes, en la mayoría de los casos, en un mismo espacio geográfico (“Las bondades de las lenguas”, Iraide Ibarretxe-Antuñano, *Archiletras*, 20/01/2022, <https://www.archiletras.com/firma/las-bondades-de-las-lenguas/>).
- (2) El respeto a la diversidad cultural y lingüística es, sin duda, un patrimonio cultural que hay que cuidar y proteger; sin embargo, si bien es cierto que, desde una dimensión sociopolítica, institucionalmente se reconoce el multiculturalismo y el multilingüismo, no siempre se concibe la convivencia de lenguas y de culturas como un valor positivo. (“Los mixtecos, más allá de ‘Roma’”, Azucena Palacios, *Archiletras*, 10/07/2019, <https://www.archiletras.com/firma/los-mixtecos-mas-alla-de-roma/>).
- (3) Fantaseo con cómo de diferente sería este país (y cómo mejoraría nuestra capacidad para entender idiomas) si el multilingüismo impregnase los distintos ámbitos de la vida cotidiana, también en las regiones supuestamente monolingües: que la tele, la ficción audiovisual, los letreros en las calles o los discursos de las Cortes reflejasen la diversidad de lenguas que se hablan en el territorio. (“La fantasía de la España monolingüe”, Elena Álvarez Mellado, *elDiario.es*, [https://www.eldiario.es/opinion/zona-critica/fantasia-espana-monolingue\\_129\\_8530274.html](https://www.eldiario.es/opinion/zona-critica/fantasia-espana-monolingue_129_8530274.html)).
- (4) A través de la diversidad de las lenguas podemos intentar diferenciar qué aspectos de las lenguas son principios universales del lenguaje comunes a todas los idiomas y qué valores son variables y están sujetos a cambios. (“¿Para qué sirven las lenguas?”, Elena Álvarez Mellado, *elDiario.es*, [https://www.eldiario.es/opinion/zona-critica/sirven-lenguas\\_129\\_7283273.html](https://www.eldiario.es/opinion/zona-critica/sirven-lenguas_129_7283273.html)).
- (5) El tiempo vuela, sinceramente, y este mes celebramos, de nuevo, el día de la Lengua Materna. Hace un año os intenté convencer aquí de la naturaleza íntima y trascendente de esta lengua, que nos hace ser como somos, pensar como pensamos y sentir como sentimos. Entonces os animé a defender el derecho de los hablantes nativos a usar su lengua en los momentos más complicados de la vida. (“El derecho a tener lengua materna”, Mamen Horno, *Letras Libres*, 08/02/2021, <https://letraslibres.com/ciencia-y-tecnologia/el-derecho-a-tener-lengua-materna/>).

La referencia a la lengua materna también se hace en artículos de corte cognitivista que abordan su adquisición, el aprendizaje de las lenguas extranjeras y, también, las diferencias



en los procesos psicolingüísticos cuando se hablan unas u otras:

- (1) Otro de los beneficios con los que se relaciona el uso del área de Broca del hemisferio izquierdo por parte de los bilingües equilibrados tiene que ver con el cambio de código. Tal y como veíamos antes, el cambio de una lengua a otra supone un coste atencional que, en general, resulta mayor cuando pasamos de la L2 a la lengua materna que viceversa (“Bilingüismo, cognición y ciencia”, Mamen Horno, *Letras Libres*, 17/05/2016, <https://letraslibres.com/revista-espana/bilinguismo-cognicion-y-ciencia/>).

Por otro lado, términos como *presuposición*, *enunciado performativo*, *máxima de Grice*, *máxima cooperativa* muestran que las cuestiones pragmáticas tienen, asimismo, relevancia:

- (1) De hecho, quizá donde mejor se vea operar las máximas de Grice es cuando no se respetan, porque su incumplimiento tiende a producir comunicaciones insatisfactorias (o, como se dice en términos pragmáticos, infelices): la violación de las máximas de Grice explica la risa que nos produce cuando en un paquete de nueces se anuncia solemnemente que “contiene nueces” o la frustración que nos causa cuando alguien intenta hacerse el gracioso y responde con un escueto “sí” a la pregunta “¿Tienes hora?” (“Un cerdo a la izquierda”, Elena Álvarez Mellado, *elDiario.es*, 23/08/2020, [https://www.eldiario.es/opinion/zona-critica/cerdo-izquierda\\_129\\_6178302.html](https://www.eldiario.es/opinion/zona-critica/cerdo-izquierda_129_6178302.html)).
- (2) La elección de este titular (cuyo contenido es en apariencia obvio, pero periodísticamente atractivo) es en realidad un buen ejemplo de lo que en Lingüística se conoce como la teoría de la relevancia, una teoría formulada por los lingüistas Sperber y Wilson en 1986 y que supone una vuelta de tuerca sobre las máximas de Grice (de las que ya hablamos aquí en su día). (“El arte de escribir titulares relevantes”, Elena Álvarez Mellado, *elDiario.es*, 18/09/2021, [https://www.eldiario.es/opinion/zona-critica/arte-escribir-titulares-relevantes\\_129\\_8307960.html](https://www.eldiario.es/opinion/zona-critica/arte-escribir-titulares-relevantes_129_8307960.html)).

Así pues, la divulgación de la lingüística, en nuestro país, en español y en la última década, está especialmente enfocada a hablar sobre las cuestiones de las que más se habla también por parte de los no especialistas y que son, generalmente, los temas que el destinatario no especializado suele asociar con la lengua: lo académico (la RAE, el *DLE* –que son, sin duda, el referente y el contacto, respectivamente, más cercanos para el lector no especializado con el lenguaje y su lengua–) y los diccionarios en general, así como las nuevas palabras que se incluyen en ellos. Se habla sobre esto con una perspectiva descriptivista y con la intención de desterrar mitos, específicamente. El otro tema principal es el sexismo lingüístico y las estrategias morfológicas de creación de un género neutro, en el que sí hay posturas enfrentadas entre los autores de las columnas: desde quienes lo rechazan frontalmente hasta quienes acogen las propuestas con, incluso, entusiasmo, aunque también existen posturas intermedias. Otros temas están ya más vinculados al autor que escribe la columna, pues, como es de esperar, cada autor suele escribir sobre los temas de los que es especialista. En las columnas de Lola Pons es habitual encontrar artículos sobre historia del español o sobre variedades del español; en la de Mamen Horno, sobre cuestiones cognitivas y adquisición aprendizaje de lenguas y, en el caso de Elena Álvarez Mellado, aunque sus temas son más variados, son frecuentes las cuestiones sobre pragmática.

### 3.2. Corpus de textos digitales y terminografía. Establecimiento de un estatus terminológico propio de la divulgación y de contextos divulgativos

El análisis de la temática del corpus confeccionado pone de manifiesto que las cuestiones relacionadas con la semántica léxica (salvo en lo que se refiere a la lexicografía y, en menor medida, cuando se tratan cuestiones de semántica cognitiva) no tienen una presencia muy habitual en la divulgación de la lingüística. No obstante, se detecta un número suficiente de términos que nos permite abrir la puerta al establecimiento, en las fichas terminológicas, de un estatus terminológico propio de la divulgación para algunos términos tan habituales que aparecen con suficiente frecuencia incluso en un corpus no centrado en cuestiones semánticas.

Los términos propios de la semántica léxica presentes en el corpus y que hemos podido extraer con el análisis de palabras clave son los siguientes:

Orden	Término	FAce	FRce	FRcc	Score
18	<b>diccionario</b>	268	1004,61	13,03	71,69
28	<b>neologismo</b>	27	101,21	0,67	61,22
34	<b>lexicográfico</b>	18	67,47	0,24	55,07
35	<b>lexicógrafo</b>	16	59,98	0,12	54,57
51	<b>prototípico</b>	13	48,73	0,32	37,64
61	<b>lexicón</b>	9	33,74	0,09	31,82
67	<b>eufemismo</b>	21	78,72	1,61	30,57
69	<b>metafórico</b>	20	74,97	1,61	29,06
71	<b>polisemia</b>	9	33,74	0,22	28,41
73	<b>despectivo</b>	20	74,97	1,70	28,16
92	<b>significado</b>	226	847,17	33,49	24,59
97	<b>acepción</b>	28	104,96	3,56	23,24
180	<b>lexicografía</b>	5	18,74	0,19	16,65
315	<b>metonimia</b>	4	14,99	0,27	12,62

Tabla 2. Términos propios de la semántica léxica presentes en el corpus con una frecuencia significativa (FAce: frecuencia absoluta en el corpus de estudio, FRce: frecuencia relativa por millón de palabras en el corpus de estudio, FRcc: frecuencia relativa por millón de palabras en el corpus de comparación).

Aportamos, para cada uno de ellos, tres contextos metalingüísticos lo más definitorios, representativos y adecuados posible, de entre los disponibles en el corpus, para determinar los usos terminológicos de estos términos propios de la divulgación. Es cierto que no siempre se pueden encontrar contextos que definan el término, pero sí aportan, en todos los casos, información relevante del uso de estos términos y, también, del vínculo que presentan con otros términos de la lingüística o bien muestran información aclaratoria sobre el concepto manejado que no siempre encontramos en textos escritos por especialistas para especialistas.

- diccionario:

- (1) Los diccionarios no incluyen solo las formas que consideramos prestigiosas sino también otras que están marcadas por ser restringidas, por ejemplo, porque están en desuso, se limitan a una zona hispanohablante o tienen un empleo restringido a determinados contextos muy formales o muy informales. (“Toda la verdad sobre ‘almóndiga””, Lola Pons, *Verne*, 10/03/2018, [https://verne.elpais.com/verne/2018/03/06/articulo/1520325761\\_808313.html](https://verne.elpais.com/verne/2018/03/06/articulo/1520325761_808313.html)).
- (2) La lengua se crea porque los hablantes la usan y la función de los diccionarios y gramáticas es (o debería ser) recoger fielmente cómo nos expresamos los hablantes, no decirnos cómo debemos hablar. (“Iros’: el bueno, el feo y el malo”, Elena Álvarez Mellado, *elDiario.es*, 18/07/2017, [https://www.eldiario.es/opinion/zona-critica/iros-bueno-feo-malo\\_129\\_3270024.html](https://www.eldiario.es/opinion/zona-critica/iros-bueno-feo-malo_129_3270024.html)).
- (3) Es exactamente lo mismo que debe hacer el diccionario: no eliminar palabras, sino recogerlas, definir las e indicar siempre la connotación y el uso que tienen. (“Dejad de pedirle a la RAE que elimine palabras”, Elena Álvarez Mellado, *elDiario.es*, 20/05/2017, [https://www.eldiario.es/opinion/zona-critica/dejad-pedirle-rae-elimine-palabras\\_129\\_3381428.html](https://www.eldiario.es/opinion/zona-critica/dejad-pedirle-rae-elimine-palabras_129_3381428.html)).

- metáfora / metafórico:

- (1) Son las metáforas: el mecanismo lingüístico por el que nos referimos a algo utilizando palabras que, en principio, usamos para denominar a otro objeto simplemente porque creemos ver entre ellos una cierta semejanza o paralelismo. (“Metáforas peligrosas: el cáncer como lucha”, Elena Álvarez Mellado, *elDiario.es*, 23/10/2017, [https://www.eldiario.es/opinion/zona-critica/metáforas-peligrosas-cancer-lucha\\_129\\_3108779.html](https://www.eldiario.es/opinion/zona-critica/metáforas-peligrosas-cancer-lucha_129_3108779.html)).
- (2) Generalmente, lo que sucede cuando se emplea una metáfora es que los hablantes ponemos en relación dos conceptos que aparentemente no tienen nada que ver entre sí, pero que sentimos que comparten algunas características, a veces físicas (la estructura familiar basada en las raíces y las ramas de un árbol), a veces funcionales (la medición de la potencia de un coche según el número de caballos). (“El poder de la metáfora en tiempos de pandemia”, Iraide Ibarretxe-Antuñano, *Archiletras*, 04/11/2020, <https://www.archiletras.com/firma/el-poder-de-la-metáfora-en-tiempos-de-pandemia/>).
- (3) Lo interesante reside en que, en este tipo de juegos metafóricos, la creatividad funciona en el sentido inverso al que suele hacerlo en la lengua: en lengua, el escenario habitual suele consistir en una situación o realidad novedosa a la que le asignamos una imagen que la representa y que utilizamos para nombrar: los grupos de convivencia aislados frente a la COVID-19 pasan a ser burbujas, porque están cerrados como las burbujas. (“600 formas de mirar una pandemia”, Elena Álvarez Mellado, *elDiario.es*, 08/04/2021, [https://www.eldiario.es/opinion/zona-critica/600-formas-mirar-pandemia\\_129\\_7378281.html](https://www.eldiario.es/opinion/zona-critica/600-formas-mirar-pandemia_129_7378281.html)).

- prototípico:

- (1) Lo que se propone desde la psicolingüística es que cuando los hablantes manejamos categorías y conceptos tenemos inconscientemente en la cabeza uno o varios ejemplares concretos que consideramos prototípicos, es decir, ejemplos que los hablantes consideran buenos representantes de esa categoría y que son por lo general compartidos por los hablantes. (“Hambur-

guesa de espinacas, leche de soja”, Elena Álvarez Mellado, *elDiario.es*, 07/11/2020, [https://www.eldiario.es/opinion/zona-critica/hamburguesa-espinacas-leche-soja\\_129\\_6394604.html](https://www.eldiario.es/opinion/zona-critica/hamburguesa-espinacas-leche-soja_129_6394604.html)).

- (2) Así, por ejemplo, si hablamos de aves, un ruiseñor o un cóndor probablemente serán ejemplos más prototípicos en la cabeza de los hablantes que un pingüino. (“Hamburguesa de espinacas, leche de soja”, Elena Álvarez Mellado, *elDiario.es*, 07/11/2020, [https://www.eldiario.es/opinion/zona-critica/hamburguesa-espinacas-leche-soja\\_129\\_6394604.html](https://www.eldiario.es/opinion/zona-critica/hamburguesa-espinacas-leche-soja_129_6394604.html)).
- (3) Es decir, no se trata de que un objeto cumpla con una serie de requisitos o definiciones abstractas para definirlo como sopa, sino que es la comparación respecto al ejemplo prototípico de la clase (en este caso, la sopa de fideos) lo que establece su sopidad. (“Hamburguesa de espinacas, leche de soja”, Elena Álvarez Mellado, *elDiario.es*, 07/11/2020, [https://www.eldiario.es/opinion/zona-critica/hamburguesa-espinacas-leche-soja\\_129\\_6394604.html](https://www.eldiario.es/opinion/zona-critica/hamburguesa-espinacas-leche-soja_129_6394604.html)).

- neologismo:

- (1) Cualquier innovación que se registra en una lengua pasa por dos fases: una es la propia innovación y otra, la verdaderamente clave, es la difusión. Innovar lingüísticamente no es tan complejo, pero esa creación queda en nada si no es adoptada por otros hablantes y se generaliza más allá de la persona que inventa. Esa nueva palabra que se crea y, si tiene suerte, se extiende, recibe el nombre de neologismo. (“‘Covidiota’, ‘balconazis’, ‘cuarenpena’... los neologismos que nos ha traído la pandemia”, Lola Pons, *Verne*, 08/04/2020, [https://verne.elpais.com/verne/2020/04/07/articulo/1586246728\\_179666.html](https://verne.elpais.com/verne/2020/04/07/articulo/1586246728_179666.html)).
- (2) O en uno de mis neologismos favoritos: el brillibrilli o el efecto del uso (o abuso, según se mire) de purpurina o cualquier elemento que brille y cause destellos. Neologismo, por cierto, muy convencionalizado ya en diversas redes sociales y que sigue el patrón reduplicativo e ideofónico tan característico de otras lenguas con grandes dosis de este tipo de iconicidad como el japonés, donde se dice kira-kira, o el euskera, donde se dice dizdiz o briz-briz. (“Ser o no ser en lingüística”, Iraide Ibarretxe-Antuñano, *Archiletras*, 14/06/2021, <https://www.archiletras.com/firma/ser-o-no-ser-en-linguistica/>).
- (3) Todos estos neologismos se pueden usar en la conversación, los hablantes saben crearlos sin necesidad de que nadie les enseñe y sus interlocutores los entienden. Está claro que no todos tendrán una larga vida, no muchos se convertirán en palabras de uso generalizado y solo unos pocos entrarán en el diccionario, pero todos ellos, con más o menos éxito, son palabras de nuestro idioma. (“Cread, cread, benditos. Desterrando el mito del diccionario”, Mamen Horno, *Letras Libres*, 07/12/2020, <https://letraslibres.com/ciencia-y-tecnologia/cread-cread-benditos-desterrando-el-mito-del-diccionario/>).

- lexicón:

- (1) Así, por ejemplo, se ha comprobado en laboratorio que los hablantes multilingües activan simultáneamente todos sus lexicones con independencia de la lengua en la que estén hablando, incluso en entornos marcadamente monolingües. (“Bilingüismo, cognición y ciencia”, Mamen Horno, *Letras Libres*, 17/05/2016, <https://letraslibres.com/revista-espana/bilinguismo-cognicion-y-ciencia/>).
- (2) Este efecto de interferencia apunta, como vemos, a que en esta prueba de reconocimiento de palabras no solo está activado el lexicón de la lengua que se está utilizando, sino que se

activan todos los lexicones al mismo tiempo. (“Bilingüismo, cognición y ciencia”, Mamen Horno, *Letras Libres*, 17/05/2016, <https://letraslibres.com/revista-espana/bilinguismo-cognicion-y-ciencia/>).

- (3) Esta activación de todos los lexicones a un mismo tiempo en este tipo de pruebas se da tanto cuando la prueba se realiza en la L2 (y se activa a la vez la lengua materna) como cuando la prueba se realiza en la L1 (y se activan las lenguas segundas). (“Bilingüismo, cognición y ciencia”, Mamen Horno, *Letras Libres*, 17/05/2016, <https://letraslibres.com/revista-espana/bilinguismo-cognicion-y-ciencia/>).

- eufemismo:

- (1) Si el suavizar es la función del eufemismo y lo políticamente correcto, el disfemismo juega con la intensificación que provoca el término interdicto. [...] de igual manera, se conservan las ideas que guardan las palabras aunque usemos otras voces sustitutas que las encubran... (“Eufemismo de antes y tabúes de ahora: dos maneras de decir lo mismo”, Mercedes de la Torre García, *Archiletras*, 25/10/2021, <https://www.archiletras.com/firma/eufemismo-de-antes-y-tabues-de-ahora-dos-maneras-de-decir-lo-mismo/>).

- (2) El contexto y el paso del tiempo obligan a una continua sustitución, por ejemplo, los místicos usaron en sus obras el giro “retrete de amor”, donde retrete tiene el significado de ‘cuarto pequeño en la casa o habitación destinada para retirarse’(a orar), y a la postre se resemantiza para referirse al ‘lugar donde vamos a hacer aguas menores y aguas mayores’(otro eufemismo) que sustituye, a su vez, a letrina, y así sucesivamente (urinario, aseos, servicios, W.C., etc.). (“Eufemismo de antes y tabúes de ahora: dos maneras de decir lo mismo”, Mercedes de la Torre García, *Archiletras*, 25/10/2021, <https://www.archiletras.com/firma/eufemismo-de-antes-y-tabues-de-ahora-dos-maneras-de-decir-lo-mismo/>).

- (3) Pero *kelly* ha acabado funcionando también como eufemismo.[...] Así, la palabra sustituye no solo a «camarera», sino a otras denominaciones tradicionales y más claras, como «limpiadora», «señora de la limpieza» o «fregona». [...]. Es curioso, porque «camarera», en su sentido actual, ya es un eufemismo (y muy enfático), sacado de la terminología cortesana, como sucede también en España con la palabra «azafata»: en ambos casos se alude a oficios desemeñados antiguamente por grandes damas que tenían el privilegio de servir a la reina, en un caso organizando el servicio femenino en el palacio y, en el segundo, custodiando su joyero. (“Limpian, fijan y dan esplendor: las *kellys*”, Óscar Esquivias, *Archiletras*, 25/09/2020, <https://www.archiletras.com/firma/limpian-fijan-y-dan-esplendor-las-kellys/>).

- lexicografía / lexicográfico / lexicógrafo

- (1) Pero el caso es que la labor lexicográfica (esto es, el noble arte de hacer diccionarios) no funciona así. Las palabras no pertenecen a la RAE ni a los diccionarios, pertenecen a los hablantes. (“El mito de las palabras que no están en la RAE”, Elena Álvarez Mellado, *elDiario.es*, 13/03/2017, [https://www.eldiario.es/opinion/zona-critica/mito-palabras-rae-diccionario-drae\\_129\\_3529181.html](https://www.eldiario.es/opinion/zona-critica/mito-palabras-rae-diccionario-drae_129_3529181.html)).

- (2) Esta respuesta tan poco entusiasta le ha granjeado a la institución fama de antipática y carca, pero es, en realidad, bastante prudente y está plenamente justificada desde el punto de vista técnico: las definiciones deben confeccionarse atendiendo al uso (es decir, a datos empíricos), no a intuiciones o preferencias personales, así que conviene ser cautelosos para garantizar el rigor del trabajo lexicográfico. (“La RAE y el lujo de definir”, Elena Álvarez Mellado, *elDiario.es*, 12/03/2018, [https://www.eldiario.es/opinion/zona-critica/rae-lujo-definir\\_129\\_2225819.html](https://www.eldiario.es/opinion/zona-critica/rae-lujo-definir_129_2225819.html)).

- (3) La misión de los lexicógrafos es decirnos qué palabras se usan, con qué sentido, en qué contextos y con qué consecuencias; nos alertan sobre si una palabra se considera actualmente vulgar o si se usa poco, o la usa solo un determinado sector de la población. (“Cread, cread, benditos. Desterrando el mito del diccionario”, Mamen Horno, *Letras Libres*, 07/12/2020, <https://letraslibres.com/ciencia-y-tecnologia/cread-cread-benditos-desterrando-el-mito-del-diccionario/>).

- polisemia:

- (1) El motivo por el que estos principios son fundamentales en la conversación reside en que, cuando nos expresamos, casi nada de lo que decimos es unívoco o plenamente explícito, sino que está cuajado de ambigüedad, de polisemia o de información implícita que se presupone. (“Las trabajadoras y la cooperación”, Elena Álvarez Mellado, *elDiario.es*, 25/06/2018, [https://www.eldiario.es/opinion/zona-critica/trabajadoras-cooperacion\\_129\\_2055070.html](https://www.eldiario.es/opinion/zona-critica/trabajadoras-cooperacion_129_2055070.html)).
- (2) La polisemia y la ambigüedad son el pan nuestro de cada día en lengua. De hecho, la mayoría de palabras de un idioma acaparan varios significados. Y tiene sentido: al fin y al cabo, una lengua que asignase un único significado a cada palabra acumularía un repertorio de vocabulario inabarcable. Habitualmente, la polisemia no genera demasiados conflictos a los hablantes porque sabemos discernir sin mucho esfuerzo el significado con el que nuestro interlocutor está usando una palabra polisémica, ya sea a través de las palabras que la acompañan o de la situación que nos rodea. (“De qué hablamos cuando hablamos de consentir”, Elena Álvarez Mellado, *elDiario.es*, 20/11/2017, [https://www.eldiario.es/opinion/zona-critica/hablamos-consentir\\_129\\_3049368.html](https://www.eldiario.es/opinion/zona-critica/hablamos-consentir_129_3049368.html)).
- (3) Estas posiciones (que recuerdan a la de los solotildistas recalcitrantes que buscan ambigüedades imposibles para justificar su nostalgia tildica) olvidan que la polisemia es una característica inherente a las lenguas naturales y que esa ausencia de distinción ya ocurre de facto en la oralidad para millones de hispanohablantes, sin que reine por ello el caos lingüístico entre ellos. (“Sobre la ortografía”, Elena Álvarez Mellado, *elDiario.es*, 12/09/2020, [https://www.eldiario.es/opinion/zona-critica/ortografia\\_129\\_6217681.html](https://www.eldiario.es/opinion/zona-critica/ortografia_129_6217681.html)).

- despectivo:

- (1) Pero los diminutivos, y esto lo sabemos bien todos los hispanohablantes, también pueden ser muy despectivos, y lejos de atenuar el significado de la palabra, acentúan lo que pueda tener de negativo. (“Queridos y temidos diminutivos”, Óscar Esquivias, *Archiletras*, 04/12/2018, <https://www.archiletras.com/firma/queridos-y-temidos-diminutivos/>).
- (2) ‘Señoro’ tiene un sentido despectivo, señala a los varones que tratan de forma condescendiente a las mujeres. (“La vida empuja a la lengua: de señora a señor”, Lola Pons, *Verne*, 05/10/2018, [https://verne.elpais.com/verne/2018/10/05/articulo/1538748293\\_739942.html](https://verne.elpais.com/verne/2018/10/05/articulo/1538748293_739942.html)).
- (3) Pongamos por caso que un extranjero viese u oyese la expresión “sexo débil” y nos preguntase por su significado. O por “sudaca”. O por “maricón”. O por cualquier otra palabra ofensiva o denigrante. ¿Acaso le responderíamos que no lo sabemos? ¿Que esas palabras no existen? No. Lo más razonable sería que le explicásemos el significado, pero advirtiéndole de que se trata de una palabra muy despectiva que suele usarse como insulto. (“Dejad de pedirle a la RAE que elimine palabras”, Elena Álvarez Mellado, *elDiario.es*, 20/05/2017, [https://www.eldiario.es/opinion/zona-critica/dejad-pedirle-rae-elimine-palabras\\_129\\_3381428.html](https://www.eldiario.es/opinion/zona-critica/dejad-pedirle-rae-elimine-palabras_129_3381428.html)).

- significado:

- (1) Lo que dota de significado a las palabras no es su etimología, ni lo que significaron en el pasado, ni siquiera su definición en un diccionario; lo que dota de significado a una palabra es el uso que los hablantes hagan de ella. Las palabras no tienen un significado natural inherente porque el significado de las palabras es pura convención entre hablantes. (“La falacia etimológica”, Elena Álvarez Mellado, *Archiletras*, 04/10/2019, <https://www.archiletras.com/firma/la-falacia-etimologica/>).
- (2) Las palabras no significan según el significado primigenio que tenían en origen, sino según el significado que le atribuye el conjunto de hablantes de una comunidad en un momento dado. (“Lenguaje inclusivo: algunas claves lingüísticas”, Elena Álvarez Mellado, *elDiario.es*, 30/07/2021, [https://www.eldiario.es/opinion/zona-critica/lenguaje-inclusivo-claves-linguisticas\\_129\\_8021559.html](https://www.eldiario.es/opinion/zona-critica/lenguaje-inclusivo-claves-linguisticas_129_8021559.html)).
- (3) La función mágica de las palabras: aunque la relación entre significante y significado sea arbitraria, a veces los sonidos que conforman una palabra tienen una carga especial, puede ser mágica o también convertirse en un tabú. (“La función mágica de las palabras”, Mamen Horno, *Letras Libres*, 09/11/2020, <https://letraslibres.com/ciencia-y-tecnologia/la-funcion-magica-de-las-palabras/>).

- acepción:

- (1) Hablamos de “bizarro”, la palabra que, tras muchos años de lucha, ha visto reconocida su largamente reclamada acepción de ‘raro o extravagante’ en el diccionario de la RAE. (“Bizarro: Las palabras no somos estancas, evolucionamos”, Elena Álvarez Mellado, *elDiario.es*, 25/12/2021, [https://www.eldiario.es/opinion/zona-critica/bizarro-palabras-no-estancas-evolucionamos\\_129\\_8601427.html](https://www.eldiario.es/opinion/zona-critica/bizarro-palabras-no-estancas-evolucionamos_129_8601427.html)).
- (2) Casi todos los diccionarios otorgan una acepción a *leer* como verbo de lengua y otra diferente como verbo de percepción, pero ninguno –si no estoy equivocado– le asigna una acepción como verbo de consumición. (“Las palabras como prismas”, Ignacio Bosque, *Archiletras*, 01/02/2021, <https://www.archiletras.com/firma/palabras-como-prismas/>).
- (3) Y hace cuatro décadas, en 1984, incluyó una nueva acepción para nombrar cualquier proceso de “aumento rápido y por lo general alarmante de algo”, ya sean precios, delitos o, en este caso concreto, contagios y muertes por una enfermedad, sin que se señale su condición de anglicismo. (“Desescalando la nueva normalidad”, Isaías Lafuente, *Archiletras*, 02/12/2020, <https://www.archiletras.com/firma/desescalando-nueva-normalidad/>).

- metonimia:

- (1) El mecanismo de creación de palabras en todos estos casos es siempre el mismo: la metonimia, es decir, cuando utilizamos el nombre de una parte (el color del uniforme, la forma del sombrero) para denominar al todo. (“‘Piolines’: el nacimiento de una palabra”, Elena Álvarez Mellado, *elDiario.es*, 09/10/2017, [https://www.eldiario.es/opinion/zona-critica/piolines-nacimiento-palabra\\_129\\_3138250.html](https://www.eldiario.es/opinion/zona-critica/piolines-nacimiento-palabra_129_3138250.html)).
- (2) La metonimia es un proceso muy vivo en español que resulta particularmente agradecido para generar motes o al menos apelativos no demasiado bienintencionados cuando lo aplicamos a personas (“el Coletas”, “Manolito Gafotas”). “Piolines” es también un ejemplo de metonimia

nacida a partir de una anécdota un tanto estrambótica y acuñada con mucha sorna. (“Piolines”: el nacimiento de una palabra”, Elena Álvarez Mellado, *elDiario.es*, 09/10/2017, [https://www.eldiario.es/opinion/zona-critica/piolines-nacimiento-palabra\\_129\\_3138250.html](https://www.eldiario.es/opinion/zona-critica/piolines-nacimiento-palabra_129_3138250.html)).

- (3) En realidad, la expresión *oír el río* contiene una metonimia, como sucede en *oír un piano* u *oír una moto*, ya que el sustantivo designa en estos casos el sonido que produce aquello a lo que se hace referencia. (“Las palabras como prismas”, Ignacio Bosque, *Archiletras*, 01/02/2021, <https://www.archiletras.com/firma/palabras-como-prismas/>).

#### 4. CONCLUSIONES

El presente trabajo vincula la lingüística de corpus, la terminografía, los textos digitales y la divulgación de la lingüística. La terminología como disciplina teórica, la terminografía como aplicación práctica y la lingüística de corpus como metodología para el análisis de grandes cantidades de texto destacan por su relevancia en el panorama investigador de la lingüística actual. Además, la importancia que la divulgación de la ciencia tiene en la sociedad actual y la relativamente poca atención que se les ha prestado a los textos de corte divulgativo como fuentes que deben ser tenidas en cuenta para la confección de bases de datos terminológicas hacen de este trabajo una propuesta original y novedosa.

El tratamiento de los textos de corte divulgativo –escritos por especialistas, no por periodistas– nos ha permitido poner de manifiesto, en primer lugar, la concentración de este tipo de textos en unos cuantos temas: la lexicografía, el papel de las academias y la ortografía, así como cuestiones etimológicas e históricas, junto con, cierto es, la pragmática y, en ocasiones y siempre dependiendo del autor de los textos, el multilingüismo o la diversidad lingüística. Creemos que el análisis que llevamos a cabo en este trabajo puede servir como punto de partida para reflexionar, como especialistas, sobre qué otros temas podríamos divulgar (al menos en el medio que aquí tratamos, la prensa escrita digital) y qué conviene que la sociedad conozca sobre nuestras investigaciones además de lo que ya estamos contando.

La segunda parte del trabajo se ha centrado en el aprovechamiento que se puede hacer desde la terminografía de un corpus digital de textos divulgativos como el que presentamos. Se han detectado doce términos propios de la semántica (en algunas ocasiones únicos; en otras hemos unido un término y sus derivados adjetivales o de la misma familia léxica): *diccionario*, *metáfora* / *metafórico*, *prototípico*, *neologismo*, *lexicón*, *eufemismo*, *lexicografía* / *lexicográfico* / *lexicógrafo*, *polisemia*, *despectivo*, *significado*, *acepción* y *metonimia*. Para estos términos proponemos, pues, que se incluya su estatus de propios de la divulgación como estatus que revela una alta frecuencia de aparición en los textos de cualquier grado de especialidad. Además, aportamos para cada uno de ellos tres contextos extraídos del corpus, que estimamos relevantes como ejemplos del uso que se hace de ellos en este tipo de textos. Aunque el corpus es susceptible de ser ampliado, no solo con textos más antiguos o de otras zonas hispanohablantes, sino también con textos en otras lenguas, consideramos que este acercamiento a la inclusión de los textos divulgativos en el trabajo terminográfico y el establecimiento de un estatus propio de la divulgación y de contextos extraídos de un corpus de textos divulgativos contribuyen a enriquecer el futuro diccionario de términos de la semántica léxica elaborado en el Instituto de Lingüística Aplicada de la Universidad de Cádiz con datos de frecuencia de uso y contextos representativos que no son tenidos en cuenta de manera habitual en la terminografía.



## 5. REFERENCIAS BIBLIOGRÁFICAS

- Álvarez Torres, V. y S. Gómez Díaz. (2022). Terminología y documentación: la confección de un corpus metalingüístico y multilingüe de semántica léxica. En M. Casas Gómez y C. Varo Varo (Coords.). *La terminología de la semántica* (en prensa). Editorial Universidad de Sevilla.
- Auger, P. y L. J. Rousseau. (1977). *Méthodologie de la recherche terminologique*. Québec: Régie de la langue française. Traducción de J. M.<sup>a</sup> Bermúdez y G. Guerrero Ramos. Universidad de Málaga, 2002.
- Baron, N. (1984). Computer Mediated Communication as a Force in Language Change, *Visible Language*, XVIII(2), 118-141.
- Belda-Medina, J. y R. Casañ-Pitarch (Eds.) (2020). *Análisis del discurso en la era digital. Una recopilación de casos de estudio*. Comares.
- Cabré, M.<sup>a</sup> T. (1993). *La terminología. Teoría, metodología, aplicaciones*. Antártida-Empúries.
- Cabré, M.<sup>a</sup> T. (2002). Textos especializados y unidades de conocimiento: metodología y tipologización. En J. García Palacios y M. T. Fuentes Morán (Eds.). *Texto, terminología y traducción* (pp. 15-36). Almar.
- Cabré, M.<sup>a</sup> T. (2004). ¿Lenguajes especializados o lenguajes para propósitos específicos?, *Revista hispánica de Flandes y Holanda*, 26, número dedicado a: Textos y discursos de especialidad: el español de los negocios, 19-33. [https://doi.org/10.1163/9789401202893\\_003](https://doi.org/10.1163/9789401202893_003)
- Cabré, M.<sup>a</sup> T. e I. da Cunha (2022). El papel de los corpus en la terminología. Una mirada específica a la terminología del español. En G. Parodi, P. Cantos-Gómez y Ch. Howe (Eds.). *Lingüística de corpus en español* (pp. 189-203). Routledge. <https://doi.org/10.4324/9780429329296-16>
- Cantamutto, L. y C. Vela Delfa. (2016). El discurso digital como objeto de estudio: de la descripción de interfaces a la definición de propiedades, *Aposta. Revista de Ciencias Sociales*, 69, 296-323. <http://apostadigital.com/revistav3/hemeroteca/cvela2.pdf>
- Casas Gómez, M. (2006). Modelos representativos de documentación terminográfica y su aplicación a la terminología lingüística, *Revista de Lingüística y Lenguas Aplicadas*, 1, 25-36. <https://doi.org/10.4995/rlyla.2006.680>
- Casas Gómez, M. (2020a). Perspectivas actuales en el estudio del léxico, *Lynx: Panorámica de estudios lingüísticos*, 19, 5-72.
- Casas Gómez, M. (2020b). Conceptual Relationships and their Methodological Representation in a Dictionary of the Terminological Uses of Lexical Semantics, *Fachsprache. Journal of Professional and Scientific Communication*, 42(1-2), 2-26. <https://doi.org/10.24989/fs.v42i1-2.1789>
- Casas Gómez, M. (2020c). *DOCUTERM: Modelo de documentación de usos terminológicos*. Patente registrada con número 202099908175522, número de expediente: CA-00256-2020 y número de asiento registral 04/2020/4177.
- Casas Gómez, M. y M. Hummel (Eds.) (2017). *Semántica léxica*. Número monográfico extraordinario de *RILCE. Revista de Filología Hispánica*, 33(3). <https://doi.org/10.15581/008.33.3.869-911>
- Casas Gómez, M. y G. Fernández Smith (2021). El hablar y la variación de especialidad. En O. Loureda y A. Schrott (Eds.). *Manual de lingüística del hablar* (pp. 715-732). Walter de Gruyter, Colección Manuals of Romance Linguistics (MLR). <https://doi.org/10.1515/9783110335224-036>
- Casas Gómez, M. y C. Varo Varo. (2022). Redes terminológicas de la semántica léxica: bases metodológicas y recursos digitales. En M. Casas Gómez y C. Varo Varo (Coords.). *La terminología de la semántica* (en prensa). Editorial Universidad de Sevilla.
- Crespo Miguel, M. (2020). Lingüística digital: revisión de su estado actual y retos en el Instituto Universitario de Investigación en Lingüística Aplicada de la Universidad de Cádiz. *Pragmalingüística*, 28, 148-165. <https://doi.org/10.25267/Pragmalinguistica.2020.i28.08>
- De Benito Moreno, C. (2022). Uso de los medios digitales de comunicación como corpus de español. En G. Parodi, P. Cantos-Gómez y Ch. Howe (Eds.) *Lingüística de corpus en español* (pp. 480-492). Routledge. <https://doi.org/10.4324/9780429329296-36>

- Dubuc, R. (1999, 3ª ed. corregida y actualizada). *Manual práctico de terminología*. Traducción de I. Cabrera. Unión Latina. RiL editores.
- Duque, E. (2014). Análisis de contenido mediante análisis de palabras clave: La representación de los participantes en los discursos de Esperanza Aguirre. *Mediaciones Sociales. Revista de Ciencias Sociales y de la Comunicación*, 13, 39-73. <https://doi.org/10.5209/revMESO.2014.n13.49432>
- Escandell Vidal, M.ª V. (Coord.) (2011). *Invitación a la lingüística*. Editorial Universitaria Ramón Areces/UNED.
- Estopà, R. (1999). *Extracció de terminologia: elements per a la construcció d'un SEACUSE (Sistema d'Extracció Automàtica de Candidats a Unitats de Significació Especialitzada)*. Tesis doctoral. IULA, Universitat Pompeu Fabra.
- Galán Rodríguez, C. (2003). "La ciencia en zapatillas": análisis del discurso de divulgación científica. *Anuario de estudios filológicos*, 26, 137-156.
- Galán Rodríguez, C. (2021). El discurso tecnocientífico. En O. Loureda y A. Schrott (Eds.). *Manual de lingüística del hablar (657-673)*. Walter de Gruyter, Colección Manuals of Romance Linguistics (MLR), 28. <https://doi.org/10.1515/9783110335224-033>
- Galán Rodríguez, C. y J. Montero Melchor (2002). *El discurso tecnocientífico: la caja de herramientas del lenguaje*. Arco/Libros.
- Gaviño Rodríguez, V. (2021). La polémica lingüística como tipología discursiva en la prensa española del siglo XIX. Los discursos polémicos de Fernando Gómez de Salazar. *Pragmalingüística*, (29), 173-189. <https://doi.org/10.25267/Pragmalinguistica.2021.i29.09>
- Guantiva Acosta, R., M.ª T. Cabré y J. M. Castellà Lidon (2008). Clasificación de textos especializados a partir de su terminología. *Íkala, revista de lenguaje y cultura*, 13(19), 15-39.
- Kilgarrif, A., M. Husák, K. McAdam, M. Rundell y P. Rychlý (2008). GDEX: Automatically finding good dictionary examples in a corpus. En *Proceedings of the 13th EURALEX International Congress (Barcelona, 15-19 July 2008)*, (pp. 425-432). IULA, Universitat Pompeu Fabra, Documenta Universitaria.
- Kilgarrif, A., V. Baisa, J. Bušta, M. Jakubíček, V. Kovář, J. Michelfeit, P. Rychlý y V. Suchomel. (2014). The Sketch Engine: ten years on, *Lexicography*, 1, 7-36. <https://doi.org/10.1007/s40607-014-0009-9>
- Llácer Llorca, E. V. y F. J. Ballesteros Roselló (2012). El lenguaje científico, la divulgación de la ciencia y el riesgo de las pseudociencias, *Quaderns de filologia. Estudis lingüístics*, 17, 51-67.
- Luzón, M. J. y C. Pérez-Llantada (Eds.) (2019). *Science Communication on the Internet. Old genres meet new genres*. John Benjamins Publishing Company. <https://doi.org/10.1075/pbns.308>
- Mancera Rueda A. y A. Pano Alamán (2013). Nuevas dinámicas discursivas en la comunicación política en Twitter. *Círculo de Lingüística Aplicada a la Comunicación*, 56, 53-80. <https://doi.org/10.5209/revCLAC.2013.v56.43867>
- Mancera Rueda A. y A. Pano Alamán (2014). Las redes sociales como corpus de estudio para el análisis del discurso mediado por ordenador. En S. López Poza y N. Pena Sueiro (Eds.). *Humanidades Digitales: desafíos, logros y perspectivas de futuro* (pp. 305-315). Janus, Anexo 1. Recuperado el 20 de marzo de 2022, de <https://www.janusdigital.es/anexos/contribucion/descargar.htm?id=28>
- Marimón Llorca, C. (Ed.) (2020). *El columnismo lingüístico en España desde 1940. Análisis multidimensional y caracterización genérica*. Arco / Libros.
- Olmedo Estrada, J. C. (2011). Educación y Divulgación de la Ciencia: Tendiendo puentes hacia la alfabetización científica, *Revista Eureka sobre Enseñanza y Divulgación de las Ciencias*, 8(2), 137-148. [https://doi.org/10.25267/Rev\\_Eureka\\_ensen\\_divulg\\_cienc.2011.v8.i2.01](https://doi.org/10.25267/Rev_Eureka_ensen_divulg_cienc.2011.v8.i2.01)
- Pano Alamán, A. y Moya Muñoz, P. (2015). CorpusRedEs. Proyecto de creación y anotación de un corpus de comunicación mediada por ordenador en español. *CHIMERA: Romance Corpora and Linguistic Studies*, 2, 117-129. <https://doi.org/10.15366/chimera2015.2>
- Parodi, G. (2008). Lingüística de corpus: una introducción al ámbito. *Revista de Lingüística Teórica y Aplicada*, 46(1), 93-119. <https://doi.org/10.4067/S0718-48832008000100006>

- Rayson, P. (2008). From key words to key semantic domains, *International Journal of Corpus Linguistics*, 13(4), 519-549. <https://doi.org/10.1075/ijcl.13.4.06ray>
- Rodríguez Tapia, S. (2016). Los textos especializados, semiespecializados y divulgativos: una propuesta de análisis cualitativo y clasificación cuantitativa, *Signa*, 25, 987-1006. <https://doi.org/10.5944/signa.vol25.2016.16926>
- Sánchez-Saus Laserna, M. (2022). Las redes sociales como corpus para el estudio lingüístico. Revisión bibliográfica y catálogo de herramientas digitales. En L. Mañas Viniegra, I. Sacaluga Rodríguez y S. Mariscal Vega (Coords.), *Manifestaciones del humanismo en el siglo XXI* (pp. 455-464). Tirant lo Blanch.
- Scott, M. y C. Tribble. (2006). *Textual Patterns: key words and corpus analysis in language education*. John Benjamins Publishing Company. <https://doi.org/10.1075/scl.22>
- Sinclair, J. (2004). *Trust the text: language, corpus and discourse*. London/New York: Routledge. <https://doi.org/10.4324/9780203594070>
- Vela Delfa, C. y L. Cantamutto (2015). Methodological approach to the design of digital discourse corpora in Spanish. Proposal of the CÓDICE project. En P. Fuertes Olivera *et al.* (Eds.). *Current work in corpus linguistics: Working with traditionally-conceived corpora and beyond. Selected papers from the 7th International Conference on Corpus Linguistics (CILC2015)* (pp. 494-499). <https://doi.org/10.1016/j.sbspro.2015.07.472>
- Vivaldi, J. (2001). *Extracción de candidatos a término mediante combinación de estrategias heterogéneas*. Tesis doctoral. IULA, Universitat Pompeu Fabra.
- Zulaica Hernández, I. (2015). Lingüística de corpus. En J. Gutiérrez Rexac (Ed.). *Enciclopedia de Lingüística Hispánica*, volumen 1 (pp. 216-224). Routledge. <https://doi.org/10.4324/9781315713441-20>

## Emotions in specialised genres: Power, manipulation and persuasion from the *Affect Spectrum Theory*

Las emociones en los géneros textuales especializados: Poder, manipulación y persuasión desde la AST, o *Teoría del Espectro Afectivo*

MARÍA ÁNGELES ORTS LLOPIS  
Universidad de Murcia, España  
mageorts@um.es  
<http://orcid.org/0000-0002-8178-9255>

*The emotional life is not simply a part or aspect of human life. It is not, as we so often think, subordinate or subsidiary to the mind. It is the core and essence of human life. The intellect arises out of it, is rooted in it, draws its nourishment and sustenance from it, and it is the subordinate partner in the human economy.*

(John Macmurray 1962/1992:42-43)

### Abstract

This paper seeks to examine professional genres from a novel perspective linking social constructivism and basic emotion theory, making the assumption that emotions are intrinsically social, and that social phenomena (such as professional communities) have an emotional nature (TenHouten, 2021). Thus, we aim to consider the constructive role of emotions in the creation and development of specialised texts (occasionally focusing specifically on the legal field), and, in turn, on the professional commu-

### Resumen

Este trabajo pretende examinar los géneros profesionales desde una perspectiva novedosa que vincula el constructivismo social y la teoría básica de la emoción, partiendo del supuesto de que las emociones son intrínsecamente de carácter social, y de que los fenómenos sociales (como son las comunidades profesionales) tienen una naturaleza emocional (TenHouten, 2021). Así, pretendemos considerar el papel constructivo de las emociones en la creación y el desarrollo de los textos especializados (en algunos casos, centrándonos específicamente en el ámbito ju-

---

**To cite this article:** Orts Llopis, M. A. (2022). Emotions in specialised genres: Power, manipulation and persuasion from the *Affect Spectrum Theory*. *ELUA*, (38), 99-115. <https://doi.org/10.14198/ELUA.21432>

Recibido: 25/11/2021 Aceptado: 06/05/2022

© 2022 María Ángeles Orts Llopis



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

nities from which these texts emerge. Our work builds on the socio-evolutionary theory of emotions, or *affect-spectrum theory* (AST), which is developed in dissonance with other Cartesian and rational-choice models that set reason and emotion in opposition and assert that society only progresses to the extent that the former can control, suppress and triumph over the latter. Specifically, it is our underlying hypothesis that specialised and professional communication is articulated around emotions –and power, manipulation and persuasion, among other negative or positive emotional expressions. Thus, textual mechanisms or genres, as communicative instruments of specialised communities reveal the tensions that occur centripetally or centrifugally, i.e., inwards or outwards, from that community. They do so from an agonistic relational model, in a centripetal tension that promotes agency through coercion and/or manipulation, or from a hedonistic relational model, in a centrifugal tension where interpersonal relationships of persuasion and dissemination are established in order to engage in communal relationships.

**KEYWORDS:** AST, professional genres, persuasion, manipulation, power, specialised communities

rídico) y, a su vez, en las comunidades profesionales de las que surgen estos textos. Nuestro trabajo se basa en la teoría socio-evolutiva de las emociones, o teoría del espectro afectivo (en inglés, AST o *affect-spectrum theory*), que se desarrolla en disonancia con otros modelos cartesianos y de elección racional que contraponen razón y emoción y afirman que la sociedad sólo progresa en la medida en que la primera puede controlar, suprimir y triunfar sobre la segunda. En concreto, nuestra hipótesis de fondo es que la comunicación especializada y profesional se articula en torno a las emociones -y al poder, la manipulación y la persuasión, entre otras expresiones emocionales negativas o positivas-. Así, los mecanismos o géneros textuales, como instrumentos comunicativos de las comunidades especializadas, revelan las tensiones que se producen de forma centrípeta o centrífuga, es decir, hacia dentro o hacia fuera de esa comunidad. Para ello, lo hacen desde un modelo relacional agonista, desplegando la autonomía en una dinámica centrípeta a través de la coerción y/o la manipulación), o desde un modelo relacional hedonista, empleando la interpersonalidad en una dinámica centrífuga que les permita entablar relaciones comunitarias a través de la persuasión y la divulgación.

**PALABRAS CLAVE:** AST, géneros profesionales, persuasión, manipulación, poder, comunidades especializadas.

## 1. INTRODUCTION

This research addresses professional genre studies from a novel perspective that links social constructivism and basic emotion theory and assumes that emotions are intrinsically social, and that social phenomena (such as professional communities) have an emotional nature (TenHouten 2021). Our aim will therefore be to analyse the constructive role of emotions in the creation and development of specialised texts (occasionally focusing specifically on the legal field), and, in turn, on the professional communities from which these texts emerge. Our work builds on the socio-evolutionary theory of emotions, or *affect-spectrum theory* (AST), which is developed in dissonance with other Cartesian and rational-choice models that set reason and emotion in opposition and assert that society only progresses to the extent that the former can control, suppress and triumph over the latter. Sociologists, anthropologists and neuroscientists have, however, in recent decades argued, to the contrary, that emotions are not opposed to reason, since they actually play a key role in formulating logical goals, rational decision-making and the pursuit of human

achievements (Damasio 1994, 2001 and 2018; Barbalet 1998, for example). In fact, both reason and emotion are foundations of social relations, processes and institutions. Damasio in particular (1994, 2020) talks about how emotions (such as anger, anticipation, surprise, fear or disappointment) can contribute to making decisions swiftly and effectively, especially when it comes to situations of risk or uncertainty where the cognitive part of the brain would need to spend a lot of time and energy to make a rational decision. In these cases, it is ‘somatic markers’ or ‘gut-feelings’ (Damasio *et al.* 1991) –a mixture of experiences (the memory of other emotions and feelings) and swift cognitive processes– which trigger these rapid responses. Thus, our study deviates from the reductionist utilitarianism of rational choice theory (Persky 1995), according to which we make decisions solely on the basis of reasoning and cost-benefit calculations (Barbalet 1998), our tenet being that reasoning, motivation and emotions interact in the context of social relations (TenHouten 2014: 2). In the Humanities and the Social Sciences, interest in language and emotion is part of the so-called ‘emotional turn’, similar to the ‘linguistic turn’ that took place in the 1970s. If the linguistic turn claims that language helps to constitute reality, the emotional, or affective, turn implies that emotions play an equally crucial role in human experience (Lemmings and Brooks, 2014). In linguistics, emotions and language have been addressed, for example, by cognitive linguistics (Schwarz-Friesel 2015; Foolen 2012 and 2016), semiotics (Lüdtke, 2012), and applied linguistics (Benítez-Castro & Hidalgo-Tenorio 2019 and 2021; Alba-Juez 2018; Alba-Juez & Mackenzie 2019), mainly within the area of appraisal theory<sup>1</sup>. This article sets out to explore the theory of specialised genres from a different, but complementary, perspective, that of AST and human evolution.

Our underlying premise is that specialised and professional communication – i.e., the communicative activity that professional communities pursue in their role as “discursive communities” (Swales, 1990)– builds upon emotions, being power, manipulation and persuasion expressions of such emotions. This is because emotions are intrinsically social, and their main adaptive function takes place through interpersonal communication (Miller *et al.* 2004), which can only occur through language. The transmission of meanings, ideological frameworks and constructs of social and professional institutions is based upon the use of language as the main means of communication and is “intrinsically related to human cognitive processes” (Salmi-Tolonen 2011: 1). As such, the genres of a specialised community are powerful tools which organize the internal communication among its members and make it possible –and effective– for them to engage with other specialised communities (Bawarshi & Reiff 2010); additionally, professional genres contain the knowledge intrinsic to a given discipline, reflecting its ideologies, power structures and social patterns (Gunnarsson, Linell & Nordberg 1997: 3). But again, emotions are central to organise the individual’s experience of reality, sense of self and competitive and cooperative orientations towards others, and, hence, in the creation of human cultural enterprises (TenHouten 2014; Damasio 2018), which is why they play a crucial role in the way in which disciplinary communities interact.

---

<sup>1</sup> Appraisal theory constitutes the reformulation of Halliday’s systemic functional grammar by the so-called Sydney School (White 2003; Martin & White 2005; Martin & Rose 2007), and is concerned with the linguistic resources for by which a texts/speakers come to express, negotiate and naturalize particular inter-subjective (attitudinal) and ultimately ideological positions.

## 2. AFFECT SPECTRUM THEORY (AST)

Emotions have a deep evolutionary history. Plutchik (1979, 1980) was the first to develop Darwin's theory (1872) and identified eight primary emotions, corresponding to the four existential problems all humans face, and their consequent adaptive reaction. *Identity* has to do with who we are and is associated with the emotions of acceptance and rejection; hierarchy deals with the way in which we organize ourselves and is associated with the emotions of anger and fear; *territoriality* establishes our physical boundaries (our property and that of others) and is associated with the emotions of anticipation and surprise. Finally, *temporality*, the fact that we are born, we reproduce ourselves and we die, is associated with joy/happiness and sadness. Later, MacLean (1990) improved this taxonomy, adding a sociological dimension that had hitherto been absent, since Plutchik considered emotions as merely adaptive stimuli in the face of the four vital problems. To this end, MacLean first established the concept of the 'triune brain', according to which the human brain is a compendium of three brains in one, the result of evolutionary processes whereby three neural systems have been created that regulate behavioural and physiological adaptation. Of these three evolutionary parts, the 'reptilian brain' is the most primitive form, and is responsible for maintaining the functions necessary for immediate survival; the 'limbic', or mammalian, brain is the system that enables mammals to achieve higher levels of motivation, emotional response and social engagement; and, the 'neocortex' is the centre of rationality in our nervous system, allowing for systematic and logical thinking, which exists independently of the emotions and behaviour programmed by our genetics. In addition, MacLean provided Plutchik's fourfold model with a communicative dimension, based on four neurobiological reactions. In MacLean's theory, identity was called 'signature' and was attributed a social dimension. 'Social identity'—key to the study of communication in professional communities—is the dimension of individuals that arises from their membership of, and identification with, the social categories and groups to which they belong. What Plutchik had called 'hierarchy' was replaced by MacLean's 'aggression/submission', which is very pertinent to the present study in terms of explaining the monogloss orientation of certain genres; that is, the exercise of power through language. Underlying hierarchy is the desire for dominance, an impossible-to-eradicate aspect of social relations, deeply rooted in the reptilian and ancestral human mind—in contrast to social identity, which belongs to the later-developing limbic brain. Finally, 'temporality' (the 'courtship' dimension for MacLean) and 'territoriality' refer, respectively, to the human lifespan, the need for the perpetuation of the species, family and groups, and to natural behaviour oriented towards the control, possession, use and defence of space which is necessary for survival.

All these neurobiological, sociological and anthropological findings were subsequently developed in Alan Fiske's theory of the four relational models (1991). Fiske claimed that humans have addressed the four existential problems discussed by Plutchik and MacLean with four fundamental adaptive relationships that have developed through gradual processes of evolution: 'communal-sharing', 'authority-ranking', 'equality-matching' and 'market-pricing'. Each model expresses a great plurality of social actions, beliefs or judgements in different cultures which are not the consequence of the experience of an individual; rather, they are universally-shared mental models (Pinker 2002). Prior to Fiske, but in the same vein, ethologist Chance (1988, in TenHouten 2014: 42), had distinguished between two forms of social interaction, or modes of thought, in his studies on primate and human

behaviour: formal or ‘agonic’ society, which is characterised by threat, power and anxiety, and informal or ‘hedonic’ society, which is categorised as hosting egalitarian and community-shared forms of relationships. Agonic society relates to Fiske’s authority-ranking and market-pricing relations; the former is reptilian and primitive, where high-status individuals exercise leadership over social prerogatives, choices and preferences and can exercise coercive power through force, threat or ideological persuasion. In contrast, market-pricing relationships are not instinctive, but result from a later stage of brain development. In this same line, rational choice (Tversky & Kahneman 1981, for example) and related theories postulate that in market-based exchange relationships individuals strive to maximise their material interest when participating in the social world. On the other hand, hedonic society is based on equality-matching relationships (reciprocity and balanced social relations) and community-based ones (equivalence, unity, collective identification without differentiation).

Individuals and social groups thus experience a double tension: community versus market, hierarchy versus equality, as opposing tendencies either towards egocentricity –which TenHouten (2014: 38) calls ‘agency’, or self-protection and material self-interest–altruism, or non-contractual cooperation. The aim of this study is to find out how these tensions influence the interactions within and between different discourse communities and how they are expressed and constructed through the different genres.

### 3. EMOTION AND SPECIALISED GENRES

The most recent trends in the study of emotion and genre give both concepts a constructivist perspective, as artifacts that generate social change. On the one hand, Barbalet (1998) writes about the double dimension of emotion: both the experiential and contextual elements of emotion are necessary to conceptualise it as a principal factor for social change. In the same vein as Barbalet, Foucault (1989), Laclau (2004) and Koschut (2018, 2020) all argue that it is through language that emotions contribute significantly to the discursive construction of social identities and power relations. Without disregarding their more physical and individual aspect, emotions are constructed at the social level and can change it, thus shaping power and status relations. Accordingly, certain discourses can manifest themselves as dominant (those of the ‘self’ or ‘sameness’) or marginal (those of ‘others’ or ‘otherness’) depending on whether they are identified with members of the established group or refer to those outside it (Koschut 2018: 510). By recognising the central role of emotions in the completion of basic cognitive processes such as learning and decision-making (Zlatev 2012: 3), we acknowledge that the natural function of language is the conceptualisation and expression of emotion (Foolen 2012: 350). Likewise, language is inextricably linked to institutions as discursive communities: it is the fibre from which such communities are made (Fairclough & Wodack 1997), and is deployed to create and impose their own discourses (Simpson & Mayr 2010) or their own genres as an expression of those discourses. In contrast to earlier and ‘flatter’ approaches to genre theory, rhetorical linguistics (Bawarshi & Reiff 2010) and Critical Genre Analysis (Bhatia, 2004, 2008, 2012), view genres as powerful, ideologically active and historically changing elements that shape texts, meanings and social actions. The multidimensional and complex understanding of genre as a dynamic concept, marked by stability and change–functioning as a form of specialised cognition, linked to ideology, power and social actions and relations, and recurrently contributing to community creation



and development—is, in my view, entirely complementary to Fiske’s and Chance’s relational models in the previous section. In line with this AST socio-evolutionary perspective, opposing human tendencies coexist in professional discourse communities: the impetus to obtain a dominant social position and interpersonal distance, creating unequal and hierarchical social relations—reflected through communicative opacity and epistemological asymmetries—versus the desire of professional groups to achieve the acceptance of others in the exercise of ideological persuasion and knowledge dissemination (Engberg, Luttermann, Cacchiani & Preite, 2018). In other words, corporativism as self-identity in professions— which requires the ability to successfully negotiate the social world— and the need to establish interpersonal relations with the rest of society, coexist side by side in institutions and occupational groups. It is both a centripetal and a centrifugal force which reflects the reality of social interactions: maintaining one’s social status and socio-economic resources and creating and sustaining positive social connections. In fact, both critical linguistics (Fowler, Hodge, Kress & Trew 1979; Kress & Hodge 1979) and critical discourse analysis (Fairclough 1989, 2014; van Dijk 2008), understand expert communities as constituting a dominant bloc that treats social hierarchies as natural and produces, maintains, and replicates underlying asymmetrical power relations (Fox & Fox 2004: 17) to achieve domination and subordinate individuals (Mumby & Clair 1997; van Dijk 1993). Opacity, distance and other rhetorical devices are, then, deployed by communities in order to impose their unquestioned authority and exercise manipulation and negative persuasion. But, at the same time, institutions are dialogical social realities where meanings must be negotiated between issuers and receivers (Fairclough & Wodack 1997; Mumby & Clair 1997); hence, specialised communities need, additionally, to work towards consensus, which is achieved through positive persuasion and the dissemination of their epistemology through genres that reveal the expert knowledge and inner workings of professional communities.

Finally, in this paper we will focus equally upon both agonic and hedonic communication channels—authority, negative persuasion and manipulation—as well as positive persuasion and dissemination. This tension between the need to impose the discursive supremacy of professional communities and the need to establish interpersonal relations is what we will discuss below as, respectively, monogloss or heterogloss discourse (Bakhtin 1981) as expressions of agonic and hedonic modes of relationship, respectively.

#### **4. AGONIC SOCIETY: THE MONOGLOSS DISCOURSE OF AUTHORITY AND MANIPULATION**

The difference between monogloss and heterogloss discourse—related to Foucault’s (1994) intertextuality and recent studies on interpersonality in systemic-functional linguistics, especially the notion of ‘engagement’ in appraisal theory (White 2003; Martin & White 2005; Martin & Rose 2007)— is the result of Bakhtin’s (1981) translinguistic theory, which gives speakers the possibility of creating multiple ways of inserting themselves into discourse and constructing different perspectives on it. While taking for granted that all discourse involves dialogism (since there is no discourse that is isolated from the multiplicity of past or future texts), Bakhtin differentiates between the position of the distant speaker who ignores the points of view of the other speakers (‘monoglossia’) and those utterances which give participation to other voices (‘heteroglossia’). From our perspective, ‘monogloss discourse’ is proto-

typical of agonistic modes of relationship whereby, in order to maintain their social dominance and status in the hierarchy, specialised communities control rhetorical resources in a monopolistic manner. Partington and Taylor indeed affirm (2018: 50) that persuasion through the acceptance of authority is the most common form of persuasion. Thus, disciplines make use of the Aristotelian *logos*, the supremacy of reason and logic, to assert themselves. For example, Weber's 'formal rationality' (1921/1978: 75-76, in TenHouten 2014: 101), in line with rational choice theories, postulates that calculability and impersonality are geared towards the achievement of social success. Law (Orts 2015, 2016), but also other disciplines such as medicine (Carter-Thomas and Rowley-Jollivet 2017; Garzone 2017) or hard sciences (Llácer & Ballesteros 2012) make use of impersonal, opaque texts that exhibit inequality or asymmetric relationships between issuers and receivers of texts. Asymmetries may be ideological (the alienation of others), material (greater government resources or wealth) or geopolitical. But in discursive communities, asymmetry entails epistemic asymmetry: a monopoly over what Engberg *et al.* call domain-internal knowledge (2018: xi). Epistemic asymmetry occurs in highly specialised, expert-to-expert texts, favouring discursive distance between issuer and receiver in order to maintain gnoseological dominance (Orts 2015, 2016). Thus, dominant groups work to remain dominant by disseminating the beliefs, practices and genres of the discipline (their expertise) through language. Furthermore, and paradoxically, as TenHouten argues (1999: 266), force, threat or ideological persuasion may also be directed "pastorally", to make lower-ranking individuals perceive that those in power are responsible for providing protection, help and support. In this section of the paper we make a series of considerations about the two mechanisms of monoglossia in professions, authority and manipulation, as negative persuasion involving the violation of epistemic trust.

#### 4.1. Monogloss discourse resources: authority in legal texts

We have chosen legal discourse to illustrate the presence of authority and power in specialised genres. According to the literature, legal discourse is quite impersonal, features of interpersonality hardly being present at all (e.g. Bhatia, 1982; Hiltunen, 1990; Tiersma 1999, in Breeze *et al.*, 2014, 13). According to Salmi-Tolonen (2014, 64); legal discourse has, in fact, three main characteristics: it is always normative in nature; always functional and/or admonitory (linked to some part of the legal order); and always professional or institutional (a specific kind of discourse). In an analysis of international legal texts in English (Orts, 2016), I used genre theory to show that the lexical-discursive complexity of some legal texts is inextricably linked to the desire of domination. But it is the concept of power distance (hereafter, PD) from social anthropology (Hofstede, 1983, 1985, in Orts, 2016, 4) which might be most helpful in studying symmetry differences between participants in legal discourse, and uncovering the ways in which power and imposition are concealed in these genres. The concept of PD is very much in line with the Hallidayan definition of 'tenor' as the role of relations between participants in terms of status, affect and contact, and constitutes a decisive factor in the formality of texts. Tenor influences interpersonal choices in the language system and, therefore, affects the structures and strategies chosen to activate verbal formality and complexity. Similarly, the PD index can be used to establish the presence or absence of a hierarchical distance between issuers and recipients of legal texts. As Table 1 shows, in order to isolate patterns of pragmatic-linguistic realisation of directives and commissives in international normative and contractual

texts, I adapted the concept of PD to Trosborg's taxonomy (1995, in Orts 2016: 5-7), on a scale ranging from speech acts of obligation, prohibition and permission, to commissive speech acts, such as those constituting contractual promises.

STRATEGY	MEANING	FULFILMENT
<b>OBLIGATION</b>	HIGHEST POWER DISTANCE (MAXIMUM IMPOSITION, BALD ON RECORD)	MUST/SHALL+VERB BE+TO, HAVE TO <i>LEXICAL VERBS:</i> OBLIGATE, OBLIGED, ORDER, DEMAND
<b>PROHIBITION</b>	HIGHEST POWER DISTANCE (MAXIMUM IMPOSITION, BALD ON RECORD)	MUST/SHALL+NEGATIVE (NOT, NOTHING, NO)
<b>OBLIGATION WITH REDRESS</b>	MEDIUM POWER DISTANCE (FACE REDRESS, OFF RECORD)	SHALL/MUST +PASSIVE BE+TO, HAVE TO+PASSIVE SHOULD
<b>PROHIBITION WITH REDRESS</b>	MEDIUM POWER DISTANCE (FACE REDRESS, OFF RECORD)	SHALL+NEGATIVE+PASSIVE CAN/MAY/WILL+NEGATIVE (NOT, NOTHING, NO)
<b>ASSIGNMENT OF RIGHTS, PERMISSION</b>	LOW POWER DISTANCE (ASSIGNMENT OF RIGHTS)	CAN, MAY <i>LEXICAL VERBS:</i> GRANT, GIVE, ALLOW
<b>PROMISES, VOLITION</b>	NEUTRAL POWER DISTANCE (SYMMETRICAL RELATIONSHIP)	WILL <i>LEXICAL VERBS:</i> AGREE, UNDERTAKE, ACCEPT, WARRANT, PROMISE, ACKNOWLEDGE

Table 1. Power distance strategies in directives and commissives (Orts, 2016).

Directives –those appearing in laws or regulations– are face-threatening acts (FTAs) with maximum illocutionary force (Leech 1983, in Orts 2016: 4). Such imposition can be conveyed ‘bald-on-record’, i.e. directly, clearly, unequivocally and concisely, or can be conveyed ‘off-record’, i.e. minimising imposition (Brown & Levinson, 1987, in Orts 2016: 4), thus reinforcing or softening the existing asymmetry between legislature and citizens. In this study I was able to observe that there is a scalarity between maximum prescription and greater PD on the one hand, and the granting of rights and less PD on the other, depending on the function of the text and the tenor between the participants: the issuers of the norm, and the users or receivers of normative texts (i.e., citizens), or as contracting parties in a commercial agreement. However, there are ‘hybrid’, reciprocal, normative texts, such as the rules of international arbitration courts, where users have the discretion to adhere to them or not in order to solve their disputes. Additionally, and paradoxically, there are also commissive texts, coming from very powerful institutions such as the Lloyd's Institute of Underwriters, which impose onerous, authoritative contractual clauses on their users. All of this is an illustration of the fact that the language of imposition appears irregularly in some legal genres, often constituting, however, a powerful mechanism of negative persuasion.

#### 4.2. Manipulation as a monogloss resource in media genres

Manipulation is here regarded as monogloss discourse insofar as it arbitrarily directs the opinion and action of receivers. Manipulation, as Cialdini points out (2001: 73), is the

ability to produce a specific type of automatic and thoughtless compliance on the part of individuals; the willingness to say “yes” without thinking first. In my view, this is an implicit, not explicit, exercise of power that deploys emotions in a mercenary and advantageous way. In Aristotelian terms, the kind of reaction that the expert community should provoke is the confidence that something is morally irreproachable and clearly trustworthy (*ethos*), that it is based on logic (*logos*) and that it can even be a source of satisfaction (*pathos*). Manipulation, then, becomes an implicit violation of these three principles; it turns into an exercise of power that deploys emotions in a mercenary and advantageous way, based upon the distortion of truth and the violation of trust, as one of the somatic markers or gut-feelings that aid decision-making (Damasio *et al.* 1991; Damasio 1994, 2018).

Alba-Juez and Mackenzie (2019) consider lying and manipulation as a type of unreliable persuasion elicited by a series of emotions that call into question reliability itself. In their study of fake news they discuss relevant concepts such as ‘bullshit’ (Frankfurt 2005, in Alba-Juez & Mackenzie 1999: 20), ‘post-truth’ (Harari 2018, in Alba-Juez & Mackenzie 1999: 20), ‘flashbulb memories’ (highly salient and extremely emotional memories, subjectively permanent in time and highly manipulable, according to Brown & Kulik 1977, in Alba-Juez & Mackenzie 1999: 29), ‘misinformation’ (the dissemination of false information) and ‘disinformation’, or fabrications (false information within real contexts, according to Stahl 2010, in Alba-Juez & Mackenzie 1999: 25). The concept of post-truth is particularly interesting as it underlies the consensus generated by hierarchies of power to establish certain ideas as true with the validation of the media (Sztanjnsrajber 2017 in Murolo 2019). The use of manipulation is carried out through external and internal textual devices: interdiscursive appropriation of other discourses and voices known to be truthful is a very important intertextual device, often exploited to establish the credibility of news and journalistic reports. Bhatia and Bhatia (2017), for example, study sensationalism in the news broadcast in the wake of the Panama Papers leaks as a type of infotainment, i.e., a hybrid genre where reliable information is mixed with the gimmickry intended for entertainment and spectacle with the ultimate aim of discrediting certain public figures (Bhatia and Bhatia 2017: 33). On the other hand, they also identify a series of internal textual resources such as the use of modals, pronouns or extraposition –which are used to modulate epistemic statements and generate an air of consensus– as well as questions with persuasive or coercive force. In questions, it is the sender who controls the information, often answering on behalf of the reader, and they become textual cues to drive discourse in the direction intended by the sender. In writing, they often reveal the author’s power over the textual material and over the reader, evidencing a clear asymmetry between discourse participants and potentially constituting image-threatening acts, especially when they are hostile (Carter-Thomas & Rowley-Jollivet 2017; Partington & Taylor 2018).

Verbal manipulation may also be regarded as “persuasion by appeal to emotions” (Partington & Taylor 2018: 40), often resulting from subtle lexical choices. Whyte (2003 in Partington & Taylor 2018: 29) coined the term ‘hooray word’ –for those words with a socially positive connotation which almost automatically arouse a feeling of approval in the audience–, and the term ‘boo word’ (for those which have a negative connotation and automatically arouse a feeling of disapproval in the listener) to illustrate such choices. Bocanegra-Valle (2017), for example, studies the annual speeches of the Secretary-General of the International Maritime Organization (IMO), and has identified four main topics that

recur in his argumentation: ‘globalisation’, ‘leadership’, ‘hegemony’ and ‘legitimation’. Globalisation emerges as a generative topic that enhances the institution’s interests and makes its leadership evident in the eyes of multiple audiences. The IMO deploys it to boost its public image as an accountable actor on the world stage.

Likewise, there are manipulative exercises, such as *ad hominem* criticism (discrediting by attacking the person, rather than their arguments or ideology), ‘euphemism’ (the neutralisation of negative concepts in order to normalise them), ‘dysphemism’ (exaggerated negative criticism) and ‘binarism’ or ‘false dichotomy’ (used for argumentative polarisation) (Partington & Taylor 2018: 81-83). Hidalgo-Tenorio and Sánchez-García (2017) carry out an interesting study on a series of discursive strategies in political language, including the above, and other conversational fallacies, such as *ad nauseam* (basing the truth of an argument on its repetition) or *ad ignorantiam* (defending a position by arguing that there are no arguments to the contrary), among others. All these studies are also included in the second type of negative persuasion, i.e. manipulation as a corruption of epistemic trust.

## 5. HEDONIC SOCIETY: PERSUASION, EMOTION, DISCLOSURE

The centrifugal force involved in establishing a dialogue between sender and receiver, which we have called ‘heteroglossia’, requires us to return to the beginning of our work and talk about AST. As already noted, while hierarchy and territorial dominance are agonistic instincts of our reptilian heritage, Chance’s hedonic mode of thought involves the later evolution of our limbic brain to regulate emotions (1988, in TenHouten 1999: 259-262). At this point, it is essential to talk about the concept of ‘homeostasis’ (Damasio 1994, in Grijalba-Uche & Echarte 2015: 89; Damasio 2018). Homeostasis is the eternal quest of human beings to achieve balance, adaptive response and well-being to ensure social survival. Professional communities, such as human social organisations, seek this balance in the continuous tension between their own interests (their internal hierarchies, and their competitiveness with other communities), and –contrarily– the need for solidarity and reciprocity with such communities. In order to positively relate to the rest of society, and to achieve homeostasis, discursive communities have to legitimise their discourses and their genres, which implies, on the one hand, the deployment of strategies of positive verbal persuasion by the issuers of these genres and discourses, and on the other, their dissemination. And in this transition towards their social immersion, the credibility of experts on the part of non-experts is necessary. Trust (the *tharsos* of Nicomachean rhetoric) is a ‘meta-emotion’ (Belli & Broncano 2017), a complex set of somatic markers that mix primary emotions such as joy and fear with conscious rational processes. Of these processes, the main one is the voluntary suspension of disbelief, which can only be obtained through the image of a coherent discursive community and the will of communicative cooperation of the lay members of society. Here we will briefly discuss trust in the context of popularising texts, where the layperson puts himself or herself in the hands of the expert to gain access to specialised knowledge (Engberg *et al.* 2018; Engberg 2020).

### 5.1. Persuasion and dissemination in specialised discourse

As was mentioned above, persuasion (which we have branded as ‘positive’, in order to distinguish it from negative persuasion, or manipulation), as well as dissemination, are the

main means for professional communities to create synergy and gain credibility in the social context. Studies on persuasive language follow two fundamental approaches in applied linguistics: metadiscourse<sup>2</sup> and evaluation<sup>3</sup>, consciously leaving metaphor aside, which would require a study of its own. Thus, on the one hand, metadiscourse, the construction of *ethos* or textual *personas*, has been extensively studied mainly in relation to journalistic and academic texts (Hyland & Tse 2004; Abdollahzadeh, 2007; Bednarek 2010; Dafouz & Nuñez 2010; González-Rodríguez 2011, in Orts 2015, among others), in fields such as tourism (Suau 2006; Suau & Labarta 2006, in Gallego-Hernández 2017: 224) and economics (Orts 2016; Gallego-Hernández 2017). Textual and interpersonal markers of metadiscourse are the set of strategies that reveal the existence of a dialogical framework between writers and readers of texts (Vande Kopple 1985; Crismore, Markkanen & Steffensen, 1993; Dahl 2004; Hyland 2005; Dafouz 2008, in Orts, 2015). It is the means by which the propositional content of utterances is made coherent, intelligible and persuasive to the receivers of texts (Hyland 2005: 39, in Orts 2015). The paradigm has been habitually neglected in fields such as law, due to the monogloss character of the texts in this field, which are allegedly written to be interpreted, rather than read and understood. However, the presence of ideational markers (those that indicate semantic and syntactic relations between sections of discourse) is essential in trying to discern patterns of difficulty or clarity in such texts or, rather, the ways in which legislators wish to organise information in a coherent and convincing way for the audience (Orts 2016, 2017).

On the other hand, persuasion is intrinsically related to evaluation (Partington & Taylor (2018: 18): we persuade by trying to convince our recipients of the goodness of our opinions over others. Evaluative language allows the author to express an attitude on an issue from both an objective and a subjective stance, where the choice of language is tactical. In fact, appraisal theory (White 2003; Martin & White 2005; Martin & Rose 2007) is one of the most powerful ways to analyse emotional language. It connects directly to the notion of heteroglossia, studying utterances that convey a positive or negative evaluation, or that can be interpreted as an invitation for the reader to provide his or her own assessment. The theory divides evaluative resources into three broad semantic domains: attitude, engagement, and gradation. Engagement places more emphasis than metadiscourse on heteroglossia by positioning the voice of the sender in a dialogue with the multiple voices that are inevitably inserted into his or her utterances, accepting them, refuting them, taking them up or distancing himself or herself from them. Despite this, and despite the fact that there are interesting studies on the heterogloss character of engagement, for example, in the legal field (Garofalo 2017, 2020), it is the domain of attitude (subdivided into affect –feelings and emotional reactions–, judgement –moral and ethical assessment– and appreciation –evaluation of objects, constructs, texts, products) which is the most connected to the realm of emotion. However, Alba-Juez (2018: 228) argues that the appraisal framework is an incomplete paradigm with which to assess emotion in discourse as there are features such as contextualisation and e-implicatures that are much more subtle than net words and their potential emotional semanticity (Alba-

2 Metadiscourse deals with the writer's/speaker's ability to make statements about the external, experiential world coherent, intelligible and persuasive to a particular audience. The most influential strand is that developed by Hyland (2018), which emphasises the interactive character of communication where writers formulate a connected discourse or express their attitude towards either what or who is addressed.

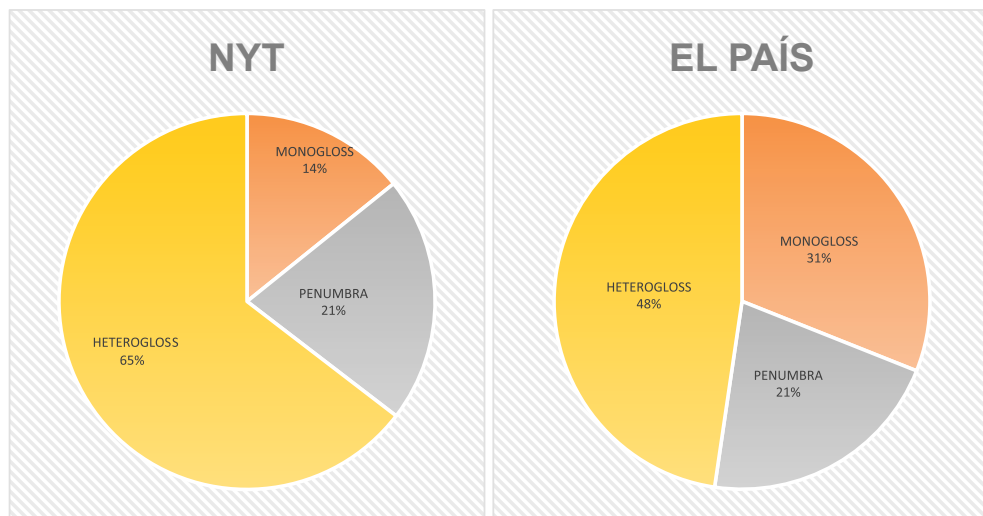
3 Evaluation (Thompson & Alba-Juez 2014) is a strand of study which is related to the expression of a speaker's attitudes and feelings about their own propositions.

Juez 2018: 247). In my view, appraisal theory is not a paradigm that always offers explicit solutions, since discerning implicit patterns is complex when trying to locate the presence of emotion and attitude indicators. However, even with all its provisos, it remains a valid –if arduous– taxonomy for measuring the expression of emotion in texts.

## 5.2. Persuasion and dissemination in legal texts

To test how evaluation is carried out in the legal field, I undertook a series of contrastive works within the framework of appraisal theory between two languages, English and Spanish, (Orts, 2018a and 2018b) on a corpus of 40 legal opinion columns, extracted from two newspapers of national and international quality and prestige, *El País* and *The New York Times*. The legal opinion columns (in this specific case, taken from the *Tribuna* section of *El País* and from *Law and Legislation* in the *NYT*) are written by specialists and combine journalistic opinion on the one hand, and current legal issues, on the other. They constitute a special case of dissemination of expert knowledge, where there is what Engberg (2020: 181) calls a “recontextualization” of specialised content; that is, its adaptation to non-expert contexts– and where two epistemic polarities exist: core knowledge (i.e. domain-internal) and peripheral knowledge. Engberg points out that, divested from its hard technical content, peripheral knowledge is central to arousing the interest of the lay public (2020: 185). Specifically, my studies attempted to trace impersonal (monogloss) and interpersonal (heterogloss) structures in legal opinion discourse, where patterns of emotion might occur irregularly, depending on the legal culture: that of the American common-law system –where the opposing parties in the courtroom act as adversaries who compete to convince the jury and the judge that their version of the facts is more convincing– and the Spanish continental law system –where the criminal procedure (called accusatory-formal) retains vestiges of the non-adversarial regime: a centralised and hierarchical structure where the judge is in charge of managing the entire judicial process. However, in the course of the analysis, a third area emerged, which I called ‘penumbra’, belonging to neither monogloss or heterogloss discourse, which eventually demonstrated that genres of legal opinion can be persuasive only insofar as they are hybrid products, the result of ‘genre colonisation’. Penumbra is expressed linguistically through two different patterns: a) ostensibly impersonal structures, such as nominalisations, which, in fact, convey some ideological message to the reader, and b) structures that appraisal theory considers trigger implicit attitudes, such as deontic and epistemic expressions, emphatic expressions and rhetorical questions.

The findings of the study, shown in Figures 1 and 2, are in line with our previous hypothesis: the degree of persuasiveness in specialised texts is related to the type of genre and its ‘purity’ or hybridity (that is, whether it is oriented inwards towards or outwards away from the discourse community), as well as to the socio-cultural context in which it emerges. The conclusion reached was that Spanish opinion columns, faithful to the system from which they originate, are more alien to the reader, more monolithic and institutional in the expression of attitude. The American ones, on the other hand, show a bias towards debate and argumentation, in accordance with the polyhedral reality that surrounds the common law adversarial system. Both corpora, in general, confirm that there is still a strong element of interaction linking the legal opinion writer to his/her audience, probably more evident in this type of legal genre than in any other where monoglossia is dominant, such as in enacted laws, ordinances or regulations.



Figures 1 and 2. Lexicogrammar in *The New York Times* and *El País* (Orts, 2018b).

Finally, and as we have discussed throughout this paper, dissemination is the second centrifugal force necessary for professional communities to exhibit their value to society. Specifically, Engberg *et al.* (2018) claim that popularisation occurs in three phases: the detection of an asymmetry of knowledge in the lay-expert relationship; the displacement of legal knowledge outside the institutional contexts in which it is normally applied (recontextualisation); the presentation of expert knowledge to the layperson (dissemination); and greater accessibility in presentation than in expert-expert communication, or popularisation (2018: XIII). Popularisation is understood as the transmission of knowledge in contexts where laypeople are communicatively dominant, and do not act as mere participants. This transmission of power is very significant, in my opinion, since it establishes a sphere of parity and communicative exchange between both parties: the expert, who has to contribute his or her specialisation in a relevant and clear way, and the layperson, who receives that communication in a context that he or she knows and dominates. As stated earlier in this paper, the necessary epistemic trust –a prerequisite for a successful dissemination process (Engberg *et al.* 2018: XVI)– can only be obtained if the expert community provides evidence that its knowledge is understandable and apprehensible, thus generating the trust and willingness of non-expert members to cooperate communicatively.

## 6. CONCLUSION

This paper has sought to consider the role professional genre communication from a novel, interdisciplinary perspective which accounts for the multifarious character of the references presented. In the plural paradigm we have constructed to regard the multidimensional character of the phenomenon of genre we have tried to reconcile a constructivist view of specialised discourse –professions and their genres as social phenomena– with a bio-social perspective which considers emotions to be at the root of social manifestations resulting



from the “interface between brain, mind and society” (TenHouten 2021: 16). Our ultimate aim was to consider professional genres as products of the emotional evolution of human beings in the context of the occupational groups, or discursive communities, that constitute part of the social fabric. In the context of professional communities, as in the life of the individual, the agonising tensions that come from our most primitive, reptilian brain, such as the desire for primacy and the drive to compete jostle with the need of our limbic system, of later development, to establish positive and hedonic social relations—to exercise empathy for others, creating bonds of social connection. These tensions coexist, as we have said, simultaneously, and are materialised in the production of genres, as gnoseological resources articulated within the community—with the egotistical drive to maintain hierarchical relationships and to orchestrate knowledge asymmetries within the group— and outside it—with the aim of legitimising the community through the dissemination of those genres, in order to ultimately favour understanding and social value. Power, manipulation and persuasion in the discourse of professions are the manifestations of the contradictory tensions in occupational groups to either impose and deceive, or to genuinely convince and engage with society at large. At a time in history when social networks and the consumption of instant, unverified knowledge blur the need for rigour, and encourage half-truths and manipulation, rigorous research and in-depth study about the communicative behaviour of disciplinary and occupational groups is more necessary than ever, just as dissemination is necessary to foster trust and faith in experts and to combat the ghosts of post-truth. After all, Damasio’s homeostasis as the search for the quintessential positive emotion (2018)—happiness—constitutes the most adaptive response for expert communities to achieve a balance between their own uniqueness and internal organisation, together with their integration into society, which inevitably encompasses the acceptance of these professions by the lay public, and the latter’s worthwhile and transparent access to specialised knowledge.

## REFERENCES

- Alba-Juez, L. (2018). Emotion and appraisal processes in language: How are they related? In M. A. Gómez González and L. J. Mackenzie L. (Eds), *The Construction of Discourse as Verbal Interaction* (pp. 227-250). Amsterdam: John Benjamins.
- Alba-Juez, Laura and L. Mackenzie (Eds.). (2019). *Emotion in Discourse*. Amsterdam: John Benjamins.
- Barbalet, J. M. (1998) *Emotion, social theory, and social structure: macrosociological approach*. Cambridge: Cambridge University Press.
- Bawarshi, A. and M. J. Reiff. (2010). *Genre: An Introduction to History, Theory, Research, and Pedagogy*. West Lafayette: Parlor Press.
- Benítez Castro, M. Á. and E. Hidalgo-Tenorio. (2019). Rethinking Martin and White’s AFFECT taxonomy: A psychologically-inspired approach to the linguistic expression of emotion. In L. Alba-Juez and L. Mackenzie (Eds), *Emotion in Discourse* (pp. 307-338). Amsterdam: John Benjamins.
- Bhatia, V. K. (2004). *Worlds of Written Discourse*. London: Continuum.
- Bhatia, V. K. (2008). Towards Critical Genre Analysis. In V.K. Bhatia, J. Flowerdew and R. Jones (Eds) *Advances in Discourse Studies* (pp. 166–177). London: Routledge.
- Bhatia, V. K. (2012). Critical Reflections on Genre Analysis. *Ibérica*, 24, 17–28.
- Bhatia, V. K (2016). *Critical Genre Analysis: Investigating Interdiscursive Performance in Professional Practice*. London: Routledge.

- Bhatia, V. K. & A. Bhatia. (2017). Interdiscursive Manipulation in Media Reporting: The Case of the Panama Papers in India. In M. A. Orts, M. Gotti and R. Breeze (Eds.), *Power, Persuasion and Manipulation in Specialised Genres* (pp. 29-50). Berna: Peter Lang.
- Belli, S. & F. Broncano. (2017). Trust as a Meta Emotion. *Metaphilosophy* 48 (4), 430-448.
- Bocanegra-Valle, A. (2017). Empowering the Discourse of Globalization in International Organizations: The International Maritime Organization as a Case in Point. In M.A. Orts, M. Gotti, M. and R. Breeze (Eds.), *Power, Persuasion and Manipulation in Specialised Genres* (pp. 51-70). Berna: Peter Lang.
- Breeze, R., Gotti, M. & C. Sancho-Guinda. (Eds.). (2017). *Interpersonality in Legal Genres*. Berna: Peter Lang.
- Carter-Thomas, S. and E. Rowley-Jollivet. (2017). Maintaining a Dominant Voice: A Syntactic Analysis of the Way Power is Welded in Medical Editorials. In M.A. Orts, M. Gotti, M. and R. Breeze (Eds.), *Power, Persuasion and Manipulation in Specialised Genres* (pp. 71-100). Berna: Peter Lang.
- Chance, M. R. A (Ed). (1988). *Introduction*. In *Social Fabrics of the Mind* (pp. 1-35). Hillsdale, N.J.: Erlbaum.
- Cialdini, R. B. (2001). Harnessing the science of persuasion. *Harvard Business Review*, 79(9), 72-81.
- Damasio, A. (1994) *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: G.P. Putnam.
- Damasio, A. (2003). *Looking for Spinoza. Joy, Sorrow and the Feeling Brain*. Orlando, FL: Hartcourt.
- Damasio, A. (2018). *The Strange Order of Things: Life, Feeling, and the Making of Cultures*. New York: Pantheon Books.
- Damasio, A. R., D. Tranel and H. Damasio. (1991). Somatic markers and the guidance of behavior: Theory and preliminary testing. In S. Harvey, H. Levin, H. M. Eisenberg and A. L. Benton (Eds), *Frontal Lobe Function and Dysfunction* (pp. 217-29). New York: Oxford University Press.
- Darwin, C. 1872. *The expression of emotion in man and animals*. London: John Murray.
- Engberg, J. Luttermann, K. Cacchiani, and S. Preite, C. (2018). Studying Popularization in Legal Communication: Introduction and Overview. In J. Engberg, Jan, K. Luttermann, S. Cacchiani, S. and C. Preite (Eds.) *Popularization and Knowledge Mediation in the Law. Popularisierung und Wissensvermittlung im Recht*. Zurich: Lit Verlag.
- Engberg, J. (2020). Institutional Dissemination of Legal Knowledge: An Instance of Knowledge Communication. In M. Gotti, S. Maci, and M. Sala (Eds.), *Scholarly Pathways: Knowledge Transfer and Knowledge Exchange in Academia* (pp. 175-205). Frankfurt: Peter Lang.
- Fairclough, N. 1989/2014. *Language and Power*. London: Longman.
- Fairclough, N. and R. Wodak. (1997). Critical Discourse Analysis. In van Dijk, T. (Ed.), *Discourse as Social Interaction* (pp. 258-284). London: Sage.
- Fiske, A. P. (1991). *Structures of social life: The four elementary forms of human relations: Communal sharing, authority ranking, equality matching, market pricing*. New York: Free Press.
- Foolen, A. (2012). The Relevance of Emotion for Language and Linguistics. In A. Foolen, U. M. Lüdtke, T. P. Racine and J. Zlatev (Eds) *Moving Ourselves, Moving Others. Motion and Emotion in Intersubjectivity, Consciousness and Language* (pp. 349–368). Amsterdam: John Benjamins.
- Foolen, A. (2016). Expressives. In N. Riemer (Ed.) *The Routledge Handbook of Semantics* (pp. 473–490). London and New York: Routledge.
- Foucault, M. (1980). The Confession of the Flesh. In C. Gordon (Ed.), *Power/Knowledge: Selected Interviews and Other Writings*. (pp. 194–228), New York: Pantheon.
- Fowler, R., R. Hodge, G. Kress and T. Trew. (1979). *Language and Control*. London: Routledge.
- Fox, R. & J. Fox. (2004). *Organizational Discourse. A Language-Ideology-Power Perspective*. Westport: Praeger.

- Gallego-Hernández, D. (2017). Persuasion in Promotional Banking Products: A Comparative Corpus-based Study. In M.A. Orts, M. Gotti, M. and R. Breeze (Eds.), *Power, Persuasion and Manipulation in Specialised Genres* (pp. 219-242). Berna: Peter Lang.
- Garofalo, G. (2017). *La insoportable levedad del acusar. El lenguaje de la valoración en las querellas de la Fiscalía Superior de Cataluña contra Carme Forcadell i Lluís*. Milano: FrancoAngeli.
- Garofalo, G. (2020). Escenarios heteroglósicos en las sentencias del Tribunal Supremo de España y del Tribunal de Justicia de la UE. El caso de los conectores condicionales complejos. *Revista de Lingüística y Lenguas Aplicadas*, 15, 57-61.
- Garzone, G. E. (2017). Persuasive Strategies on Surrogacy Websites: A Discourse-Analytical and Rhetorical Study. In M.A. Orts, M. Gotti, M. and R. Breeze (Eds.), *Power, Persuasion and Manipulation in Specialised Genres* (pp.101-130). Berna: Peter Lang.
- Goodrich, P. (1987). *Legal Discourse: Studies in Linguistics, Rhetoric, and Legal Analysis*. London: Macmillan.
- Grijalba-Uche, M. & L. E. Echarte. (2015). Homeostasis y representaciones intelectuales: una aproximación a la conducta moral desde la teoría de la emoción de Antonio Damasio. *Persona y Bioética*, 19 (1), pp.80-98.
- Gunnarsson, B.-L., Linell, P. & B. Nordberg. (Eds) (1997). *The Construction of Professional Discourse*. New York: Longman
- Hidalgo-Tenorio, E. & J. Sánchez-García. (2017). ¿Por qué los debates políticos los ganan la mentira y la frialdad? El caso de España. *RIPS. Revista de Investigaciones Políticas y Sociológicas*, 16 (1), 41-64.
- Hyland, K. (2018). *Metadiscourse: Exploring interaction in writing*. Bloomsbury Publishing.
- Koschut, S. (2020). *The Power of Emotions in World Politics*. Abingdon, Oxon; Nueva York: Routledge.
- Koschut, S. (2018). The power of (emotion) words: on the importance of emotions for social constructivist discourse analysis in IR. *Journal of International Relations and Development*, 21, 495–522.
- Kress, G. and Hodge, R. (1979). *Language as Ideology*. London: Routledge.
- Laclau, E. (2004). Glimpsing the Future. In S. Critchley and M. Marchart (Eds.), *Laclau: A Critical Reader* (pp. 279–329). New York: Routledge.
- Lemmings D. & A. Brooks. (2014). *Emotions and Social Change*. New York: Routledge.
- Llácer, E. V. & F. J. Ballesteros. (2012). El lenguaje científico, la divulgación de la ciencia y el riesgo de las pseudociencias. *Quaderns de filologia. Estudis lingüístics*, 17, 51-67.
- Lütke, U. M. (2015). Introduction: From Logos to Dialogue. In Ulrike M. Lütke (Ed.), *Emotion in Language* (pp. vii–xi). Amsterdam: John Benjamins.
- MacLean, P. D. (1990). *The triune brain in evolution: role in paleocerebral functions*. New York: Plenum Press.
- Macmurray, J. (1962/1992). *Reason and Emotion*. Amherst, NY: Humanity Books.
- Martin, J. R. and Rose, D. (2003). *Working with discourse: Meaning beyond the clause*. London: Continuum.
- Martin, J. R. & P. R. White. (2005). *The language of evaluation: Appraisal in English*. London: Palgrave Macmillan.
- Miller, D. A., Smith, R. Elliott & D. M. Mackie. (2004). Effects of intergroup contact and political predispositions on prejudice. *Role of Intergroup emotions. Group Processes and Intergroup Relations*, 7, 221-237.
- Mumby, D.K. & R. P. Clair. (1997) Organizational Discourse. In T.A. van Dijk (Ed.) *Discourse Studies*, Volume 2: *Discourse as Social Interaction* (pp. 181-205). London: Sage.
- Murolo, L. (2019). La posverdad es mentira. Un aporte conceptual sobre fake news y periodismo. In R. Aparici and M. García Marín (Eds), *La posverdad. Una cartografía de los medios, las redes y la política* (pp. 65-80). Barcelona: Gedisa.
- Orts, M.A. (2015). Power and Complexity in Legal Genres: Unveiling Insurance Policies and Arbitration Rules. *International Journal for the Semiotics of Law*, 2 (28) 485–505

- Orts, M.A. (2016). Power distance and persuasion: The tension between imposition and legitimation in international legal genres. *Journal of Pragmatics*, 92, 1-16.
- Orts, M.A. (2018a). The expression of emotion in institutionalized legal opinion. A contrastive Spanish-English pre-translational study. *RESLA, Revista Española de Lingüística Aplicada*, 30(2), 611 – 635.
- Orts, M.A. (2018b). A Bilingual, Bicultural Approach to Detachment and Appraisal in the Law: Tracing Impersonality and Interaction in English and Spanish Legal Op-Eds. *International Journal for the Semiotics of Law*, 31 (2), 805-828.
- Orts, M.A. & R. Breeze. (2017). *Introduction*. In M.A. Orts, Gotti, M. and Breeze, R. Power (Eds.), *Persuasion and Manipulation in Specialised Genres* (pp. 9-26). Berna: Peter Lang.
- Partington, A. & C. Taylor (2018). *The Language of Persuasion in Politics: An Introduction*. London: Routledge.
- Pinker, S. (2002). *The Blank Slate: The Modern Denial of Human Nature*. New York: Viking.
- Plutchik, R. (1979). Universal problems of adaptation: hierarchy, territoriality, identity and temporality. In J.B. Calhoun (Ed.), *Environment and population: problems of adaptation* (pp. 223–6). New York: Praeger.
- Plutchik, R. (1980). *Emotion, a Psychoevolutionary Synthesis*. New York: Harper and Row.
- Salmi-Tolonen, T. (2011). Introduction. In T. Salmi-Tolonen, I. Tukiainen and R. Foley (Eds) *Law and Language in Partnership and Conflict* (pp. 3-11). A Special Issue of *The Lapland Law Review*.
- Salmi-Tolonen, T. (2014). Interpersonality and Fundamental Rights. In R. Breeze, R., M. Gotti, M. and C. Sancho-Guinda (Eds.), *Interpersonality in Legal Genres* (pp. 63-86). Berna: Peter Lang.
- Schwarz-Friesel, M. (2015). Language and Emotion. The Cognitive Linguistic Perspective. In U. M. Lüdtke (Ed), *Emotion in Language* (pp. vii–xi). Amsterdam: John Benjamins.
- Simpson, P. & A. Mayr. (2010). *Language and Power: A Resource Book for Students*. London: Routledge.
- Swales, J. (2011). *Genre Analysis in Academic and Research Settings*. Cambridge: Cambridge University Press.
- TenHouten, Warren D. (2021) Basic emotion theory, social constructionism, and the universal ethogram. *Social Science Information*, 60 (4), 610-630.
- TenHouten, W. (2014). *Emotion and Reason: Mind, Brain, and the Social Domains of Work and Love*. London: Routledge
- TenHouten, W. (1999). The Four Elementary Forms of Sociality, their Biological Bases, and their Implications for Affect and Cognition. In L. Freese (Ed.), *Advances in Human Ecology*, Volume 8 (pp 253-284). Stamford: JAI Press Inc
- Thompson, G. & L. Alba-Juez. (2014) *Evaluation in Context*. Amsterdam/London: John Benjamins Publishing.
- Tverski, A. & D. Kahneman. (1981). The framing of decisions and the psychology of choice. *Science*, 211, (4481), 453-458.
- van Dijk, T. (1993). Discourse, Power and Access. In C. R. Caldas (Ed.), *Studies in Critical Discourse Analysis*. London: Routledge.
- van Dijk, T. (2008). *Discourse and Power*. New York: Palgrave Macmillan.
- Weber, M. (1921/1978). *Economy and society: an outline of interpretive sociology*. Berkeley: University of California Press
- White, P. R. (2003). Beyond modality and hedging: A dialogic view of the language of intersubjective stance. *Text*, 23(2), 259-284.
- Zlatev, J. (2012). Bodily motion, emotion and mind science. In A. Foolen, U. Lüdtke, T.P. Racine and J. and Zlatev (Eds.), *Moving Ourselves, Moving Others. Motion and Emotion in Intersubjectivity, Consciousness and Language* (pp. 1-25). Amsterdam: John Benjamins.



## Fine-tuning machine translation quality-rating scales for new digital genres: The case of user-generated content

Adaptación de las escalas de calidad de la traducción automática a los nuevos géneros digitales: el caso del contenido generado por el usuario

MIGUEL A. CANDEL-MORA

Universitat Politècnica de València, España

[mcandel@upv.es](mailto:mcandel@upv.es)

<https://orcid.org/0000-0001-8754-6046>

### Abstract

With the active participation of users in product review platforms, online consumer-generated content, and, more specifically, user-generated reviews, have become a clear reference in purchasing decision-making processes, which sometimes exceed the impact of advertising campaigns. A common feature of most tourism review platforms is the use of machine translation (MT) systems to immediately make reviews available to users in various languages. However, the quality of the MT output of these reviews varies greatly, primarily due to the subjective and unstructured nature of this digital genre. Different studies confirm that there are no universal quality rating scales. The assessment of MT output quality usually depends on factors such as the purpose of the text or the value given to the immediacy of the translation. New neural MT systems have been a revolution in the quality increase of the translated output;

### Resumen

Con la participación activa de los usuarios en las plataformas de reseñas de productos, los contenidos online generados por los consumidores, y más concretamente, las opiniones de los usuarios se han convertido en una clara referencia en los procesos de decisión de compra, que en ocasiones superan el impacto de las campañas publicitarias. Una característica común de la mayoría de las plataformas de reseñas turísticas es el uso de sistemas de traducción automática para poner inmediatamente las reseñas a disposición de los usuarios en diferentes idiomas. Sin embargo, la calidad de la traducción automática de estas reseñas varía en gran medida debido a la subjetividad y a la naturaleza no estructurada de este género digital. Diferentes estudios confirman que no existen escalas universales de valoración de la calidad y que la evaluación de la calidad del resultado de la MT suele depender de factores como la finalidad del texto o el

---

**To cite this article:** Candel Mora, M. A. (2022). Fine-tuning machine translation quality-rating scales for new digital genres: The case of user-generated content. *ELUA*, (38), 117-136. <https://doi.org/10.14198/ELUA.21900>

Recibido: 07/02/2022 Aceptado: 03/05/2022

© 2022 Miguel A. Candel-Mora



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

however, new lines of research are opening up to verify whether the quality of this new paradigm of MT can be assessed with the existing scales, mainly from previous rule-based systems and statistical translation, or whether it is necessary to develop new quality metrics specifically for these new intelligent systems. On the other hand, one of the questions that remain to be resolved in this new context of neural MT is whether the use of large amounts of textual data in the training of these systems is as effective as the use of less data but of higher quality and better-adjusted to the specialty and type of text for which it is used. Based on the hypothesis that each genre requires specific quality rating scales, this work identifies the error patterns and textual characteristics of online user reviews from a corpus-based approach analysis that will contribute to adapting quality rating scales to this specific digital genre.

**KEYWORDS:** machine translation; post-editing; quality assessment; user-generated content; online reviews

valor que se da a la inmediatez de la traducción. Los nuevos sistemas de traducción automática neuronal han supuesto una revolución en el incremento de la calidad del texto traducido, sin embargo, se abren nuevas líneas de investigación para verificar si la calidad de este nuevo paradigma de traducción automática se puede valorar con las escalas existentes, procedentes en su mayoría de los anteriores sistemas basados en reglas y traducción estadística, o si es necesario desarrollar nuevas métricas de calidad acordes con estos nuevos sistemas inteligentes. Por otro lado, una de las cuestiones que quedan por resolver en este nuevo contexto de traducción automática neuronal es si la utilización de grandes cantidades de datos textuales en el entrenamiento de estos sistemas es igual de eficaz que un uso de menos cantidad de datos pero de mayor calidad y más ajustados a la especialidad y el tipo de texto en el que se utiliza. Partiendo de la hipótesis de que cada género requiere escalas de valoración de la calidad específicas, este trabajo identifica patrones de error y características textuales de las reseñas de usuarios en línea a partir de un análisis basado en un corpus que contribuirá a adaptar las escalas de valoración de la calidad a este género digital específico.

**PALABRAS CLAVE:** traducción automática; posesición; evaluación de la calidad; contenido generado por el usuario; reseñas en línea

## 1. INTRODUCTION

With Web 2.0 and the active participation of users, online consumer-generated reviews have become a widespread reference in purchasing decision-making processes, which on occasions, exceed the impact, reliability and authenticity of advertising campaigns. According to Schemmann (2011: 1), "...seven in every ten Internet users worldwide trust consumer opinions and peer recommendations posted online".

Following Ricci and Wietsma (2006: 297), "product reviews can be described as a subjective piece of non-structured text describing the user's product knowledge, experiences and opinions, together with a final product rating". However, Vásquez (2012: 107) acknowledges the constraints of carrying out research on user reviews solely based on the language used since other nonlinguistic cues also play an essential role.

User reviews have already been studied to a large extent from the point of view of marketing, business, tourism and information technology (Schemmann 2011; Popović 2021) in areas such as their influence on decision-making (Ricci and Wietsma, 2006) or the

characteristics of the textual genre (Vásquez 2014). User reviews may appear in different formats and structures (Vásquez 2014): as evaluations of a product, in the form of dialogue in a forum, or, as in the case of the reviews selected for this study, as unstructured free text for the evaluation of a tourism product, and more specifically in this work, hotel reviews. Other genre-specific features include intertextuality, reference to previous comments, the personal profile of the reviewer and paralinguistic elements, mainly “orthographic strategies designed to compensate for the impersonality of written discourse” (Pollach 2006: 8), such as capitalization, spelling, and punctuation. Among other aspects that Pollach (2006) notes are emoticons, the use of capital letters, and overuse of exclamation marks and acronyms. However, Pollach also holds that the use of non-verbal cues was not that common in the corpus she analyzed, perhaps because reviewers mostly take their tasks seriously and use neutral, non-emotive language.

A common feature of most hotel review platforms is the use of machine translation systems to immediately make that review available to as many users as possible in different languages, which has also been studied in recent years by different authors (Castilho *et al.* 2018; Gerlach *et al.* 2013; Gorög, 2014; Jiang, Way and Haque 2012; Lommel 2018). These studies have focused mainly on improving the translation engines and language resources used in order to optimize the MT output (Aranberri 2014; Koby *et al.* 2014; Specia, Raj and Turchi, 2010; Temnikova, 2010), but so far it has not been studied in depth from the point of view of identifying error patterns in specific types of texts to fine-tune MT output quality-rating scales.

From the point of view of MT output quality assessment and keeping in mind the purpose of the target text, post-editing (PE) or work to improve MT output consists in repairing and accommodating the MT output, to varying degrees - to the appropriate target language linguistic conventions (O’Brien 2005), or the expectations of users (Allen 2003; Castilho *et al.* 2018: 11). The discussion seems to be focused on the different definitions of translation quality assessment (TQA); as Castilho *et al.* (2018) put it, “the meaning of quality can vary considerably for different individuals, groups, and contexts” (Castilho *et al.* 2018: 11).

Several studies confirm that there are no universal MT quality-assessment scales (Allen 2003; Lommel 2018; Popović 2018; TAUS 2010), and according to Castilho *et al.* (2018: 30), there is “a serious lack of standardization in TQA for both HT and MT”. In sum, each type of text requires specific quality rating scales, and the most important factor in translation is, therefore, purpose; quality is related to the user’s judgement (Castilho *et al.* 2018: 14). Thus, the objective of this work is to identify the standard textual conventions of Consumer-Generated Reviews (CGR) and determine the level of quality of their MT output, and then propose a specific classification of error patterns common to this genre. One potential use of this classification would be educational purposes, where it could serve as guidance and practical training material for future post-editors. According to different authors (Gaspari, Almaghout, and Doherty 2015; Kenny and Doherty 2014; Mellinger 2017), such training is imperative to be current with industry requirements. In particular, Mellinger (2017: 281) highlights that “the incorporation of machine translation (MT) into translation curricula is a growing trend, as demonstrated by several recent scholarly works on MT pedagogy”.

As Mellinger (2017: 285) also points out, “it would be prudent to expose students to machine-translated texts in domain-specific translation courses so that the evaluation stages can be addressed appropriately”. The high availability of CGR on different products, set-



tings and fields of specializations such as consumer electronics, tourism products, or film reviews poses an excellent opportunity to reflect on translation quality assessment and the specificity of error classifications.

## 2. METHODOLOGY

In order to achieve the objective of this study, two corpora were designed: CorpusCGR-EN, which consists of 100 consumer-generated hotel reviews originally written in English, and posted online on TripAdvisor. The second corpus, CorpusCGR-MT-ES, is formed by the Spanish MT output of the reviews in CorpusCGR-EN, that is, the 100 consumer-generated hotel reviews mentioned before exactly as they appeared in the platform after their MT processing by Google Translate, which was the default MT engine provided by the platform at the time, September 2020.

The two corpora, CGR-EN and CGR-MT-ES, with a total size of 14.511 and 14.790 words, respectively (see Table 1), were then segmented into sentences and aligned to facilitate processing and manual assessment tasks, totaling 812 segments.

Finally, the quality assessment of the Spanish MT output was determined following the TAUS Dynamic Quality Framework (DQF) (TAUS, 2016; Van der Meer et al., 2017) and its typology of errors: Accuracy, Fluency, Terminology, Style, Design, Locale conventions and Verity, which concentrate on more specific items such as syntax errors, word agreement, word order, use of articles, or mistranslations due to ambiguity or omission, among others.

### 2.1. Corpus analysis

Both corpora were processed using Wordsmith Tools 7 in order to identify characteristic elements in these types of texts, such as underlying patterns, sentence length, wordlists and keyword lists.

text file	CorpusCGR-EN.txt	CorpusCGR-MT-ES.txt
tokens (running words) in text	14,511	14,790
types (distinct words)	1,949	2,200
type/token ratio (TTR)	13.61	15.06
standardized TTR	41.21	42.91
standardized TTR std.dev.	53.51	52.16

Table 1. Corpus statistics.

TAUS DQF (2016) distinguishes two levels of quality determined by two main criteria: the quality of the MT raw output and the expected end quality of the content. These levels are “good enough” quality, and quality “similar or equal to human translation”. TAUS’s “good enough” level is defined as comprehensible and accurate but not very convincing concerning style. Based on these levels proposed by TAUS, for this work, we propose a third level labelled “Unacceptable”, in which the message is not accurate due to incorrect grammar or lexical usage, or unusual syntax or mistranslation, among other issues. In order

to differentiate our quality classification from TAUS's proposal, for this work TAUS's "good enough" level was labelled "Acceptable" (accurate but not fully convincing or with minor errors), and TAUS's "similar or equal to human translation" was labelled "Correct" (without any errors). Table 2 shows the distribution of the initial MT output quality estimation of the segments in the corpus with the proposal of labels to classify them throughout this work.

Correct	Acceptable	Unacceptable	Total segments
305	324	183	812
38%	40%	22%	

Table 2. Distribution of preliminary MT output quality estimation.

Secondly, the categories labelled as Unacceptable and Acceptable underwent a thorough manual revision and analysis by expert linguists to identify specific recurrent error patterns, which will be described in detail in the analysis and discussion section below and that will serve as the basis for the proposed items to adapt quality assessment scales in line with the objective of this study.

### 3. MT QUALITY ASSESSMENT SCALES

The concept of quality in translation has always been a topic of interest for research in translation, both human and machine translation, for the industry and the academia alike (Castilho et al., 2018). In the case of translation quality assessment of MT, several variables determine the approach to its assessment. It seems almost impracticable to find common ground that serves as a starting point for proposing universal quality evaluation criteria. According to Valli (2015: 128), among the reasons for this situation are the "lack of transparent evaluation criteria, the difficulty of finding the right metrics, the lack of standardization and the need for different quality levels".

Other variables include the increase in the quality of MT output from recent MT engines, especially since the emergence of neural machine translation systems, their widespread use in professional settings (Torres-Hostench, Presas, and Cid-Leal 2016), the restricted number of methodologies and criteria on how to train post-editors or perform post-editing tasks (Kenny and Doherty, 2014). Finally, some other constraints encountered during the course of this research include the lack of consensus on and diverse definitions of translation quality assessment (Castilho *et al.* 2018) or different definitions of error categories (Popović 2018), which are not always accessible for research purposes for confidentiality reasons, which hinders the possibility of conducting a more general overview of existing post-editing guidelines.

Some authors note that quality is conditioned by the purpose of the MT product (Allen 2003; O'Brien, 2011; TAUS, 2010), i.e., whether the translation is intended to be published and disseminated or if the translation is only aimed at guiding the reader as to the text's overall meaning. All this leads to reflecting on the changing nature of MT output quality assessment and post-editing, and the obstacles to proposing a universal tool that is applicable in any context. As Allen (2003) points out, aspects such as the specifications of the client, the volume of documentation expected to be processed, or expectations with regard to the level of quality,

among others, might influence the exhaustiveness of a post-editing project since “differing percentages of MT accuracy have even been found when applied to different subdomains and different document types within the same technical domain” (Allen 2003: 303).

Among the most common categories of errors are terminology errors, lexical ambiguity, syntax, omission, word agreement and punctuation errors. What is more, different types of metrics also assign a different weight to each error type (Guzmán 2007; Mitchell *et al.* 2014; SAE International 2001; TAUS 2010), in addition to more general criteria such as readability and acceptability of MT output. However, metrics especially consider whether the objectives of the text type are met (Stymne and Ahrenberg 2012; TAUS 2016; Van der Meer *et al.* 2017).

For this work, the TAUS Dynamic Quality Framework (TAUS 2016; Van der Meer *et al.* 2017) was chosen because it is a widespread and consolidated quality metrics and its thoroughness and flexibility facilitate achieving the aim of this study in that it offers a global vision of the translation process.

In this context of CGR, it is especially significant Allen’s (2003: 300) consideration of the use of MT in the Web 2.0 and user participation context, which has led to a “change in expectations with regard to the type and quality of translated material”. In addition to translation as a high-quality text product for important documents on user safety or commercial information, for example, there is an increased demand for gisting translation where users simply need to understand the main idea of the text in their own language.

Thus, having concluded that there is no universal quality assessment scale and since MT quality-assessment scales cannot be used directly on all types of texts, TAUS DQF seemed to be the most flexible and appropriate instrument to carry out manual translation quality assessment and the identification of the most common error patterns in the corpus of consumer-generated reviews because it takes into account the changing landscape taking into account different content types. However, limitations of this manual methodology, such as time-consuming processes or the subjectivity of assessors, should be pointed out here. In addition, based on Valli’s findings (2015: 132) and due to the distinctive features of the genre being examined here and the user’s specific needs, the error typology can be more or less granular. Finally, the pass/fail threshold is flexible and depends on content type.

### 3.1. TAUS Dynamic Quality Framework (DQF)

TAUS DQF proposes seven types of errors: Accuracy, Fluency, Terminology, Style, Design, Locale conventions, and Verity. Since TAUS DQF allows a more or less complete post-editing task depending on the context for which the content is translated, this study has focused on the first four error types: accuracy, fluency, terminology and style, given the characteristics of the reviews and the fact that the objective is to propose specific error categories for this digital genre and help translators in identifying them and becoming familiar with them in post-editing. To this end, TAUS DQF is also enriching since it offers a global vision of the most common types of errors and is a reference in the language industry.

In brief, the description of the categories (TAUS 2016:11) and the specific errors they incorporate are as follows: Accuracy is defined as the situations when “the target text does not accurately reflect the source text, allowing for any differences authorized by specifications”. We find specific errors within this type of error: addition, omission, mistranslation, over-

translation, under-translation, and untranslated text. Fluency considers formal or content aspects, not necessarily related to translation, but rather to the use of punctuation, spelling and syntax, as well as the register and coherence of the text. After these two factors are assessed, the MT output of Terminology is evaluated, the main error identified here being the inconsistent use of terminology. Design refers to problems relating to design aspects (vs linguistic aspects) of the content, while Style rates the error as awkward style, in the case of not fulfilling the genre-specific characteristics.

As noted above, the Locale conventions category seems to be more oriented to the translation of more specialized texts and considers error types such as date format, currency format, or measures. Finally, Verity and the Culture-specific reference category assess whether the reference will be understandable to the intended audience.

For the assessment of consumer reviews processed with MT, it is necessary to keep in mind from the beginning that the users of this translation might be less demanding in terms of quality, therefore, the type of error Locale conventions has been disregarded as it targets much more specialized texts. From the category Verity, we have selected some illustrative examples from the *Culture-specific references* section to highlight the most severe errors since, in the translation of tourist texts, the cultural load associated would significantly exceed the scope of this study.

#### 4. RESULTS AND DISCUSSION

Conventional translation quality scales include error annotation and calculation of the proportion of errors with the total amount of words in the translated text. However, in the case of consumer reviews, with an average of 128,05 words per review (see Table 3), the error proportion would be higher, and low-quality translation would be more noticeable.

	CorpusCGR EN	CorpusCGR MT-ES
Average number of words	128.05 words	137.02 words

Table 3. Average number of words per review.

With the design of specific guidelines to assess the MT quality of CGR in mind, the most relevant findings are presented below. Firstly, Figure 1 shows an overview of the most frequent types of error according to the TAUS DQF scale, which will help post-editors anticipate and plan their PE strategy. Next, a detailed analysis is presented for each of the classifications selected: Accuracy, Fluency, Terminology and Style, concluding with a singular contribution which consists in the manual identification of errors in the original text not included in TAUS DQF, which, in turn, produce errors in the MT output which are not common in specialized MT environments.

##### 4.1. Most common error categories (based on TAUS DQF)

Figure 1 presents the overall percentages of the most common error categories, which will provide the basis for an approach to the planning of post-editing tasks in consumer reviews. After that, the main recurrent error patterns of this genre will be detailed and illustrated with examples.

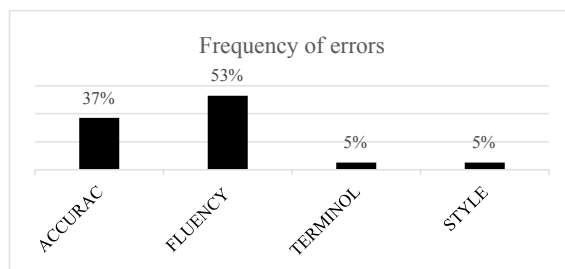


Figure 1. Frequency of error types.

As can be seen, the most recurrent type of error in the Spanish MT output corresponds to the Fluency category, with 509 errors, 53% of the total, followed by errors in Accuracy (354 errors), Terminology (50) and Style (49). Given the amount of text of the sample analyzed, 100 reviews comprising 14,511 words, this finding reveals which error typology would require more effort to be input in a future post-editing project and the skills required from the translator willing to specialize in this type of texts.

#### 4.2. Detailed analysis of recurrent error patterns in CGR

Each of the sections of the following detailed analysis starts with a global view of the distribution of errors from the TAUS DQE, and then, each type of specific error is analyzed and illustrated with examples taken from the corpus. Although errors have been classified following the TAUS typology, on many occasions, the same segment contains different types of errors and has been computed in other sections.

##### 4.2.1. Accuracy

As can be seen below (Figure 2), within this type of error, the first place is occupied by mistranslations - when the translated content does not correspond exactly to the original - with a total of 280, which represents 79% of the total, followed by 47 omission errors (13%), and 13 under-translation errors (4%). Untranslated, text addition and over-translation errors are practically negligible, with 6, 6, and 2 instances, respectively.

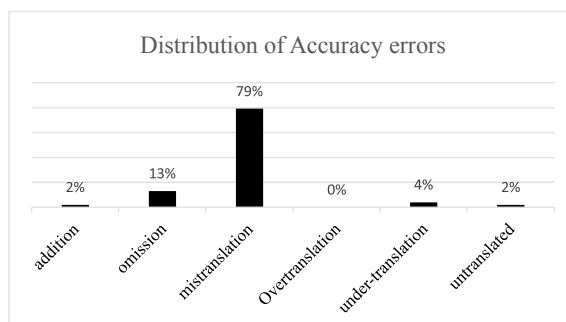


Figure 2. Distribution of Accuracy errors.

#### 4.2.1.1. Addition

Regarding addition errors, only six instances were identified, and as shown below (Examples 1 to 3), they do not affect the content but rather contribute to the naturalness of the reviews.

	Original English	Spanish MT output
Ex. 1.	No complaints all happy and fine.	No tengo ninguna queja todos contentos y <b>muy</b> bien.
Ex. 2.	The hotel is conveniently located within walking distance of Buckingham Palace, Big Ben, Parliament and many other <b>sites</b> .	El hotel está convenientemente ubicado a poca distancia a pie del Palacio de Buckingham, el Big Ben, el Parlamento y muchos otros <b>lugares de interés</b> .
Ex. 3.	It is located in Belgravia just steps from Buckingham Palace.	Está situado en la <b>elegante zona de</b> Belgravia a sólo unos pasos del Palacio de Buckingham.

The addition of content in the target text, not present in the source, is not, therefore, common, and in this digital genre does not seem to have the severity it would in other genres such as legal or technical texts.

#### 4.2.1.2. Omission

Overall, 47 errors were identified in which content is missing from the translation but is present in the source text. However, as can be seen in the examples below, the textual elements omitted are, in most cases, modifiers which do not affect the overall content of the translation. However, in the case of tourist product reviews, these modifiers have a substantial role in the user's assessment. Thus, for instance, in example 4, the omission of "as well" in a very favorable opinion of one of the hotel services ("Breakfast was superb") removes the referent and the consequent favorable opinion of the other services valued within the same review.

	Original English	Spanish MT output
Ex. 4.	Breakfast was superb <b>as well</b> .	El desayuno era excelente.
Ex. 5.	could not see any <b>issues</b> or problems.	no pudimos ver ningún problema.
Ex. 6.	We recently spent 3 nights at Hotel 41, before a long haul flight <b>home</b> .	Recientemente pasamos 3 noches en el hotel 41, antes de un largo vuelo.
Ex. 7.	I cannot praise this hotel and the brilliant staff <b>enough</b> .	No puedo alabar este hotel y el personal estupendo.

Some authors trace the origin of "online consumer reviews" or "electronic word of mouth" precisely to the traditional word of mouth, previous to computer-mediated communication, when users communicated orally their experience of a tourism product (Ricci and Wietsma 2006: 297). Holgado and Redio (2013: 94) and Yus (2011: 19), when referring specifically to written language in an electronic medium, refer to "oralized written text" due to its hybrid nature and the use of oral and written features in the same medium, and highlight the use of capitals and repetitions as strategies for the compensation of the loss of nonverbal

features. In this study, the approach to capitalization and repetitions within the *omission* category was also identified as a fundamental feature of this digital genre. As shown below, for a representative number of cases from the corpus, both capitalization and repetition have been neutralized in the Spanish MT output, which in fact seems to have an effect on acceptability, which according to Castilho *et al.* (2018: 20) “In the context of TQA, it refers to the degree to which the target or output text meets the needs and expectations of its reader(s) or user(s)”.

	Original English	Spanish MT output
Ex. 8.	Hotel 41 played a <b>HUGE</b> part.	Hotel 41 jugó un papel muy importante.
Ex. 9.	I really can't fault this Hotel, it's perfect in every way and delivers on <b>ALL</b> levels and exceeds expectations regularly.	Realmente no puedo quejarme de este hotel, es perfecto en todos los sentidos y entrega en todos los niveles y sobrepasa las expectativas con frecuencia.
Ex. 10.	Everything is <b>so So SO</b> amazing.	Todo es tan increíble.
Ex. 11.	we did the 'historic London' bus tour which was <b>HORRIBLE</b> .	hicimos la excursión en autobús de 'Londres histórico que era horrible.

#### 4.2.1.3. Mistranslation

As mentioned above, this category concentrates the highest number of errors (280). One of the main difficulties during this research work was their inclusion in the error classification, since sometimes when the target content does not accurately represent the source content, it could be due to ambiguity or polysemy (as in examples 12, 13, 14, 15, 16 and 17), word order, or a combination of several errors (see examples 18 and 19).

	Original English	Spanish MT output
Ex. 12.	Weather was perfect and London was absolutely a <b>gas</b> .	El tiempo era perfecto y Londres era absolutamente un <b>gas</b>
Ex. 13.	<b>Home</b> baked biscuits.	<b>Inicio</b> galletas horneados
Ex. 14.	We <b>WILL</b> be back for Christmas and have started <b>saving</b> already.	Volveremos para Navidad y ya hemos comenzado a <b>salvar</b> .
Ex. 15.	We dined <b>too</b> .	Cenamos <b>demasiado</b> .
Ex. 16.	The rooms are cosy but space in London is a <b>premium</b> .	Las habitaciones son acogedoras pero el espacio en Londres es una <b>prima</b> .
Ex. 17.	I had selected a few <b>plays</b> (from internet) and they booked them for me.	Había seleccionado algunos <b>juega</b> (desde Internet) y que había reservado para mí.
Ex. 18.	Warm, welcoming personnel.	Dando la bienvenida al personal cálidos.
Ex. 19.	They provided excellent restaurant recommendations and tips on sightseeing.	Nos dieron excelentes recomendaciones y consejos sobre restaurantes en turismo.

#### 4.2.1.4. Over-translation

The errors in this category are practically negligible compared to the rest of the error types. As shown in the example below (Example 20), the target text is more specific than the source text, but it does not have a significant effect on the quality of the text in Spanish.

	Original English	Spanish MT output
Ex. 20.	The breakfast <b>included</b> in the room was HUGE - we actually couldn't eat it all.	El desayuno incluido en el <b>precio de la habitación</b> era enorme, de hecho no podía comer todo.

#### 4.2.1.5. Under-translation

In this category, a total of 13 instances were identified, some of which were also labelled as omission, although they also correspond to TAUS's definition of under-translation (2016): "The target text is less specific than the source text". The cases in which the exclamations and capital letters are omitted, thus removing the emphasis added by the author of the original, have also been included within this category because, in fact they have an effect on the acceptability of the target text and entail a loss of specificity.

	Original English	Spanish MT output
Ex. 21.	We recently spent 3 nights at Hotel 41, before a long haul flight <b>home</b> .	Recientemente pasamos 3 noches en el hotel 41, antes de un largo vuelo.
Ex. 22.	but <b>all</b> the other thing were great!	pero la otra cosa era genial!
Ex. 23.	<b>Very</b> poor.	Pobre
Ex. 24.	Enjoy!!!!!!!!	Disfrute
Ex. 25.	Cheap and <b>NOT WORTH IT</b> .	Barato y no vale la pena.

#### 4.2.1.6. Untranslated

In this category, three scenarios were identified: The first one, the most common, consists in the non-translation of certain elements that contain proper names or proprietary services of a tourist establishment which, when capitalized, could be interpreted as a proper name. In the second category, parts of the text remain untranslated, as in examples 26 and 27. The last scenario refers to the vocabulary of tourism since certain terms in English sometimes have a greater degree of acceptance when used as a loan word or Anglicism than they do in Spanish (Examples 28, 29).

	Original English	Spanish MT output
Ex. 26.	The Executive lounge is a brilliant concept and works so well as the ' <b>hub</b> ' of this hotel.	El Executive Lounge es un brillante concepto y funciona tan bien como la ' <b>hub</b> ' de este hotel.
Ex. 27.	The <b>Executive Lounge</b> is a lovely place to sit and it's very hard not to be tempted by the many complimentary treats that await the weal willed (such as myself).	El <b>Executive Lounge</b> es un lugar encantador para sentarse y resulta muy difícil no ser tentado por las muchas delicias gratuitas que le esperan el bienestar querida (como yo).
Ex. 28.	Service, champagne at <b>check-in</b> , the pantry to plunder - all perfect.	El servicio, champán en el <b>check-in</b> , la despensa para que saquearan - todo perfecto.



	Original English	Spanish MT output
Ex. 29.	On check in you are led to a beautiful Hogwarts like library and ask which on of four <b>champaignes</b> you would like with canapes.	En el registro llevado a una bonita Hogwarts como biblioteca y pide que en cuatro <b>champaignes</b> desea con canapés.

#### 4.2.2. Fluency

This category takes into consideration punctuation, spelling and grammatical errors. It also includes a section called internal inconsistency, but this has not been included in the analysis given the shortness of the reviews studied. However, to study the behavior of the machine translation engine, a minor revision is presented as an illustration of examples of internal inconsistency within the whole corpus to assess the different choices for the same translation issue.

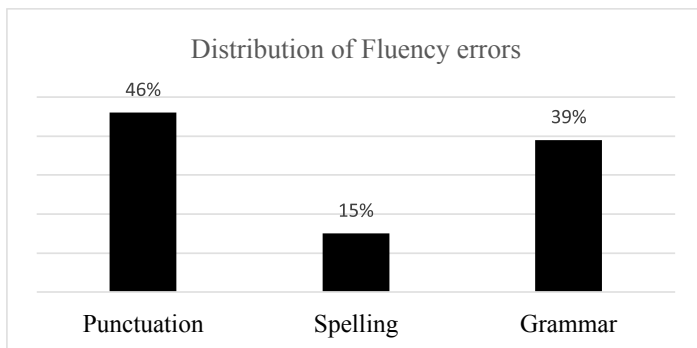


Figure 3. Distribution of Fluency errors.

It should be noted that there are no instances of spelling errors in the Spanish MT output in this category. However, it has also been considered the spelling errors in the original, which in many cases has had an effect on the translated text and which, therefore, should be part of the post-editing strategy for this type of text.

##### 4.2.2.1. Punctuation

Exclamation marks are a widespread feature of this genre, used to compensate for the impersonality of reviews and express emotions. There are several scenarios: exclamation marks in the Spanish MT output are left as in English (with a single exclamation at the end); exclamation marks are adapted to the norms of Spanish and, therefore, correct (with exclamation at the beginning and at the end); Several exclamation marks are used - which is ultimately incorrect in both languages, but characteristic of this digital genre.

Only in 10 examples of the 363 identified has MT correctly processed the use of exclamation marks in Spanish (see examples 33 and 34). In most cases, the use of exclamations reproduces the original text in English.

	Original English	Spanish MT output
Ex. 30.	but all the other thing were great!	pero la otra cosa era genial!
Ex. 31.	I'm going back!	Voy a volver!
Ex. 32.	We loved it!	¡Nos encantó!
Ex. 33.	I will definitely be back!	¡Sin duda volveré!
Ex. 34.	WOW!	¡GUAU!
Ex. 35.	Again, thanks to all of you!!!!	Una vez más, gracias a todos!!!!
Ex. 36.	Thanks to the management and the staff for an unforgettable 2nd stay. !!!!	Gracias a la dirección y el personal para una segunda estancia inolvidable .!!!!

#### 4.2.2.2. Grammar

Among the grammar and syntax errors of the text, the most common errors identified are word agreement (99 instances), word order (57), use of articles (54), collocations (49), pronouns (46), and verb tenses (37).

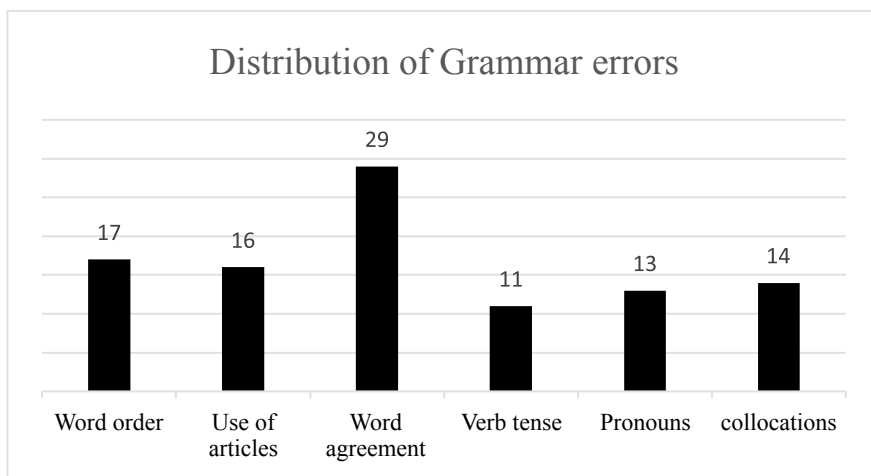


Figure 4. Distribution of grammar errors.

Editing translations made by a human translator is a common practice in professional translation processes. Although this type of error is much less frequent in human translation than in machine translation, it is one of the most laborious tasks and requires more thoroughness. In sum, grammar errors are usually less frequent in human translation but more easily identifiable.

As in the previous cases, the accumulation of errors in the same segment has hindered the inclusion of errors in a single specific category. However, for illustration purposes, some examples of the most frequent types of grammatical errors found in this work are presented below:

	Original English	Spanish MT output	Type of error
Ex. 37.	a very personal, small hotel.	una muy personal, un pequeño hotel.	Word order
Ex. 38.	My husband travels to England every so often for business.	Mi marido viaja a Inglaterra cada tan a menudo por negocios.	Word order
Ex. 39.	Hotel 41 has very good service.	Hotel 41 tiene un muy buen servicio.	Use of articles
Ex. 40.	It's London centre after all.	Es centro de Londres después de todo.	Use of articles
Ex. 41.	There's no getting round it, a stay here is not cheap.	No hay que ir, una estancia aquí no es barato.	Word agreement
Ex. 42.	We felt comfortable, relaxed but with luxury too.	Nos sentimos cómodos, relajado pero con el lujo.	word agreement
Ex. 43.	We were never left waiting and the staff really couldn't do enough for us.	Nunca nos estará esperando y el personal no podía hacer lo suficiente para nosotros.	Verb tenses
Ex. 44.	Will definitely book to stay here again	Sin duda reservar aquí de nuevo	Verb tenses
Ex. 45.	They were genuinely <b>pleased to be</b> part of our special occasion.	Estaban verdaderamente <b>satisfechos a ser</b> parte de nuestra ocasión especial.	Collocations
Ex. 46.	Simply put, the best hotel I've ever had the pleasure of visiting.	Sencillamente, el mejor hotel en el que he tenido el placer de visitar.	Collocations

#### 4.2.2.3. Grammatical register

The only instances of incorrect grammatical register consist of using informal pronouns and verb forms when their formal counterparts are required (see examples 47 to 49) or the combination of a formal and informal register within the same segment as in example 50.

	Original English	Spanish MT output
Ex. 47.	The rooms although small have all the luxury that you need, with luxury items better than a 5* hotel.	Las habitaciones, aunque pequeñas tienen todo el lujo que usted necesita, con artículos de lujo mejor que un hotel de 5 *.
Ex. 48.	You are always greeted with an appetizer.	Usted siempre son recibidos con un aperitivo.
Ex. 49.	Treats are also served in your room during the late afternoon.	También se sirven Treats en su habitación durante la tarde.
Ex. 50.	Don't bother going elsewhere , it's almost like having your own private stylish London residence in town.	No te molestes ir a otro lugar, es casi como tener su propio Londres elegante residencia privada en la ciudad.

## 4.2.2.4. Inconsistency

This item has not been considered since the length of the reviews is excessively short. However, for illustrative purposes, several inconsistencies have been identified in the corpus, such as the different translations of snacks as “tentempiés, bocadillos or aperitivos”; check-in as “registro, check-in”, Champagne as “champaña, champagne”, Buckingham Palace as “Palacio de Buckingham, Buckingham Palace”.

	Original English	Spanish MT output
Ex. 51.	some delicious <b>snacks</b> are available for all hotel guests.	algunos deliciosos <b>tentempiés</b> están disponibles para todos los huéspedes del hotel.
Ex. 52.	The small lounge area is so lovely and the complimentary <b>snacks</b> are gorgeous.	La pequeña zona del salón es tan encantadora y los <b>bocadillos</b> de cortesía son preciosas.
Ex. 53.	The atmosphere was relaxed and we had another glass and some <b>snacks</b> before being taken to our room.	El ambiente era relajado y tuvimos otra copa y algunos <b>aperitivos</b> antes de ser llevado a nuestra habitación.
Ex. 54.	Every aspect of our stay was managed with care and attention, from check-in to <b>check-out</b> .	Cada aspecto de nuestra estancia fue manejado con cuidado y atención, desde la llegada hasta la <b>salida</b> .
Ex. 55.	Service, champagne at <b>check-in</b> , the pantry to plunder - all perfect.	El servicio, champán en el <b>check-in</b> , la despensa para que saquearan - todo perfecto.

## 4.2.3. Terminology

This is not a common type of error, as only 50 were found (5% of the total). In this section, the analysis paid special attention to the vocabulary related to the hotel industry and the usual lexicon of this sector. The main types of error found were due to polysemy of some words, such as “bar” (establishment / counter / candy), “play” (sport / theatre), “ticket” (train / theatre), “glass” (receptacle / material), or common English verbs that have two forms in Spanish: “to be” (*ser* or *estar*), “to have” (*tener* o *tomar*), “to miss” (*perder* o  *echar de menos*) (see Examples 56 to 59).

On occasions, the term in English remains untranslated, however, the language of tourism seems to be quite receptive to importing Anglicisms and loan words. Thus, the resulting Spanish MT output does not seem completely unacceptable, for example, in the translation of “amenities”, “check-in”, “check-out”, “lobby”, “lounge”, and “staff” (see example 60).

	Original English	Spanish MT output
Ex. 56.	The hotel also booked theatre <b>tickets</b> for me.	El hotel también reservamos <b>billetes</b> de teatro para mí.
Ex. 57.	After a long day of travel to London it was delightful to be offered a <b>glass</b> of chilled champagne while our bags were taken to our room.	Después de un largo día de viaje a Londres era una delicia ofrecerá un <b>vaso</b> de champaña helada mientras nos llevaron las maletas a la habitación.

	Original English	Spanish MT output
Ex. 58.	There were no <b>wardrobe</b> or <b>cupboard</b> doors nor any drawers as there was simply not enough room to open them.	No había <b>armario o armario</b> puertas ni ningún cajones, ya que simplemente no había suficiente espacio para abrirlos.
Ex. 59.	<b>Check in, lobby</b> and dining are all combined in one area centrally located among the rooms.	El <b>registro</b> , el <b>vestíbulo</b> y los restaurantes están combinados en una zona con una ubicación céntrica entre las habitaciones.
Ex. 60.	You <b>check in</b> on the fifth floor as these are the only two floors in the hotel and you are greeted with a glass of champagne.	El <b>check-in</b> en el quinto piso, ya que son las únicas dos plantas del hotel y te dan la bienvenida con una copa de champán.

#### 4.2.4. Style

The majority of errors in this section affect the naturalness of the texts and, therefore, the purpose of the genre: reliability and credibility. In sum, the style errors identified are mainly due to awkward choice of words or the addition of articles and pronouns which show a lack of naturalness, although the segments are apparently grammatically acceptable.

	Original source text	Spanish MT output
Ex. 61.	Quite unique and refreshing.	Muy <i>único</i> y refrescante.
Ex. 62.	We love it.	¡A nosotros nos apasiona!
Ex. 63.	I felt at home, and welcome, even though I was alone.	Me sentí como en casa, y bienvenido, aunque yo estaba solo.
Ex. 64.	This is the secret to this hotel: detail.	Este es el secreto para este hotel: los detalles.

#### 4.2.5. Verity: culture specific references

The only culture-specific references consist in the translation of currency, measures and on occasions, food, which can be considered as acceptable within the context of the hotel industry because they add an exotic flavor to the travel experience. However, in professional translation and localization processes, the lack of adaptation to the locale would be labelled as incorrect. In addition, the target reader of the review platform is not clearly identified with the Latin America or Spain versions of Spanish, which would be the other source of errors. Overall, verity errors do not need a detailed revision as they contribute to the characteristic style of the language of tourism.

	Original source text	Spanish MT output
Ex. 65.	Worth every cent - and the executive lounge is a great place to sit and have a drink after a day out in the city.	Vale la pena cada centavo - y el salón <b>Executive</b> es un gran lugar para sentarse y disfrutar de una bebida después de un día en la ciudad.
Ex. 66.	Thoroughly recommend this place and worth every penny	Recomendaría este lugar y vale la pena cada <b>centavo</b>

#### 4.2.6. Errors in original text

This category is not included in the TAUS DQF scale since its use is mainly intended for specialized professional translation processes where it is not common to find errors in the original text. However, one of the characteristics of consumer-generated reviews lies in the spontaneity and the variety of different user profiles.

In the corpus analyzed, a total of 315 segments contained errors in the original which consequently affected the quality of the Spanish MT output. Considering the pedagogical implications of the post-editing tasks suggested in this work, it is necessary that the future post-editor takes into consideration the potential effect of an incorrect text in the source language and its subsequent improvement in the post-editing process. Here we also find different scenarios: most of the time the MT output directly reproduces the misspelt word (Examples 67 to 73). In contrast, in others, it corrects the error in the original (Examples 74 and 75).

	Original English	Spanish MT output
Ex. 67.	All the staff <b>wen</b> up and above to get you what ever you needed.	Todo el personal <b>Wen</b> y por encima de todo lo que necesitamos.
Ex. 68.	the staff <b>where</b> exceptionally professional.	el personal <b>donde</b> excepcionalmente profesional.
Ex. 69.	Is one of those hotels that you don't mind what you'll pay at the end <b>if</b> the day.	Es uno de esos hoteles que no te importa lo que tendrás que pagar al final <b>si</b> el día.
Ex. 70.	<b>Breaks aft</b> did not try and other facilities most of the time in city and entertaing guests.	<b>Escapadas aft</b> no probamos y otras instalaciones más del tiempo en la ciudad y tumbonas.
Ex. 71.	Would definitely stay here <b>ahain!</b>	Sin duda me alojaría aquí <b>ahain!</b>
Ex. 72.	All the <b>staff</b> I cam across were very helpful and friendly.	Todo el personal me <b>leva</b> a través era muy servicial y amable.
Ex. 73.	<b>Prob</b> nicest and friendliest staff I've come across.	<b>Prob</b> nicest and friendliest staff I've come across.
Ex. 74.	We had a very <b>quite</b> room facing the courtyar	Teníamos una habitación muy <b>tranquila</b> con vistas al patio.
Ex. 75.	We give this <b>ony</b> 4 stars because of the size of the room, and the lack of a lobby at street level,	Debemos dar este <b>solamente</b> 4 estrellas debido al tamaño de la habitación, y la falta de un vestíbulo en el nivel de la calle,

## 5. CONCLUSIONS

The extensive production of content generated by user reviews posted on platforms of all kinds (tourism, restaurants, consumer electronics) and the use of machine translation to make this content available in different languages have paved the way for new research opportunities. This work has explored the exploitation of a comparable corpus of consumer-generated reviews written originally in English and processed into Spanish in order to identify the most important textual conventions and error patterns of this new digital genre. In the first place, post-editors should be aware that there is no universal translation quality

scale or set of guidelines that apply to all scenarios. The decision as to whether a more or less detailed post-editing effort is appropriate depends on the use and purpose of the translated document. Thus, the characteristics of the textual genre should be considered and the quality-rating scale should be adapted accordingly. Therefore, a comprehensive analysis of the textual genre must be taken into consideration when training future post-editors, along with training in common post-editing techniques and guidelines. For example, in the case of reviews of a tourist product, apart from language, other genre-specific features include naturalness, reliability and credibility.

The nature of user reviews conditions, to a large extent, the need to propose specific quality-rating scales. Most translation quality scales include error annotation and the calculation of the proportion of errors as a function of the total amount of words in the translated text; however, in the case of consumer reviews, which consist of free text of reduced dimensions, the error proportion would be higher, and low-quality translation would be more visible. The results of this work will help to fine-tune MT quality assessment scales according to other parameters, such as the length of the text, for example.

The training of future post-editors would benefit from the identification and extraction of MT-related data obtained from a comparable corpus of the same text type and domain. Among the MT-related data extracted from comparable corpora are collections of translation equivalent fragments of text, such as terminological expressions, frequent multi-word expressions, or usual content words and collocations.

One of the most valuable resources for the development and improvement of machine translation systems is parallel texts; however, in the case of consumer-generated content, it is rather complicated to find a parallel corpus since most of the consumer-generated content is translated by MT systems without any further post-editing. Thus, further exploitation might include using that translation knowledge and adding bilingual dictionaries of previously extracted translation pairs to the parallel corpus of the MT system to improve its overall performance, a well-known practice in domain adaptation and training MT engines.

Among the future research lines identified during this work is the application of this methodology of genre-specific features in user-generated content and contrast it with other types of texts, such as social media content, reviews and testimonials, blog posts, video content, or Q&A forums, which are also susceptible to being processed by machine translation systems, or with other combinations of languages, especially low resource languages.

## REFERENCES

- Allen, J. (2003). Post-editing. In H. Somers (Ed.), *Computers and Translation. A translator's guide* (pp. 297-317). John Benjamins.
- Aranberri, N. (2014). Posedición, productividad y calidad. *Tradumàtica: Tecnologies de la Traducció*, 12, 471-477. <https://doi.org/10.5565/rev/tradumatica.62>
- Castilho, S., S. Doherty, F. Gaspari & J. Moorkens. (2018). Approaches to Human and Machine Translation Quality Assessment. In J. Moorkens, S. Castilho, F. Gaspari and S. Doherty (Eds.). *Translation Quality Assessment: From Principles to Practice* (pp. 9-38). Springer.
- Gaspari, F., H. Almaghout & S. Doherty. (2015). A survey of machine translation competences: Insights for translation technology educators and practitioners. *Perspectives*, 23(3), 333-358. <https://doi.org/10.1080/0907676X.2014.979842>

- Gerlach, J., V. Porro Rodriguez, P. Bouillon & S. Lehmann. (2013). Combining pre-editing and post-editing to improve SMT of user-generated content. In S. O'Brien, M. Simard and L. Specia (Eds.). *Proceedings of MT Summit XIV Workshop on Post-editing Technology and Practice*, (pp. 45-53). Nice (France) 2 Sept., 2013. <https://archive-ouverte.unige.ch/unige:30952>
- Guzmán, R. (2007). Manual MT Post-Editing: If it's not Broken, don't Fix it. *Translation Journal*, 11(4). <http://translationjournal.net/journal/42mt.htm>
- Holgado Lage, A. & A. Recio Diego. (2013). La oralización de textos digitales: usos no normativos en conversaciones instantáneas por escrito. *Caracteres: estudios culturales y críticos de la esfera digital*, 2(2), 92-108. <http://hdl.handle.net/10366/124906>
- Jiang, J., A. Way & R. Hague. (2012). Translating user-generated content in the social networking space. In *Proceedings of the Tenth Biennial Conference of the Association for Machine Translation in the Americas*, (pp. 1-9). <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.299.5601&rep=rep1&type=pdf>
- Kenny, D. & S. Doherty. (2014). Statistical machine translation in the translation curriculum: overcoming obstacles and empowering translators. *The Interpreter and Translator Trainer*, 8(2), 276-294. <https://doi.org/10.1080/1750399X.2014.936112>
- Koby, G., P. Fields, D. Hague, A. Lommel & A. Melby. (2014). Defining Translation Quality. *Tradumàtica: Tecnologies de la Traducció*, 12, 413-420. <https://doi.org/10.5565/rev/tradumatica.76>
- Lommel, A. (2018). Metrics for Translation Quality Assessment: A Case for Standardising Error Typologies. In J. Moorkens, S. Castilho, F. Gaspari, and S. Doherty (Eds.), *Translation Quality Assessment: From Principles to Practice*, (pp. 109-127). Springer.
- Mellinger, C. (2017). Translators and machine translation: knowledge and skills gaps in translator pedagogy. *The Interpreter and Translator Trainer*, 11(4), 280-293. <https://doi:10.1080/1750399X.2017.1359760>
- Mitchell, L., S. O'Brien & J. Roturier. (2014). Quality evaluation in community post-editing. *Machine Translation*, 28, 237-262. <https://doi.org/10.1007/s10590-014-9160-1>
- O'Brien, S. (2011). Towards Predicting Post-Editing Productivity. *Machine Translation*, 25, 197-215. <https://doi:10.1007/s10590-011-9096-7>
- Pollach, I. (2006). Electronic word of mouth: A genre analysis of product reviews on consumer opinion web sites. In *Proceedings of the 39th Annual Hawaii International Conference on System Sciences-Volume 03. IEEE Computer Society*, (pp. 51-53). IEEE. <https://doi.org/10.1007/s10590-014-9160-1>
- Popović, M. (2018). Error Classification and Analysis for Machine Translation Quality Assessment. In J. Moorkens, S. Castilho, F. Gaspari, & F. Doherty (Eds.) *Translation Quality Assessment: From Principles to Practice*, (pp. 129-15). Springer.
- Ricci, F. & R. Wietsma. (2006). Product reviews in travel decision-making. In M. Hitz, M. Sigala and J. Murphy (Eds.), *Information and communication technologies in tourism* (pp. 296-307). Springer.
- SAE International. (2001). *SAE J2450: Translation Quality Metrics*. Society of Automotive Engineers.
- Schemmann, B. (2011). A Classification of Presentation Forms of Travel and Tourism-Related Online Consumer Reviews. *e-Review of Tourism Research*, 2. <http://ertr.tamu.edu/enter-2011-short-papers>
- Specia, L., D. Raj & M. Turchi. (2010). Machine Translation Evaluation Versus Quality Estimation. *Machine Translation*, 24, 39-50. <https://doi.org/10.1007/s10590-010-9077-2>
- Stymne, S. & L. Ahrenberg. (2012). On the practice of error analysis for machine translation evaluation. In *Proceedings of the LREC 2012 Conference. Istanbul: European Language Resources Association* (pp. 1785-1790). [http://www.lrec-conf.org/proceedings/lrec2012/pdf/717\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2012/pdf/717_Paper.pdf)
- TAUS. (2010). *Error Typology Guidelines*. <https://www.taus.net/academy/best-practices/evaluate-best-practices/error-typology-guidelines>
- TAUS. (2016). *Quality Dashboard White Paper*. <https://www.taus.net/think-tank/reports/evaluate-reports/taus-quality-dashboard-white-paper>



- Temnikova, I. (2010). Cognitive Evaluation Approach for a Controlled Language Post-Editing Experiment. In *Proceedings of the LREC 2010 Conference*, (pp. 17-23). Valletta: European Language Resources Association. [http://www.lrec-conf.org/proceedings/lrec2010/pdf/437\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2010/pdf/437_Paper.pdf)
- Torres-Hostench, O., M. Presas & P. Cid-Leal. (2016). *El uso de la traducción automática y la posesión en las empresas de servicios lingüísticos españolas: Informe de investigación ProjectA* [The Use of Machine Translation and Post-editing among Language Service Providers in Spain]. UAB. [http://ddd.uab.cat/pub/estudis/2016/148361/usotraaut\\_2016.pdf](http://ddd.uab.cat/pub/estudis/2016/148361/usotraaut_2016.pdf)
- Valli, P. (2015). The TAUS Quality Dashboard. In *Proceedings of the 37th Conference Translating and the Computer*, 127–136. London, UK, November 26-27.
- Van der Meer, J., Görög, A., Dzeguze, D., & Koot, D. (2017). *TAUS Quality Dashboard White Paper*. <https://www.taus.net/insights/reports/taus-quality-dashboard-white-paper>
- Vásquez, C. (2012). Narrativity and involvement in online consumer reviews. The case of Tripadvisor. *Narrative Enquire*, 22 (1), 105-121. <https://doi.org/10.1075/ni.22.1.07vas>
- Vásquez, C. (2014). *Online consumer reviews*. Bloomsbury.
- Vilar, D., J. Xu, L. F. d'Haro & H. Ney. (2006). Error analysis of statistical machine translation output. In *Proceedings of the LREC 2006 Conference* (pp. 697-702). Genoa: European Language Resources Association, [http://www.lrec-conf.org/proceedings/lrec2006/pdf/413\\_pdf](http://www.lrec-conf.org/proceedings/lrec2006/pdf/413_pdf).
- Yus, F. (2011). *Cyberpragmatics*. John Benjamins.

## Análisis de la comunicación experto-lego por medio del consentimiento informado: legibilidad y relación médico-paciente

Analysis of expert-patient communication through informed consent: legibility and the doctor-patient relationship

SOFÍA ZAKHIR PUIG  
Consejo Superior de Investigaciones Científicas, España  
[sofia.zakhir@cchs.csic.es](mailto:sofia.zakhir@cchs.csic.es)  
<https://orcid.org/0000-0001-5919-5859>

### Resumen

El presente trabajo de investigación tiene como objetivo analizar la legibilidad y la opinión y experiencia de sujetos no especializados sobre el Consentimiento Informado (CI), un género médico-jurídico que los pacientes, generalmente legos en la materia, leen y firman antes de iniciar un tratamiento médico. Para poder realizar dicho análisis, hemos contado con 40 sujetos de la Comunidad Valenciana con edades comprendidas entre los 18 y los 30 años y sin formación médico-sanitaria. La investigación empírica consta de dos estudios: en primer lugar, un estudio de legibilidad, en el que los sujetos leyeron un CI previo a un proceso de hormonación y subrayaron aquellos elementos que no comprendían. En segundo lugar, entrevistas abiertas semiestructuradas, con tres bloques principales relacionados con la opinión de los sujetos acerca del texto previamente leído, su experiencia

### Abstract

The aim of this article is to analyse the legibility and the beliefs and experience of lay subjects on informed consent forms, a medical-legal genre addressed to laypeople. The empirical data for this study has been obtained from 40 subjects from the Valencian Community with no medical training between the ages of 18 and 30. The article consists of two studies: a legibility study, in which the subjects were asked to read an informed consent form prior to a hormone process and then highlighted those elements that they did not understand; and a number of semi-structured, open-ended interviews with three main blocks related to the subjects' beliefs, their experience with doctor-patient communication and their evaluation of this type of document. From these studies, we found that the most important problem with the informed consent form stems from the terminology used. Besides, 45%

---

**Para citar este artículo:** Zakhir Puig, S. (2022). Análisis de la comunicación experto-lego por medio del consentimiento informado: legibilidad y relación médico-paciente. *ELUA*, (38), 137-164. <https://doi.org/10.14198/ELUA.22373>

Recibido: 30/03/2022 Aceptado: 13/05/2022

© 2022 Sofia Zakhir Puig



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

con la comunicación médico-paciente y su valoración del tipo de documento. De estos estudios, obtuvimos que el mayor problema que presenta el CI es la terminología empleada sin explicación o adaptación para personas legas. Además, un 45 % de los sujetos afirma que no pregunta sus dudas en consulta o que solo preguntaría en determinadas circunstancias. Finalmente, de todas las personas entrevistadas, solamente un 17,5 % considera que la comunicación se produce satisfactoriamente por medio del tipo de documento que nos ocupa. Los resultados de este trabajo nos llevan a pensar que el CI podría necesitar un sustancial proceso de adaptación al público lego, por ser un documento de vital importancia para los pacientes que se enfrentan a un tratamiento médico.

**PALABRAS CLAVE:** consentimiento informado; comunicación experto-lego; comunicación médico-paciente; legibilidad; lenguaje especializado.

of the subjects stated that they do not ask their doubts to doctors, or that they only do it under specific circumstances. Finally, only 17.5% of the respondents consider that laypeople can successfully understand this genre in particular. The results of this work lead us to think that informed consent forms might need a significant process of adaptation for the lay public, as it is a document of vital importance for patients.

**KEYWORDS:** informed consent; expert-lay communication; doctor-patient communication; legibility; specialized language.

## 1. INTRODUCCIÓN

La presente investigación versa sobre la comunicación médico-paciente, en concreto, sobre el acto de comunicación que se produce a través del género textual *consentimiento informado* (en adelante, CI). La motivación para realizar este trabajo nace de la preocupación por la legibilidad del género textual que nos ocupa, en tanto que género dirigido a audiencias legas. Esta no es una cuestión baladí, pues aborda la problemática de que ciertos consentimientos informados podrían transmitir información sin hacer uso de un lenguaje comprensible para personas no expertas. Que la comunicación médico-paciente no se produzca de manera satisfactoria no deja de ser un riesgo y, a la vez, ejemplifica un fallo del sistema. El punto de partida del trabajo que presentamos, pues, es que la información especializada debería ser accesible a personas legas por medio de un proceso de adaptación lingüística (quizá incluso multimodal) y, por ende, de divulgación del conocimiento experto. Alrededor de esta premisa hemos dado forma al trabajo que presentamos a continuación, que se basa en dos estudios paralelos en los que se contó con 40 sujetos legos, gracias a los cuales podemos indagar en esta cuestión, utilizando como punto de partida un CI real extraído de un proceso de hormonación por medio de testosterona.

Los objetivos de este trabajo son, en primer lugar, investigar sobre la naturaleza y las características de la comunicación médico-paciente en tanto que acto de comunicación experto-lego. Por tanto, nos ha resultado fundamental revisar la bibliografía respectiva a la comunicación especializada. En esta línea, surge como objetivo abordar tanto la traducción especializada como la transmisión de conocimiento experto a personas no especialistas para entender la adaptación a legos como un acto de traducción intralingüística. Además, tratamos de caracterizar el consentimiento informado y los requisitos y partes del documento.

En segundo lugar, tenemos como objetivo contestar a la pregunta que motiva la investigación que presentamos: ¿Se produce la comunicación médico-paciente por medio del CI? De esta pregunta inicial, surgen otros propósitos: obtener resultados sobre la experiencia y opinión de los sujetos en lo que respecta a la comunicación médico-paciente y al CI en cuestión y conocer los principales problemas que presenta el documento para la comprensión, indagando en las barreras comunicativas de la comunicación experto-lego.

A lo largo de este artículo, revisamos brevemente las bases teóricas y administrativas que sustentan este trabajo en el apartado 2; abordamos la parte empírica en el apartado 3 por medio de la metodología; en el apartado 4, por medio de los resultados; y en el apartado 5, por medio de la discusión. Finalmente, en el apartado 6, concluimos el trabajo.

## 2. MARCO TEÓRICO

### 2.1. Comunicación experto-lego

La cuestión de los lenguajes especializados ha suscitado gran interés, tanto en el ámbito de los estudios de la lengua como en los ámbitos a los que pertenece cada lenguaje o jerga especializada. A continuación, abordaremos brevemente la cuestión de los lenguajes de especialidad, por tratarse el género que nos ocupa de un texto médico-jurídico.

En primer lugar, cabe preguntarse qué define a los lenguajes de especialidad. Cabré (1993) describe la lengua como un conjunto de subcódigos que se adscriben a unas unidades y reglas comunes para toda la comunidad de hablantes, y que estos emplean en función de las necesidades del contexto dado.

Lehrberger (1986: 22) expone los seis factores establecidos que podrían ayudar a la caracterización del lenguaje de especialidad:

- (1) temática específica;
- (2) restricciones léxicas, sintácticas y semánticas;
- (3) reglas gramaticales anómalas;
- (4) alta frecuencia de aparición de determinadas construcciones;
- (5) estructura textual;
- (6) uso de símbolos específicos.

Es decir, encuentra características comunes a los lenguajes de especialidad que podrían generalizarse a los diferentes tipos de jerga de ámbito experto.

Además, respecto a la relación entre el lenguaje de especialidad y el lenguaje general, existe un debate con diferentes posiciones en el seno de los estudios lingüísticos. Diversos autores, como Juste (2016) consideran la relación entre el lenguaje de especialidad y el lenguaje general como un espectro con diferentes grados de especialidad. Según este punto de vista, el lenguaje general y los de especialidad no serían categorías estancas, sino que la lengua se movería en un espectro que tendría el lenguaje general en un polo y el de especialidad en el otro.

Como hemos visto, existe cierta diversidad en la conceptualización de los textos que usan lenguajes especializados. Se contempla, además, la aparición de diferentes lenguajes en un mismo texto, es decir, la existencia del texto híbrido. El género textual que nos ocu-

pa, el CI, como se explica más adelante, se trata de un género híbrido de carácter médico-jurídico, ya que conforma un acto jurídico que tiene lugar en el ámbito médico y por el cual se produce una autorización por parte del paciente (Gallego-Borghini 2015); es decir, como hemos anticipado, se trata de una cuestión ética y jurídica en la que las partes deberían estar correctamente informadas.

Cabe destacar, en la línea de lo previamente mencionado acerca de la comunicación experto-lego, que la diferencia en el conocimiento genera una marcada asimetría en la comunicación, cuyo grado dependerá del género textual al que nos refiramos y del registro empleado. La comprensión de documentos relacionados con el derecho, por ejemplo, se plantea como una necesidad, ya que suelen caracterizarse por ser prescriptivos (Harvey 2002). En estos casos, por tanto, la asimetría en la comunicación debe paliarse, no sólo por la democratización del conocimiento en sí, sino por las implicaciones jurídicas de tales documentos. Dicha asimetría también se materializa, como se ha anticipado, en el CI.

En esta línea, destacaremos los tres tipos de traducción que reconoce Jakobson (1959): la traducción interlingüística (la interpretación y traslación de un mensaje de una lengua a otra), la intersemiótica (la interpretación y traslación de un mensaje desde un sistema de signos a otro) y, finalmente, la traducción intralingüística (la interpretación de un mensaje que se ha escrito con unos símbolos y la expresión de dicho mensaje por medio de otros símbolos que forman parte de la misma lengua). Nos centraremos concretamente en la traducción intralingüística, ya que es la que se produce cuando se adapta un texto, así como en otras situaciones, como la adaptación de un dialecto a otro dentro de una misma lengua o las modificaciones que se aplican en las adaptaciones para un público infantil de novelas dirigidas a un público adulto.

Cuando el objetivo es divulgar el conocimiento experto, se debe seleccionar de aquellos fragmentos que sean de interés para el público general. Además, se debe reformular para explicar el contenido de especialidad en términos sencillos. Por último, se debe recontextualizar para tratar de conciliar la precisión de la ciencia con la adaptación del contenido para personas legas (Mapelli 2004). Todo esto está necesariamente relacionado con el concepto de «diseño de información», esto es, gestionar la relación entre la persona que recibe el mensaje y la información en cuestión para convertir una información de difícil acceso en una información que pueda entender el público general (Sless 1992). Durante este proceso, como hemos visto, la persona que escribe debe tomar una serie de decisiones relativas a la elección del lenguaje y del contenido y, como explica Wright (1999), el proceso podría ser más eficiente si se recibe respuesta con sugerencias o impresiones por parte del tipo de público receptor del texto.

Cabe preguntarse, llegados a este punto, si la divulgación del conocimiento experto en los actos comunicativos que se producen entre personas expertas y legas se practica realmente en el ámbito médico. Cobos (2019), en referencia a la transmisión del conocimiento científico en medicina, pone de manifiesto que no siempre los especialistas expresan la información de manera clara, concisa y directa. No obstante, la democratización de la información ha supuesto últimamente una tendencia hacia la adaptación en el ámbito médico por medio de la reformulación (Muñoz- Miquel 2016).

Una comunicación poco fructífera entre el médico y el paciente puede llevar a consecuencias negativas para el paciente en lo que respecta a su salud y, para el médico, en términos legales. Esto se relaciona necesariamente con el uso de una terminología propia de la especialidad y que el paciente no puede entender (Tena 2002). Según el autor, este es

uno de los factores que lleva al paciente a demandar al médico que lo ha tratado, ya que informar al paciente correctamente es una exigencia de la *lex artis* médica.

Aplicado al contexto jurídico, destaca la campaña Plain English, que surge en 1979 en Reino Unido con el objetivo de simplificar el uso del lenguaje especializado de los organismos del Estado (concretamente en el ámbito jurídico, aunque con el tiempo se ha extendido a otros ámbitos) para que sea más accesible al público general; es decir, más claro y conciso (Cobos 2019).

De manera similar, pero en el ámbito de la traducción biosanitaria, destacamos también el proyecto ONCOTRAD, que surge para paliar la dificultad de los pacientes con cáncer y sus familiares de entender plenamente las novedades sobre investigación con respecto a la enfermedad, por medio de una traducción que cuenta con un proceso de adaptación al público general (Cobos 2019).

Otro resultado del esfuerzo que se ha hecho en las últimas décadas en lo que respecta a la divulgación del conocimiento experto en el ámbito médico es el *e-consent*: el consentimiento informado online, que, si bien presenta ciertos beneficios, no se ha implementado aún en la mayoría de los centros médicos. Se trata de un tipo de consentimiento interactivo y multimedia que está adaptado al nivel de especialización del paciente, por lo que la comprensión se podría producir de manera más satisfactoria. Se ha señalado que el *e-consent* tiene beneficios tales como que el paciente pueda acceder al contenido del consentimiento todas las veces que lo desee (Coughlin 2015), ya que uno de los objetivos de este formato es que el paciente pueda leer y visualizar todo el contenido a su propio ritmo. Por otro lado, en lugar de una firma para todo el documento, algunos formatos de *e-consent* requieren que el paciente clique sobre todos los posibles riesgos para asegurar que ha leído todo el texto. Además, el *e-consent* permite que el paciente tenga acceso a más información que en formato papel, ya que se puede recomendar cómo acceder a información adicional, que puede encontrar, por ejemplo, en páginas web revisadas y enlazadas. A pesar de las ventajas que presenta este formato, la autora (*ibid.*) considera que puede suponer una reducción en el tiempo de interacción entre el paciente y el médico, lo cual podría llevar a que los médicos no reconocieran correctamente la situación en la que se encuentra el paciente. Por otro lado, algunos autores consideran que el coste de implementación y mantenimiento de este formato supera los beneficios que podría aportar (Eloff y Keshnee 2006). Sin embargo, cabe tener en cuenta que este formato podría suponer una comunicación más fructífera y adecuada entre los participantes de la comunicación en el futuro, lo cual no debería, en principio, ir en detrimento de la relación entre el médico y el paciente.

Estos son algunos ejemplos del fenómeno que se ha producido en las últimas décadas para lograr una transmisión del conocimiento experto más eficiente, pero, si bien se debe reconocer el esfuerzo en una mejor comunicación, aún queda mucho trabajo para conseguir una comunicación satisfactoria en el ámbito médico.

## 2.2. Caracterización del consentimiento informado

En los últimos años, el género consentimiento informado ha despertado cierto interés entre investigadores de diferentes campos del saber: el jurídico, en relación con las implicaciones legales y a la protección del personal médico y de los pacientes; el médico, por la necesidad del paciente de entender y consentir un procedimiento; o el lingüístico, en la caracterización

del género textual y el análisis de la comunicación médico-paciente. En este trabajo, pese a que abordamos la cuestión desde un punto de vista lingüístico, es conveniente que en la caracterización del CI expliquemos su relevancia en la medicina y en el derecho para poder profundizar en el análisis del acto comunicativo. Por ello, a continuación, situaremos el nacimiento del CI y las tendencias a nivel jurídico y bioético que ha sufrido el género para llegar a nuestros días. Explicaremos, además, el derecho del paciente al CI en el ordenamiento jurídico español. Concluiremos esta parte refiriéndonos a la legibilidad del CI y a la necesidad de que sea comprensible para el paciente.

Como hemos anticipado en la parte anterior, el CI se encuentra entre el ámbito médico y el jurídico y, por ende, se considera un género híbrido médico-jurídico. Casado (2008) define este tipo de comunicación como «todas aquellas actuaciones escritas que utiliza el médico en sus relaciones profesionales con las autoridades, los organismos y las instituciones o con cualquier persona». Borja (2012), por su parte, aporta la siguiente definición, más cercana al punto de vista del presente trabajo: «Aquellos [textos] en los que aparecen combinados conceptos propios de la medicina y conceptos legales [...], o bien textos médicos que pueden surtir efectos legales».

Una vez situado el CI, ilustraremos su naturaleza a partir de la definición de Gallego-Borghini (2015), que se refiere al género como acto comunicativo tanto oral como escrito: «El consentimiento informado es el acto jurídico por el cual la persona accede a participar en una investigación después de haber recibido una información exhaustiva, oralmente y por escrito, sobre todos los aspectos que son relevantes para su decisión», lo cual, a pesar de que se refiera principalmente a los CI en la investigación clínica, también podría aplicarse al ámbito de la medicina asistencial.

A grandes rasgos, el consentimiento informado es la prueba del consentimiento otorgado por el paciente para tres casos distintos (Gallego-Borghini 2015):

- (1) Una investigación clínica.
- (2) Asistencia sanitaria habitual.
- (3) Un contrato (CI para donaciones remuneradas).

Asimismo, tal y como explica el autor, el acto comunicativo que permite al paciente consentir un procedimiento médico incluye también la comunicación oral, pese a que este es un rasgo que, por extensión del presente trabajo, no abarcamos. Además, no en toda intervención queda constancia del consentimiento porque en ocasiones como, por ejemplo, un examen médico, el consentimiento es tácito o se transmite de manera oral (Cadenas 2018), lo que complica la caracterización de CI en tanto que género oral. Por tanto, en este trabajo nos centraremos en el CI como género textual escrito que firman tanto médico como paciente y que responde a unas necesidades concretas de carácter médico y jurídico.

### *2.2.1. Historia y legislación del CI*

El nacimiento del consentimiento informado suele situarse durante la celebración de los juicios de Núremberg, tras la derrota del Tercer Reich (Cadenas 2018). En estos juicios surge, entre otras cosas, el Código de Núremberg, que establece diez puntos relativos a la experimentación médica para el cumplimiento de los principios del consentimiento informado,

la libre voluntad de los sujetos participantes en el experimento y la obtención de beneficio por parte de dichos sujetos (Mainetti 1989). Destacamos, por ser de gran relevancia para este trabajo, el primer punto, en el que se estima esencial que los sujetos participantes en un experimento posean el conocimiento suficiente como para poder consentir una intervención. No obstante, cabe destacar que el código de Núremberg, si bien se convirtió en un paradigma de la bioética, no tenía como objetivo la estandarización de prácticas a nivel global, sino la persecución de crímenes de guerra, por lo que dichos puntos no eran vinculantes para los estados participantes (Cadenas 2018).

En los años 70, el consentimiento informado ya se recogía en diversos tratados internacionales, como la declaración de Helsinki (1964), y en diferentes países de todo el mundo, ya que nos encontramos en una época fuertemente marcada por el respeto a los derechos de los pacientes (Cadenas 2018).

Con respecto a la evolución en el ordenamiento jurídico español, se reconoce por primera vez en 1972 en el Reglamento General para el Régimen, Gobierno y Servicio de las Instituciones Sanitarias de la Seguridad Social el derecho de los pacientes a autorizar, por sí mismos o por medio de sus allegados, aquellos procedimientos que supongan riesgos (*ibid.*). De entre los procesos en el seno del ordenamiento jurídico español que se sucedieron a lo largo de los años 70 y 80, destacamos la Ley General de Sanidad (LGS), promulgada en 1986, por ser la primera norma jurídica de carácter general en la que se reconoce el derecho a la información asistencial, al consentimiento informado y a la negación del paciente a participar en una actuación médica. Actualmente, se encuentra sustituida por la Ley 41/2002, de 14 de noviembre, básica reguladora de la autonomía del paciente y de derechos y obligaciones en materia de información y documentación clínica-

Todos estos esfuerzos se materializan en el CI en el ordenamiento jurídico español, cuyas bases las establecen los diferentes parlamentos autonómicos. El uso del término *consentimiento informado* se produce por primera vez en el ordenamiento jurídico español hasta el Real Decreto 561/1993, que introduce lo que se considera por algunos autores una traducción literal del inglés que se ha naturalizado en nuestra lengua por su difusión (Gallego-Borghini 2015).

La ley que rige actualmente en España el CI y el derecho a la información asistencial es la Ley 41/2002, de 14 de noviembre, básica reguladora de la autonomía del paciente y de derechos y obligaciones en materia de información y documentación clínica, que entró en vigor con fecha de 16/05/2003. Esta norma reconoce el derecho del paciente a autorizar o denegar su participación en un procedimiento médico tras recibir la información pertinente y establecer varias excepciones.

### 2.2.2. Partes del CI

El género de consentimiento informado está compuesto por la Hoja de Información para el Paciente (HIP) y la declaración de consentimiento informado o formulario de consentimiento informado (Gallego-Borghini 2015). A continuación, expondremos brevemente los puntos de la Guía para la correcta elaboración de un modelo de HIP y consentimiento informado de la AEMPS. Pese a que los siguientes puntos sean relativos al documento propio de las pruebas de una investigación clínica, las recomendaciones sobre la información que se debe proporcionar con respecto al tratamiento y el lenguaje que se debe usar se aplica a todos los tipos de CI.



En primer lugar, la HIP debe informar sobre el título del estudio, su código y el nombre del investigador o investigadora principal y del centro en el que se realiza el ensayo, seguido de una introducción y un párrafo en el que se expliciten la voluntariedad en la participación del paciente en el estudio. Los puntos que dicta la AEMPS para la HIP como obligatorios son los siguientes:

- (1) Objetivo del estudio: se recomienda una descripción breve y adaptada a los conocimientos del paciente.
- (2) Descripción del estudio: según la AEMPS se debe «utilizar directamente la descripción o explicación no técnica».
- (3) Actividades del estudio: se recomienda no incluir las tablas del protocolo clínico, sino elaborar otras más sencillas y adaptadas a personas legas.
- (4) Riesgos y molestias derivados de su participación en el estudio: se recomienda no repetir molestias que ya se hayan mencionado con anterioridad.
- (5) Posibles beneficios: explicitar que es posible que no se obtenga ningún beneficio.
- (6) Advertencia relativa al embarazo: se debe informar de los riesgos a personas de sexo femenino en edad fértil o personas de sexo masculino con parejas en edad fértil.
- (7) Tratamientos alternativos: explicar brevemente la existencia de alternativas terapéuticas.
- (8) Protección de datos personales: en la explicación sobre la situación legal de la Protección de Datos, la AEMPS afirma que se debe explicare con la mayor brevedad posible.
- (9) Gastos y compensación económica y otra información relevante.
- (10) Qué tratamiento recibirá el paciente cuando finalice el ensayo clínico.
- (11) Contacto en caso de dudas.

En segundo lugar, la hoja de consentimiento del participante o declaración de consentimiento informado debe cumplir las siguientes directrices: no repetir información, incluir una lista de autorizaciones acompañadas de casillas que se deben marcar una por una, el consentimiento para recibir información adicional sobre el estudio y la firma del paciente y del médico. Respecto a la macroestructura, la AEMPS establece lo siguiente:

*Nota aclaratoria: este documento no deberá exceder las 15 páginas (incluidos los CI) y debe estar redactado en Arial o Times New Roman, de mínimo 11 puntos, con un espaciado interlineal mínimo 1,5, sin reducir los márgenes del documento. Para cada apartado se indica el contenido mínimo indispensable en negro a modo de recomendación, en cursiva y entrecomillado el texto propuesto y en negrita aquello que no es aceptable.*

### 2.2.3. Legibilidad

Como hemos podido observar en la guía previamente expuesta, el objetivo de algunas de estas directrices es que el paciente tenga acceso real a la información para poder consentir. Es decir, que, en la línea de lo explicado con anterioridad sobre la divulgación, el CI se adapte a la comprensión de una persona que puede ser lego tanto en el ámbito jurídico como en el médico.

Gallego-Borghini (2015) explica que, si bien este tipo de documento debe redactarse con un tono «neutro y formal», el uso de un lenguaje sencillo y asequible no supone una contradicción con la formalidad que requiere la situación.

De este modo, cabe destacar que las aclaraciones de los comités éticos de investigación clínica resultan en una serie de peticiones relativas al estilo para evitar confusión, además de una mejora en la redacción del documento (*ibid.*: 5). Por otro lado, tal y como explica el

autor, es común que los comités pidan que el CI sea más breve, o bien que se resuma en su totalidad o que se resuma algún apartado concreto, y que se eviten repeticiones. Gallego-Borghini ilustra dichas peticiones con comentarios como los siguientes: «los consentimientos informados son demasiado farragosos», «la redacción de las 3 hojas es confusa», «debería explicarse qué significa “doble ciego”» y objeciones similares relativas a la terminología y al estilo de redacción (*ibid.*).

En definitiva, cabría preguntarse si las recomendaciones de la AEMPS se aplican realmente a los consentimientos informados y si, por ende, se produce la comunicación entre especialista y lego o si el CI tiene un objetivo meramente probatorio para evitar conflictos legales entre pacientes y médicos.

En nuestra opinión, consideramos necesario que la comunicación médico-paciente no tenga reparos en realizar un ejercicio de traducción intralingüística cuando el objetivo último de la comunicación sea, precisamente, la divulgación de conocimiento experto para personas legas.

### 3. METODOLOGÍA

El estudio que planteamos a continuación emplea como punto de partida la bibliografía revisada en torno a la traducción intralingüística, la comunicación especializada y el género CI con la finalidad de evidenciar la problemática que plantea la legibilidad y comprensión de documentos de carácter especializado entre audiencias legas.

El principal objetivo de este estudio es contestar a la siguiente pregunta: ¿Se produce correctamente la comunicación entre médico y paciente por medio del CI? De este objetivo central se desengrana toda una serie de objetivos parciales que exponemos a continuación:

- Conocer algunos de los problemas de legibilidad respecto al género textual que nos ocupa.
- Explorar si existe una barrera lingüística de cara al acceso a discursos especializados por parte de personas legas. En caso afirmativo, profundizar sobre qué tipo de barrera.
- Contrastar los datos y contextualizarlo con la opinión de usuarios y su experiencia en la comunicación médico-paciente.

Los objetivos que hemos planteado previamente y sobre los que desarrollamos esta investigación nos llevaron a considerar las siguientes hipótesis con respecto a los resultados del estudio:

- Ciertos elementos del texto que nos ocupa pueden dificultar su comprensión para personas legas, especialmente la terminología empleada.
- Algunos sujetos pueden considerar que el consentimiento informado no tiene como objetivo proporcionar una información accesible, sino proteger al personal médico-sanitario de posibles complicaciones médicas a raíz de un tratamiento o una intervención.
- Si bien puede que algunos sujetos consideren que la comunicación médico-paciente se produce de manera adecuada según su experiencia, otros podrían sentirse cohibidos y no preguntar en consulta o no sentirse suficientemente informados.

#### 3.1. Población y muestra

En lo que respecta al muestreo, cabe destacar que la elección de los sujetos de estudio se vio determinada, principalmente, por la extensión del presente trabajo. La inclusión de sectores

de la sociedad más amplios de los que detallamos a continuación, pese a que ampliaría los resultados y ciertamente otorgaría más valor a los mismos, sería inabarcable para un estudio de estas características. Por esto, la muestra que presentamos a continuación se llevó a cabo partiendo de los siguientes parámetros de inclusión:

- (1) Sujetos con un rango de edad entre los 18 y los 30 años, ambos inclusive.
- (2) Sujetos que vivan en la Comunidad Valenciana.
- (3) Sujetos con un nivel socioeconómico medio.
- (4) Sujetos que no tengan formación relacionada con el ámbito médico-sanitario.
- (5) Sujetos que no se hayan sometido previamente al proceso médico que describe el CI.

En total, se obtuvo una muestra formada por 40 sujetos (como hemos anticipado antes, 20 hombres y 20 mujeres) que residían en los siguientes municipios de la Comunitat Valenciana: València, Burjassot, Beneixama, Villena, Silla, Carcaixent, Alzira y Castelló.

La edad media de los sujetos era de 21,67 años, teniendo una mayor representación los sujetos con 20, 21 y 22 años, con una desviación estándar de 2,15.

Un 23% de los sujetos (9 en total) confirmó no tener estudios superiores, mientras que el 77% restante tenía formación en los siguientes campos: Comunicación audiovisual, Química, Ingeniería del diseño e Ingeniería mecánica, Periodismo, Ciencias políticas, Economía, Publicidad y Ciencias de la actividad física y del deporte. Un 28% del total de sujetos entrevistados (es decir, 11 sujetos) tenía formación lingüística de algún tipo (Traducción e Interpretación, Filología inglesa e hispánica y Lenguas modernas).

### **3.2. Instrumento**

Con el propósito de alcanzar los objetivos expuestos anteriormente, hemos trabajado con dos instrumentos principales: la legibilidad de un consentimiento informado (Estudio 1) y una entrevista semiestructurada (Estudio 2).

Para el estudio de legibilidad, hemos empleado un consentimiento informado previo al proceso de hormonación por medio de la administración de testosterona para un hombre trans. Debido a las limitaciones del estudio, decidimos reducir el documento, que inicialmente tenía seis páginas, a la hoja informativa del consentimiento para el paciente, de tres páginas.

#### **Estudio 1: análisis de legibilidad**

En primer lugar, como hemos anticipado, realizamos un análisis de legibilidad para conocer aquellos aspectos textuales, como la terminología o la sintaxis, que sean difíciles de comprender para las personas legas.

#### **Estudio 2: entrevistas semiestructuradas de respuesta abierta**

El segundo instrumento fue una entrevista semiestructurada de respuesta abierta con la intención de recibir diferentes opiniones sobre todo aquello que las personas entrevistadas

podieran considerar como un obstáculo para la comunicación experto-lego, así como aquellos elementos que los sujetos de estudio no habían podido subrayar durante el Estudio 1 pero que igualmente identificaban como obstáculo. El método empleado en la formulación de preguntas tiene como objetivo que las personas entrevistadas puedan expresar todas sus valoraciones o ideas sobre el consentimiento informado sobre el que trabajamos sin estar limitadas por preguntas cerradas.

Las preguntas de la entrevista se organizaron alrededor de los siguientes bloques temáticos:

- ¿Qué elementos consideran complejos los sujetos que no hayan podido subrayar?
- ¿Qué opinión tienen la comunicación experto-lego y el género *consentimiento informado*?
- ¿Plantean o plantearían sus dudas al personal sanitario en una situación real? ¿hasta qué punto firman los documentos a sabiendas de no entender toda la información que se incluye en los mismos?

### 3.3. Método

La metodología empleada para llevar a cabo este estudio se puede resumir en varios puntos: En primer lugar, con respecto al estudio 1:

- La primera fase consiste en que los sujetos del estudio lean el consentimiento informado, en tanto que género perteneciente a un campo de conocimiento especializado, y subrayen aquellas partes que consideren más difíciles para su comprensión.
- En una segunda fase, se codificó y se analizaron las respuestas obtenidas.
- En tercer lugar, se identificaron aquellos aspectos textuales o extratextuales que pudieran dificultar la comprensión del CI.
- Finalmente, se valoraron los resultados obtenidos a lo largo de la realización del estudio de legibilidad.

Con respecto al estudio 2:

- Tras la lectura del CI, se realizaron las preguntas abiertas siguiendo los tres bloques mencionados previamente.
- Se codificaron las respuestas y se analizaron las experiencias y opiniones de los sujetos sobre el documento y sobre la comunicación médico-paciente.
- Finalmente, se valoraron los resultados del análisis llevado a cabo tras realización de entrevistas abiertas.

Ambos estudios tuvieron lugar en el año 2020, desde el 12 de abril hasta el 10 de mayo. Se realizaron de manera síncrona, por medio de videoconferencias que fueron grabadas, previo consentimiento de los participantes, con el objetivo de recoger toda la información proporcionada por las personas entrevistadas.

Este método, junto los instrumentos y la muestra descritos previamente, nos llevaron a los resultados que exponemos a continuación.

## 4. RESULTADOS

Presentaremos, en primer lugar, los resultados del análisis de legibilidad del fragmento de CI que hemos seleccionado para este trabajo y, en segundo lugar, los resultados de las entrevistas realizadas.

### 4.1. Estudio 1: Legibilidad

En este primer estudio analizamos los resultados relacionados con la legibilidad del consentimiento informado y aquellos elementos que los sujetos encontraron particularmente complejos de entender durante su lectura.

De entre los elementos subrayados por parte de los sujetos, se han extraído un total de siete categorías que representan aquellos aspectos textuales que dificultan la comprensión del texto para personas legas.

Los 40 sujetos del estudio subrayaron un total de 388 elementos. Cabe destacar el hecho de que todos los sujetos subrayaran algún elemento y que, por tanto, todos encontrarán, en mayor o menor medida, problemas en la legibilidad del texto. En la Tabla 1, se puede observar cada una de las categorías entre las cuales hemos dividido los aspectos subrayados por los sujetos, la frecuencia con la que se ha subrayado cada aspecto y el porcentaje que representa cada uno.

Aspectos subrayados	Frecuencia absoluta	Porcentaje
Terminología propia del campo de especialidad	230	59%
Fraseología propia del campo de especialidad	59	15%
Sintaxis	56	14%
Falta de especificidad, estilo abstracto	18	5%
Estructura del documento	13	3%
Siglas, acrónimos y abreviaturas	8	2%
Léxico no especializado	4	1%
Total	388	

Tabla 1. Aspectos subrayados por los sujetos y la frecuencia con la han sido indicados.

A continuación, comentaremos los aspectos más destacados e ilustraremos cada categoría con ejemplos correspondientes.

En primer lugar, la terminología propia del campo de especialidad, en tanto que el conjunto de unidades de conocimiento especializado, supondría la mayor barrera comunicativa para los sujetos, representando el 59% de frecuencia de elementos subrayados (230 ocasiones totales). Aquellos términos con mayor recurrencia entre los subrayados son los siguientes:

- (1) *Incongruencia de género*
- (2) *Antiandrógenos*
- (3) *Tromboembolismos (cerebrales, pulmonares y periféricos)*
- (4) *Análogos de GnRh*
- (5) *Tumores hipofisarios*
- (6) *Apnea del sueño*
- (7) *Transaminasas hepáticas*

Además, algunos sujetos señalan que conocen ciertos términos, pero no son conscientes del riesgo que puede suponer lo descrito en el documento sobre los mismos, como es el caso de *aumento de glóbulos rojos o efectos indeseables sobre la densidad mineral ósea*.

Muy ligadas a la terminología específica del campo, la aparición de siglas propias del ámbito médico supone el 2% del total de elementos destacados. Las siglas, subrayadas en 8 ocasiones, son las siguientes: en primer lugar, *ECR06*, código de la hoja informativa que aparece en la cabecera (probablemente por una cuestión de comunicación entre el personal médico-sanitario). En segundo lugar, *WPATH*, que alude a *World Professional Association for Transgender Health* (Asociación profesional mundial para la salud transgénero). No obstante, cabe destacar que en el documento en cuestión no se aclara a qué se refieren dichas siglas, por lo que el paciente tendría que buscarlo por su cuenta si lo considerara oportuno.

A continuación, el léxico no especializado, en referencia a aquellas palabras del texto que no forman parte de la terminología específica del campo pero que pueden suponer cierta dificultad. El número de casos recogidos representaría el 1% del total (4 elementos subrayados). Concretamente, nos referimos a *de forma anónima* y *de forma específica*, que los sujetos en cuestión subrayaron porque, si bien conocían el significado de las palabras, no estaban seguros de qué implicaba cada palabra en el contexto en el que aparece.

La sintaxis, como categoría que agrupa aquellas oraciones que podrían considerarse complejas (frases largas, voz pasiva, dobles negaciones, incisos largos, etc.), representa el 14% de los casos (56 frases subrayadas). Aquellas oraciones que se presentan como más recurrentes entre las subrayadas son las siguientes:

*Continuamente se van publicando en la literatura médica algún efecto indeseable, no siempre atribuible los tratamientos, pero tampoco pudiendo descartarse.*

*Es imposible recogerlas todas en este consentimiento porque continuamente se están publicando otras nuevas, aunque son muy poco frecuentes.*

*Con independencia del fármaco y la dosis que se emplee, los riesgos variarán según las personas y sus individuales características.*

Por otro lado, la fraseología médica, referida a aquellas frases que no presentan complejidad por su sintaxis pero que pueden afectar a la legibilidad del documento por no ser propias del ámbito general, representa un 15% (59 frases subrayadas). Aquellas oraciones que se han subrayado en mayor medida son las siguientes:

*Se aportan datos de series publicadas por varios grupos de trabajo, pero son de series cortas y sin estar adecuadamente controladas.*

*(...) para descartar algún proceso acompañante que pudiera distorsionar el procedimiento. El acompañamiento hormonal pretende conseguir reducir o eliminar la producción de sus actuales hormonas sexuales.*

La estructura del texto, por su parte, representa el 3% (13 sujetos), y su aparición se concentra exclusivamente en el apartado **riesgos del procedimiento**, en el que se agrupan todos los riesgos que acarrearán las tres hormonas, especificando entre paréntesis qué hormona puede ocasionar cada riesgo. Los sujetos que consideran que la estructura no es clara han subrayado las aclaraciones entre paréntesis o, en su defecto, el párrafo entero. Como veremos en el siguiente estudio, algunos de estos sujetos proponen ideas para paliar el problema y conseguir, por medio de una estructura más sencilla, que el apartado de riesgos, siendo tan relevante para el consentimiento, sea más legible.

Finalmente, la falta de especificidad, señalada por los sujetos cuando manifiestan que podrían llegar a entender ciertos elementos, pero que necesitarían más información sobre los mismos. Este obstáculo representa el 5% de los casos totales (es decir, se ha subrayado en 18 ocasiones). Los sujetos que subrayan elementos por este motivo expresan, como veremos en el siguiente estudio, que necesitarían más datos para sentirse informados. Algunas de las oraciones más señaladas por este motivo son las siguientes:

*Hay muchas variaciones individuales.*

*Pueden producirse efectos secundarios adversos en el organismo de diferentes tipos.*

*La mayor o menor respuesta depende de condiciones genéticas, edad de inicio de los tratamientos y de otras circunstancias no bien conocidas.*

Este problema, como veremos a continuación, no solo supondría un riesgo en lo que respecta a la comprensión del contenido del texto, sino que podría llevar a algunos sujetos a sentir desconfianza y confusión tras la lectura del documento.

## **4.2. Estudio 2: análisis cualitativo de entrevistas**

Con respecto a la segunda parte del estudio, se llevaron a cabo 40 entrevistas semiestructuradas, tal y como se detallan en la sección de metodología arriba. Las entrevistas estuvieron guiadas por tres preguntas principales relacionadas con la opinión y experiencia de las personas entrevistadas con el tipo de documento que nos ocupa y con la comunicación entre el personal médico-sanitario y los pacientes. Con el objetivo de mantener el anonimato de los sujetos durante el análisis de los datos, asignamos la letra H (para los hombres) y M (para las mujeres), y un número del 1 al 20 para cada participante en función del orden en que se hicieron las entrevistas.

### **Aspectos textuales: ¿Qué dificulta la comunicación?**

A la pregunta inicial sobre qué elementos consideraban complejos del primer estudio que no hubieran podido subrayar, obtuvimos cierta diversidad de respuestas porque, como hemos mencionado, no había opciones a las que los sujetos se pudieran ajustar, sino que se les pedía que contestaran desde la subjetividad, con el objetivo de paliar algún rasgo de la investigación que hubiéramos pasado por alto en el primer estudio del presente trabajo.

En cualquier caso, se puede observar que hubo ciertas coincidencias. A continuación, expondremos las respuestas agrupando aquellos aspectos y opiniones que comentaron las personas entrevistadas e incluiremos, además, las transcripciones de aquellas declaraciones que puedan resultar de interés para este estudio.

En primer lugar, cabe destacar que 10 personas, un 25% de los sujetos, manifiestan que el texto, en general, no presenta demasiados problemas para su comprensión. A este respecto, H18 considera lo siguiente:

*El texto parece estar bien estructurado de cara a la información que aporta. (...) Teniendo en cuenta que quien se esté planteando la situación de querer hacer un tratamiento para hacer el tránsito de género probablemente haya buscado todos los detalles posibles que haya podido encontrar, y que analice lo que significa cada una de las cosas de los efectos secundarios si su vocabulario médico no abarca el del texto, creo que es comprensible para todo aquel que se presente a la situación de recibir este texto.*

En esta línea, coincide con M11 en que la responsabilidad recae sobre el paciente que va a someterse al tratamiento, que debe haberse informado por su cuenta, y no tanto sobre el personal médico-sanitario:

*Para el grupo al que en teoría va dirigido está bien porque esas personas ya tienen un mínimo de conocimiento o se están empezando a informar porque están requiriendo esto.*

No obstante, de entre los 10 sujetos que consideran que el texto no presenta mayor problema, cabe destacar que 6 comentan que, aunque el texto se entienda a grandes rasgos, sí puede resultar compleja cierta terminología.

De este modo, con mayor grado de coincidencia, 20 sujetos, un 50% del total de la muestra, manifiesta que la terminología empleada en el documento es compleja y, o bien no la entienden, o bien no entienden el porqué del uso de términos especializados en este tipo de documento. En esta línea, M19 dice lo siguiente:

*El texto en general se entiende, pero se emplean conceptos técnicos que, si no se tiene ningún tipo de conocimiento sobre ese tema, puede que no se entiendan. Creo que los médicos deberían explicar la parte de riesgos de procedimiento con un lenguaje más adaptado a una persona que no entiende de medicina.*

Del mismo modo, 8 sujetos (20%) expresan que el texto tiene, en general, un grado alto de complejidad que dificulta la comprensión para personas legas, especialmente el apartado de riesgos del procedimiento. Así lo expresa M1:

*Hay mucha información, y sobre todo bastante compleja. Porque claro, [el texto] habla de todos los riesgos que puede acarrear la intervención (...) Creo que se podría adaptar un poco más.*

Con respecto a las propiedades textuales, algunos sujetos observan falta de coherencia (2) y de cohesión (2). Sobre la coherencia, tanto M6 (uno de los sujetos con formación lingüística) como M12 coinciden en lo siguiente:

*Veo que en algunas partes la terminología y el léxico es muy específico y en otras es más genérico. [M6]  
[...] no sé, es muy raro. Porque por una parte es cero técnico y está redactado súper raro y por otra parte es como que no te explica los tecnicismos. [M12]*



Sobre la falta de cohesión, M9 y H2, sujeto que también tiene formación lingüística, consideran que faltan conectores que hagan el discurso más legible. Ambos coinciden en la siguiente idea:

*Yo creo que hay fallos de cohesión, en el sentido de que faltan muchos conectores y a veces las frases sólo se separan por un punto y seguido y va saltando de tema en tema sin ningún conector y uno a veces se pierde leyendo el discurso de párrafos que son muy grandes. [H2]*

Del mismo modo, 6 sujetos señalan que encuentran en el texto diversos errores ortotipográficos. Concretamente, los sujetos mencionan errores de puntuación, de paréntesis, tildes, mayúsculas, etc. M9, a este respecto, comenta lo siguiente:

*No encuentro los signos de puntuación que tocaría, es decir, sí que hay puntos, pero las comas... es como que han desaparecido. Cuando lees, tú sabes dónde parar, o sea, sabes dónde poner los signos de puntuación, aunque no estén visibles, pero aun así se puede llegar a hacer farragoso, porque son párrafos muy grandes.*

Asimismo, 11 sujetos (28%) consideran que hay un exceso de repeticiones. De manera generalizada, manifiestan que algunas frases o ideas se repiten de forma innecesaria, lo cual hace la lectura del texto más pesada. Podemos observar a continuación la opinión de M12 sobre este aspecto:

*Empezando por ahí, hay cosas que se repiten (...) básicamente en todos los párrafos, como por ejemplo en el primer trocito y en el último. Y es que además pone como lo mismo o casi lo mismo. Y, por ejemplo, en la segunda hoja me parece [que] está copiado y pegado lo mismo en la línea 1 y en la línea 3. Y es como: me lo acabas explicar, ¿por qué me lo estás explicando otra vez?*

Igualmente, 11 sujetos manifiestan que el texto presenta un exceso de información acumulada y que, quizá, debería distribuirse de otro modo. M4 y M7 coinciden en lo siguiente:

*Además, en cuanto a la estructura, me parece que es un texto que está todo muy acumulado, muy denso, y dificulta la lectura, porque es difícil poner atención todo el rato a un bloque de texto, cuando se podría dividir en párrafos o no sé, algo mucho más visual. [M4] Y, sobre todo, corregiría el formato y haría los resaltes necesarios, para que la estructura visual ayude a la comprensión. [M7]*

Además, M16, H6 y H15 consideran que no debería haber un único consentimiento para testosterona, estrógenos y antiandrógenos, y que sería más apropiado que hubiera un consentimiento distinto para cada una de las hormonas mencionadas para así no abrumar al paciente. En esta línea, podemos leer a continuación la opinión de H6:

*Creo también que cuando habla de los riesgos del procedimiento y pone cosas tipo «Se han descrito casos de tromboembolismos (estrógenos)». Creo que esa clase de frases no son muy claras. [...] O incluso tener un consentimiento para cada tipo de hormona sería una buena idea. Eso permitiría que cuando dice a lo largo del texto que no se pueden recoger todos los riesgos o todas las condiciones y eso... Si el consentimiento*

*fuera solo sobre testosterona o solo sobre estrógenos o antiandrógenos, creo que eso permitiría que se añadiera más información.*

Igualmente, 6 sujetos se refirieron a la estructura e hicieron constar que podría mejorarse, por medio, como comenta H6, de dividir párrafos para que sea más legible o, como sugiere H3, de fusionar aquellos que estén relacionados. Los otros 4 sujetos simplemente consideran que la estructura no es clara, es poco visual y que debería revisarse y adaptarse.

Por otro lado, 9 sujetos (un 23% del total) señalan que en el texto falta información y que, por tanto, no sienten que la información aportada sea la suficiente como para poder consentir un tratamiento. H2, H9 y M9, por su parte, consideran concretamente que el texto debería presentar un mayor grado de especificidad con respecto a los riesgos. En esta línea, M9 comenta lo siguiente:

*Luego, no te dan datos concretos. ¿Cómo que la posibilidad es mínima? Dame datos. ¿Qué tienes que hacer, un trabajo de investigación para informarte?*

Además, podemos observar a continuación la opinión de M6:

*Creo que esto sería como un resumen de todo lo que hay detrás, entonces el paciente o la paciente tendría que informarse muchísimo más de todo lo que hay por su cuenta.*

Por el contrario, 5 sujetos consideran que hay un exceso de información o una falta de síntesis. En este sentido, M12 expresa *que son tres páginas y la información útil de estas tres páginas es como media página*. Desde un punto de vista distinto, H12 dice lo siguiente sobre la decisión de incluir datos sobre la literatura científica relacionada con el proceso al que se refiere el CI:

*Cuando se habla un poco de lo que es el estudio, la literatura más científica... Pues yo no sé, creo que la mayoría de la gente no sabrá lo que es un estudio académico, lo que es una revista académica o eso, entonces, claro, decirle que hay datos de series publicadas, que son series cortas o que están revisando y tal, pues igual puede sonar un poco a chino.*

Del mismo modo, 8 sujetos consideran que el estilo de redacción es inadecuado, lo cual, según expresan, dificulta que se produzca la comunicación. Por ejemplo, M18 considera que «se hace complicado de leer y es enrevesado», M17 considera que las frases son demasiado largas y M12 que «está mal escrito». M7, por su parte, expresa lo siguiente:

*La redacción parece más bien una transcripción de algo expresado de manera oral sobre la marcha, sin atender a las pausas, sino más bien en lo que se dice y en anotar lo.*

En esta línea, 10 sujetos (de los cuales, 5 destacaron una sintaxis demasiado densa) señalan que el texto está redactado con un estilo pesado que, según apuntan, hace que sea cansado de leer, lo cual podría llevar a que el paciente no lo leyera. Precisamente esto nos comenta H3:

*Me ha costado demasiado, he tardado demasiado en leerlo y no me ha gustado, no sé explicarlo. Hace que no quieras que te lo leas. Yo me lo he leído porque tú me lo has pedido, pero a mí me dan esto y yo no sé si me lo leo.*

Además, como respuesta a la pregunta de este bloque, algunos sujetos expresan sus impresiones o aquello que les transmite el documento. Concretamente, 8 sujetos consideran que el texto genera desconfianza o miedo. Coinciden en que la información se expone de modo que resulta poco alentadora. Algunas personas, de hecho, mostraron preocupación por este hecho. Un ejemplo de esto es la declaración de H8:

*Es que no te aseguran nada. En plan, hay una parte que dice que igual no se consiguen los objetivos que la persona pretende conseguir, ¿sabes? Y es en plan... Vamos a ver, ¿cómo me voy a someter a un tratamiento que igual no me ofrece lo que yo quiero? ¿Sabes? No sé, lo veo muy confuso. O sea, ¿muy poco fiable? Es que es como: esto es lo que hay, esto es lo que hay que hacer y si tú firmas no nos hacemos responsables de lo que pase, porque tú ya has firmado.*

Otros 5 sujetos manifiestan que el texto transmite frialdad y falta de cercanía. Si bien aclaran que entienden que este tipo de documento suele estar caracterizado por ser impersonal, consideran que no debería serlo en una situación que podría resultar tan emocionalmente impactante como la que se describe en el CI en cuestión. Por ejemplo, M10 considera lo siguiente:

*Considero que es un tipo de comunicación un poco fría, más de tipo contractual, para que simplemente quede constancia de que el paciente o la paciente consiente todas aquellas acciones médicas que tengan que ver con el tratamiento.*

Asimismo, M15 expone lo siguiente:

*Se tendría que utilizar igual un lenguaje un poco más cercano, no en plan demasiado, pero no dejarlo tanto como las instrucciones de un ordenador.*

Además, 6 sujetos comentan que perciben el documento como una salvaguarda jurídica del profesional sanitario y no como un verdadero consentimiento con fines informativos. Es decir, que sienten que el objetivo de este documento es proteger al profesional que va a aplicar el tratamiento y no tanto informar al paciente. Por ejemplo, M2 y M13 consideran lo siguiente:

*Parece que esté todo en tus manos y si quieres someterte a esto es tu responsabilidad, nosotros no nos hacemos cargo de nada. [M2]*

*Parece que se haya redactado un poco rápido y corriendo, en plan: esto lo tengo que redactar porque está por ley, pero no me interesa que lo hayas entendido. [M13]*

Por otra parte, 5 sujetos expresan que este documento genera confusión. Comentan que dicha confusión podría llevar al paciente a no poder hacer preguntas o a no querer leer el texto. En esta línea, H3 reconoce lo siguiente:

*[...] las cosas están un poco puestas así soltando información sin más y eso lo hace un poco confuso y como que te apetece no leerlo.*

Asimismo, algunos sujetos señalan aspectos concretos del documento que podrían perjudicar la relación médico-paciente.

En primer lugar, 2 sujetos sugieren que se debería evitar el uso de marcas no inclusivas y optar por un lenguaje inclusivo por el tipo de tratamiento que describe el documento. M6 y H4 coinciden en la siguiente idea:

*Teniendo en cuenta que es un texto de unos papeles, digamos, para hormonarse, confirmar tu identidad de género, veo que existen todavía muchas marcas no inclusivas, sobre todo lo que viene siendo al inicio y al final, sobre todo lo que hemos visto de «los padres», «el médico». O sea, no tiene sentido poner «el o la interesada» y luego poner «los padres», «el médico», es que no tiene sentido. [M6]*

En segundo lugar, M12 considera, además, que en el texto hay marcas particularmente peyorativas que podrían afectar psicológicamente a la persona que lo recibe, como podemos ver a continuación:

*Que suena muy despectivo lo de «incongruencia de género» y algunas cosas que me han parecido un poco despectivas. Yo como persona que está leyendo esto que no lo ha recibido [el documento, en una situación real], pues vale. Pero una persona que ha recibido esto y que realmente le afecta...*

En tercer lugar, M4 considera que el principal error del texto es que es elitista y que ese es el motivo por el cual no se produce la comunicación.

*Esta comunicación, la verdad, me parece, si te soy política, un poco elitista. En plan, siempre se ha dado esto de que un médico es como una figura superior, así respetable, que lo son, pero... O sea, el objetivo de un médico no es demostrar que tiene conocimientos, sino de verdad ayudar a la persona y esta persona no tiene por qué haber estudiado cuatro años medicina, ¿sabes?*

Si bien M4 es el único sujeto que califica al texto de elitista, otros 10 sujetos consideran también que la comunicación no se produciría de manera satisfactoria con personas sin estudios.

Finalmente, 4 sujetos observan que el texto está desactualizado y que algunos términos o expresiones no son apropiados actualmente. Concretamente, se refieren al uso de *transsexual*, al uso indistinto de *género* y *sexo*, así como *identidad sexual* en lugar de *identidad de género*, y a la expresión *género al que siente pertenecer* en lugar de *género al que pertenece*. A continuación, podemos ver algunas de las opiniones a este respecto:

*También me parece que está un poco desactualizado, que en algunos momentos utiliza unos términos y en otros, otros. Y en algunos momentos utiliza términos que directamente no tendría que utilizar, como la palabra «transsexual». A ver, que tienes que estar puesto en el tema para hacer esto. [M13]*

*Da la sensación de que podría estar utilizándose desde hace 10 años y que de aquí otros 10 se siga usando el mismo formulario. [M14]*

### **Aspectos vitales: ¿cuál es la experiencia de los sujetos en cuanto a comunicación médico-paciente?**

A continuación, analizaremos desde un punto de vista cuantitativo y cualitativo las respuestas de los sujetos a la cuestión sobre si preguntan todas sus dudas cuando van al médico. En la Tabla 2, podemos observar la frecuencia total y porcentual de las respuestas de los sujetos:

Respuestas	Frecuencia absoluta	Porcentaje
Sí	22	55%
Depende	8	20 %
No	10	25%

Tabla 2. Respuestas a la pregunta: ¿Expresas todas tus dudas en la consulta médica?

En primer lugar, cabe destacar que los sujetos que han contestado que sí preguntan todas las dudas coinciden en que consideran importante preguntar siempre por tratarse de un tema de salud. Gran parte de ellos, además, enfatizan que sí realizarían todas las preguntas necesarias en una situación como la que proponemos en este estudio, debido al riesgo que podría suponer no hacerlas.

De hecho, M18 comenta que incluso llama al médico para formular todas las preguntas que se le hayan ocurrido fuera de la consulta y, de este modo, no tener que recurrir a buscar información por internet. Por otro lado, M6 y H1 explican que sí formulan todas las preguntas en general, pero que cabe la posibilidad de que en ocasiones no pregunten cosas que no consideran relevantes.

Por otro lado, algunos de los sujetos que sí preguntan señalan que no siempre obtienen las respuestas que necesitan. Es decir, que el personal médico-sanitario no siempre resuelve sus dudas o que, si lo hacen, emplean un lenguaje que no comprenden. En esta línea, H8 y H12 coinciden en la siguiente idea:

*Yo creo que la principal barrera que hay es cuando a lo mejor te dicen un tecnicismo, ¿no? «Pues usted tiene un tal, tal, tal» y claro, tú no sabes qué es eso y le preguntas «¿Y qué es un tal, tal, tal?» y te responden con otros tecnicismos. Entonces, claro, responder con más tecnicismos, pues no ayuda, que [es algo que] a veces los médicos lo hacen. [H12]*

Además, H14 comenta que, al no tener formación en el ámbito médico-sanitario, no tiene claro qué preguntas debería formular:

*El problema es que cuando voy a tratarme de algo obviamente no tengo los conocimientos que tiene esa persona y yo, aunque sé más o menos cuál puede ser mi problema que quiero curar, la cosa es que no sé; y como no sé, no puedo preguntar. Entonces, agradezco que se me deje claro todo (...), para que yo no tenga que pensar si esto lo sé o no lo sé o qué tengo que preguntar.*

Asimismo, cabe destacar que algunas de los sujetos que contestan afirmativamente aclaran que son conscientes de que, en general, mucha gente no lo hace. De hecho, 9 personas, como podemos observar en la tabla, preguntan dependiendo de diversos factores. Dichos factores son la cantidad de dudas que tengan, la actitud del personal médico- sanitario o su situación personal, entre otros que veremos a continuación.

Algunos sujetos cuentan que depende de la actitud del personal. H5 coincide con M13, H17 y H9 en la siguiente idea:

*Dependería de la actitud del personal sanitario, que muchas veces corta el rollo bastante. [H5]*

M15, por su parte, cuenta que el factor de la edad del personal que le atiende es importante:

*A veces sí que digo: pues mira, no pregunto más porque, total, no me lo va a contestar. Si eso ya lo busco en casa y mejor. Me da la impresión de que cuando son jóvenes parece que intentan acercarse más a ti y utilizar un lenguaje que digas «vale, lo puedo entender». Pero cuando son mayores es como (...): «yo hago mi trabajo, que es médico, pero ya comunicar da un poco igual».*

Aunque este aspecto solo haya sido comentado por M15, parece conveniente preguntarse si esta es una sensación generalizada y, en caso afirmativo, si se debe a la evolución en la importancia de la comunicación médico-paciente en la formación del personal médico-sanitario en los últimos años o si es, quizá, una cuestión generacional.

Por otro lado, M9 comenta que siente que en ocasiones molesta si hace muchas preguntas:

*Al principio sí hago todas las preguntas, pero luego llega un punto en el que te cohibes y buscas la respuesta por otros medios. Y encima sientes que pasa el tiempo, que hay gente fuera esperando, y dices «Mira, abandono el barco. Si tiene que pasar algo, que pase. Hasta luego».*

H6, además, considera, desde su experiencia personal, que el principal problema es recordar todas las dudas que surgen, ya que muchas veces el CI no se lee junto con el médico:

*Las veces que yo he tenido que firmar uno realmente el médico me ha dicho «bueno, ¿te has traído el consentimiento informado leído de casa? ¿Tienes alguna duda?». No hemos ido conjuntamente revisando todo, entonces, claro, es un poco difícil acordarse de todas las dudas que puedan surgir, sobre todo con términos específicos que literalmente no entiendes.*

Finalmente, 10 sujetos contestan que no suelen preguntar al personal médico-sanitario cuando tienen dudas. Han expuesto diversos motivos, como vergüenza y miedo, tanto a no entender la respuesta como a la actitud del personal que atiende en cada caso. Destacamos que M4, H2 y H13 cuentan que les da miedo ser juzgados por su desconocimiento. H2, por ejemplo, lo expresa del siguiente modo:

*Yo creo que no haría todas las preguntas de cosas que no entiendo porque como no he entendido tantas cosas quedaría como que no entiendo nada. O que mi capacidad de entendimiento es reducida.*

Del mismo modo, M12 comenta que no pregunta porque no siente que sus dudas se vayan a resolver. Además, explica que la situación que se produce en una consulta no resulta cómoda para que el paciente exponga sus dudas por la frialdad del encuentro:

*Por mucho que yo pregunte es como que luego tengo muchas más dudas y son dudas que ellos no aclaran. Es como que informan lo justo y necesario y con sus palabras. Es como que están más tecleando en el ordenador y haciendo sus cosas que prestándote atención a ti.*

M14, por su parte, explica que no formula sus preguntas porque siente que no va a entender las respuestas del médico («no preguntaría por los tecnicismos porque pienso que no lo voy a entender»).

Además, M3 y M11, que a la primera pregunta habían contestado que el texto no presentaba mayor problema si se formulaban todas las preguntas que surgieran al personal médico-sanitario, se percatan con esta pregunta de que, de hecho, ellas no exponen sus dudas y que, por tanto, el texto debería adaptarse para que la responsabilidad no la tuviera el paciente.

### Aspectos vitales II: ¿Qué valoración hacen los sujetos de la comunicación médico-paciente a partir del CI del estudio?

En este apartado expondremos finalmente las respuestas a la última pregunta, relativa a si se produce la comunicación médico-paciente o si, por el contrario, los documentos pertenecientes a este género, en opinión de los sujetos del estudio y a partir del CI analizado, se deberían adaptar a un público lego. A este respecto, encontramos que puede haber dos aspectos por los que la comunicación podría no ser satisfactoria: por una parte, el estrictamente lingüístico (sintaxis compleja, términos médicos, poca cohesión, etc.), y el aspecto pragmático (el miedo o rechazo que puede generar en el paciente la no comprensión de algún elemento). En la Tabla 3 podemos observar las frecuencias total y porcentual de las respuestas de los sujetos:

Respuesta	Frecuencia absoluta	Porcentaje
Se produce la comunicación	7	17,5%
Se produce parcialmente la comunicación	16	40%
No se produce la comunicación	17	42,5%

Tabla 3. Valoración de los sujetos sobre la comunicación mediante el consentimiento informado utilizado en el estudio.

De entre los 7 sujetos que consideran que sí se produce la comunicación, se pueden reconocer dos tendencias. Por un lado, 5 sujetos (13% del total) consideran que no habría que adaptar el consentimiento informado a un público lego. Manifiestan que el texto es correcto y que está escrito con un registro apropiado a este tipo de comunicación y a sus convenciones. En esta línea, M18 expone lo siguiente:

*Siendo un documento oficial creo que está bien que sea así de formal y utilice ese lenguaje. El médico en consulta es quien debe ocuparse de que el paciente lo comprenda todo.*

M5, por su parte, supone que el paciente se informa previamente y que, por tanto, la adaptación no sería necesaria.

Por otro lado, M11 y H1 consideran que, si bien se produce la comunicación a su parecer habría que adaptar el texto por si acaso alguien no llegara comprenderlo en su totalidad. Ambos sujetos coinciden en la siguiente idea:

*Yo creo que sí se produce la comunicación, al fin y al cabo, porque, aunque la persona no especializada no conozca alguna de las cosas, lo que es el mensaje global sí que lo suele captar. Lo único que a lo mejor podría cambiar es que los médicos introdujeran los tecnicismos de manera más informal para que el paciente lo pueda entender mejor. [H1]*

Por el contrario, 17 sujetos (43%) consideran que por medio de este tipo de documento no se produce la comunicación. Están de acuerdo en que el tipo de lenguaje empleado no es apropiado para la comunicación entre una persona experta y una leiga, por lo que debería adaptarse. Por tanto, manifiestan que se debería producir una adaptación lingüística, entre otras cosas. Destacamos la siguiente declaración de M2 por ser representativa del pensamiento generalizado de este grupo de sujetos:

*No creo que se esté produciendo la comunicación. Es un texto de experto para explicarse y si hay algún problema recurrir a eso. En plan: tú has firmado esto, atente a las consecuencias. [...] Tienes que adaptarlo a un público general, creo.*

Por otro lado, también se ha señalado que el texto debería ser más visual y ameno, en la línea de lo señalado en la pregunta uno sobre la distribución de la información y la estructura, tal y como lo expresa M7.

*También he encontrado pesada la lectura por el hecho de que los diferentes apartados no están bien separados de una manera visual, ya que, al utilizar el mismo tipo de letras sin resalte alguno, hace que se confunda y puedas no darte cuenta de la separación de los apartados del texto.*

Además, también se ha señalado que uno de los motivos por los que no se produce la comunicación es que falta información o especificidad. H8 considera que la comunicación médico-paciente no se produce en general por la frialdad de la situación:

*Yo no veo que haya buena comunicación entre médico y usuario. O sea, para empezar, pienso que tiene que haber empatía en general, pero en estos casos bastante más. O sea, que sé que es complicado, pero tienen que empatizar con la persona.*

Finalmente, 16 sujetos (40% del total) expresan que la comunicación se produce de manera parcial o que podría producirse o no dependiendo de la situación. Algunos de estos sujetos han señalado que consideran que esto se debe a que es probable que el paciente pueda comprender algunas palabras o fragmentos y otros no, ya que, como se ha señalado anteriormente, el párrafo sobre los riesgos del procedimiento podría tener una dificultad mayor al resto del texto y coinciden en esta idea, expresada por H10:

*Algunas frases son un poco complejas y a lo mejor si le cambias en algunas palabras y lo dijeras de otra manera pues la gente lo entendería mejor; no sería tan denso de leer.*

Por otro lado, algunos consideran que la comunicación se podría llegar a producir dependiendo de la formación del paciente. Explican que, si bien podría no suponer mayor dificultad para aquellas personas que tengan formación académica, el resto podría no comprender el texto. En esta línea, H20 comenta lo siguiente:



*La comunicación se producirá mejor o peor dependiendo de la persona, pues alguien sin ningún tipo de formación sí que podría llegar a encontrarse con dificultades*

De este modo, de entre los que consideran que la comunicación se produce parcialmente, 17 sujetos (43%) manifiestan que se debería adaptar. Por su parte, 7 de ellos proponen que la adaptación debería de producirse por medio de un cambio en la situación en la que se lee el tipo de documento que nos ocupa. Sugieren que quizá lo más apropiado sería que se leyera junto con el médico y no que, como suele ser habitual, el paciente se lleve el documento para leerlo sin la presencia de personal médico-sanitario.

Por otro lado, H7 propone, de forma intuitiva, una alternativa más novedosa, que podría recordarnos a lo explicado en el marco teórico sobre el *e-consent*:

*Seguro que hay muchas formas de agilizar. Imagínate que con una animación 3D o así te dicen cómo se va a hacer, qué se va a hacer, qué riesgos tiene, la recuperación... Y todo de una forma didáctica, que no sea tan compleja.*

Por otra parte, 13 sujetos (33% del total de la muestra) piensan que la adaptación debería producirse en el tipo de lenguaje empleado en este género. Así lo expresa M17:

*A lo mejor la manera de expresarse también el texto, aunque la idea general sí que se entienda podría expresarse de otra manera como con frases más cortas o con explicaciones más sencillas, para que parezca más fácil de leer, que no sea tan pesado. Pienso que debería haber más contexto y que esté explicado más fácil.*

H18 y H20, por su parte, reconocen que la comunicación se produce parcialmente e inciden en el hecho de que las personas sin formación podrían no llegar a entenderlo todo, pero aun así manifiestan que no es necesario adaptarlo, como podemos ver a continuación:

*Sí que creo que aquella persona que no tenga ninguna formación de ningún tipo, ni siquiera [la] ESO o primaria, tendría una dificultad por la falta de la comprensión de cierto vocabulario. (...) En esa situación creo que el texto podría intimidar un poco [sobre todo] la sensación de volver a preguntar y aclarar las dudas que le hayan podido surgir al paciente, ya que habría quizá grandes porciones que no se habrían entendido (...). Sobre todo, me refiero a la parte 5 [riesgos del procedimiento], ya que esta es la que más abarca términos que una persona con una educación inferior quizá no comprenda en absoluto. Esto, no obstante, no implica que crea que se deba alterar algo del texto. [H18]*

Finalmente, nos gustaría aportar la experiencia que cuenta M9 con este tipo de género textual, cuando lo recibió en Valencia (España) antes de someterse a una operación. El CI, sin embargo, no estaba en su idioma:

*Pero sí que tengo que decir que sí que hubo un documento en concreto que leímos juntos [refiriéndose al médico] en la cita que tuvimos porque estaba relacionado con una decisión que tenía que tomar. Sí que es verdad que el texto estaba en inglés. Él me dijo «¿Sabes inglés?». Y yo le dije que sí, que no tengo un C2, pero tengo un B2.*

Resulta curioso saber que un paciente puede llegar a recibir un consentimiento informado en inglés y cabría preguntarse si bien podría suponer un problema incluso en los casos en los que

el paciente afirma entender el inglés, ya que, por motivos como vergüenza o miedo, puede no llegar a comprender el CI y no manifestarlo, como hemos visto a lo largo de las entrevistas.

## 5. DISCUSIÓN

A continuación, realizaremos la evaluación de los resultados obtenidos y propondremos las futuras líneas de investigación con respecto al CI y a la comunicación experto-lego en el ámbito médico.

Cabe destacar, antes de abordar esta evaluación, que la situación a la que se enfrenta el personal médico (por ejemplo, el poco tiempo que puede dedicar a cada paciente) dificulta una comunicación satisfactoria y, de momento, no podemos proponer una solución inmediata. No obstante, sí consideramos que la mejora de la comunicación por medio del CI podría paliar algunas carencias comunicativas.

En primer lugar, respecto a los elementos indicados por los informantes, como elemento más señalado en ambos estudios, la terminología se presenta como la mayor barrera comunicativa que presenta el CI. Especialmente, destacamos los términos empleados en el apartado «riesgos del procedimiento» por haber resultado especialmente problemáticos.

En segundo lugar, podría considerarse —tal y como han señalado algunos sujetos— que el CI no supone un problema para la comunicación debido al hecho de que el paciente puede formular preguntas. Así pues, nos parece importante destacar que un 45% de los sujetos entrevistados no tienen reparos en afirmar, por una parte, que no pregunta sus dudas (25%) y, por otra, que solo lo haría en circunstancias determinadas (20%), por lo que no parece conveniente relegar la responsabilidad de que se produzca la comunicación al paciente.

Finalmente, con respecto a la última parte, cabe destacar que un 42,5% de los sujetos considera que no se produce la comunicación en sí por medio del documento presentado a los informantes, mientras que un 40% considera que potencialmente se podría producir, dependiendo de diferentes factores. Es decir, solo un 17,5% manifiesta que la comunicación médico-paciente se produce de manera satisfactoria por medio del CI.

Algunos sujetos manifiestan que sería necesario que el documento se leyera junto con el personal médico, lo cual no siempre tiene lugar. En esta línea, algunos sujetos expresan que la presión que genera el poco tiempo en consulta puede afectar a la comunicación. No obstante, cabría preguntarse si el personal médico puede permitirse, teniendo tan poco tiempo para cada consulta médica, leer el documento con el paciente o dedicar parte de su tiempo a la comunicación.

Cabe destacar también que se ha señalado en el estudio que presentamos el hecho de que el personal médico-sanitario más joven parece mostrarse más preocupado por la comunicación con el paciente. Podríamos pensar, quizá, que la evolución en la importancia con la que se concibe la comunicación médico-paciente y, en general, experto-lego, ha supuesto que se haga más hincapié en este aspecto en la formación del personal médico-sanitario.

Si bien dicho cambio sería positivo para la cuestión que nos ocupa, algunos sujetos señalan que entienden el CI como un documento para proteger al personal médico-sanitario de posibles consecuencias legales, en lugar de para que el paciente pueda autorizar una actuación médica con la información pertinente.

Finalmente, destacamos la propuesta de algunos sujetos de adaptar el CI a un público lego por medio del uso de un lenguaje más claro y sencillo, mejorando, asimismo, la

estructura y presentación del documento. Uno de los sujetos incluso propone, en la línea con lo explicado en el marco teórico sobre los consentimientos multimodales, un tipo de consentimiento que utilice la animación 3D para exponer de forma visual y didáctica el procedimiento al que se va a someter el paciente. Vemos cómo, efectivamente, de forma intuitiva y sin conocer los últimos avances respecto al CI, algunos de los sujetos están de acuerdo con los expertos respecto al futuro del género en cuestión y de la comunicación médico-paciente.

Sobre las limitaciones encontradas, reconocemos que la extensión del presente trabajo ha constreñido el alcance de su análisis empírico, sobre todo en las circunstancias experimentadas durante el estudio relacionadas con la crisis de salud pública y el consecuente confinamiento. Así, la muestra de la que parte la investigación ha estado compuesta por personas de 18 a 30 años de la Comunidad Valenciana, por lo que las conclusiones, quizá no sean aplicables a otros grupos de edad o a personas de otras comunidades. Además, hemos trabajado únicamente con un ejemplo de CI. Por ello, tras la realización del presente trabajo, sugerimos la posibilidad de continuar con futuras líneas de investigación que trabajen con una muestra de población diferente o con otros CI correspondientes a otros procedimientos y condiciones médicas, cabiendo la posibilidad de crear un corpus con una mayor variedad de textos pertenecientes al género que nos ocupa, para obtener conclusiones más generalizables. Además, proponemos abordar las posibles soluciones a la problemática planteada, que podrían pasar por opciones como, por ejemplo, las siguientes: formar al personal médico-sanitario en la cuestión de la comunicación experto-lego, contar con profesionales de la lengua con habilidades para la transmisión del conocimiento especializado, o, como hemos mencionado, crear nuevas propuestas de CI multimodales que presenten la información de una forma accesible. En general, entendemos que la respuesta a este problema no es sencilla y requiere de investigación que aborde diferentes soluciones para paliar las necesidades comunicativas entre el personal médico-sanitario y los pacientes.

## 6. CONCLUSIONES

En este apartado, expondremos las conclusiones que podemos extraer de este trabajo y su relación con el objetivo del estudio.

A lo largo de estas páginas hemos estudiado la complejidad de los actos de comunicación médico-paciente en tanto que un caso muy particular de comunicación especializada. Para ello, hemos analizado la legibilidad del consentimiento informado y hemos conocido la experiencia y opinión de personas legas sobre el texto, el género y la comunicación médico-paciente en su conjunto. El método empleado ha sido, tal y como exponemos arriba, la lectura y subrayado de aquellos elementos no comprendidos por parte de los sujetos y la realización de entrevistas semiestructuradas con el objetivo de contestar a la siguiente pregunta: ¿Se produce la comunicación entre persona experta y persona lega por medio del CI?

Las hipótesis de las que partíamos eran que algunos elementos del texto podrían llegar a dificultar su legibilidad, especialmente la terminología empleada, por lo que algunos sujetos podrían tener la sensación de que el objetivo del documento no es informar, sino proteger al personal médico-sanitario. Además, considerábamos la posibilidad de que, si bien algunos sujetos pregunten todas sus dudas, otros no lo hagan, lo que haría un fracaso del consentimiento informado y la situación de comunicación que representa.

Tras la realización de los estudios previamente explicados, hemos podido confirmar nuestras hipótesis y concluir que existen problemas relacionados con la redacción del texto que pueden dificultar su comprensión para personas no especializadas. Además, respecto a la posibilidad de que el paciente formule todas las preguntas que le surjan la lectura del documento, hemos confirmado que una parte considerable de los sujetos no expone sus dudas. En definitiva, gran parte de los sujetos consideran, a partir de la lectura de este documento, que podría no producirse la comunicación por medio del CI.

Como hemos explicado, el proceso de adaptación debería llevarse a cabo siempre que se transmite conocimiento experto a personas legas, lo cual debe incluir —no solo por motivos jurídicos, sino también por motivos éticos— al género híbrido médico-jurídico que es el consentimiento informado. Como hemos podido comprobar, parece que tal y como está configurado en la actualidad no se puede afirmar que la comunicación se produzca siempre de manera completamente satisfactoria por medio de dicho género. Según este trabajo preliminar, parece que existe una problemática que se debería paliar desde los centros de salud e instituciones para que los pacientes pudieran tomar las decisiones que consideren oportunas habiendo sido adecuadamente informados.

## REFERENCIAS BIBLIOGRÁFICAS

- Agencia española de medicamentos y productos sanitarios (AEMPS). (Actualización de 2018). *Guía para la correcta elaboración de un modelo de hoja de información al paciente y consentimiento informado (HIP/CI)*. Departamento de medicamentos de uso humano. Anexo VIII A. <https://www.aemps.gob.es/investigacionClinica/medicamentos/docs/anexo8a-Ins-AEMPS-EC.pdf>
- Borja, A., I. García-Izquierdo y V. Montalt. (2009). Research Methodology in Specialized Genres for Translation Purposes. *The Interpreter and Translator Trainer*, 3(1), 55-77. <https://doi.org/10.1080/1750399X.2009.10798781>
- Cabré, M. T. (1993). *La terminología: teoría, metodología y aplicaciones*. Editorial Empúries.
- Cadenas, D. (2018). *El consentimiento informado y la responsabilidad médica*. BOE.
- Casado Blanco, M. (2008). *Manual de documentos médico-legales*. Consejería de Sanidad y Dependencia de la Junta de Extremadura.
- Cobos, I. (2019). Traducir para el paciente: acercamiento y adaptación como modalidad de traducción. *Quaderns de Filologia: Estudis Lingüístics*, 24, 214-215. <https://doi.org/10.7203/qqf.24.16307>
- Coughlin, C. (2015). E-consent: Can Informed Consent be Just a Click Away? *Wake forest law review*, 50, 380-393.
- Eloff, J., y K. Padayachee. (2006). An aspect-oriented implementation of e-consent to foster trust. *SAICSIT '06*, 164-169. <https://doi:10.1145/1216262.1216280>
- Gallego-Borghini, L. (2015). *La Traducción inglés-español del consentimiento informado en investigación clínica*. Fundación Dr. Antonio Esteve.
- García Izquierdo, I. (2009). *Divulgación médica y traducción: El género Información para pacientes*. Perer Lang AG.
- García-Izquierdo, I., y A. Muñoz-Miquel. (2015). Los folletos de información oncológica en contextos hospitalarios: la perspectiva de pacientes y profesionales sanitarios. *Tremédica, Pance@XVI*(42), 225.
- Harvey, M. (2002). What's so special about legal translation? *Meta: Journal des traducteurs*, 47, 177-185. <https://doi.org/10.7202/008007ar>
- Jakobson, R. (1959). On linguistic aspects of translation. En L. Venuti (Ed.), *The Translation Studies Reader*, 2, pp. 230-232. Routledge. <https://doi.org/10.4159/harvard.9780674731615.c18>
- Juste, N. (2016). *Análisis empírico-descriptivo de los géneros notariales: el caso del "poder de representación" y su traducción*. Tesis doctoral. Universitat Jaume I.

- Lehrberger, J. (1986). Sublanguage Analysis. En R. Grishman y R. Kittredge (Dir.) *Analyzing Language in Restricted Domains: Sublanguage Description and Processing* (pp. 19-38).
- Mainetti, J. (1989). Ética médica. Quirón.
- Mapelli, G. (2004). Estrategias lingüístico-discursivas de la divulgación científica. *AISPI. Actas XXII*, 165-169.
- Muñoz-Miquel, A. (2016). La traducción médica como especialidad académica: algunos rasgos definitorios. *Hermeneus*, 18.
- Sless, D. (1992). Literacy: An Ecological View. *ACAL National Conference, II*, 9-20.
- Tena, C. (2002). La demanda como efecto de la mala comunicación médico-paciente. *CONAMED*, 7(3), 16-18.
- Williams, C. (2004). Legal English and Plain language: an introduction. *ESP Across cultures*, 1(1), 111-124.
- Wright, P. (1999). Writing and information design of healthcare materials. En N. Candlin y K. Hyland (Eds.), *Writing: Texts, Processes and Practices* (pp. 85-87). Routledge. <https://doi.org/10.4324/9781315840390-5>

### **Legislación referenciada**

- (2002). *Ley 41/2002, de 14 de noviembre, básica reguladora de la autonomía del paciente y de derechos y obligaciones en materia de información y documentación clínica*. «BOE» núm. 274, de 15/11/2002. <https://www.boe.es/eli/es/l/2002/11/14/41/con>
- (2017). *LEY 8/2017, de 7 de abril, de la Generalitat, integral del reconocimiento del derecho a la identidad y a la expresión de género en la Comunitat Valenciana*. DOGV núm. 8019 de 11.04.2017. <https://www.dogv.gva.es/es/eli/es-vc/l/2017/04/07/8/>

### **Consentimiento informado del estudio**

<https://drive.google.com/file/d/1vffHSbP7CRnuApaj0RZuNWuA4cU-xOrs/view?usp=sharing>

## ¿Quién es quién? Análisis de corpus de los procesos de desfocalización del agente en español y en francés en dos géneros digitales

Who is who? A corpus-based study of agent-defocussing mechanisms in Spanish and French in two digital genres

ÉMELINE PIERRE

Université catholique de Louvain, Belgique  
emeline.pierre@uclouvain.be  
<https://orcid.org/0000-0003-0073-9633>

### Resumen

En este trabajo analizamos tres estrategias de desfocalización del agente en español y en francés, basándonos en dos géneros digitales. Pese a que estas estrategias ya han sido examinadas, carecemos de estudios empíricos realizados en corpus digitales. No obstante, los géneros digitales, por presentar características propias, son de gran interés en el estudio de los fenómenos sintácticos, especialmente en una era que se vuelve cada vez más digital. En el presente trabajo trataremos de demostrar la influencia del género y, en cierta medida, de la lengua en el uso de los participantes involucrados en tres estrategias de desfocalización: la pasiva perifrástica, la construcción con *se* – ambas disponibles en español y en francés– y el pronombre impersonal *on* en francés. Para llevar a cabo esta investigación, se han analizado datos procedentes de dos corpus: el foro

### Abstract

This paper analyses three agent-defocussing strategies in Spanish and French, based on two digital genres. Although this type of strategy has already been extensively explored, the literature lacks empirical studies on digital corpora. Nevertheless, digital genres, as they exhibit their own characteristics, are of great interest in the study of syntactic phenomena, especially in an era that is becoming increasingly digitalised. In the present paper we try to demonstrate the influence of genre and, to a certain extent, of language on the use of the participants involved in three agent-defocussing strategies: the periphrastic passive, the construction with *se*, available both in Spanish and French, and the impersonal pronoun *on* in French. In order to carry out this research, data from two corpora were analysed: the Yahoo digital forum and the online encyclopaedia Wikipedia. While the

---

**Para citar este artículo:** Pierre, E. (2022). ¿Quién es quién ? Análisis de los procesos de desfocalización del agente en español y en francés en dos géneros digitales. *ELUA*, (38), 165-185. <https://doi.org/10.14198/ELUA.22382>

Recibido: 31/03/2022 Aceptado: 20/05/2022

© 2022 Émeline Pierre



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

digital Yahoo (De Smet 2009) y la enciclopedia en línea Wikipedia (Reese et al. 2010). Mientras el corpus Yahoo, constituido por interacciones preguntas-respuestas, presenta un grado menor de formalidad y rasgos de la oralidad (Placencia 2012), la Wikipedia proporciona acceso a un lenguaje de tipo enciclopédico, y se considera muestra de un género formal. Asimismo, estos dos corpus digitales incluyen situaciones discursivas que propician el uso de una estrategia de desfocalización. Se propone, por un lado, la protección de la imagen de los interlocutores en el caso del foro digital, en el cual se plantean preguntas o se piden consejos, una situación contextual sensible tanto para el emisor de la pregunta como para el interlocutor que le contesta, y, por otro lado, un fin de objetividad en el caso de la enciclopedia en línea. El análisis ha demostrado que el género y la lengua influyen no solo en la distribución de los mecanismos y en la presencia de participantes, sino también en la categoría a la cual pertenecen esos últimos, es decir, animado (humano o no) o inanimado (organización, país/ciudad u otro). Por consiguiente, este trabajo pretende abrir vías de estudio en los mecanismos de desfocalización del agente y mostrar la importancia de tomar en cuenta factores externos al fenómeno sintáctico, como el género y el idioma de la producción textual.

**PALABRAS CLAVE:** lingüística de corpus; géneros digitales; formalidad; estrategias de desfocalización; pasiva perifrástica; construcciones con *se*; *man-impersonal on*

Yahoo corpus, consisting of question-answer interactions, exhibits a lower degree of formality and features of orality (Placencia 2012), Wikipedia provides access to specialised language, considered as a sample of a formal genre. Likewise, these two digital corpora include discursive situations prone to the use of agent-defocussing strategies. The digital forum includes situational contexts, which involve the protection of the interlocutors' image and the consequent use of an agent-defocussing mechanisms to reduce to prominence of agents while the online encyclopaedia Wikipedia would resort to this type of mechanism to attempt to achieve objectivity. The analysis has revealed that genre and language influence not only the distribution of mechanisms and the presence of participants, but also the category to which the latter belong, i.e. animate (human or not) or inanimate (organisation, country/city or other). This paper therefore aims to open up avenues of study in the mechanisms of agent-defocussing mechanisms and to show the importance of considering factors that are external to the syntactic phenomenon, as the genre and the language of the textual production.

**KEYWORDS:** corpus-based study, digital genres; formality; agent-defocussing mechanisms; participants, contrastive linguistics.

## 1. INTRODUCCIÓN

Algunos géneros digitales son más propensos a recurrir a estrategias de desfocalización del agente para omitir o disminuir la presencia de la persona que lleva a cabo una acción en el discurso, por distintos motivos. Las estrategias de desfocalización del agente disminuyen la prominencia de este último y confieren diversos grados de impersonalidad a la producción lingüística. La impersonalidad suele recibir dos lecturas (Siewierska 2008): una estructural, asociada a la ausencia total de sujeto canónico, y otra funcional, vinculada a la desfocalización del agente. Según Siewierska (2008: 7), el proceso de desfocalización puede realizarse a través de: (i) la no elaboración o la sub-elaboración del agente, (ii) la democión del agente de su posición argumental prototípica, (iii) ambos procesos. El lenguaje ha desarrollado una amplia gama de mecanismos sintácticos para desfocalizar el agente o, dicho de otro modo, para proporcionar al discurso un carácter impersonal o «faceless» en términos de Achard (2015: ix).

El presente trabajo explora tres de estas estrategias de desfocalización del agente en español y en francés: la pasiva perifrástica, la construcción con *se*, –ambas construcciones disponibles en las dos lenguas–, y el pronombre *on* en francés, que, como se explica en 2.3, se presenta a menudo como alternativa a las dos primeras construcciones. Como el español y el francés son dos lenguas romances que disponen de una pasiva perifrástica y una construcción con *se* formalmente similar, la comparación de los datos contribuye a completar la descripción del uso de estas estructuras y sus posibles variaciones según el género digital.

Aunque se ha escrito abundantemente sobre el tema de las estrategias de desfocalización del agente (ver Cabañas Maya 2006; Fløttum *et al.* 2007; Dieck 2016; Devís Márquez 2017; De Cock 2020), aún queda mucho por estudiar. De hecho, la mayoría de los estudios de los que disponemos hasta ahora, bien son de tipo descriptivo, bien carecen de base empírica o bien no contemplan la variación de género. No obstante, como se ha comprobado en una investigación anterior (Pierre 2021), el género y, por consiguiente, el grado de formalidad del lenguaje, es una variable crucial en el estudio de fenómenos sintácticos como los que nos interesan en este trabajo.

Para contribuir de manera novadora al análisis de estas estrategias, proponemos llevar a cabo un estudio de corpus basado en dos géneros digitales: el foro digital Yahoo y la enciclopedia en línea Wikipedia. Teich *et al.* (2016: 1669) sostienen que los contextos situacionales particulares tienen correlatos lingüísticos en el nivel léxico-gramatical. Centrándonos en los participantes involucrados en las estrategias, es decir, el agente cuando es elaborado y el paciente, postulamos que la distribución y las características de los participantes varían en el foro Yahoo y la Wikipedia. Estos dos géneros textuales, que presentan características diferentes, son susceptibles de contener estrategias de desfocalización, pero, *a priori*, con fines bastante distintos. Entre las situaciones discursivas que favorecen el uso de una estrategia de desfocalización, podemos plantear la protección de la imagen de los interlocutores y la objetividad. Ya se ha subrayado la importancia de las actividades de imagen y sus manifestaciones en interacciones en línea (Mancera Rueda 2015). En el foro digital se plantean preguntas o se piden consejos. Además, el foro contiene situaciones contextuales sensibles que implican tanto la imagen del emisor de la pregunta como la del internauta que le contesta (Ehrhardt 2014), y, por consiguiente, es propenso al uso de mecanismos de desfocalización del emisor. Por otro lado, postulamos que la Wikipedia, al ser una enciclopedia, utiliza más bien esos mecanismos con el fin de conservar objetividad. Ambos corpus se pueden calificar de géneros participativos: la Wikipedia es editada de manera colaborativa, mientras el foro Yahoo permite a cualquier usuario realizar preguntas y contestarlas. Por estas razones resultan de gran interés, ya que cabe suponer que todo esto influirá en la distribución de los participantes en las estrategias, tanto en el tipo de participantes incluidos como en su frecuencia de uso. Por lo tanto, en este estudio queremos vislumbrar hasta qué punto los participantes involucrados en los mecanismos de desfocalización varían según el género y la lengua.

La estructura del presente artículo es la siguiente: en la segunda sección delimitamos los fenómenos estudiados; en la tercera presentamos el tipo de datos analizados y la metodología empleada; en las secciones cuatro y cinco exponemos, de manera cuantitativa y cualitativa, los principales resultados y una discusión de los mismos. En la sexta, ofrecemos las conclusiones y proponemos perspectivas futuras de investigación.



## 2. DELIMITACIÓN DE LOS FENÓMENOS ESTUDIADOS

### 2.1. Pasiva perifrástica

El término *pasiva perifrástica* puede abarcar varias definiciones, dependiendo de la posición que adopte el investigador. Por ello, es preciso delimitar esta noción. El presente estudio considera como *pasiva perifrástica* una construcción que cumple los siguientes requisitos sintácticos, léxicos y morfológicos (basados en Jalenques 2016): (i) promoción del segundo argumento del verbo, tradicionalmente el objeto directo en el equivalente activo, en posición sintáctica de sujeto, (ii) uso de una forma finita o no finita del auxiliar *ser* en español y *être* en francés, (iii) seguida de un participio pasado que concuerda en género y número con el sujeto. Aunque se pueden distinguir varias categorías de verbos considerados como *ser/être* + participio pasado, nos centraremos en las pasivas perifrásticas prototípicas y en las impersonales, en las cuales un agente (explícito o no) realiza una acción.

### 2.2. Construcción con *se*

El mecanismo de desfocalización basado en el clítico *se* se emplea tanto en español como en francés. Varios autores se han centrado en el estudio de esta construcción y han aportado sus contribuciones respecto a la clasificación de este pronombre polivalente (Gómez Torrego 1992; Mendikoetxea 1999 o Devís Márquez 2003). No obstante, la descripción detallada de los distintos tipos de clasificación propuestos en la literatura queda fuera del alcance del presente trabajo. Por lo tanto, tomamos como referencia la categorización propuesta por Mendikoetxea (1999), quien sugiere tres construcciones con *se* con efecto despersonalizador: la pasiva (1), la media (2) y la impersonal (3). Dieck (2016) las ilustra de la siguiente manera:

- (1) *Los artículos se mandaron al editor*
- (2) *Este libro se vende bien*
- (3) *Se criticó a todos los asistentes*

[Ejemplos de Dieck (2016: 152-153)]

En las construcciones llamadas pasivas, “el objeto nocional, o paciente en términos semánticos [...] es promocionado a sujeto” (Dieck 2016: 152) y el verbo concuerda con él. El ejemplo (1) ilustra este fenómeno. La flexión verbal concuerda con el sintagma nominal plural *los artículos*. Cuando se usa de forma impersonal, el objeto nocional no es promocionado a sujeto y, por lo tanto, no concuerda con el argumento verbal, como se ilustra en (3). Las construcciones medias (2) se acercan a las construcciones pasivas, pero se diferencian de ellas en que, en las medias, “se predica una cualidad inherente al sujeto gramatical” (Dieck 2016: 153). Estas tres construcciones comparten la propiedad de incluir una cláusula verbal en la que un paciente sufre una acción realizada por un agente desfocalizado. Dado el enfoque funcional-discursivo de nuestro estudio, adoptamos una visión no discreta del fenómeno y consideramos las tres construcciones como un contínuum.

A pesar de que las construcciones en francés y en español compartan varios rasgos, ya que ambas se construyen con el clítico *se* y pueden servir para desfocalizar al agente, difieren en varios aspectos. Lamiroy (1993: 67) distingue tres diferencias mayores: (i) el uso de *se* en francés suele aparecer mayoritariamente con cláusulas nominales inanimadas en tercera persona, mientras en español se aceptarían cláusulas con sustantivos animados; (ii) el *se* francés se limita a los verbos transitivos directos, mientras que el *se* español acepta verbos transitivos e intransitivos indirectos (1993: 68); (iii) el último punto por destacar está relacionado con el dominio aspectual. En la mayoría de los casos, el francés no permite el uso de *se* con tiempos perfectos mientras que el español lo admite (1993: 69).

### 2.3. El pronombre *on* en francés

El pronombre francés *on*, que se relaciona etimológicamente con la palabra latina *homo* (Siewierska: 60), pertenece a las construcciones impersonales con referencia humana, conocidas también como *man-impersonal*. Esas construcciones ya han sido ampliamente atestigüadas en una gran variedad de lenguas europeas (ver Giacalone Ramat y Sansò 2007; Siewierska 2010; Coussé y van der Auwera 2012, por citar solo algunos). Una construcción impersonal con referencia humana se compone típicamente de un sujeto humano no referencial expresado por un término derivado de la palabra *humano*. Centrándose en el pronombre francés *on*, Fløttum et al. (2007) sintetizan las diferentes acepciones de *on* en dos categorías: indefinidas o personales, como se indica en la Tabla 1.

Indéfini	personnel				
	impersonnel	générique	spécifique	neutre	stylistique
<i>Il, ce</i>	<i>Tous hommes, chacun</i>	<i>les</i>	<i>Ils, quelqu'un</i>	<i>nous</i>	<i>Je, tu, vous, il elle, ils, elles</i>

Tabla 1. Taxonomía de los distintos tipos de *on* en francés (Fløttum et al. 2007).

En su uso indefinido, *on* puede interpretarse como impersonal (4), genérico (5) o específico (6):

(4) *On était le 15 octobre.*

(5) *On peut toujours rêver.*

(6) *On frappe à la porte.*

[Ejemplos de Fløttum et al. (2007: 30)]

Cuando se utiliza como indefinido, las propiedades pronominales típicas del pronombre desaparecen y *on* ya no se refiere a ningún otro elemento del discurso. Koenig califica este fenómeno de *discourse inertness i.e.*, ‘inercia del discurso’ (1999: 243), que hace eco del comportamiento de otros mecanismos de desfocalización del agente como el de las pasivas perifrásticas. *On* en su uso indefinido es considerado como una estructura considerablemente influenciada por el modo y el grado de formalidad del discurso (ver Fløttum et al. 2004;

Coveney 2000; Siewierska 2010), con una tendencia a asociarse con el modo oral y con los registros con un menor grado de formalidad (Pierre 2021).

El pronombre francés *on* no tiene una construcción equivalente, basada en la forma, en español. Yao (2016), sin embargo, sugiere que, según las intenciones morfosintácticas y pragmáticas del hablante, hay varios mecanismos disponibles para cumplir la función de los *man-impersonal*, entre los cuales se encuentra la construcción *se*. Además, Gaatone señala que las construcciones con el pronombre *on* indefinido y las pasivas perifrásticas suelen tener significados (parcialmente) equivalentes y, por tanto, pueden utilizarse como alternativas (1998: 243).

El presente trabajo pretende determinar hasta qué punto estas estrategias comparten rasgos comunes, centrándose específicamente en los participantes incluidos en las tres construcciones investigadas.

### 3. DATOS Y METODOLOGÍA

#### 3.1. Datos

El presente estudio está basado en datos procedentes de dos corpus: el *Yahoo-based Contrastive Corpus of Questions and Answers* (De Smet 2009) y el *Wikicorpus V.1.0* (Reese et al. 2010). Estos dos corpus tienen la particularidad de contener datos semejantes en español y en francés, lo que facilita el análisis contrastivo de este estudio. El *Yahoo-based Contrastive Corpus of Questions and Answers* ha sido compilado a partir de preguntas y respuestas del foro digital *Yahoo* publicadas por los usuarios entre 2006 y 2009. El corpus contiene aproximadamente 90.000 preguntas y respuestas, e incluye datos sobre una amplia variedad de temas (por ejemplo, la ciencia, la salud, los viajes, los animales etc.). La selección de preguntas y respuestas ha sido realizada de manera aleatoria. Placencia (2012: 286) postula que, dadas las características específicas de este tipo de producción textual, se puede situar el *Yahoo-based Contrastive Corpus of Questions and Answers* en la frontera de la interacción dialogal, por estar constituido de interacciones preguntas-respuestas, y de la comunicación de masas, dado que, por ser publicado en la web y accesible libremente, el contenido, y más precisamente la información y los consejos incluidos en las respuestas, son de cierta manera destinados al gran público.

El *Wikicorpus v.1.0* contiene más de 750 millones de palabras recogidas a partir de una copia de datos realizada en 2006 de la enciclopedia en línea Wikipedia. Proporciona acceso a un lenguaje especializado para una amplia variedad de temas. El contenido de esta fuente de información digital es libre y editado de manera colaborativa, es decir, ha sido creado por varias personas de manera voluntaria (Obregón Sierra y González Fernández 2019). Ya se ha debatido sobre la calidad informativa de los contenidos (Tramullas 2015; Cusot y Klein 2017), pero nos limitaremos a utilizar la Wikipedia como corpus textual y objeto de investigación científica que permite el análisis de la producción de lenguaje en un género digital de tipo colaborativo. Teniendo en cuenta los rasgos de formalidad propuestos por Briz (2010), consideramos los datos de *Yahoo-based Contrastive Corpus of Questions and Answers* como muestra de un género informal y los datos de Wikipedia pertenecientes a un registro formal.

Para llevar a cabo esta investigación, dado el tamaño total del *Yahoo-based Contrastive Corpus of Questions and Answers* y del *Wikicorpus*, se han crea-

do dos subcorpus (denominados, a continuación, Yahoo y Wikipedia) y se han seleccionado muestras de textos abarcando cuatro de los temas disponibles en los dos corpus, tanto en español como en francés: aviones, perros, botánica y hábitos alimenticios. La inclusión de temas semejantes, en proporción similar, en los cuatros subcorpus garantiza la fiabilidad en el análisis contrastivo de este estudio. El número de palabras de las muestras analizadas en esta investigación se recoge en la Tabla 2.

Lengua	Español		Francés	
Corpus	Yahoo	Wikipedia	Yahoo	Wikipedia
Nro. de palabras	46603	47523	45194	46162

Tabla 2. Número de palabras de los subcorpus.

Cabe tener en cuenta que el carácter participativo de estos dos corpus nos impide controlar la variedad del español que contienen. Por lo tanto, se puede asumir que incluyen español peninsular, pero también variedades latinoamericanas. Rasson (2017), sobre el pronombre indefinido español *uno*, otro mecanismo utilizado para desfocalizar el agente, subraya que, si bien la variedad regional tiene un impacto directo en la frecuencia de uso del *uno*, su influencia en los diferentes tipos interpretativos que puede recibir el pronombre es relativamente limitada. Siguiendo esta línea, se asume que la variedad no tiene un impacto significativo en el tipo de datos tratados en este estudio. Dado que el objetivo principal no es examinar la frecuencia y el uso de las estructuras de desfocalización centrándose en la variación geográfica, esta característica no se tendrá en cuenta.

Una vez recopilados los cuatro subcorpus, se extrajeron las ocurrencias de cada una de las estructuras que nos interesan. Este procedimiento cuenta con dos etapas: primero, se procedió a una extracción automática a través del sistema de procesamiento de corpus UNITEX; este programa ofrece una herramienta de búsqueda de textos que permite la recuperación automática de fenómenos lingüísticos en recursos electrónicos. A continuación de la extracción automática, se llevó a cabo una revisión manual de las ocurrencias, que permitió desambiguar ciertas formas y eliminar datos irrelevantes (por ejemplo, casos de *se* reflexivo, o de pasivas adjetivales).

En la Tabla 3 se incluye el número total de ocurrencias de cada estructura encontradas en los cuatro subcorpus. Estas ocurrencias forman el material objeto de estudio de la presente investigación.

	Español		Francés	
	Wikipedia	Yahoo	Wikipedia	Yahoo
Pasiva perifrástica	213	53	416	171
Construcción con <i>se</i>	337	279	46	19
<i>Man-impersonal on</i>	-	-	98	213

Tabla 3. Repartición de los datos en el corpus Wikipedia y el Yahoo.

Como se detallará más adelante, algunas diferencias significativas emergen en cuanto a la distribución de las estructuras en los dos géneros y en las dos lenguas.

### 3.2. Procedimiento de análisis

Para comprobar nuestra hipótesis, es decir, si el contexto de producción influye en la repartición y las características de los participantes involucrados en los mecanismos de despersonalización, hemos empleado la siguiente metodología. Primero, hemos examinado el número de participantes incluido en el sintagma nominal. Quirk et al. identifican la noción de participante de la siguiente manera: “Every clause describes a situation in which a number of participants are involved. By PARTICIPANTS, we understand entities realized by noun phrases, whether such entities are concrete or abstract” (1985: 740). A partir de esta definición, y siguiendo el trabajo de Hopper y Thompson (1980), consideraremos como participantes a los agentes y a los pacientes. El agente se entiende en este trabajo como el actor, instigador o iniciador (Siewierska 2008: 7) de la acción o del evento y el paciente como receptor de la acción. Siguiendo la postura adoptada por Vázquez Rozas (2006), además del objeto directo también consideraremos los objetos indirectos como participantes.

A continuación, se anotaron semánticamente los diferentes tipos de participantes. Por ello, se diseñó una clasificación tripartita basada en Fauconnier (2012). Según esa taxonomía, un participante puede ser (i) ausente (como por ejemplo en el caso de una pasiva perifrástica sin agente expresado de manera explícita), (ii) animado (humano o no) o (iii) inanimado (organización, país/ciudad u otro). La anotación ha sido realizada manualmente.

## 4. PRESENTACIÓN DE LOS RESULTADOS

En este apartado, comenzamos con una discusión sobre la distribución de las estructuras en los dos géneros digitales. A continuación, examinamos el número de participantes (4.1) y estudiamos la semántica del agente (4.2) y del paciente (4.3). Antes de adentrarnos en el estudio de los participantes, parece oportuno examinar la repartición de las estructuras en Wikipedia y Yahoo.

Los datos del Gráfico 1 muestran que, en francés, los textos producidos en el marco de la Wikipedia contienen un mayor número de pasivas perifrásticas (frecuencia absoluta de 90,1 por diez mil palabras) y de construcciones con *se* (frecuencia absoluta de 9,9) en comparación con las interacciones recogidas en Yahoo (37,8 y 4,2 respectivamente).

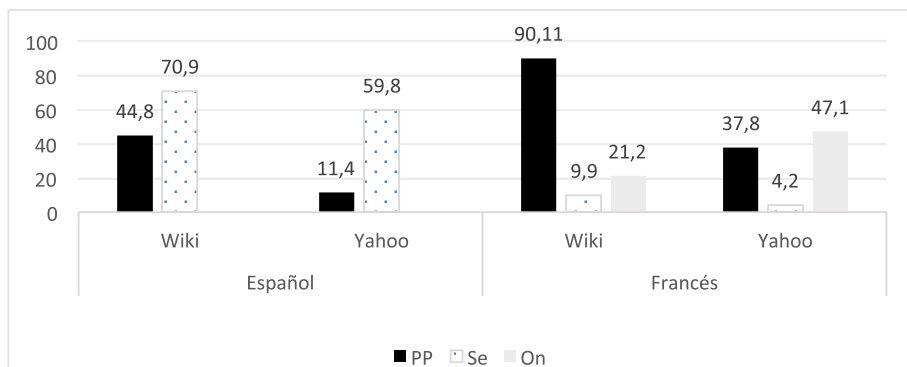


Gráfico 1. Número de ocurrencias de las estrategias en la Wikipedia y el Yahoo (por diez mil palabras).

Pese a este uso más frecuente en la Wikipedia, cabe mencionar que la proporción de pasivas perifrásticas y de construcciones con *se* es relativamente semejante en ambos corpus. Así, se usa aproximadamente diez veces más pasivas perifrásticas que construcciones con *se*, sea en la enciclopedia en línea o en el foro digital. Mientras que las dos estructuras que acabamos de comentar se usan con mayor frecuencia en los textos de la Wikipedia, el *man-impersonal on* parece más bien típico de Yahoo, considerado como un género con más rasgos de la coloquialidad.

Veamos ahora la distribución de las estructuras en español: se puede constatar que el uso de la construcción con *se* domina en ambos corpus, lo que no era el caso en francés. Este resultado corrobora varios estudios sobre la construcción con *se* que presentan esta construcción como el mecanismo más frecuente para desfocalizar el agente en la lengua española, suplantando otros mecanismos impersonales como, por ejemplo, la pasiva perifrástica (Calvo Montoro 1983; Quesada 1999; Mendikoetxea 1999; Cabañas Maya 2006; Laslop y Díaz 2010). La gran versatilidad de la estructura puede ser un factor que contribuya a su uso especialmente productivo. Así, aunque existe una tendencia generalizada a vincular las construcciones con *se* con géneros más formales y distanciadores (Jisa y Tolchinsky 2009; De Cock 2020), el mecanismo aparece también en gran medida en las situaciones informales (Quesada 1999; De Cock 2020; Pierre 2021), lo que coincide con nuestros resultados. Cabe añadir que nuestros resultados muestran una proporción de construcciones con *se* relativamente estable entre los dos géneros: 70,9 en la Wikipedia y 59,8 en Yahoo. Así, pese a la tendencia a asociar el *se* con un mayor grado de formalidad, nuestros resultados parecen indicar un efecto de género bastante limitado respecto a esa estructura. Por último, conviene destacar la mayor presencia de pasivas perifrásticas en la Wikipedia, fenómeno ya observado en francés. La marcación estilística de esta estructura y su uso para cumplir una serie de requisitos del discurso formal (Weiner & Labov 1983; Siewierska 1984) pueden contribuir a explicar el uso mayor de pasivas perifrásticas en este género.

A modo de conclusión intermedia, hemos visto que el efecto de género varía no solo en cuanto a la estructura, sino también en cuanto a la lengua, a pesar de que el español y el francés son lenguas afines que disponen de mecanismos de desfocalización semejantes. En el apartado siguiente, examinaremos si el número de participantes y sus características semánticas pueden explicar este fenómeno de variación.

#### 4.1. Participantes

En esta sección nos centramos en el estudio de los participantes que, como se ha mencionado en la metodología, se dividen entre agentes y pacientes. Empezamos por estudiar la variación del número de participantes en las diferentes estructuras, los dos géneros y las dos lenguas antes de examinar su semántica.

En cuanto al análisis de los datos en español, como se puede apreciar en el Gráfico 2, la pasiva perifrástica siempre va acompañada de dos participantes: el que hace la acción (desfocalizado en el ejemplo (7) o no desfocalizado en el ejemplo (8)) y el que sufre la acción. En (7), la oración consta de un paciente, el avión Alca, y un agente desfocalizado que realiza la acción de utilizar el Alca<sup>1</sup>:

---

1 Todos los ejemplos aparecen en su forma original y, por tanto, pueden contener fallos gramaticales u ortográficos.

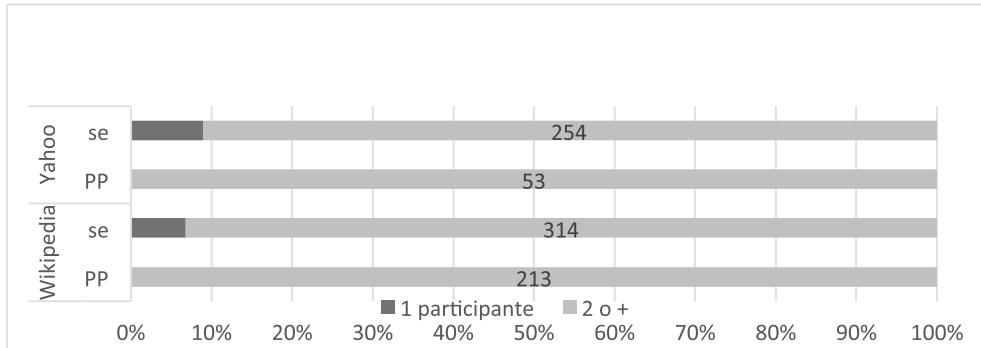


Gráfico 2. Número de participantes en la pasiva perifrástica y la construcción con *se* en español.

- (7) *Debido a su alta capacidad de maniobra al volar a bajas velocidades, el Alca puede ser **utilizado** también contra helicópteros.* (WIKI ESP)
- (8) *La nueva aerolínea **fue guiada por** el último vicepresidente y jefe de operaciones de Pan Am, Marty Shugrue, quien también ayudo en la creación del programa de viajero frecuente WorldPass [...].* (WIKI ESP)

En (8), además del paciente, la pasiva perifrástica incluye un agente explícito, expresado en el complemento proposicional *por el último vicepresidente y jefe de operaciones*.

Ese uso exclusivo de oraciones con dos participantes se explica por la arquitectura sintáctica de la estructura: como en la pasiva perifrástica el paciente se promueve en la posición que suele ocupar el agente (es decir, la posición del sujeto), en la mayoría de las pasivas perifrásticas, hay, al menos, un paciente y un agente.

Del mismo modo, la arquitectura específica de la construcción con *se* que sitúa un objeto en posición de sujeto es un factor que contribuye a explicar el predominio de las cláusulas con dos participantes. Aunque prevalecen las oraciones con dos participantes, también se observan construcciones con un participante, lo que no ocurría con la pasiva perifrástica.

Una breve revisión cualitativa de las construcciones con *se* con un participante indica que se construyen con una cláusula subordinada, como en (9). En este estudio, dado que no se trata de sintagmas nominales, las cláusulas subordinadas no se consideran participantes:

- (9) *Pero mi pregunta es, si ya tienen una licencia de piloto de aerolínea o comercial, porque tienen que hacer el curso si **se supone** que ya saben pilotar un avión?* (YAHOO ESP)

Por tanto, el ejemplo (9), construido con el verbo *se supone que*, solo implica a un participante, el que realiza la acción de *suponer*. Se puede observar en el Gráfico 2 que este tipo de construcción aparece con mayor frecuencia en el foro digital. La diferencia, sin embargo, no es significativa ( $\chi^2 = 0.9689$ ,  $p < .05$ ). Por la arquitectura sintáctica de estas dos construcciones, encontramos escasa variación entre los géneros respecto al número de participantes en las estructuras.

El panorama es muy similar en el caso del francés (Gráfico 3). De hecho, predominan las oraciones con dos o más participantes, tanto en el caso de la pasiva perifrástica como en el de la construcción *se*. Una diferencia mayor entre las dos lenguas es la utilización de la

pasiva perifrástica con un solo participante, fenómeno que no se produce en el español. Las oraciones con un participante parecen ocurrir exclusivamente cuando la pasiva perifrástica se construye con el expletivo *il*, ilustrado en (10), donde el único participante en la cláusula es el agente desfocalizado que realiza la acción, en este caso el verbo *noter* ('señalar').

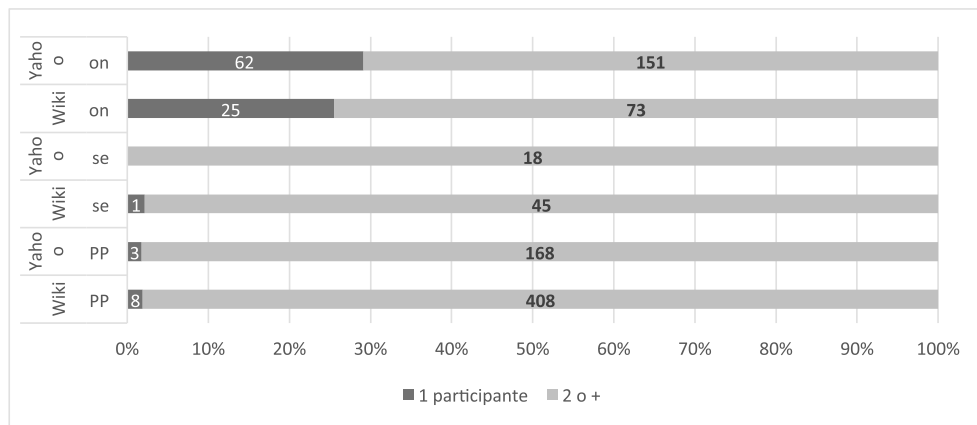


Gráfico 3. Número de participantes en las tres estrategias de desfocalización en francés.

- (10) *Il doit cependant être noté qu'une étude japonnaise a montré l'existence d'une corrélation entre la consommation accrue de certains produits animaux et la baisse de certaines formes de maladies cérébrovasculaires (...)* (WIKI FR)

Este fenómeno, sin embargo, es poco común y, aunque se podría haber supuesto que este giro es específico del registro más formal, los resultados muestran que su proporción de uso es similar en ambos géneros (aproximadamente un 2% de las pasivas perifrásticas).

Mientras la pasiva perifrástica se utiliza ampliamente en francés, la construcción con *se* es mucho menos productiva (65 ocurrencias en este estudio) y, por lo tanto, es difícil proporcionar tendencias generales al respecto. No obstante, en nuestros datos parece que se utiliza principalmente con dos o más participantes, como en el caso del español.

En comparación con las dos estructuras analizadas hasta ahora, la construcción con el *man-impersonal on* en francés presenta una mayor flexibilidad en cuanto al número de participantes. Los resultados muestran que esta estrategia de desfocalización del agente es la que acepta la mayor proporción de oraciones con un solo participante, como se ilustra en el ejemplo (11):

- (11) *Pour ce qui est des humains...je pense que c'est la même chose on dit que quand on aime quelqu'un on doit accepter de le laisser partir...c'est un peu le problème de l'euthanasie* (YAHOO FR)
- (12) *On gère le risque que l'on a le plus de « chances » de voir se produire.}S} Ainsi, la neige est présente chaque hiver en abondance dans les pays nordiques, ils sont donc équipés pour [...].* (YAHOO FR)

El ejemplo (11) cuenta con un participante, el agente desfocalizado mediante el uso del *man-impersonal on* que realiza la acción de *dit* 'dice'. Comparemos con el ejemplo (12),



que cuenta con dos participantes, el *on*, que actúa como agente y *le risque* ‘el riesgo’, en el papel semántico del paciente.

Por otro lado, se observa que hay un número ligeramente superior de oraciones con un solo participante en el corpus Yahoo. Para explicar este fenómeno, postulamos que las interacciones del foro digital, en el que se intercambian consejos, informaciones, opiniones, etc., es un género que se aproxima a la conversación (por ejemplo, el uso de marcadores orales (Placencia 2010, p. 10) o la estructura dialogal, es decir, la toma de turnos conversacional). En las conversaciones, los interlocutores a menudo aluden a lo que han escuchado, usando verbos como *dire*, *penser*, *observer*, etc. Estos verbos propician el uso de una proposición subordinada y a la formulación de oraciones con un participante. Aunque se puede apreciar una diferencia de frecuencia de uso entre el corpus Yahoo y la Wikipedia en el Gráfico 2, no es significativa en nuestros datos ( $\chi^2 = 0,511406$ ,  $p < 0,05$ ). Un estudio centrado en el *on* que incluya más datos podría, sin duda, aportar más aclaraciones sobre este fenómeno.

A modo de resumen, la mayoría de las oraciones incluyen dos participantes, tanto en español como en francés. La diferencia principal entre las dos lenguas reside en la presencia de oraciones con un participante en la perifrástica pasiva en francés, que no se da en español. Mientras que la frecuencia de uso entre géneros varía, como se ha observado anteriormente, el número de participantes, vinculado con la arquitectura del propio mecanismo, muestra poca variación entre registros. Se sospecha, sin embargo, que existe una influencia de género en el caso del *on* francés, una estructura mucho más flexible que la pasiva perifrástica y la construcción con *se*. Esa variación estaría vinculada con el tipo de verbos asociados con cada uno de los géneros.

Con el fin de precisar nuestra comprensión del uso de participantes involucrados en las estrategias de desfocalización, estudiaremos a continuación la semántica de los participantes centrándonos primero en el agente (4.2) y, a continuación, en el paciente (4.3).

#### 4.2. Análisis de los agentes

Esta sección se centra en la pasiva perifrástica, que permite mucha más variación en cuanto a la presencia de agente. Efectivamente, el *man-impersonal on* en francés siempre implica un agente humano animado en posición sujeto en nuestros datos. Del mismo modo, como se suele señalar en la biografía (ver Gómez Torrego 1992; Cabañas Maya 2006; Fernández 2007; Pierre 2021), la construcción con *se* incluye raramente un agente expresado. Por lo tanto, hay poca variación posible en cuanto al tipo de agente para estas dos estructuras. En cuanto a la presencia de un agente, el francés presenta una tendencia al equilibrio entre los dos géneros, mientras que el español se caracteriza por una tendencia a usar más pasivas perifrásticas largas, es decir, con agentes, en la Wikipedia (Gráfico 4).

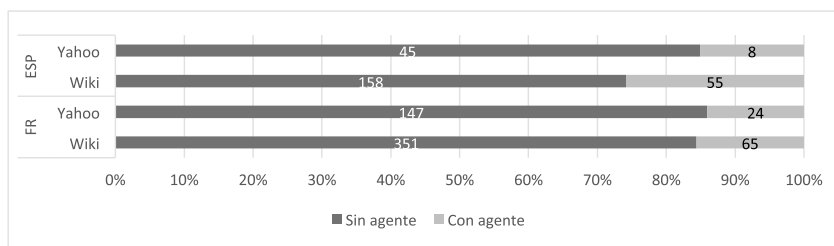


Gráfico 4. Repartición de pasivas perifrásticas con y sin agentes.

Comparemos (13) y (14): el ejemplo (13) ilustra un caso de pasiva perifrástica sin agente, mientras (14) ejemplifica un caso de pasiva perifrástica larga, con un agente explícito introducido por la preposición *por*.

- (13) *Un pitbull tambn seria una buena eleccion, es un perro atleta, **fue creado** para eso, es muy amigable con las personas y con otros perros, (...) (YAHOO ESP).*
- (14) *El Anóbido y el Bostrícido atacan a las ramas muertas de la madera húmeda pero también atacaran a la madera no tratada. El daño **es empeorado por** los adultos que emergen reinfectando el mismo trozo de madera. (WIKI ESP)*

Se observa entonces una tendencia hacia la explicitación de la persona que realiza la acción en la Wikipedia. Una explicación de esta tendencia podría residir en el hecho de que el lenguaje formal tiende a ser más específico y complejo, mientras que en las interacciones del foro se suelen utilizar mecanismos menos complejos sintácticamente. Así, se puede observar que la Wikipedia se acerca, a este respecto, a la formalidad de la enciclopedia clásica. Dado que esta tendencia a especificar el agente no se observa en francés, también podría significar que esta lengua utiliza un método diferente para especificar el agente en este tipo de texto (p.ej. proporciona información mediante el uso de una preposición de lugar o tiempo) o simplemente lo especifica menos que el español.

Analicemos ahora las tendencias respecto a los tipos de agentes en el corpus Yahoo y en la Wikipedia. En las dos lenguas, se puede observar que, en Yahoo, aunque las cifras son limitadas (ver Tabla 4), predomina la categoría de agente *inanimado otro*, como se ha ilustrado en (15) con el sintagma nominal *el suelo o sustrato en el que reside* y en (16), en francés, con el sintagma *un moteur*.

- (15) *Te aconsejo que sí. Cualquier resto vegetal sufre virulentos ataques de lombrices, bacterias, hongos y otros descomponedores hasta un nivel que **es asimiliado** por el suelo o sustrato en el que reside. (YAHOO ESP)*
- (16) *En combinant des pièces déjà existantes \_ l'engin **est propulsé** par un moteur ordinaire. (YAHOO FR)*

		Francés		Español	
		Wikipedia	Yahoo	Wikipedia	Yahoo
<b>Animado</b>	Humano	15	5	23	0
	No humano	2	0	7	1
<b>Inanimado</b>	Organización	12	4	7	1
	País/Ciudad	2	0	1	0
	Inanimado otro	34	15	17	6
	Total	65	24	55	8

Tabla 4. Clases de agentes en las pasivas perifrásticas en español y en francés.

Dado este predominio de agentes inanimados, se puede deducir que, cuando el agente es animado, los interlocutores prefieren omitirlo usando perifrásticas cortas u otras estrategias

de desfocalización como, por ejemplo, la tercera persona del plural no anafórica. Esta estrategia permite mantener un agente animado en posición sujeto a la vez que se realiza el proceso de desfocalización.

Las tendencias generales que se observan en la Wikipedia son las siguientes: se registra un mayor número de agentes humanos en español que en francés. El ejemplo (17) ilustra un uso de agente humano en español. En (17) se indica que fue el *inversionista de Nueva York Richard Hoyt* quién estableció la compañía:

- (17) *La compañía Atlantic, Gulf and Caribbean Airways fue establecida el 11 de octubre de 1927 por el inversionista de Nueva York Richard Hoyt, quien actuó como presidente. (WIKI ESP)*

Este es el tipo de agente más frecuente en español, junto con la categoría *inanimado otro*. Esta tendencia a incluir menos agentes humanos en francés puede ser compensada por el hecho de que este idioma cuenta con el mecanismo *on* con referencia humana, que, aunque se asocia a un registro más informal, está ampliamente presente en contextos considerados más formales, como los textos producidos en el marco de la enciclopedia en línea.

#### 4.3. Análisis de los pacientes

Ahora que se ha establecido la descripción general de los agentes, procedemos a analizar los pacientes.

Veamos primero la repartición de los pacientes en la pasiva perifrástica: como muestra el Gráfico 5, la categoría de paciente *inanimado otro*, predomina en los dos géneros, tanto en español como en francés. El sintagma nominal *los usos de la madera* (ejemplo 18) y *les suppléments de B12 tels que les pilules de vitamines* (ejemplo 19) ilustran esa categoría:

- (18) *En la actualidad y desde principios de la revolución industrial muchos de los usos de la madera han sido cubiertos por metales o plásticos, [...] (WIKI SP).*

- (19) *Les suppléments de B12 tels que les pilules de vitamines sont souvent préparés à partir de déchets d'abattoir et ne conviennent donc pas aux végétaliens (...). (WIKI FR)*

A pesar de esta semejanza, el tipo de paciente varía según el género y la lengua en los aspectos desarrollados a continuación.

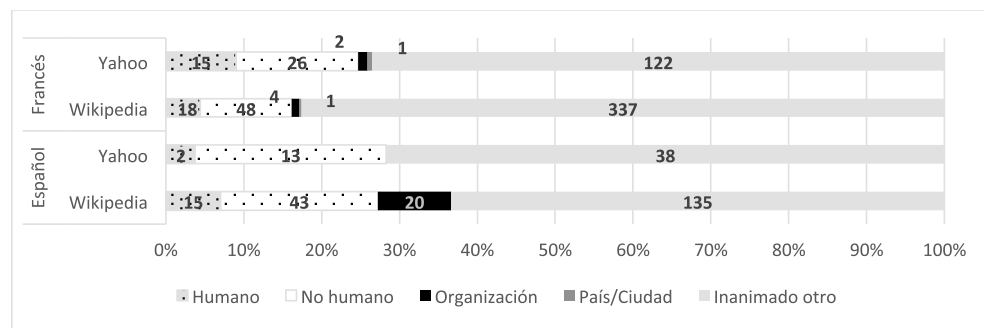


Gráfico 5. Repartición de los pacientes en la pasiva perifrástica.

En francés, se utiliza un mayor número de pacientes humanos en el foro Yahoo que en la Wikipedia (significativo  $\chi^2 = .030933$ ,  $p < 0,05$ ). Esto significa que, en las interacciones del foro, hay una mayor tendencia a individualizar a los pacientes, mientras que, en la Wikipedia, los seres humanos y, más generalmente, los seres animados, desempeñan con menor frecuencia este papel semántico. Esto puede explicarse por el mayor grado de inter-subjetividad que existe en el foro (Rasson 2016). Sin embargo, esta mayor frecuencia de pacientes animados no se observa en los datos españoles. En estos últimos, cabe destacar la mayor presencia de organizaciones en el papel de pacientes, como se ilustra en (20):

(20) *Pan American Airways Incorporated fue fundada el 14 de marzo de 1927, por Henry H. «Hap» Arnold y asociados.* (WIKI ESP)

No se observa este uso de organizaciones como la compañía *Pan American Airways* (20) en el foro digital. Las organizaciones pueden estar, hasta cierto punto, relacionadas con colectivos de personas y, por lo tanto, están conectadas con pacientes humanos. Como se mostrará más adelante, tal número elevado de pacientes clasificados como *organización* no se encuentra en los otros mecanismos de desfocalización.

En las construcciones con *se*, se aprecia un predominio de pacientes de tipo *inanimado otro*, sea en francés o en español (ver Gráfico 6). Un paciente de tipo *inanimado otro* puede ser, por ejemplo, *sulfato sódico*, como en (21), o *le moteur* en (22):

(21) *Después, en ebullición, se añade sulfato sódico que posteriormente pasa a sulfuro sódico y se elimina.* (WIKI ESP)

(22) *Le moteur se lance à la main, même si beaucoup sont aujourd'hui équipés d'un démarreur électrique.* (WIKI FR)

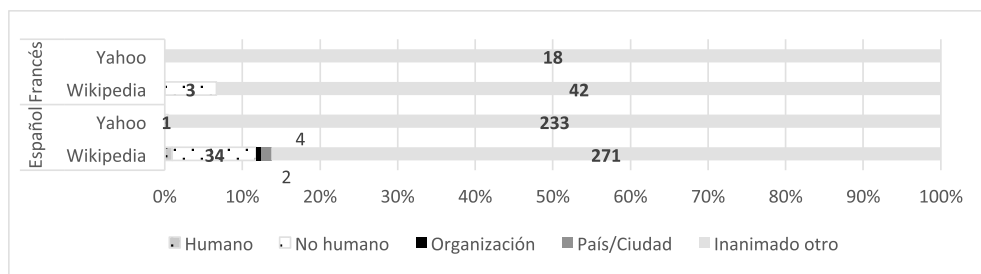


Gráfico 6. Repartición de los pacientes en la construcción con *se*.

No obstante, si nos fijamos en el español, que cuenta con un *se* mucho más productivo que el del francés, se puede observar una mayor variación en la Wikipedia, donde se nota el uso de animados no humanos, es decir, en nuestro estudio, animales. El ejemplo (23) ilustra este tipo de paciente con *el Can de los Fenicios*, una antigua raza de perros:

(23) *Dentro de las 20 razas antiguas, el Can de los Fenicios puede ser el perro más evolucionado e inteligente que se conoce.* (WIKI ESP)

El uso de estos pacientes no humanos no se encuentra en Yahoo, que, sin embargo, también incluye un conjunto de datos que tratan el tema de los animales. Por lo tanto, se puede suponer que, en los registros más informales, se utiliza un tipo de mecanismo de desfocalización diferente cuando el paciente es un animal, o que el animal no recibe el papel de paciente. Con respecto a este último, sería a favor, por ejemplo, de una posición sujeto, dándole al animal el papel de actor en vez de receptor. Así, el foro Yahoo como género digital informal sería más personalizado, individualizado, mientras que un género formal como la Wikipedia abordaría el tema de forma más distanciada.

Analicemos ahora los pacientes usados en la construcción impersonal con referencia humana, el *man-impersonal on* en francés. Como se ha comentado anteriormente, el *man-impersonal on* es una estructura que se suele considerar característica de los registros informales. Sin embargo, como muestran nuestros resultados, también se utiliza a menudo en la enciclopedia en línea, aunque esta estructura es dos veces menos productiva en este género. Es de interés constatar que, a pesar de su tendencia a utilizarse en situaciones con menor grado de formalidad, no hay influencia de género en el tipo de paciente incluido en la oración (ver Gráfico 7), a diferencia de las dos estructuras anteriormente analizadas.

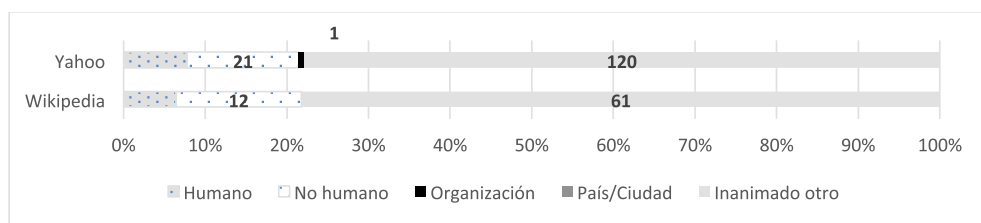


Gráfico 7. Repartición de los pacientes en el *on* francés.

En ambos géneros se observa el siguiente comportamiento: una prevalencia de pacientes clasificados como otros inanimados, ejemplificados en (24) con el sustantivo *une estimation* ('una estimación'), seguidos de pacientes de tipo *inanimado* no humano, como los pájaros en el ejemplo (25) (13,6%, 21 ocurrencias, en Yahoo y 15,4%, 12 ocurrencias, en la Wikipedia) y, por último, pacientes humanos (7,8% en Yahoo y 6,4% en la Wikipedia) ilustrado en (26) con el pronombre personal *vous* que refiere al interlocutor:

(24) *Avec la main d'oeuvre, on aura une estimation de X millions de Dollars pour remettre l'avion en état. (Yahoo FR)*

(25) *On rencontre des oiseaux jusqu'à 6000 mètres parfois plus... et à cette altitude, en plus ce sont de très gros... (Yahoo FR)*

(26) *si un jour on vous demande de signer une pétition signer le se la peut peut être changer quelque chose. (Yahoo FR)*

La bibliografía menciona el *on* como posible recurso para traducir el *se* español. No obstante, conviene señalar que existe una notable diferencia en la distribución de los pacientes utilizados en estas dos estructuras: mientras que la construcción con *se* tiende a utilizarse

casi exclusivamente con pacientes inanimados, especialmente en Yahoo, se encuentra un panorama más matizado con el *on* francés. Así, entre otros, además de una frecuencia mayor de pacientes animados no humanos, se observa, en las oraciones con *on*, la presencia de pacientes humanos, casi ausentes en la construcción con *se*.

## 5. DISCUSIÓN

Los resultados del análisis de tres estructuras de desfocalización del agente en español y en francés destacan la influencia del género en la productividad y en el tipo de participante involucrado en dichos mecanismos. El foro digital y la enciclopedia en línea presentan características específicas que influyen en las estructuras en cuanto a su productividad, el número de participantes, la presencia de un agente expresado en la pasiva perifrástica el tipo de participantes involucrados en la oración.

Los resultados han mostrado, en línea con los estudios sobre el tema (Fernandez 2007; Shibatani 1985; Siewierska 1984), que, de manera general, la pasiva perifrástica cuenta con pocos agentes expresados. Dado su papel de desfocalización, parece de interés comentar las situaciones en las que sí aparece el agente. Según Shibatani (1985), cuando se expresa el agente en la pasiva perifrástica, la ocurrencia se aleja de la instancia prototípica de la pasiva. Las pasivas perifrásticas con agentes explícitos se denominan a veces *non-basic* (Keenan 1975), en oposición a las pasivas sin agente consideradas como el prototipo o la pasiva perifrástica básica. Aunque el paciente está en posición tópica y recibe más atención que en una oración activa, la presencia de un agente explícito impide que el proceso de desfocalización se considere completo. La prominencia del paciente es, por tanto, menor en comparación con una pasiva perifrástica sin agente prototípica. Cuando se expresa el agente, las aportaciones en el campo de la agentividad indican que se suele tratar de una figura conocida o de una persona que ocupa un cargo específico. También se señala que el agente tiende a expresarse en situaciones presentadas como negativas (Davison 1980; Fernández 2007). Los ejemplos (14) y (17) de nuestro estudio, reproducidos a continuación en (27) y (28) para la facilidad del lector, ilustran estos fenómenos:

(27) *La compañía Atlantic, Gulf and Caribbean Airways fue establecida el 11 de octubre de 1927 por el inversionista de Nueva York Richard Hoyt, quien actuó como presidente. (WIKI ESP)*

(28) *El Anóbido y el Bostrícido atacan a las ramas muertas de la madera húmeda pero también atacaran a la madera no tratada. El daño es empeorado por los adultos que emergen reinfectando el mismo trozo de madera. (WIKI ESP)*

En (27), *Richard Hoyt*, un inversionista de Nueva York, actuó como presidente de la aerolínea, ocupando así una posición laboral importante. En (28), la situación puede considerarse adversativa, debido a la semántica del verbo incluido en la pasiva perifrástica, es decir, *empeorado*. Una aproximación cualitativa a los datos confirma que la mayoría de los agentes explícitos aparecen en una de estas dos situaciones.

Los resultados han mostrado que esta tendencia a especificar los agentes se da particularmente en la Wikipedia. Este fenómeno puede explicarse, en parte, por las características de este género que, al ser una enciclopedia, fuente reconocida de informaciones, requiere mayor precisión.

En cambio, el foro digital Yahoo es una plataforma en la que los internautas piden consejo e intercambian información, entre otras cosas, de manera informal. Cabe suponer que se trata de un contexto menos propicio para la explicitación del agente, especialmente en situaciones negativas. Se pueden avanzar motivos de atenuación del compromiso del hablante o de protección de la imagen del emisor y del interlocutor como posibles explicaciones a esa presencia menor de agentes. Convendría analizar, sin embargo, otros temas, considerados más controvertidos (por ejemplo, la política o la religión), para observar si se obtienen resultados semejantes. Esto permitiría destacar el impacto del género, por un lado, y del tema tratado, por otro. Otro factor que contribuiría a explicar la presencia mayor de agentes en la Wikipedia es la complejidad sintáctica que se puede atribuir a los géneros de nuestro estudio. En efecto, en una investigación anterior, se mostró que, en los registros informales y en el modo oral, los mecanismos de desfocalización presentan un menor (o distinto) grado de complejidad sintáctica (Pierre 2021). Yahoo es un género híbrido que presenta no solo rasgos del escrito, sino también de la oralidad (Placencia 2010, p. 9) y características del registro coloquial. Por lo tanto, cabe de esperar una complejidad sintáctica diferente, con, entre otros, el uso de pasivas perifrásticas cortas.

Con respecto a los pacientes, los resultados indican un predominio de los pacientes de tipo *inanimado otro* con distintos grados de variación de las otras categorías de paciente según el género y la lengua. El foro digital Yahoo español contiene más pacientes humanos involucrados en las pasivas perifrásticas, mientras la Wikipedia contiene más pacientes de tipo *organización*. Estos resultados muestran que cada género y cada idioma presentan sus peculiaridades, que varían según los objetivos del género (por ejemplo, presentación de información en el caso de la Wikipedia o intercambio de consejos en situaciones que implican varios individuos en el caso de Yahoo), pero también según las características de la lengua y de los recursos de los que dicha lengua dispone para llevar a cabo esos objetivos.

Terminamos la presente discusión con una reflexión sobre el pronombre impersonal *on*: el estudio ha mostrado que los pacientes involucrados en oraciones que llevan a cabo el proceso de despersonalización con el pronombre *on* presentan una escasa variación entre los dos géneros. Una posible explicación reside en el trabajo de Creissel (2008: 13), quien demuestra que, en su acepción impersonal, *on* ha sufrido un proceso de gramaticalización y se ha transformado en un marcador impersonal. De este modo, este proceso de gramaticalización, que fija su uso, podría explicar, en parte, la ausencia de variación entre los géneros.

## 6. CONCLUSIÓN

La finalidad de este estudio de corpus ha sido mostrar la variación de los participantes involucrados en tres mecanismos de desfocalización del agente en español y en francés, basándonos en dos géneros digitales distintos. El análisis ha demostrado que el género, por una parte, y la lengua, por otra, influyen no solo en la repartición de los mecanismos y la presencia de participantes, sino también en la categoría a la cual pertenecen estos últimos.

El examen detallado de los datos de corpus ha destacado que la variación en la productividad de las estructuras de desfocalización del agente depende del género y de la lengua. También ha permitido mostrar que, aunque las dos lenguas poseen estructuras equivalentes de un punto de vista de la forma, no las emplean con la misma frecuencia, y que esta frecuencia varía en función del registro.

Al indagar en los distintos tipos de participantes, los resultados han resaltado lo siguiente: respecto a los agentes, el análisis ha mostrado que la enciclopedia en línea cuenta con una proporción más alta de pasivas perifrásticas con agentes, principalmente en español. Una aproximación cualitativa a los datos ha vislumbrado dos situaciones en las que aparecen los agentes: (i) personas conocidas u ocupando un cargo prestigioso y (ii) una situación considerada como negativa. Otro elemento particularmente relevante que este análisis ha puesto de relieve es la mayor presencia de agentes humanos en la Wikipedia, especialmente en español. Sin embargo, esta tendencia debe confirmarse, dado que una de las principales limitaciones del presente trabajo es que no cuenta con un número elevado de perifrásticas largas.

En cuanto a los pacientes, los resultados muestran un predominio general de los pacientes de tipo *inanimado otro*. Sin embargo, esa proporción varía según la estructura, el género y el idioma. Esta variación es particularmente evidente en el caso de la pasiva perifrástica, cuyo análisis ha puesto de manifiesto las dos siguientes particularidades: (i) una frecuencia mayor de pacientes de tipo *organización* en la Wikipedia en español y (ii) una presencia menor de pacientes animados en los datos francés de la Wikipedia. En la construcción con *se*, esa preponderancia de pacientes de tipo *inanimado otro* es aún más marcada. Se ha observado, no obstante, cierta variación en la Wikipedia, especialmente en español. Respecto al pronombre francés *on*, los dos géneros presentan un comportamiento bastante similar, lo que resulta sorprendente, visto que se asocia la estructura con situaciones de menor grado de formalidad.

A lo largo de este artículo, hemos documentado la variación de la presencia y del tipo de participantes involucrados en tres estrategias de desfocalización. Un factor clave de esta variación es el género textual en el cual aparece la estructura. En efecto, los objetivos discursivos del género, así como las características sintácticas que se suelen atribuir a las producciones textuales con mayor y menor grado de formalidad influyen en el tipo de participante usado. Además del género, el análisis ha confirmado que la lengua es otro elemento pertinente que cabe considerar a la hora de examinar fenómenos sintácticos. El estudio ha demostrado que, aunque dos idiomas cercanos tienen recursos gramaticales similares, estos se emplean de forma distinta.

A la vista de los resultados prometedores obtenidos en este trabajo, una pista para enriquecer la comprensión de la influencia del género en la producción textual sería replicar este estudio en otros géneros digitales propensos a recurrir a estrategias de desfocalización del agente. Sugerimos, por ejemplo, comentarios sobre noticias digitales, que son contextos propicios a interacciones controvertidas, o páginas web comerciales, que, mientras tienen que convencer a consumidores de comprar productos, deben promocionar y mantener una imagen positiva de la marca.

## REFERENCIAS BIBLIOGRÁFICAS

- Achard, M. (2010). Fields and settings: French *il* and *ça* impersonals in copular complement constructions. De Gruyter. <https://doi.org/10.1515/COGL.2010.016>
- Briz, A. (2010). El registro como centro de la variedad situacional. Esbozo de la propuesta del grupo Val.Es.Co. sobre las variedades diafásicas. Fonte, I. & Rodríguez Alfano, L. 331 (compiladoras), *Perspectivas dialógicas en estudios del lenguaje*, 21-56. México: Universidad Autónoma de Nuevo León, Facultad de Filosofía y Letras.
- Cabañas Maya, R. (2006). La animicidad como factor de variación entre la pasiva perifrástica y la pasiva refleja. *Anuario de Letras*, (44): 125-146.



- Calvo Montoro, M. J. (1983). *La voz pasiva*. Madrid. Coloquio.
- Creissels, D. (2008). *Impersonal and related constructions: a typological approach*. Université de Lyon.
- Coussé, E. y J. Van der Auwera. (2012). Human impersonal pronouns in Swedish and Dutch: A contrastive study of man and men. *Languages in contrast*, 12(2), 121-138. <https://doi.org/10.1075/lic.12.2.01cou>
- Coveney, A. (2000). Vestiges of “Nous” and the 1st Person Plural Verb in Informal Spoken French. *Language Sciences*, 22(4), 447-481. [https://doi.org/10.1016/S0388-0001\(00\)00014-0](https://doi.org/10.1016/S0388-0001(00)00014-0)
- Cusot, G. y C. Klein, C. (2017). Redes colaborativas: Wikipedia ¿Es confiable?. *ESTRATEGIAS, Investigación en Comunicación ISSN: 2550-6870*, 2.
- Davison, A. (1980). Peculiar passives. *Languages*, 56. <https://doi.org/10.2307/412642>
- De Cock, B. (2020). Intersubjective Impersonals in Context: A Multivariate Analysis of Spanish “Uno” and “Se” in Spoken Language. *Catalan Journal of Linguistics*, 93-114. <https://doi.org/10.5565/rev/catjl.310>
- Devís Márquez, P. (2003). La impersonalidad y las denominadas construcciones impersonales en español. *Zeitschrift für romanische Philologie (ZrP)*, 119(3): 393-442. <https://doi.org/10.1515/zrph.2003.119.3.393>
- Devís Márquez, P. (2017). Las denominadas construcciones con se y verbos de alternancia causativa en español. *Neuphilologische Mitteilungen*, 118(2), 405-446.
- Dieck, M. (2016). ¿Y qué más hace uno, pues?: la expresión de la impersonalidad en el español de Medellín. *Linguística y Literatura*, 69. <https://doi.org/10.17533/udea.lyl.n69a06>
- Ehrhardt, Claus, 2014. Politeness and Face Work in German Forum Communication. En Bedijs, Kristina; Held, Gudrun y Maaß, Christiane (Eds.) (pp. 83-107). *Face Work and Social Media*. Berlin: Lit.
- Fauconnier, S. (2012). Constructional effects of involuntary and inanimate Agents: A crosslinguistic study. Doctoral dissertation. Leuven.
- Fernández, S. (2007). *La voz pasiva en español: hacia un análisis discursivo*. Peter Lang.
- Fløttum, K., Jonasson, K., & Norén, C. (2007). *On: pronom à facettes*. De Boeck/Duculot. <https://doi.org/10.3917/dbu.jonas.2007.01>
- Gaetone, D. (1998). *Le passif en français*. Paris / Bruxelles: Editions Duculot, Champs Linguistiques.
- Giacalone-Ramat, A. & Sansò, A. (2007). The spread and decline of indefinite man-constructions in European languages: An areal perspective. Europe and the Mediterranean as Linguistic Areas : Convergencies from a historical and typological perspective. Ramat P. y Roma E. (Eds). *Studies in Language Companion*, 88, 95-131. <https://doi.org/10.1075/slcs.88.07gia>
- Gómez Torrego, L. (1992). *La impersonalidad gramatical: descripción y norma*. Madrid. Arco Libros.
- Hopper, P. & Thompson, S. (1980). Transitivity in grammar and discourse. *Language*, 56(2), 251-299. <https://doi.org/10.1353/lan.1980.0017>
- Jalenques, P. (2016). Le passif en français et le statut référentiel du sujet. SHS Web of Conferences, 27, 14001. EDP Sciences. <https://doi.org/10.1051/shsconf/20162714001>
- Jisa, H. & Tolchinsky, L. (2009). Developing a depersonalized stance through linguistic means in typologically different languages: Written expository discourse. *Written Language & Literacy*, 12(1), 1-25. <https://doi.org/10.1075/wll.12.1.01jis>
- Keenan, E. L. (1975). Some universals of passive in relational grammar. Papers from the 11th Regional Meeting, *Chicago Linguistic Society*, 340, 52
- Lamiroy, B. (1993). Pourquoi il y a deux passifs. *Langages*, 27(109), 53-72. <https://doi.org/10.3406/lgge.1993.1091>
- Laslop, M. E. & Díaz, A. H. (2010). Impersonalidad y pasividad en las normas sobre derechos humanos: El uso de “se” y de pasivas perifrásticas en las constituciones mexicanas (1824–2001). *Revista Internacional de Lingüística Iberoamericana*, 203-231.
- Mancera Rueda, A. (2015). Los estudios sobre (des) cortesía y actividades de imagen en las redes sociales. *Texts in Process*, 1(1), 50-70. <https://doi.org/10.17710/tep.2015.1.1.3man>
- Mendikoetxea, A. (1999). Construcciones con se: medias, pasivas e impersonales. En I. Bosque y V. Demonte (Eds), *Gramática descriptiva de la lengua española* (pp. 1631-1722). Madrid. Espasa Calpe.

- Obregón Sierra, Á. y N. González Fernández. (2019). Las universitarias en la Wikipedia en español. *Pixel-Bit: Revista de Medios y Educación*, 54, 145-164. <https://doi.org/10.12795/pixelbit.2019.i54.08>
- Placencia, M. E. (2010). Yahoo!Respuestas como columna de consejos: algunos rasgos de un género híbrido. *Tonos Digital - Revista Electrónica de Estudios Filológicos*, 20, s.p.
- Placencia, M. E. (2012). Online peer-to-peer advice in Spanish Yahoo!Respuestas. En H. Limberg y M. Locher (Eds.) *Advice in discourse* (pp. 281-305). Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/pbns.221.17pla>
- Quesada, J. D. (1998). Transitivity, voice, and the middle: Spanish se revisited. *Romanische Forschungen*, 110(1), 1-36.
- Quirk, R., S. Greenbaum, G. Leech y J. Svartvik. (1985). *A comprehensive grammar of the English language*.
- Pierre, E. (2021). *Agent-defocussing structures in Spanish, English and French: A contrastive corpus-based account of the interaction between different passive-like structures*. Doctoral dissertation, Université catholique de Louvain. Belgium.
- Rasson, M. (2016). Interpretación y tipología del pronombre indefinido uno a partir de tres géneros discursivos. *CÍRCULO de Lingüística Aplicada a la Comunicación*, 67, 227. <https://doi.org/10.5209/CLAC.53484>
- Rasson, M. (2017). Interpretaciones y funciones del pronombre indefinido uno Análisis de tres variedades geográficas (Chile, España y México) y de tres géneros discursivos (conversaciones, foro digital y redacciones). Université Catholique de Louvain, Belgique. <https://doi.org/10.5209/CLAC.53484>
- Siewierska, A. (1984). *The passive. A comparative linguistic analysis*. Londres. Croom Helm.
- Siewierska, A. (2008). Introduction: Impersonalization from a subject-centred vs. agent-centred perspective. *Transactions of the Philological Society*, 106(2), 115-137. <https://doi.org/10.1111/j.1467-968X.2008.00211.x>
- Siewierska, A. (2010). From third plural to passive. Incipient, emergent and established passives. *Diachronica*, 27(1), 73-109. <https://doi.org/10.1075/dia.27.1.03sie>
- Shibatani, M. J. (1985). Passives and related constructions: A prototype analysis. *Language*, 61(4), 821-848. <https://doi.org/10.2307/414491>
- Teich, E., S. Degaetano-Ortieb, P. Fankhauser, H. Kermes y E. Lapshinova-Koltunski. (2016). The linguistic construal of disciplinarity: A data mining approach using register features. *Journal of the Association for Information Science and Technology*, 67(7), 1668-1678. <https://doi.org/10.1002/asi.23457>
- Tramullas, J. (2015). Wikipedia como objeto de investigación. *Anuario ThinkEPI*, 9, 223-226. <https://doi.org/10.3145/thinkepi.2015.50>
- Vázquez Rozas, V. (2006). Gustar-type verbs. Functional approaches to Spanish syntax, 80-114. Springer. NY. [https://doi.org/10.1057/9780230522688\\_4](https://doi.org/10.1057/9780230522688_4)
- Weiner, E. J. & Labov, W. (1983). Constraints on the Agentless Passive. *Journal of Linguistics*, 19(1), 29-58. <https://doi.org/10.1017/S0022226700007441>
- Yao, K. (2016). Referencias y equivalencias del pronombre on francés en el español. *Revista de Lenguas Modernas*, 25, 115-130. <https://doi.org/10.15517/rlm.v0i25.27685>

## Referencias de los corpus

- De Smet. (2009). *Contrastive Corpus of Questions and Answers*. Compiled by Hendrik De Smet. Department of Linguistics, University of Leuven.
- Reese, S. et al. 2010. Wikicorpus: A Word-Sense Disambiguated Multilingual Wikipedia Corpus. Proceedings of the 7th Language Resources and Evaluation Conference (LREC2010). La Valleta. Malta.



# Los marcadores del discurso en la narrativa financiera: análisis de las cartas a los accionistas<sup>1</sup>

Discourse markers in the financial narrative: analysis of letters to shareholders

ANA GARCÍA TORO  
Universidad Autónoma de Madrid, España  
[ana.garciatoro@estudiante.uam.es](mailto:ana.garciatoro@estudiante.uam.es)  
<https://orcid.org/0000-0002-3773-4679>

## Resumen

Los marcadores del discurso (MD) son elementos de cohesión textual que guían las inferencias del discurso a nivel intra y extraoracional. Entre sus inferencias se registran la adición, la contraargumentación, el refuerzo, la atenuación o la reformulación, que intervienen constantemente en la argumentación y en la construcción del discurso. Por su parte, las cartas a los accionistas, redactadas por el primer ejecutivo de la empresa, son un resumen y evaluación del ejercicio anual de la compañía, donde se exponen los principales datos u objetivos conseguidos en ese año, con el propósito de mantener la confianza de sus accionistas y/o atraer potenciales inversores. En este trabajo se combinan un estudio cuantitativo, con la obtención de las frecuencias de los marcadores; un estudio cualitativo, porque se presenta un análisis de casos donde se podrán observar ejemplos

## Abstract

Discourse markers (DMs) are elements of textual cohesion that guide discourse inferences at the sentence level and outside of it. Their inferences include addition, counter-argumentation, reinforcement, attenuation, or reformulation, which are constantly involved in argumentation and discourse construction. For their part, letters to shareholders, written by the company's chief executive, are a summary and evaluation of the company's annual financial year, in which the main data or objectives achieved that year are set out, with the aim of maintaining the confidence of its shareholders and/or attracting potential investors. This work combines a quantitative study, with the obtaining of the frequencies of the markers; a qualitative study, because a case analysis is presented where can be observed examples and extrapolate the data to the text; and a contrastive study between

---

<sup>1</sup> Esta investigación ha sido financiada por el Ministerio de Ciencia e Innovación, dentro del marco del Proyecto CLARA-FINT. Código: PID2020-116001RB-C31

**Para citar este artículo:** Los marcadores del discurso en la narrativa financiera: Análisis de las cartas a los accionistas. *ELUA*, (38), 187-214. <https://doi.org/10.14198/ELUA.22421>

Recibido: 06/04/2022 Aceptado: 25/05/2022

© 2022 Ana García Toro



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

que extrapolen los datos al texto; y un estudio contrastivo entre dos corpus de carácter financiero: los textos de empresas con ganancias y los de empresas con pérdidas. Así, se pretende exponer la distribución de los marcadores del discurso en un corpus financiero con el fin de determinar si el empleo y la frecuencia de algunas partículas discursivas en las cartas están condicionadas por los resultados de la empresa. Los principales hallazgos tras este análisis pueden resumirse en que en las ganancias van a primar las inferencias de adición de la información y de refuerzo argumentativo, mientras que, en las pérdidas se utilizan marcadores contraargumentativos y operadores de modalidad emotiva. Además, la apariencia de seguridad y esperanza se busca a través de los operadores modales, y encontramos un uso generalizado de la partícula *adicionalmente*, potencialmente por la influencia anglosajona. La originalidad de este trabajo recae en que se realiza en un campo inexplorado como es la narrativa financiera en español. En esencia, nuestro estudio combina herramientas propias de la lingüística computacional, con una metodología empírica de lingüística de corpus, para acercarnos un poco más al estudio de la argumentación y de las estrategias utilizadas en el discurso financiero.

**PALABRAS CLAVE:** Discurso; Informes financieros; Argumentación; Lingüística informática.

## 1. INTRODUCCIÓN

El presente trabajo tiene como objetivo realizar un estudio cuantitativo, cualitativo y, a su vez, contrastivo, de la distribución de los marcadores del discurso (MD) en textos de carácter financiero. Los dos corpus de los que nos hemos servido para la investigación están compuestos, por un lado, por las cartas a los accionistas de las empresas con ganancias y, por otro, por las correspondientes a las empresas con pérdidas. Serán de nuestro interés, principalmente, los MD más frecuentes y los que destaquen en un corpus, pero no en el otro.

El análisis llevado a cabo tiene como base, en primer lugar, las consideraciones teóricas de autores en español que han trabajado sobre la existencia, delimitación y comportamiento de estas partículas y el análisis del discurso. Tomamos como base las clasificaciones de Zorraquino y Portolés (1999), con su posterior actualización en Portolés (2011), así como el trabajo descriptivo que realiza Fuentes Rodríguez (2009) en su *Diccionario de conectores*

two financial corpora: the texts of companies with profits and those of companies with losses. Thus, the aim is to expose the distribution of discourse markers in a financial corpus in order to determine whether the use and frequency of some discursive particles in the letters are conditioned by the company's results. The main findings of this analysis can be summarised as follows: in the case of profits, information addition and argumentative reinforcement inferences are predominant, whereas in the case of losses, counterargumentative markers and emotive modality operators are used. Moreover, the appearance of security and hope is sought through modal operators, and we find a generalised use of the particle *additionally*, potentially due to Anglo-Saxon influence. The originality of this work lies in the fact that it is carried out in the unexplored field of financial narrative in Spanish. In essence, our study combines tools from computational linguistics with an empirical methodology of corpus linguistics to get a little closer to the study of argumentation and the strategies used in financial discourse.

**KEYWORDS:** Discourse; Financial Reports; Argumentation; Computational Linguistics.

y *operadores del español*. Estas referencias han sido el pilar de nuestro marco teórico y las bases para la aplicación de nuestro objeto de estudio. Asimismo, han facilitado la clasificación de las partículas, también en los casos en los que un mismo marcador podía pertenecer a dos categorías diferentes. Si bien es cierto que nos hemos servido de otras perspectivas para la definición y la delimitación de la función de marcador discursivo y sus características, con trabajos como los de Garcés (2008), Montolío (2001), Loureda y Acín (2010) o Llamas *et al.* (2012).

Más allá de la base teórica de los marcadores del discurso, construida en los últimos 30 años, estas partículas no se han tratado de manera empírica y práctica en el ámbito del análisis del discurso. Lo que proponemos en esta investigación es ofrecer un estudio exhaustivo de los marcadores discursivos en un panorama concreto de los lenguajes de especialidad, como son las cartas de los presidentes a los accionistas en los informes anuales de las empresas.

Por su lado, las cartas del presidente o del CEO<sup>2</sup> a los accionistas son textos incluidos en la primera parte de los informes financieros anuales de las empresas. En ellas se resumen, se contextualizan y se valoran los resultados que se detallan en el informe anual y, a la vez, se exponen propuestas para el futuro crecimiento de la empresa. El principal fin en la redacción de estas cartas es argumentar y convencer a los actuales y potenciales accionistas de la solidez y los buenos resultados de la empresa. Lógicamente, no todas las empresas tienen resultados positivos, por lo que en el caso de las compañías que hayan obtenido unos resultados más pobres, deberán buscar la manera, en este texto epistolar, de minimizar los aspectos desfavorables, los defectos en la gestión y resaltar cualquier buena práctica para convencer a los inversores de que esa situación se revertirá.

Las hipótesis de esta investigación tienen en cuenta los resultados correspondientes de las empresas para el análisis de su argumentación. En primer lugar, creemos que los grupos de marcadores con más frecuencias y partículas serán los conectores y los operadores. En segundo lugar, de los conectores, los aditivos tenderán a prevalecer en las empresas con ganancias, y los contraargumentativos en las pérdidas, por sus respectivas inferencias de añadir información (positiva) y contraargumentar información (negativa); por su lado, estimamos que los operadores argumentativos y de refuerzo argumentativo, así como los operadores de modalidad (que actúan en el plano del grado de seguridad del hablante) se presentarán con mayor asiduidad en las empresas con ganancias; mientras que en las empresas con malos resultados podrían destacar el subgrupo de los operadores de atenuación o los marcadores de modalidad emotiva negativa. Por último, creemos, acorde con la literatura al respecto, que los reformuladores serán un recurso empleado mayoritariamente por las empresas con pérdidas, ya que se verán obligados a dar una explicación o una reformulación de sus resultados.

En el propósito último de esta investigación subyace la pregunta a la pretendemos responder: ¿son los marcadores del discurso un posible indicador de los resultados de una empresa?

---

2 Se corresponde a las siglas en inglés *Chief Executive Officer*, y se refieren a lo que se corresponde en español como *primer ejecutivo*. En caso de que se conozca su labor, se especifica como consejero delegado, director general o presidente ejecutivo, entre otros (el «CEO» es el primer ejecutivo de una empresa | FundéuRAE (fundeu.es).

## 2. ESTADO DE LA CUESTIÓN

### 2.1. Los marcadores del discurso

Los marcadores del discurso o marcadores discursivos (MD) son un grupo amplio y heterogéneo de unidades lingüísticas invariables que constituyen un nexo intra y supraoracional para la coherencia y la cohesión textuales. Su función, a grandes rasgos, es la de marcar el discurso y definir la relación que se establece entre dos partes o segmentos del discurso. Relacionan desde palabras, grupos de palabras, a oraciones y párrafos.

Esta clase es definida por algunos autores como “una categoría pragmática de unidades lingüísticas que son los engranajes del discurso (oral y escrito), ya que procesan su cohesión, coherencia, adecuación y eficacia” (Landone 2012: 431). Con esta definición se inclinan hacia una concepción intencionalmente genérica con la que pretenden abarcar los rasgos funcionales comunes a varias definiciones de los marcadores del discurso. En este artículo se habla también de la problemática que plantea “la imposibilidad de ceñir un marcador a una categoría única” (p. 431), algo con lo que nos hemos encontrado en el curso de la investigación (*absolutamente* como operador argumentativo, pero también como operador modal). Otros autores se han aproximado a su definición desde un punto de vista más pragmático y procedimental (Zorraquino y Portolés 1999; Pons 2000; Montolío 2001; Briz *et al.* 2008; Fuentes 2009; Portolés 2011), y los consideran los encargados de guiar las inferencias de los enunciados para que sean interpretados correctamente de acuerdo con unas propiedades morfosintácticas, semánticas y pragmáticas. Concretamente, Portolés (2011), establece una definición de marcador del discurso partiendo de la noción de *comunicación inferencial*, *implicatura conversacional* y *significado de procesamiento*:

[S]on unidades lingüísticas invariables, no ejercen una función sintáctica en el marco de la predicación oracional y poseen un cometido coincidente en el discurso: el de guiar, de acuerdo con sus distintas propiedades morfosintácticas y pragmáticas, las inferencias que se realizan en la comunicación (pp. 25-6).

En conclusión, cada autor dar diferentes matices a las funciones de los marcadores dependiendo de su perspectiva de estudio y, por tanto, de lo que entiende por MD.

#### 2.1.1. Límites y características

La definición de estos elementos, como hemos visto, es conflictiva y los límites del término marcador del discurso quedan difusos, establecidos de una manera u otra nuevamente dependiendo de la perspectiva adoptada por cada autor y conforme a su estudio. Se ha intentado definir y unificar este grupo heterogéneo atribuyendo características comunes a sus componentes. Dichas características se pueden organizar en tres criterios siguiendo el modelo de Portolés (2015): un criterio semántico, un criterio sintáctico y un criterio morfológico, que hemos adaptado según nuestro objeto de estudio.

En primer lugar, el criterio semántico se basa en que estas unidades tienen un significado procedimental y no conceptual, ya que guían las inferencias, pero no representan la realidad. En ocasiones, con este significado de procesamiento reflejan la actitud, el grado de segu-

ridad, la opinión o la intención del hablante; asimismo, realizan funciones de coherencia y cohesión textuales a través de inferencias.

En segundo lugar, sintácticamente se dice que son elementos al margen de la oración y que no forman parte de la predicación oracional. Por ello, en ocasiones van seguidos de una coma y suelen aparecer al comienzo de una oración (estos marcadores son conocidos como intratextuales o extraoracionales). Sin embargo, otros marcadores cumplen funciones dentro de la oración, cohesionando proposiciones, orientando ciertos elementos según la intención, seguridad u opinión del hablante, o siguen instrucciones de finalidad, causa, consecuencia o contraargumentación, sin que esto sea al comienzo de un párrafo (estos son los llamados marcadores intraoracionales). Se trata de diferentes niveles de argumentación y organización del texto. Esta distinción ya era aportada por Montolío (2001), que distinguía entre conectores o marcadores parentéticos, y los integrados en la oración.

Por último, el criterio morfológico establece que son unidades invariables, algo que se debe a su fijación gramatical y a que provienen de distintas categorías, especialmente de conjunciones (*pero*), adverbios (*finalmente*, *adicionalmente*), locuciones adverbiales (*de esta manera*) o, incluso de la unión entre sintagmas y preposiciones (*desde luego*); todas ellas también de carácter invariable. La proveniencia de otras categorías genera ambigüedad en la determinación y en los límites de qué es y qué no un MD, derivando en problemas a la hora de tenerlos en cuenta para nuestro análisis. El criterio morfológico está estrechamente relacionado con el concepto de gramaticalización: el paso de un significado más referencial o conceptual (gramatical) a otro menos referencial o procedimental (discursivo).

Aun así, el proceso de gramaticalización de algunos MD se encuentra en estadios intermedios y son sensibles al contexto, lo que hace que se sitúen en una frontera: en ocasiones funcionan como marcador discursivo, pero en otras ocasiones mantienen su naturaleza gramatical.

### 2.1.2. Clasificación

En torno a la clasificación de estas unidades se genera la polémica de establecer una taxonomía suficientemente exhaustiva, pero, a su vez, precisa y representativa de la tipología de estos elementos. De hecho, muchas de las clasificaciones de los marcadores discursivos se basaban en listas muy extensas que pretendían recoger y diferenciar los matices de cada marcador discursivo. El resultado que se obtenía era una clasificación con muchos tipos y variantes, pero poco operativa y bastante compleja. En esta investigación hemos simplificado esta taxonomía y la hemos adaptado a nuestro objeto de estudio. Para ello, hemos tomado como referencia los estudios de Zorraquino y Portolés (1999), actualizados por Portolés (2011); la clasificación aplicada a textos académicos y profesionales de Llamas Saíz, *et al.* (2012); y la información sobre conectores y operadores en la obra de Fuentes Rodríguez (2009). El problema generalizado de la hipercategorización reside en la gran variedad de los elementos considerados marcadores discursivos y la ausencia de una definición cerrada que establezca unas características comunes. Por nuestra parte y con el fin de evitar confeccionar una lista interminable e inmanejable de categorías, hemos descartado los marcadores orales y los reformuladores conversacionales porque carecían de interés para un estudio basado en un corpus formal y escrito. Por otro lado, si hemos ampliado las subcategorías contempladas en la clasificación de 1999 y las hemos completado con criterios y subgrupos que aporta



Fuentes Rodríguez en su diccionario, de manera que se pudiera cubrir todo el espectro de nuestro corpus, sin excedernos en la categorización. En la página que sigue se muestra el resultado de esta clasificación, con los MD correspondientes (Tabla 1).

CATEGORÍA	SUBCATEGORÍA	MD
<b>Estructuradores de la información</b>	Tematizadores	<i>a este respecto, al hilo de, con relación a, con respecto a, desde el punto de vista (de), desde esta perspectiva, desde la perspectiva, en cuanto a, en el ámbito, en el caso de, en el plano, en esta línea, en lo que corresponde a, en lo que respecta a, en lo referente a, en lo referido a, en materia de, en relación con, en otra dimensión, en otro orden de cosas, en términos</i>
	Ordenadores	<i>en primer lugar, en primer término, en primera instancia, en segundo lugar, en tercer lugar, en último término, finalmente, para concluir, para finalizar, para terminar, por fin, por finalizar, por lo que respecta a, por lo que se refiere a, por otra parte, por otro lado, por su parte, por último, por un lado, por una parte, primero, segundo, en este sentido</i>
	Comentadores	<i>pues, pues bien</i>
	Digresores	<i>por cierto</i>
	Ejemplificadores	<i>como ejemplo, como muestra, ejemplo(s) de ello, entre otras cosas, muestra de ello, por ejemplo, tales como, prueba de ello</i>
<b>Conectores</b>	Aditivos	<i>a su vez, además, adicionalmente, al igual que, al respecto, aparte (de), así como, asimismo, de la misma forma, del mismo modo, igualmente, incluso, por añadidura, al margen de, de otro modo, de otro ángulo, en paralelo, paralelamente, simultáneamente, igualmente, también, tampoco, a la vez, al mismo tiempo (que), al propio tiempo, al tiempo que</i>
	Causales	<i>dado que, porque, puesto que, ya que</i>
	Concesivos	<i>a pesar de ello, a pesar de todo ello, aun, aun así, aunque, pese a (todo), si bien</i>
	Consecutivos	<i>a consecuencia de, así pues, así que, así(,), como consecuencia, como fruto de, como resultado, consecuencia de, consecuentemente, de ahí que, de esta manera, de esta forma, de este modo, de forma que, de manera que, de modo que, de tal manera que, en consecuencia, fruto de ello, por consiguiente, por ello(,), por ende, por esa razón, por ese motivo, por eso(,), por esta razón, por este motivo, por lo tanto, por tanto, por todas estas razones, por todo ello, resultado de, gracias a ello, pues</i>

<b>Conectores</b>	Contraargumentativos	<i>ahora bien, con todo(,), con todo ello, en cambio, en contraste, no obstante, pero, por contra, por el contrario, por el lado contrario, sin embargo,</i>
	Finalidad	<i>a efectos de, a este efecto, a este fin, a fin de, a este fin, a tal fin, al objeto de, con el fin de, con el objetivo de, con el propósito de, con este fin, con la finalidad de, con motivo de, con vistas a, para ello(,), para este fin, para todo ello</i>
<b>Reformuladores</b>	Explicativos	<i>dicho de manera sencilla, dicho de otra forma, dicho de otro modo, en otras palabras, es decir, esto es(,), o lo que es lo mismo, o sea</i>
	Rectificativos	<i>mejor dicho</i>
	De distanciamiento	<i>en cualquier caso, en todo caso,</i>
	Recapituladores	<i>como conclusión, en definitiva, en resumen, en síntesis, en suma, resumiendo, pues bien</i>
<b>Operadores</b>	Argumentativos	<i>absolutamente, ante todo, antes que nada, básicamente, en buena medida, , en parte, en especial, en gran medida, en gran parte, esencialmente, especialmente, exclusivamente, fundamentalmente, mayoritariamente, particularmente, por encima de todo, preferiblemente, principalmente, sobre todo, únicamente,</i>
	Atenuativos	<i>de alguna manera,</i>
	Concreción	<i>concretamente, en concreto, en particular, específicamente, particularmente, sin ir más lejos,</i>
	Enunciativos	<i>según, a tenor de, de acuerdo con, de manera general, en general, en líneas generales, francamente, generalmente, prácticamente, sencillamente, simplemente, sinceramente, singularmente, todo sea dicho</i>
	Modales	<i>afortunadamente, ciertamente, con toda probabilidad, efectivamente, en absoluto, en efecto, evidentemente, desafortunadamente, desgraciadamente, indudablemente, lamentablemente, lógicamente, naturalmente, obviamente, por supuesto, posiblemente, probablemente, qué duda cabe, seguramente, sin duda alguna, sin duda, sin dudarlo, sin lugar a dudas, tal vez, supuestamente, realmente, verdaderamente, significativamente</i>
	Refuerzo argumentativo	<i>de hecho, desde luego, en absoluto, en el fondo, en verdad, es más, más aún, precisamente</i>

Tabla 1. Clasificación de los MD de acuerdo con nuestro objeto de estudio.

## 2.2. El lenguaje financiero: las cartas a los accionistas

### 2.2.1. El lenguaje financiero

La distinción entre el lenguaje general y el especializado se ha tenido siempre en cuenta desde una perspectiva lexicográfica y terminológica, a través del léxico. Sin embargo, desde la lingüística de corpus se ha advertido que las diferencias actúan a todos los niveles de la lengua y distinguen los tipos de discurso. Por consiguiente, Biber y sus colaboradores demostraron con una extensa bibliografía que en todos los registros, géneros y estilos textuales se presentan rasgos fonológicos, gramaticales, léxicos y discursivos diferentes (Biber 1988; Biber *et al.* 1998; Biber, *et al.* 1999; Biber y Conrad, 2009; Biber y Reppen 2015). Estos rasgos, además, quedan claramente distinguidos a través de una aproximación estadística y cuantitativa, que será la técnica que utilizaremos en este estudio.

Entre los aspectos distintivos del lenguaje de especialidad y siguiendo las directrices de Mateo (2007) y Herrero (2010), podemos destacar su función exhortativa (frente al lenguaje de la economía, que es más informativo); la tendencia a un estilo coloquial, sobre todo en las cartas, para acercar la información a un lector lego; una terminología específica y, en algunos casos, técnica del dominio de especialidad; unidades fraseológicas y colocaciones determinadas, así como la nominalización de verbos para dar riqueza lingüística al texto y evitar redundancias; la gran presencia de un tipo de metáfora concreta como lo es la metáfora terminológica, que son imágenes que facilitan la comprensión de los conceptos y crean unión terminológica, acercan el discurso al receptor y refuerzan el carácter divulgativo y persuasivo de estos textos; la incorporación constante de anglicismos, préstamos o el uso de abreviaturas y acrónimos (Alcaraz *et al.* 2012; Román 2015), propios del sincretismo inglés, además de calcos de otras lenguas. Otro aspecto significativo de este tipo de lenguaje es la adaptación del discurso a un oyente semi-experto, por lo que es común encontrar perífrasis, paréntesis o frases que explican o reformulan el contenido anterior para ganar claridad. Esta reformulación abarca a los marcadores discursivos de reformulación, presentes en nuestro estudio, pero en contraste con la característica que se acaba de mencionar, no encontramos gran variedad ni frecuencia de reformuladores.

En el ámbito de la lingüística computacional y del procesamiento del lenguaje natural (PLN), las investigaciones se han encaminado al tratamiento automático de grandes colecciones de textos financieros, con el objetivo de estudiar su contenido, extraer información, redactar resúmenes automáticos, medir su claridad o grado de legibilidad, y analizar su polaridad positiva o negativa (Loughran y MacDonald 2016). El-Haj, *et al.* (2019) aporta un estado de la cuestión en el plano del análisis computacional relacionado con el discurso financiero inglés. El desarrollo de investigaciones por la Universidad de Lancaster (grupo UCREL) han inspirado la aplicación de la lingüística computacional en el ámbito financiero español, como se recoge en los trabajos de Moreno-Sandoval *et al.* (2019) y Moreno-Sandoval *et al.* (2020).

### 2.2.2. Las cartas a los accionistas

Por su lado, las cartas del presidente son textos que se encuentran integrados al comienzo de los informes anuales de las empresas, y en las que se exponen, resumidos, los principales avances y resultados (positivos y negativos) de la empresa durante ese último ejercicio

anual y son la parte más leída de estos informes (Bujaki y McConomy, 2012). Este tipo de narrativa es descrita así por algunos autores:

Located at the beginning of the annual report, the letter to shareholders provides a contextualized analysis and an over-view of the company's performance throughout the year. In addition, it is a complete narrative document, free of tables and figures, with "enormous rhetorical importance" for building credibility and trust in the company (Moreno-Sandoval, *et al.* 2020: 97).

Bhatia (2008) presenta las principales partes o movimientos en los que se pueden dividir las cartas a los accionistas, de manera que pueda entenderse la estructura argumentativa de estas: 1. Se hace una revisión general del ejercicio; 2. Se identifican temas importantes; 3. Se desarrollan dichos temas; 4. Se describen las expectativas y se hacen promesas; 5. Se expresa gratitud a los trabajadores y a los accionistas (opcional). 6. Se habla sobre el futuro. 7. Se cierra de manera positiva y educada.

Las cartas comparten características con los textos financieros, al ser un subgénero de estos, y oscilan entre un lenguaje semitécnico y divulgativo, de acuerdo con la clasificación de los géneros de especialidad presentada por Martínez (2007). Las cartas son un texto subjetivo de índole instructiva y argumentativa/exhortativa: instructiva por su inclusión en los informes anuales de las empresas; argumentativa/exhortativa porque siguiendo a Alejandro Gisbert (2017), creemos que estos textos pretenden persuadir a través de la argumentación y de la exposición de sus avances y resultados, el resumen de su año o la gran gestión y toma de dediciones empresariales de ese ejercicio, para convencer a los accionistas de que deben seguir invirtiendo en esa empresa. De esta manera, a diferencia de los informes que contienen los datos financieros, las cartas a los accionistas están consideradas una especie de "márquetin ideológico" con un carácter persuasivo:

Rhetoric is understood here as persuasive discourse, and rhetorical strategies are the choices a writer makes to achieve particular social purposes. In the genre of the CEO's letter this purpose is essentially persuasive, seeking to gain the reader's acceptance for the particular definition of reality preferred by the writer (Hyland 1998: 235).

Finalmente, este género epistolar pretende hacer retórica para exponer y convencer, pero también pretende unificar su público de especialistas y semiexpertos: "[s]o while the quality of information provided in the letters may vary enormously, it is likely to play a more rhetorical role in company communication". (Hyland 1998: 228), acercando el emisor al receptor (Bujaki y McConomy 2012).

### 2.2.3. *La argumentación y los marcadores del discurso en las cartas a los accionistas*

Para entender la intención y el objetivo persuasivo mediante estrategias argumentativas (e inferencias discursivas) de este tipo de discurso, partimos del esbozo que realiza Hyland (1998) sobre este discurso epistolar:

Generally written as a signed personal letter, the document has enormous rhetorical importance in building credibility and imparting confidence, convincing investors that the company is pursuing sound and effective strategies (p. 224).

Basándonos en una idea explicada en el trabajo de Redondo (2018): “[e]l discurso siempre tiene un propósito o intención comunicativa, aunque no siempre está implícito” (p. 22). En el caso de las cartas del presidente, tienen una intención argumentativa y persuasiva, pero esta última camuflada a través de un carácter expositivo y divulgativo. De hecho, en ocasiones se puede priorizar la imagen de la empresa por encima de la información útil para sus oyentes (Hyland, 1998). Los marcadores discursivos nos informarán de esa argumentación y persuasión, guiando los propósitos del emisor hacia una conclusión y fijando la cohesión y la coherencia textuales. Ya hay trabajos de hace un par de décadas que hablan de ellos en las cartas de los presidentes, como es el caso mencionado de Hyland, donde se explican las funciones del metadiscurso y encontramos elementos que funcionan como marcadores discursivos: los *sequencers* (secuenciadores), relacionados con los ordenadores de información de la clasificación de Portolés (2011); *emphatics* (refuerzo argumentativo), con los que pretenden aumentar su credibilidad y poder de convención; *hedges* (digresores), que se relacionan con los malos resultados de las empresas (en detrimento de los de refuerzo argumentativo), y *connectives* (conectores) que, afirma, son los más frecuentes, tendencia que hemos corroborado con nuestro corpus.

Por último, la persuasión en un discurso también depende de la credibilidad del redactor o emisor. A esta credibilidad también contribuyen los marcadores discursivos, seleccionados apropiadamente, y los pronombres personales para ganar la simpatía de los oyentes, verbos epistémicos (Biber, Conrad y Leech 2002), con lo que pretenden construir un vínculo más personal dentro de ese ámbito:

[N]oted that the narrative sections of corporate annual reports, including the LTS [Letters To Shareholders], are those sections most likely to reveal embedded cultural beliefs and values and most likely to provoke interpretive and emotional responses in readers (Bujaki y McConomy 2012: 1114).

Y, a esta credibilidad, contribuyen, asimismo, estrategias basadas en marcadores de actitud, que reflejan la actitud, opinión o sentimientos del hablante (especialmente son operadores modales, como veremos), junto con otras con las que busca formar un diálogo con el oyente con el fin de hacerle participe de su discurso, para lo que es frecuente la utilización de la primera persona del plural (Bhatia 2008).

### 3. METODOLOGÍA

#### 3.1. Corpus

Este trabajo tiene como base el corpus del Proyecto FinT-Esp<sup>3</sup> (Moreno-Sandoval, *et al.*, 2020), confeccionado con el fin de contribuir a la comprensión del discurso y comunicación financiera en el contexto de empresas españolas; algunas de ellas forman parte del conjunto de empresas del Íbex-35<sup>4</sup>. Para nuestro caso hemos dividido dicho corpus, además, en dos subcorpus: uno de ellos con las cartas de los presidentes pertenecientes a empresas que han tenido ganancias,

3 *FinT-Esp: Financial Text Analytics in Spanish: Tools and language resources*, Laboratorio de Lingüística Informática (uam.es) | Consulta Corpus Financero (uam.es).

4 El IBEX-35 es el principal índice bursátil de referencia de la bolsa española. Está elaborado por Bolsas y Mercados Españoles, y está formado por las 35 empresas con más liquidez que cotizan en el Sistema de Interconexión Bursátil Español.

y el otro con las cartas de las empresas que han tenido pérdidas. Así, nuestro corpus en total contaba con un total de 383 cartas de empresas comprendidas en el período de 2014-2017: 331 pertenecientes al corpus de ganancias y tan solo 52 en el de pérdidas. La diferencia de tamaño entre ambos corpus tiene su razón lógica, puesto que la economía de ningún país podría sobrevivir si las empresas con pérdidas igualaran o se acercaran en número a las empresas con ganancias. Por ello, y como veremos más adelante (Cf. 3.3.), tuvimos que trabajar el análisis de nuestros resultados de marcadores discursivos con la frecuencia normalizada (FN). De esta manera, podemos comparar los resultados de los MD de ganancias y de pérdidas, sobre una base normalizada de un millón de palabras. Los datos generales del corpus financiero y los concretos de ganancias y de pérdidas se encuentran en la tabla siguiente (Tabla 2):

	<b>Empresas</b>	<b>Palabras</b>	<b>MD</b>	<b>Porcentaje de MD</b>
<i>Corpus general</i>	383	461199	7939	1,72%
<i>Corpus de ganancias</i>	331	407923	7029	1,72%
<i>Corpus de pérdidas</i>	52	53276	910	1,71%

Tabla 2. Datos generales del corpus y los subcorpus.

Los marcadores del discurso suponen menos del 2% del total de las palabras del corpus. Esto puede sorprender en el sentido de que son los elementos encargados de cohesionar un texto a todos sus niveles y de guiar las inferencias del interlocutor. Podemos advertir su importancia si, aun contando con tan poca presencia cuantitativa, siguen siendo las piezas de engranaje imprescindibles en cualquier discurso.

Como vemos, y comprobaremos en el análisis, ya advertimos que, al menos el uso de los MD, es similar en las cartas de las empresas beneficios de las que no los han tenido. Sus porcentajes son prácticamente los mismos, lo que parece indicarnos que su distribución no va a darnos pistas sobre si una carta pertenece a una empresa con buenos o con malos resultados. Tendremos que esperar a analizar los grupos y, sobre todo, particularmente algunos MD y sus inferencias, para advertir diferencias entre los dos subcorpus.

### 3.2. Proceso de selección de MD

Los MD son un conjunto amplio de elementos que pueden funcionar en un plano intraoracional o extraoracional/discursivo. Para llevar a cabo la selección de MD que se pretendían filtrar y analizar, tomamos como base las partículas incluidas en la guía de anotación del detector de marcadores del discurso entrenado con Transformers (García Toro *et al.* 2022), y corroborando las inferencias de dichas partículas con la información que aporta Fuentes (2009) en su detallado diccionario. En esta guía de anotación se siguen criterios generales, positivos y negativos de los MD en el contexto de las cartas a los accionistas, aunque el corpus con el que trabajó el citado estudio incluía una cantidad de textos mayor, ampliado por cartas de años anteriores y posteriores al período de 2014-2017. En la guía de dicho estudio se contemplaron casi 300, de los cuales para nuestro objeto de investigación se simplificaron y se redujeron, teniendo en cuenta los marcadores que generaban ambigüedades y no se podían procesar, y aquellos que no podían incorporarse en nuestra clasificación. Tras esta selección, contamos para nuestro análisis con 231 marcadores del discurso diferentes (donde se aúnan las variantes en la medida de lo posible: *además/además de; por esa razón/por esta razón*, etc.).

Por otra parte, el principal problema con el que se topa cualquier estudio computacional de MD es que estos elementos son ambiguos en algunos contextos. Como ya hemos mencionado, son partículas que funcionan en un plano discursivo y desde una aproximación inferencial o procedimental, no gramatical. Esto significa que, al margen de cumplir una función discursiva en el discurso, muchos de ellos no terminan de desprenderse, en algunos contextos de su origen gramatical (normalmente adverbial). Por ello, hay que tener en cuenta que no todas las apariciones de las formas de un MD en el corpus son verdaderos MD. Para poder descartar estos falsos positivos, trabajamos sobre nuestros corpus con la herramienta LYNEAL<sup>5</sup>. De esta forma, los MD que pueden dar lugar a ambigüedades fueron buscados y filtrados. Pondremos un ejemplo: adverbio *así* funciona como conector consecutivo en algunos contextos (se equivale a *como consecuencia, de este modo, etc.*). Empero, son numerosos los contextos en los que mantiene su naturaleza como adverbio de modo, por lo que fue necesario establecer los parámetros de que se encontrara al inicio de un enunciado y seguido de una coma (.). Los contextos que resultados de esa búsqueda filtrada fueron exitosos y en todos ellos funcionaba como MD. Asimismo, tuvimos en consideración MD que comparten forma con otras palabras del español y que podían llevar a falsos positivos. Es el caso de *aparte* (*un capítulo aparte*, donde funciona como adjetivo), *con todo* (limitado a inicio de enunciado y seguido de coma, para evitar interferencias en que forme parte de otros segmentos sintácticos: *con todo a favor, podremos conseguir sanar esta situación*), *sobre todo* (limitamos su final de palabra, para no encontrarnos secuencias como: *destacamos sobre todos los sectores*), entre otros.

Por fin, una vez procesados por LYNEAL, para prescindir de las formas que no cumplieran con la función de marcador, nos dispusimos a extraer las estadísticas por marcador, grupos y subcategorías, con el fin de llegar a las conclusiones pertinentes mediante un análisis cuantitativo y cualitativo.

En el corpus analizado resultaban de especial interés los marcadores que tuvieran una frecuencia normalizada elevada, que podían indicarnos ciertas tendencias en la argumentación de las cartas. Para establecer el umbral de frecuencias altas, vimos que los MD con más de 100 ocurrencias por millón podían considerarse partículas recurrentes, teniendo en cuenta el porcentaje tan bajo que constituyen los marcadores del discurso en el corpus, puesto que es una estimación aproximada en relación con la frecuencia media que pueden tener estas partículas en el discurso. Esta ratio de 100 ocurrencias por millón es el que vamos a tener en cuenta en el análisis cuantitativo, y comentaremos en el análisis cualitativo algunos MD que cumplan con esta FN.

### 3.3. Tipo de análisis

Como acabamos de mencionar, en este estudio se ha tenido en cuenta una análisis cuantitativo y cualitativo de los datos. Consideramos que, para tener una certeza acerca de un tema, debemos basarnos en datos y, a la vez, que los datos deben explicarse y analizarse conforme a unos objetivos. Por ello, hemos combinado el análisis a través de las frecuencias absoluta (FA) y normalizada (FN), y el estudio de casos. En cuanto a las frecuencias,

5 <http://shimoda.lluf.uam.es/ueda/lyneal/lli-uam.htm>. LYNEAL (Letras y Números en Análisis Lingüísticos) “es un sistema de análisis de textos en español que facilita los procesamientos de datos textuales tanto de los archivos almacenados en el servidor como los propios del usuario” (Ueda 2018).

aunque se realizó la búsqueda de ambas, prescindimos de la FA porque para nuestro caso no es de utilidad, puesto que no nos permite comparar corpus de diferentes tamaños. La FN sí sirve para esta función y es la que hemos empleado en todas las búsquedas, de manera que los resultados obtenidos fueran lo más objetivos posibles. Para entender la frecuencia normalizada, presentamos su fórmula que explicaremos a continuación:

$$FN = \frac{\text{N.º casos de X}}{\text{N.º total de palabras del corpus}} \cdot \text{Base de normalización (1 millón)}$$

Figura 1. Fórmula de la Frecuencia Normalizada (FN).

La FN parte de dividir la FA de X (número de casos de un fenómeno), entre el total de las palabras del corpus y multiplicarlo por una base de normalización, en nuestro caso, un millón, siguiendo los parámetros que facilita la herramienta de LYENAL.

#### 4. RESULTADOS

Los resultados obtenidos tras el proceso de anotación y clasificación por tipos de estos marcadores del discurso con sus respectivas frecuencias han derivado en un análisis y las posteriores conclusiones acerca del comportamiento de algunos grupos en general y de algunos marcadores en particular. Para emprender dicho análisis, lo haremos con un procedimiento deductivo: iremos de lo general de los grupos a lo concreto de los marcadores en ganancias y en pérdidas.

Para construir una imagen general del corpus, antes de analizar las peculiaridades del comportamiento de los MD en ganancias y pérdidas, veremos cómo se distribuyen los grupos en todo el conjunto de los textos (Gráfico 1).

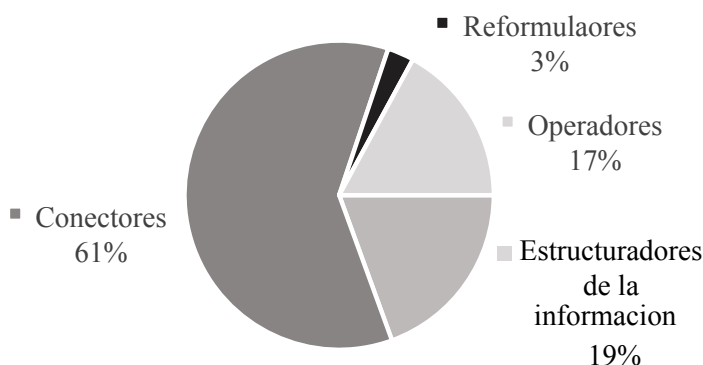


Gráfico 1. Porcentajes de grupos de MD en el corpus.

Según el *gráfico 1*, destacan ampliamente los conectores como el grupo de marcadores más frecuente en el corpus general. Estos marcadores relacionan semántica y pragmáticamente un



miembro del discurso con un miembro anterior de manera explícita o inferencial (Montolío 2001) y constituyen una pieza clave para el engranaje del discurso en cuanto a sus funciones de adición, contraargumentación, contraste o concesión, procesos de causa-consecuencia o de finalidad; en todas ellas se establecen relaciones muy productivas para este tipo de discurso. Los estructuradores de la información y los operadores son dos grupos similarmente frecuentes, si bien cumplen funciones muy distintas. En cuanto a los primeros, son unidades sin significado argumentativo y se limitan a “regular la organización informativa” (Portolés 2011: 137), es decir, cómo se presentan o se integran los miembros en el discurso; por su parte, los operadores, son el grupo que cuenta con más variedad de marcadores que cumplen con un amplio abanico de inferencias. Se estimaba que su presencia en el corpus fuera mayor, dado que se tiene en cuenta el tipo de enunciación, la modalidad, inferencias de atenuación o refuerzo argumentativo, etc., que están presentes asiduamente en cualquier tipo de discurso.

#### 4.1. Estructuradores de la información

Los estructuradores de la información, tanto en ganancias como en pérdidas, son engranajes del discurso y permiten organizar, secuenciar y comentar la información que se incluye en él. Los tematizadores introducen un nuevo tópico al discurso y, para esta inferencia son elegidos más MD en ganancias que en pérdidas (hay más variedad de marcadores); asimismo, las FN de los marcadores en ganancias son significativamente más altas que en pérdidas (Tabla 3):

<i>GANACIAS</i>		<i>PÉRDIDAS</i>	
<b>MD</b>	<b>FN</b>	<b>MD</b>	<b>FN</b>
Con respecto a	151.69	Con respecto a	268.76
Desde el punto de vista	106.19	En cuanto a	268.76
En cuanto a	326.14	En materia de	134.38
En el ámbito	235.12	En términos	134.38
En materia de	270.52		
En términos	313.50		

Tabla 3. Tematizadores.

Esto puede deberse a que se pretende introducir más recurrentemente la información en los textos de ganancias que en los de pérdidas, puesto que se aprovechan los buenos resultados para exponer los progresos y atraer o mantener la confianza de los inversores, como podemos ver en los siguientes ejemplos de ganancias en (1):

- (1) a. **Desde el punto de vista** de la Sociedad matriz, durante el Ejercicio 2015 ha mantenido el volumen de actividad con respecto al año pasado, cerrando el año con una cifra de negocios de 43,8 millones de euros, lo que le ha permitido obtener un resultado antes de impuestos de 1 millón de euros y un EBITDA de 2,4 millones de euros.
- b. **En cuanto a** la evolución de las principales magnitudes, la cifra de negocio del Grupo INYPSA ha alcanzado un importe de 19,2 millones de euros, lo que supone un aumento del 18% respecto al ejercicio anterior.

- c. **En el ámbito** de la gestión de personas, nos hemos adherido al Código de Principios de Diversidad Generacional del Observatorio Generación y Talento; hemos sido reconocidos por la entidad norteamericana WorldDataWork, por nuestra estrategia en el ámbito del desarrollo humano; y hemos obtenido el Certificado EFR Global, que nos acredita como empresa familiarmente responsable.

Los ordenadores nos aportan más información de la estructura del discurso y su organización. Por ejemplo, vemos que la secuencia de orden lógico se sigue con elementos relacionados como *en primer lugar* que cuenta con casi el doble de FN que sus continuadores, *en segundo lugar* y *en tercer lugar*. Sin embargo, esta lógica no se cumple con los distribuidores, del tipo *por un lado / por otro lado* o *por una parte / por otra parte*. En estos casos lo que hemos podido observar es que el segundo elemento de la distribución cuenta con más frecuencia que el primero (sobre todo en el par *por una parte / por otra parte*). Así mismo pasa con los ordenadores de cierre como *por último* o *finalmente*: lo entenderíamos como último elemento ordenador de una serie, pero comprobamos que no son necesarias partículas de ordenación anteriores, sino que se encarga de cerrar cualquier discurso independientemente de que se hayan numerado o no, previamente y de manera explícita, elementos de él. Sí es cierto que los ordenadores de cierre (*finalmente, para concluir, para finalizar, para terminar, por fin, por último*) albergan más variantes en el corpus de ganancias. Algunos de ellos, junto a sus frecuencias en ambos corpus, se exponen en la Tabla 4:

GANANCIAS		PÉRDIDAS	
MD	FN	MD	FN
En primer lugar	116.30	En primer lugar	172.77
En segundo lugar	60.68	En segundo lugar	38.39
En tercer lugar	25.28	En tercer lugar	19.20
Finalmente	136.52	Finalmente	191.97
Por otra parte	116.30	Por otra parte	230.37
Por otro lado	199.73	Por otro lado	115.18
Por último	240.18	Por último	326.35
Por un lado	126.41	Por un lado	115.18
Por una parte	25.28	Por una parte	38.39
Primero	20.23		
Segundo	5.06		

Tabla 4. Ordenadores de la información.

Con esto comprobamos la independencia que cobran las partes de una estructura formada por elementos distributivos, algo que sería un error de coherencia textual en una secuencia ordinal (tener *en segundo lugar*, sin haber tenido anteriormente una referencia a un primer elemento). Este fenómeno se cumple en igual medida en ganancias y en pérdidas, aunque cada subcorpus parece tener preferencia por partículas diferentes.

Es importante mencionar las frecuencias que nos encontramos en los ejemplificadores, aquellos marcadores que se encargan de aportar datos específicos y evidencias de la infor-

mación general que se está comentando, lo que es una marca de refuerzo argumentativo, realmente. En el caso de las cartas de ganancias, *por ejemplo* (2) destaca con una FN por encima de 100; como esperábamos, es la forma más escogida para representar este tipo de inferencias que permiten especificar y explicitar los buenos resultados o tendencia obtenidas:

- (2) a. Asimismo, hemos sido capaces de poner en valor activos en situaciones complejas, como **por ejemplo** la adquisición de Guadalmina a través de la compra de la deuda (...).
- b. La incorporación de las últimas tecnologías digitales y de gestión de datos nos están permitiendo encontrar, producir y suministrar recursos energéticos de forma rentable y respetuosa con el entorno, mejorando **por ejemplo** la tasa de éxito de nuestros proyectos exploratorios.

Por lo demás, lo que hemos podido observar es que el resto de los marcadores cuentan con una FN escasa y que, en las cartas correspondientes a las empresas con malos resultados, este tipo de marcadores apenas tienen ocurrencias. Como decíamos, los ejemplificadores muestran evidencias, explicitan datos y ejemplifican (valga la redundancia) la información; por ello, podemos pensar que la ausencia en el corpus de pérdidas puede deberse, precisamente, a que estas empresas no quieren incidir en ejemplificar los resultados desfavorables que han obtenido.

Para terminar de comentar el grupo de los estructuradores de la información, cabe señalar la ausencia de digresores en pérdidas, y la única ocurrencia en ganancias. Se podría conjeturar que las empresas con pérdidas harían un mayor uso de esta subcategoría, puesto que los digresores permiten abordar de una manera más indirecta los datos; por la razón que fuere, las empresas con malos resultados buscan de nuevo parecerse o exponer los resultados de la misma manera que lo hacen las empresas con resultados óptimos, y la digresión no tiene cabida en ninguno de los dos corpus.

#### 4.2. Conectores

Los conectores en el corpus son la categoría más variada y abundante. Algunos de ellos cuentan con frecuencias especialmente altas, pero, ante todo, cabe destacar las diversas formas con las que cuenta en las diferentes subcategorías. En los aditivos destacan soberanamente *además* y *también*, que pueden deber esta alta frecuencia a la combinación de una función argumentativa con una función aditiva, que constituyen dos inferencias de carácter argumentativo fuerte. Ambos introducen un nuevo argumento y, al mismo tiempo, añaden fuerza a la escala argumentativa; en los contraargumentativos destaca *pero*. Con este MD hay que tener que cuenta que puede cumplir una función aditiva, aunque son menos los casos, y no hemos podido tenerlos en cuenta para este estudio.

Del resto de partículas, queríamos reparar en *adicionalmente*, que no tiene una FN especialmente alta, pero llama la atención que su frecuencia normalizada de 122 en nuestro corpus, porque es un marcador que no se ha estudiado por los autores en el discurso español, ni se ha tenido en cuenta para las clasificaciones de MD en este idioma. Este fenómeno parece deberse a que es un calco de la correspondiente partícula en inglés, el adverbio *additionally*, proveniente, a su vez, del aditivo anglosajón *in addition*. Esta transferencia del inglés al español no sorprende en este tipo de textos donde las plantillas y los modelos los encontramos en los discursos escritos por las empresas inglesas; además, la terminología

financiera cuenta con una gran influencia inglesa y, por consiguiente, con gran cantidad de anglicismos, como comentamos en las características del discurso financiero. Los ejemplos encontrados en nuestro corpus se corresponden casi en su totalidad a partículas al margen (3ª y 3b) o en el comienzo de la oración (3c), aunque existen casos en los que se encuentran en la mitad de ella (3d). Aun así, es un marcador con un alto grado de gramaticalización y parece estar bastante fijado sintácticamente.

- (3) a. **Adicionalmente**, se han mantenido negociaciones con las distintas entidades financieras para rebajar los tipos de interés de nuestra financiación, así como el alargamiento de los plazos de vencimiento, principalmente en las áreas de negocio de Agua y Medio Ambiente.
- b. **Adicionalmente**, y gracias al reducido apalancamiento de estos activos, hemos refinanciado la deuda de proyecto de los parques en el pasado mes de diciembre, lo que ha supuesto una nueva financiación más eficiente y que ha incrementado la liquidez del Grupo.
- c. **Adicionalmente** a estos hitos alcanzados durante el año, quiero destacar la buena marcha de los trabajos actuales que avanzan a paso firme.
- d. Inversión a la que **adicionalmente** se han sumado numerosos propietarios, al constatar la capacidad de generación de valor de la nueva propuesta de NH Hotel Group.

El grupo más destacado de las empresas con buenos resultados han sido los aditivos. No es extraño que este grupo de marcadores acapare un porcentaje tan alto, puesto que aúnan las funciones de conexión y adición. Estas partículas hacen avanzar el texto, la argumentación e introducen nueva información, algo especialmente beneficioso para estas empresas que tienen tantos aspectos buenos y mejoras que destacar. Son igualmente imprescindibles en las empresas con pérdidas, si bien es cierto que, por lo general, no alcanzan frecuencias tan altas, ni cuentan con tanta variedad de MD (Tabla 5).

<i>GANANCIAS</i>		<i>PÉRDIDAS</i>	
<b>MD</b>	<b>FN</b>	<b>MD</b>	<b>FN</b>
A la vez	118.83	Además	883.07
Además	1079.55	Adicionalmente	153.58
Adicionalmente	118.83	Al igual que	115.18
Al mismo tiempo	204.79	Así como	806.28
Así como	614.36	Asimismo	191.97
Asimismo	348.89	Incluso	153.58
Igualmente	108.7	También	1574.17
Incluso	164.33		
También	2194.49		

Tabla 5. Conectores aditivos.

Los causales cuentan con pocos representantes de MD, por lo que las frecuencias se concentran más para cubrir todo el espectro de contextos en los que se pretende inferir causa

(Tabla 6) Hay una clara preferencia por *porque*, como MD causal por antonomasia. En cuanto al contraste entre los dos corpus, los causales no son un grupo de marcadores en el que se puedan apreciar realmente las diferencias entre la estructura argumentativa desde una perspectiva positiva u otra negativa. En cuanto a los consecutivos (Tabla 7), inferen una consecuencia que en el caso de las ganancias será destacable positivamente y en el caso de las pérdidas se centrarán en los aspectos logrados, por pocos que sean, e intentarán camuflar las consecuencias no tan buenas de su ejercicio.

<i>GANANCIAS</i>		<i>PÉRDIDAS</i>	
<b>MD</b>	<b>FN</b>	<b>MD</b>	<b>FN</b>
Dado que	25.28	Dado que	19.20
Porque	394.40	Porque	364.75
Pues	121.36	Ya que	287.96
Puesto que	30.34		
Ya que	230.07		

Tabla 6. Conectores causativos.

<i>GANANCIAS</i>		<i>PÉRDIDAS</i>	
<b>MD</b>	<b>FN</b>	<b>MD</b>	<b>FN</b>
Así,	192.14	Así,	422.34
Como consecuencia	235.12	Como consecuencia	422.34
Como resultado	98.60	De esta forma	134.38
Por ello,	139.05	Por tanto	307.15
Por tanto	176.98		

Tabla 7. Conectores consecutivos.

De esta apreciación podemos concluir la recurrencia a los procesos de causa-consecuencia en estos textos.

La inferencia de una causa conlleva implícitamente una consecuencia, y en las cartas encontramos gran cantidad de marcadores consecutivos por los cuales se lleva a cabo este proceso de causa-consecuencia. Son un subgrupo bastante numeroso tanto en ganancias como en pérdidas (siempre con más variantes en el primero, debido a las diferencias obvias de tamaño entre los corpus), pero, en cualquier caso, podemos inferir la recurrencia con la que aparecen los procesos de causa-consecuencia en nuestros textos, y que dicha relación impera en ambas direcciones. Aun así, priman las ocasiones en las que el foco no está puesto tanto en la causa de los acontecimientos pasados<sup>6</sup>, como en la consecuencia de lo que ha ocurrido, puesto que son los resultados de ello los que se presentan en cada informe

6 Los tiempos de pretérito son más frecuentes que los tiempos de futuro o condicional por el tipo de información que se aporta en este tipo de texto. Por su parte, el tiempo verbal de presente sigue siendo el imperante en las cartas a los accionistas (Carbajo 2021).

anual que se resume y se valora en las cartas. Algunos ejemplos de marcadores causales y consecutivos se muestran en (4), para el corpus de ganancias:

- (4) a. El año pasado por estas fechas, prometíamos a nuestros grupos de interés que lo mejor estaba por llegar y tengo el gran placer de afirmar que acertamos al hacer esa predicción optimista, **ya que** 2017 fue muy buen año para IAG y todas sus compañías operadoras.
- b. Quiero felicitar a los más de 200.000 profesionales que trabajan en el Grupo Santander, **puesto que** los logros alcanzados en 2017 no hubieran sido posibles sin la aportación de cada uno de ellos.
- c. **Como consecuencia** de los objetivos alcanzados en 2017, el Banco se sitúa en una posición sólida para encarar el Plan Director 2018-2020.
- d. Cerramos, **por tanto**, un ejercicio intenso en operaciones y negociaciones, en el que hemos conseguido conciliar y equilibrar los intereses de los accionistas de referencia con los de los minoritarios.

Como vamos a ver en (5), los ejemplos que encontramos en el corpus de pérdidas suelen centrarse en causas y consecuencias de aspectos que han mejorado o de los que van a mejorar, bajo el mensaje de “lo mejor está por venir”. Se acompañan con palabras como *saneamiento*, que actúa en un plano metafórico, como hemos comentado anteriormente, en el que se ve la empresa como un paciente herido a la que el presidente o primer ejecutivo va a sanar.

- (5) a. Me centraré en este último aspecto, **ya que** es indicativo de la solvencia y de la salud real del negocio.
- b. El saneamiento más importante corresponde a la inversión en el proyecto de plataforma tecnológica **porque** los nuevos responsables consideran más eficiente, más seguro y más rentable para la sociedad introducir modificaciones del diseño que son posibles por la evolución de la tecnología disponible.
- c. Pongamos, **por tanto**, amigos, el contador a cero porque comienza una nueva etapa, prometedora y emocionante.
- d. Eso, unido a una subida del precio medio del aceite superior al 50%, muestra la evolución subyacente de nuestra generación de caja. **Como consecuencia** de lo anterior, la deuda financiera neta ha subido en €17 mm con respecto al 31 de diciembre de 2013.

Los conectores concesivos han sido ubicados por muchos autores dentro del grupo de los conectores contraargumentativos. En este estudio, hemos podido comprobar que sus inferencias no son exactamente las mismas. Como bien apuntaba Montolío (2001), los concesivos, a los que la autora llamaba ‘contraargumentativos débiles’, ofrecen una oposición menos fuerte que los considerados ‘contraargumentativos fuertes’ (los contraargumentativos por excelencia). En su presencia en el corpus de ganancias, frente al de pérdidas, vemos que por primera vez en este análisis las cartas con pérdidas albergan un mayor número de variantes de este tipo de MD; no obstante, las frecuencias se advierten más bien similares (Tabla 8):

<i>GANANCIAS</i>		<i>PÉRDIDAS</i>	
<b>MD</b>	<b>FN</b>	<b>MD</b>	<b>FN</b>
Aun	409.57	A pesar de ello	19.20
Aun así	17.70	Aun	383.94
Aunque	326.14	Aun así	57.59
Pese a	111.24	Aunque	326.35
Si bien	96.07	Pese a	115.18
		Si bien	153.58

Tabla 8. Conectores concesivos.

Para ver cómo se comparten en el corpus, atendemos a los ejemplos de (6) para ganancias y de (7) para pérdidas:

- (6) a. **Aunque** ha sido un año positivo para todas las aerolíneas de IAG, las cuatro continúan firmemente orientadas a lograr un crecimiento rentable en el futuro impulsado por los ambiciosos objetivos que hemos fijado para ellas.
- b. **Pese a todo**, la fortaleza de la demanda interna y la propia inercia de la economía, me hacen albergar fundadas esperanzas de que durante 2017 continuaremos en la senda de crecimiento económico y creación de empleo.

En las empresas con ganancias también hay puntos débiles, y estos marcadores ofrecen un contraste menos fuerte de los argumentos negativos. Se establece una condición que ha ocurrido, pero hay un argumento que se sobrepone a ello, manteniendo el balance en positivo. En pérdidas, vemos cómo estas partículas se utilizan para llevar a cabo una justificación de los malos resultados y, a la vez, albergan la esperanza de mejorarlos.

- (7) a. Por todo ello, quisiera aprovechar la oportunidad para transmitirles mi entera confianza en el resto de proyectos que tenemos en marcha en la compañía **pese a** estas noticias que lamentablemente no han salido como esperábamos.
- b. Con todo, el enlace será suficiente, **aunque** la volatilidad de precios hasta octubre será muy fuerte.

En cuanto a los conocidos contraargumentativos (contraargumentativos fuertes), podemos observar que destacan las mismas variantes en ganancias y en pérdidas, aunque en este segundo corpus alcanzan unas FN más altas:

<i>GANANCIAS</i>		<i>PÉRDIDAS</i>	
<b>MD</b>	<b>FN</b>	<b>MD</b>	<b>FN</b>
A pesar de	401.99	A pesar de	403.14
No obstante	91.02	No obstante	115.18
Pero	1074.49	Pero	1343.80
Sin embargo	232.60	Sin embargo	345.55

Tabla 9. Conectores contraargumentativos.

Los datos que se exponen en la Tabla 9 muestran la preferencia de los conectores contraargumentativos por las empresas con pérdidas. Esto puede acercarnos a la conclusión de que las empresas con pérdidas, que hasta el actual punto del análisis habían asemejado sus FN a las empresas con ganancias, no pueden ocultar los malos resultados ni su tratamiento a los accionistas en lo relativo a este grupo de MD. Las construcciones contraargumentativas son necesarias para exponer dos argumentos, de los cuales prevalece el segundo, anulando con ello el primero. Para ganancias encontramos ejemplos como en (8), donde el segundo segmento de la contraargumentación, que es el que prevalece, muestra un impacto positivo en los resultados. En los casos de pérdidas (9) podemos ver que, aunque las empresas que no han tenido beneficios presentan unos resultados optimistas para convencer a los accionistas y potenciales inversores de la posibilidad de revertir la situación negativa o de explicar las causas (que no deben repetirse) por las que, desgraciadamente, se ha llegado a ella.

- (8) a. **Sin embargo**, gracias al esfuerzo y dedicación del gran equipo humano de ENDESA, hemos superado este difícil contexto y, por ello, podemos presentar unos resultados anuales muy satisfactorios.
- b. **No obstante, pese a** cierto escepticismo inicial, los objetivos han adquirido una credibilidad mayor a medida que nuestros planes se han puesto en marcha.
- (9) a. Esto, sin duda, ha supuesto un contratiempo para la compañía. **No obstante**, me gustaría destacar algunos puntos a este respecto que creo son importantes.
- b. Descontado este factor, las ventas bajaron un 3%, que, en todo caso, ha supuesto ganar cuota de mercado. La situación de Francia no es fácil, **sin embargo**, hemos conseguido batir la tendencia general del mercado.

No destacan en demasía los conectores orientados a inferir una finalidad. Los conectores de finalidad no parecen tener una función clave en el discurso de las empresas con ganancias, pero sí cobran algo más de importancia en las cartas de las empresas con pérdidas. Podemos destacar, por encima de todos los demás, la FN de *con el objetivo de* y *con el fin de* (Tabla 10), aunque este último pasa más desapercibido en compañías con beneficios. En resto de MD siguen el comportamiento esperado y las diferencias entre ganancias y pérdidas, en este grupo, son relativamente tenues.

GANANCIAS		PÉRDIDAS	
MD	FN	MD	FN
Con el fin de	65.73	Con el fin de	95.99
Con el objetivo de	214.90	Con el objetivo de	172.77

Tabla 10. Conectores de finalidad.

Estos dos marcadores son los encargados de guiar las inferencias de finalidad en el discurso y, parece ser, que la finalidad tiene ligeramente más relación con aspectos positivos que con aspectos negativos.

Eso sí, la finalidad empleada en uno y otro corpus no conlleva realmente las mismas inferencias. Podríamos decir que, en su caso de las ganancias (10), las medidas tomadas



no tienen que ser justificadas, porque se consideran buenas, y los aspectos a mejorar son escasos, por lo que no se presentan muchos aspectos cuyo fin deba ser la solución a un problema con el que se ha encontrado la empresa. Sí ocurre esto en las pérdidas (11), donde la finalidad va ligada a aspectos negativos que deben explicarse o aspectos futuros que se pretenden conseguir para mejorar:

- (10) **Con el fin de** consolidar esta mejoría estamos acometiendo importantes inversiones en nuestros talleres a fin de hacerlos aún más competitivos en sus respectivas áreas de negocio.
- (11) Tras más de 70 años de trayectoria, nuestra compañía se enfrentó a una gran crisis que obligó a poner en marcha una reestructuración financiera y de nuestro modelo de negocio **con el objetivo de** lograr la Abengoa del futuro que por fin vislumbramos hoy.

### 4.3. Reformuladores

Los reformuladores son el grupo menos numeroso en cuanto a variedad y frecuencias. Este grupo lo constituyen marcadores que recuperan una parte del discurso o el discurso al completo, para explicarlo, rectificarlo, distanciarse de la información o recapitular lo expuesto. No cuentan en ningún caso con una frecuencia alta, a excepción del recapitulador *en definitiva* y el explicativo *es decir* (Tabla 11):

GANANCIAS		PÉRDIDAS	
MD	FN	MD	FN
En todo caso	30.34	En todo caso	134.38
Es decir	111.24	Es decir	134.38
En definitiva	171.92	En definitiva	172.77

Tabla 11. Reformuladores.

Los reformuladores son un grupo restringido de marcadores y guían inferencias muy concretas. En ninguno de los dos corpus tienen especial protagonismo, pero cabe mencionar la variedad de marcadores que encontramos dentro de los reformuladores de ganancias, aunque sus frecuencias sean bajas (16 MD en total), frente a las pocas variantes de pérdidas (tan solo 7 MD). Los marcadores de este grupo se comparten de manera muy similar en ganancias y pérdidas, pero querríamos destacar la alta frecuencia normalizada de *en todo caso* (12) en el corpus de pérdidas, un marcador de distanciamiento (*tabla 11*). Esto nos hace reparar en la función del distanciamiento textual por parte de los interlocutores que se encuentran en la tesitura de minimizar o de distanciarse, todo lo posible, de los datos negativos de su corporación. En general, los reformuladores parecen tener algo más de presencia en aquellas cartas que deben “rectificar” información por sus datos defectuosos:

- (12) Los datos disponibles, basados en las existencias y en las salidas mensuales, hacen prever un verano muy volátil. **En todo caso**, desde aquí quiero llamar la atención de las autoridades para que vigilen que los datos aportados a la Agencia de Información y Control Alimentarios sean veraces y que no se conviertan en una herramienta de especulación; hemos visto declaraciones en las que aparecen incongruencias y, como consecuencia, se han producido subidas de precio.

#### 4.4. Operadores

Al contrario de los reformuladores, los operadores constituyen una categoría de marcador muy amplia y heterogénea, donde sus inferencias actúan sobre diversos aspectos del discurso. La función de los operadores discursivos comprende mayoritariamente resaltar algún segmento del discurso del que se infiere o se aporta alguna información: o bien un grado en la escala argumentativa (desde la atenuación, al refuerzo argumentativo), un ejemplo o concreción, la perspectiva del hablante o su grado de seguridad. Los que tienen más peso parecen ser los argumentativos y los modales, que cuentan con la mayor diversidad y las frecuencias más altas.

En los argumentativos destacan *especialmente*, *principalmente*, *fundamentalmente* y *sobre todo* (Tabla 12). Estos marcadores, aparentemente, cumplen una misma función de destacar un segmento del discurso. Además, algunos de ellos reparten sus inferencias entre la de destacar un elemento para enfatizarlo (*especialmente favorable*); y su función de concretar, dentro de un grupo, una parte de él (*especialmente en Europa*) donde, de todas las opciones que se proponen, se concreta en una.

GANANCIAS		PÉRDIDAS	
MD	FN	MD	FN
Especialmente	437.38	Especialmente	460.73
Fundamentalmente	121.35	Fundamentalmente	153.58
Principalmente	212.37	Principalmente	383.94
Sobre todo	189.62	Sobre todo	115.18

Tabla 12. Operadores argumentativos.

Este subgrupo de marcadores parece ser clave para los engranajes argumentativos. Por otra parte, es relevante la ausencia de MD de atenuación en el corpus de pérdidas. Precisamente se esperaría marcar de alguna manera la atenuación para reducir la fuerza ilocutiva con la que se transmiten algunos de los malos resultados. No parece que tengan necesidad de hacer uso de estas marcas de atenuación, lo que nos acerca un poco más a pensar que, o bien omiten o camuflan estos resultados, o bien enmarcan los resultados en estructuras contraargumentativas, dándole más fuerza argumentativa al dato o conclusión positiva, que a la parte negativa.

En cuanto a los operadores de concreción, por su parte, no hay ninguno que sobrepase las 100 ocurrencias de FN, pero el marcador que más se acerca a ello es *en particular*, seguido de *en concreto* (Tabla 13). Parecen ser los elementos elegidos por la mayoría de las empresas para expresar concreción. Las empresas con pérdidas vuelven a contar con menos variantes en este grupo, pero se repiten los MD que destacan en ganancias.

GANANCIAS		PÉRDIDAS	
MD	FN	MD	FN
Concretamente	48.04	Concretamente	19.20
En concreto	80.90	En particular	76.79

GANANCIAS		PÉRDIDAS	
MD	FN	MD	FN
En particular	98.60	En concreto	57.59
Específicamente	12.64		
Particularmente	63.21		
Sin ir más lejos	2.53		

Tabla 13. Operadores de concreción

En los enunciativos se emplea mayoritariamente el enunciativo *en general*, que marca lo contrario a la concreción que veíamos en los operadores de concreción; y *prácticamente* (13), que es un aproximativo al decir e indica “casi con exactitud” de la información aportada. Esta cercanía o casi coincidencia con el elemento de referencia al que acompaña puede entenderse como una escala de intensificación o cuantificación o, incluso, como un atenuativo del decir, que encontramos casi con el doble de frecuencia en pérdidas (230 de FN) que en ganancias (131 de FN):

- (13) De esta forma la compañía ha alcanzado **prácticamente** el punto de equilibrio en el resultado de explotación (EBIT), comparado con las pérdidas de 131 millones del año anterior.

Entre ellos, las FN más altas en ambos subcorpus atienden a locuciones adverbiales: *por supuesto* y *sin duda*. Los modales pierden presencia en el subcorpus de pérdidas, parece ser que la evidencia, la certeza o la seguridad del hablante disminuyen en estos textos, donde el hablante no tiene tantas buenas noticias que comunicar, ni puede reafirmarse en la información que aporta. Estimamos que, los ejemplos a los que remiten este tipo de marcadores en el corpus con pérdidas (14) son, precisamente, segmentos en los que se pretende convencer, dar confianza y persuadir a los accionistas para que sigan invirtiendo:

- (14) a. Y, por último, queremos ejecutar una serie de medidas de eficiencia operativa que ya tenemos preparadas y que impactarán muy positivamente tanto en la eficiencia como en el servicio que ofrezcamos, y, **por supuesto**, en nuestra rentabilidad.
- b. **Sin duda**, unos logros que no hubieran sido posibles sin la dedicación y compromiso de nuestros equipos profesionales y la confianza que, como accionista, ha depositado en nosotros.

Otros se relacionan con el grado de seguridad del hablante y actúan en el plano de la modalidad, también en la modalidad emotiva, aunque encontramos pocos ejemplos de *afortunadamente*, *desafortunadamente*, *desgraciadamente* y *lamentablemente* (Tabla 14).

GANANCIAS		PÉRDIDAS	
MD	FN	MD	FN
Afortunadamente	15.17	Afortunadamente	95.99
Desafortunadamente	2.53	Desgraciadamente	19.20

GANANCIAS		PÉRDIDAS	
MD	FN	MD	FN
Desgraciadamente	2.53	Lamentablemente	19.20
Lamentablemente	10.11		

Tabla 14. Operadores enunciativos.

Aun así, sorprende la alta frecuencia con la que aparece el marcador de modalidad emotiva *afortunadamente* (15) en las empresas con pérdidas. Esta partícula expresa alegría, reconocimiento de un sentimiento positivo, y vemos que su comportamiento en las cartas de pérdidas tiene esa función de balance, contraste y equilibrio de cara a los resultados negativos que haya podido tener la compañía. De esta manera, aunque los datos que se presenten no sean muy favorables, marcando ciertos segmentos del discurso con *afortunadamente*, se alberga un sentimiento de alegría, esperanza y tranquilidad que disminuye, o incluso anula, la información negativa expuesta.

- (15) a. **Afortunadamente** la cosecha esperada en los países productores del hemisferio sur va a ser record, unas 100.000 tns.
- b. **Afortunadamente**, y con gran satisfacción, puedo decir que el año 2014 ha sido un año de tranquilidad tras los duros años que hemos pasado.

El resto de los MD con modalidad emotiva negativa pasan desapercibidos en ambos corpus. En definitiva, a una empresa, haya tenido ganancias o haya tenido pérdidas, le interesa mostrar seguridad y esperanza en sus resultados de cara a sus accionistas.

Los operadores de refuerzo argumentativo pueden compartir inferencias con los mencionados operadores de modalidad. En algunos casos como en *desde luego*, se habla de un operador que, efectivamente, reafirma un segmento del texto. Esto refuerza argumentativamente dicho segmento, pero lo hace desde un plano de evidencia y seguridad expresada por el hablante. Aun así, los dos MD que cuentan con más FN de este grupo, pese a que no alcancen ni las 100 ocurrencias, son *de hecho* y *precisamente*, los reforzadores argumentativos por excelencia. Vemos que el refuerzo argumentativo no se expresa a través de MD como se esperaría, pero debe estar explicitado de otra manera en el texto (como ocurre en el discurso político), puesto que consideramos que es una inferencia importante en un texto de esta naturaleza.

## 5. CONCLUSIONES

En este trabajo se han analizado dos corpus de manera paralela atendiendo a los marcadores discurso. Primero se han abordado algunos problemas y se han tenido en cuenta las consideraciones teóricas para llevar a cabo un análisis estadístico útil y formal. A su vez, el estudio cuantitativo se ha complementado en algunas ocasiones con un estudio de casos. Con los resultados obtenidos de ambos corpus y su contraste, podemos esbozar las siguientes conclusiones:

- a. Los marcadores son unidades que constituyen un porcentaje muy pequeño del discurso, pero se advierten imprescindibles para la construcción de este.
- b. Los conectores y los operadores son los grupos más frecuentes y con mayor variedad de marcadores, y parecen ser la clave para encontrar las mínimas diferencias entre ambos discursos.
- c. Las ganancias están marcadas por conectores aditivos, que hacen avanzar el discurso, y los operadores de refuerzo argumentativo, así como los enunciativos, puesto que son textos en los que se pretenden reforzar los buenos resultados.
- d. Las cartas de las empresas con pérdidas, pese a su esfuerzo por maquillar los resultados en el discurso u omitirlos, deben guiar inferencias de acuerdo con sus datos, por lo que destacan los conectores contraargumentativos (también los concesivos, aunque en menor medida, puesto que conforman una oposición más débil), y los operadores de modalidad emotiva, destacando *afortunadamente* que no sobresale en las empresas con ganancias.
- e. Los conectores causales y consecutivos reflejan con sus altas frecuencias la importancia de los procesos causa-consecuencia en este discurso. El foco apunta principalmente a las consecuencias, siendo mucho más numerosos los consecutivos. Para las empresas con ganancias el empleo de este tipo de marcadores supone presentar las consecuencias de su buen ejercicio; sin embargo, las empresas con pérdidas deben buscar la manera de desviar esas consecuencias negativas, y el recurso más empleado para ello es la metáfora.
- f. Los reformuladores, en contra de nuestro pronóstico, no cuentan con gran protagonismo en ninguno de los dos corpus, cuando sería esperable encontrarlos estos marcadores en el corpus de pérdidas, principalmente.
- g. Los operadores argumentativos, ligados a la fuerza en una escala argumentativa como los de refuerzo argumentativo, son más partícipes en las empresas con pérdidas. La razón de ello puede residir en la necesidad de focalizar la fuerza argumentativa en determinados elementos, pero, a diferencia de los MD de refuerzo argumentativo, estos argumentativos focalizados no pueden limitar los demás argumentos presentados como desencadenantes de posibles conclusiones. Por ello, de cara a la argumentación y la confección de conclusiones, no son tan determinantes y pueden utilizarse para “maquillar” el discurso.
- h. Hemos comprobado la influencia anglosajona en este tipo de discurso al recoger una frecuencia alta de *adicionalmente*, un marcador no incluido en las clasificaciones de marcadores del discurso en español, cuya procedencia parece estar en el adverbio y marcador del inglés *additionally* y se fija mayoritariamente al comienzo de la oración.
- i. En definitiva, a las compañías les interesa mostrar seguridad y esperanza independientemente de sus resultados. Esto lo consiguen a través de operadores modales, que son frecuentes tanto en ganancias como en pérdidas.

Podemos concluir que el objetivo final de este análisis era ver si podríamos dilucidar diferencias entre el discurso financiero de una empresa con ganancias de otra con pérdidas a través de los marcadores discursivos empleados, considerándolos engranajes imprescindibles del discurso y su argumentación. Desafortunadamente, aunque hemos alcanzado conclusiones interesantes a nivel lingüístico, se ha observado que los resultados y las frecuencias entre

los dos subcorpus, salvando su diferencia de tamaño, son muy similares. Esto nos acerca a pensar que, o bien siguen una estructura discursiva muy parecida en la que las empresas con pérdidas intentan asemejar la estructura de sus cartas a la de las empresas con ganancias para camuflar en la medida de lo posible sus resultados; o bien que, efectivamente y como ya apuntaban otros autores (Carbajo, 2021), las diferencias entre los textos de los que han generado beneficios y los que han generado pérdidas no se encuentran a un nivel léxico o gramatical, sino que probablemente se hallen a un nivel discursivo. Como es realmente complicado dilucidar las diferencias entre los textos de ganancias y de pérdidas a través de rasgos lingüísticos, creemos que estas diferencias deben hallarse a un nivel discursivo mediante la omisión de información y datos contables, como podemos ver por la poca frecuencia de los ejemplificadores en las empresas con pérdidas. Por ende, la argumentación y la disposición o presentación de los argumentos en este tipo de textos es crucial para analizarlos.

Los resultados del presente estudio ponen de relieve la necesidad de una ampliación de las investigaciones en argumentación y análisis de la narrativa financiera, en concreto de las cartas a los accionistas. Asimismo, se pretende para un trabajo futuro llevar a cabo un clasificador automático de marcadores del discurso, que mejore el actual detector de MD, con el fin de facilitar la extracción de las inferencias que marcan el texto y, con ello, ayudar a la obtención de información del discurso, la realización de resúmenes automáticos o a la profundización en el estudio de la argumentación.

## REFERENCIAS BIBLIOGRÁFICAS

- Alcaraz Varó, E., B. Hughes y J. Mateo Martínez. (2012). *Diccionario de términos económicos, financieros y comerciales inglés-español, Spanish-English* (6 ed.). Barcelona: Ariel.
- Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511621024>.
- Biber, D., S. Conrad y G. Leech. (2002). *Student Grammar of Spoken and Written English*. Edimburgo: Longman.
- Biber, D., S. Conrad y R. Reppen. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511804489>.
- Biber, D. y S. Conrad. (2009). *Register, genre, and style*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511814358>.
- Biber, D., S. Johansson, G. Leech, S. Conrad y E. Finegan. (1999). *The Longman Grammar of Spoken and Written English*. London: Longman.
- Biber, D. y R. Reppen (eds.). (2015). *The Cambridge Handbook of English Corpus Linguistics*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139764377>.
- Bhatia, V. (2008). Towards critical genre analysis. En V. Bhatia, J. Flowerdew y R. H. Jones (Eds.). *Advances in Discourse Studies*. Londres/Nueva York: Routledge, 166-177. <https://doi.org/10.4324/9780203892299>.
- Briz, A., S. Pons y J. Portolés (coords.) (2008). *Diccionario de partículas discursivas del español*. En línea [www.dpde.es](http://www.dpde.es).
- Bujaki, M. y B. McConomy. (2012). Metaphor in Nortel's letters to shareholders 1997-2006. *Accounting, Auditing & Accountability Journal*, 25, 7, 1113-1139. <https://doi.org/10.1108/09513571211263211>.
- Carbajo Coronado, B. (2021). *¿Ganancias o pérdidas? Indicadores léxicos y morfosintácticos en cartas de empresas a los accionistas*. En Servicio de Publicaciones de la Universidad Autónoma de Madrid, curso 2020-2021 <Publicaciones | UAM>.

- El-Haj, M., P. Rayson, M. Walker, S. Young y V. Simaki. (2019). *Search of meaning: Lessons, resources and next steps for computational analysis of financial discourse*. *Journal of Business Finance y Accounting*, 46, 265-306. <https://doi.org/10.1111/jbfa.12378>.
- Fuentes Rodríguez, C. (2009). *Diccionario de conectores y operadores del español*. Madrid: Arco Libros.
- Garcés Gómez, M.P. (2008). *La organización del discurso: marcadores de ordenación y de reformulación*. Madrid: Iberoamericana / Frankfurt am Main: Vervuert. <https://doi.org/10.31819/9783865278661>.
- García Toro, A., J. Porta Zamorano y A. Moreno-Sandoval. (2022). A Discourse Marker Tagger for Spanish using Transformers. En *Procesamiento Del Lenguaje Natural*, 68, 123-132. Recuperado de: <http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/6411>.
- Gisbert Llorca, A. (2017). *La traducción de textos económico-comerciales: el género informe anual*. Alicante: Universidad de Alicante.
- Hyland, K. (1998). Exploring Corporate Rhetoric: Metadiscourse in the CEO's Letter. *The Journal of Business Communication* (1973), 35 (2), 224-244. <https://doi.org/10.1177/002194369803500203>.
- Landone, E. (2012). La clasificación de los marcadores del discurso y su valor operativo. En *Cassol, A., XXIV Congreso AISPI*. Roma: AISPI Edizioni, 431-444.
- Llamas Saiz, C., C. Martínez Pasamar y C. Taberner Sala. (2012). La comunicación académica y profesional. Usos, técnicas y estilo. Pamplona: Thomson Reuters / Aranzadi, 140-141.
- Loughran, T. y B. McDonald. (2016). Textual Analysis in Accounting and Finance: A Survey. *Journal of Accounting Research*, 54(4). <https://doi.org/10.1111/1475-679X.12123>.
- Loureda, Ó. y Acín, E. (coords.) (2010), *Los estudios sobre marcadores del discurso en español hoy*. Madrid: Arco Libros.
- Martín Zorraquino, M.A. y J. Portolés. (1999). Los marcadores del discurso. En *Gramática descriptiva de la lengua española* (pp. 4051-4214). Madrid: Espasa Calpe.
- Mateo, J. 2007. El lenguaje de las ciencias económicas. En E. Alcaraz, J. Mateo y F. Yus (Eds.). *Las lenguas profesionales y académicas* (pp. 191-203). Barcelona: Ariel.
- Montolio Durán, E. (2001). *Conectores de la lengua escrita*. Barcelona: Ariel.
- Moreno-Sandoval, A., Gisbert, A., y Montoro, H. (2020). FinT-esp: A corpus of financial reports in Spanish. En *Multiperspectives in analysis and corpus design*. Granada: Comares, 93-106.
- Moreno-Sandoval, A., A. Gisbert, P. A. Haya, M. Guerrero y H. Montoro. (2019). Tone analysis in Spanish financial reporting narratives. En M. El-Haj, P. Rayson, S. Young, H. Bouamor y S. Ferradans (Eds.), *Proceedings of the Second Financial Narrative Processing Workshop (FNP 2019)* (pp. 42-50).Turku: Linköping University Electronic Press.
- Pons, S. (2000). Los conectores. En A. Briz y Val.Es.Co (Eds.), *¿Cómo se comenta un texto coloquial?* (pp. 193-220). Barcelona: Ariel.
- Portolés, J. (2011). *Marcadores del discurso*. Barcelona: Ariel. <https://doi.org/10.31819/9783865278760-003>.
- Portolés, J. (2015). Los marcadores del discurso. En Gutiérrez-Rexach, J. (ed.), *Enciclopedia Lingüística Hispánica*. Vol. 1 (pp. 689-699). Londres / Nueva York: Routledge. <https://doi.org/10.4324/9781315713441-61>
- Redondo Redondo, A. (2018). *Análisis del discurso económico en el ámbito político de la crisis del euro: un enfoque semántico-pragmático*. TFM. Soria: Universidad de Valladolid.
- Ueda, H. (2018). *Cómo usar LYNEAL*. En línea <<http://shimoda.llif.uam.es/ueda/lyn153/doc/how-to-es.pdf>>
- Ueda, H. y A. Moreno-Sandoval. (2017). *LYNEAL: Letras y Números en Análisis Lingüísticos*. Recuperado de <http://shimoda.llif.uam.es/ueda/lyneal>.

## Bloque II

*Corpus y sus aplicaciones*





## El corpus Iberia como fuente para el estudio del léxico especializado

The IBERIA corpus as a source for the study of technical Spanish vocabulary

LUIS PABLO NÚÑEZ  
Universidad de Granada, España  
luispablo@ugr.es  
<https://orcid.org/0000-0002-6491-6873>

### Resumen

Este texto expone las posibilidades de búsqueda y recuperación de información del corpus IBERIA, el corpus sincrónico del español creado por el Consejo Superior de Investigaciones Científicas para el estudio del léxico técnico del español. Tras describir sus características, realizamos búsquedas de un número cerrado de términos de los ámbitos astronómico y médico con mayor y menor grado de especialización; posteriormente buscamos esos mismos términos en los corpus de la Real Academia Española y del IULA y comparamos sus resultados. De acuerdo con ellos, mostramos el alcance y posibles limitaciones de estos corpus para el estudio del léxico especializado.

**PALABRAS CLAVE:** Léxico especializado; vocabulario científico-técnico; corpus IBERIA; Consejo Superior de Investigaciones Científicas; CSIC

### Abstract

The search options and retrieval possibilities of the IBERIA corpus for the study of technical Spanish vocabulary are described in this text. After analysing the characteristics of this synchronous corpus created by the Spanish National Research Council (Consejo Superior de Investigaciones Científicas, CSIC), we undertook a number of searches for a determinate number of terms from the fields of astronomy and medicine with a higher and lower degree of specialisation. We then compared the results obtained from the same searches using the corpora of the Institute of Applied Linguistics (Institut Universitari de Lingüística Aplicada, IULA, UPF) and the Royal Spanish Academy (Real Academia Española, RAE). Accordingly, we demonstrate the scope and possible limitations of these corpora for the study of specialised lexicons.

**KEYWORDS:** Specialised lexicon; technical vocabulary; IBERIA corpus; Spanish National Research Council; CSIC

---

**Para citar este artículo:** Pablo Núñez, L. (2022). El corpus Iberia como fuente para el estudio del léxico especializado. *ELUA*, (38), 217-240. <https://doi.org/10.14198/ELUA.22524>

Recibido: 21/04/2022 Aceptado: 17/05/2022

© 2022 Luis Pablo Núñez



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

## 1. INTRODUCCIÓN

### 1.1. El uso de los corpus hoy

Desde la aparición de la Lingüística de corpus hace unos sesenta años (corpus de Brown, corpus del inglés británico LOB, 1963, *cf.* Halliday *et al.* 2004: 50-58), la compilación informática de datos para su posterior explotación no ha dejado de aumentar, bien desde el punto de vista cuantitativo como cualitativo.

Si esto ha sido así es debido a su utilidad: la recolección y tratamiento de textos ha permitido llevar a cabo todo tipo de investigaciones lingüísticas, ya sean normativas o descriptivas, desde la combinación y frecuencia de formas léxicas a estudios sincrónicos y diacrónicos sobre morfología, sintaxis, pragmática, sociolingüística y traducción, plasmados luego en diccionarios y gramáticas. Las disciplinas más beneficiadas por los estudios de corpus son, de hecho, la lexicografía y la traducción, y de aquí las estrechas interrelaciones que se producen entre los corpus, los bancos de datos terminológicos y los vocabularios especializados.

### 1.2. La evolución de los corpus

Más allá de las utilidades de la lingüística de corpus, ya reconocidas (Martín Herrero 2009), para llegar a nuestro análisis nos interesa resaltar cómo han evolucionado los corpus.

Los avances en la Lingüística como disciplina permiten considerar nuevos planteamientos metodológicos, y de la misma manera ocurre con los corpus: mejoran cada vez más su desarrollo y sus parámetros técnicos y permiten considerar nuevos enfoques, ya sea desde el punto de vista cuantitativo (la manipulación de más datos), ya cualitativos (inclusión de marcados y desarrollos gráficos). Así, tras la aparición de los corpus etiquetados y la integración paulatina del audio en corpus sincronizados, ha surgido la modalidad de corpus basados en internet, que se basan en porciones de la red —como los corpus basados en Wikipedia del IULA<sup>1</sup> o los incluidos en IntelliText de la Universidad de Leeds<sup>2</sup>— o permiten buscar directamente en las páginas web a través de los metadatos —como la versión 2 del corpus de Davies<sup>3</sup>—.

Frente a los corpus cerrados, creados con un conjunto de textos sobre los que realizar consultas, limitados por un periodo cronológico o una temática, los corpus de internet suelen ser corpus abiertos, que se actualizan de forma continua e incorporan el marcado a los contenidos recientes de forma automática. Es en buena medida el caso del corpus IBERIA que veremos a continuación, si bien su origen parte de un corpus específico de textos.

---

1 Universitat Politècnica de Catalunya. Research Group on Natural Language Processing; Gemma Boleda; Universitat Pompeu Fabra. Institut Universitari de Lingüística Aplicada (IULA): *GrAF version of Spanish portions of Wikipedia Corpus*, 2012. El corpus, de más de 150 millones de palabras, toma 257 019 artículos del año 2006 de la Wikipedia en español.

2 IntelliText 2.6: Página electrónica de consulta de corpus del Centre for Translation Studies, University of Leeds: <http://corpus.leeds.ac.uk/itweb/htdocs/Query.html>.

3 Mark Davies, *Corpus del español*: <https://www.corpusdelespanol.org/>.

## GrAF version of Spanish portions of Wikipedia Corpus



The image shows a digital repository interface. On the left, there is a file icon labeled 'RAR'. Below it, the document type is 'Otros' and the date is '2012-05-16'. There is a link to 'Ver Estadísticas de Ítem' and social media icons for Facebook, Twitter, Email, and YouTube. On the right, there is a 'Descripción' panel with a green button 'Consultar el documento'. The description panel contains the following text:

**Citación:**  
 Universitat Politècnica de Catalunya.  
 Research Group on Natural Language  
 Processing; Gemma Boleda; Universitat  
 Pompeu Fabra. Institut Universitari de  
 Lingüística Aplicada (IULA). GrAF  
 version of Spanish portions of Wikipedia  
 Corpus. 2012  
<http://hdl.handle.net/10230/20047>

Figura 1. Corpus descargable de Wikipedia en la muestra de 2006 ofrecida por el IULA (*GrAF version of Spanish portions of Wikipedia Corpus*). Fuente: <http://hdl.handle.net/10230/20047>.

Otro ejemplo de la evolución de los corpus es la actualización de interfaces y la ampliación de los campos de búsqueda. Un ejemplo significativo es el de IATE (<https://iate.europa.eu/>), antiguo *Eurodicautom*, que en su nueva interfaz incorpora nuevos criterios de búsqueda recuperación por frase, por forma corta, por nivel de lengua.

Otro aspecto interesante de los corpus para el posterior estudio lingüístico es la descarga libre que ofrecen algunos —como el BNC-British National Corpus o el mismo IATE— y la posible integración de sus textos en otros corpus. Esta característica puede dar lugar a corpus paralelos, a enriquecimientos de nuevos corpus o bien a la interrogación paralela en varios a la vez. Busca, en todo caso, fomentar la explotación libre.

En estos momentos este contenido solo existe en inglés.

## Descargar IATE

IATE is a living database, i.e. translators and terminologists are continuously updating its content. Using the IATE search interface (<http://iate.europa.eu/>) or the IATE search web service thus ensures that you are accessing the most complete and up-to-date data. However, in order to cater for specific needs, e.g. for linguistic research, you can also download a copy of some of the data contained in IATE.

The distribution consists of a zip file with the following name: IATE\_download.zip .

An extraction tool named IATEExtract is made available on this site in order to help users create subsets of the IATE download file, using a number of possible filtering criteria. Users can extract data for one or more specific languages, for a given domain or domain cluster. The subsets created by the extraction tool IATEExtract are provided in the same **TermBase eXchange (TBX)** format as the uncompressed IATE download file. For further details see: [TBXcoreStructV02.dtd](#), [TBXCS.xcs](#), [tbxcsdtd.dtd](#).

**You can download IATE and the IATE extraction tool IATEExtract by clicking on the button below.**

There is no need to unzip the IATE download file as the extraction tool IATEExtract will access the data in the zip file directly.

Users also need to download the extraction tool IATEExtract and copy it into a suitable directory on their computer. The tool IATEExtract is distributed as a Java jar file. It can run with a graphical user interface on any operating system supporting the Java runtime of version 1.7 or newer.

[Descargar IATE](#)

Tamaño del archivo sin comprimir: **1,86 gigabytes**  
 Tamaño del archivo comprimido (descargado): **107 megabytes**  
 Fecha de la última actualización: **26/02/2019**

Para obtener más información sobre la estructura y categorías de los datos que incluye el archivo de descarga, consulte: [Los campos de datos de IATE](#)

Figura 2. Descarga libre del corpus IATE. Fuente: <https://iate.europa.eu/download-iate>.

## 2. EL CORPUS IBERIA

### 2.1. Creación del proyecto

El corpus IBERIA (<http://iberia.cchs.csic.es/cgi-bin/iberia.cgi>), del Consejo Superior de Investigaciones Científicas (CSIC), es uno de los corpus más recientes surgidos para el estudio del léxico en español. Aunque sus primeros pasos se dieron hace algo más diez años, 2008-2009, su presentación en línea y uso abierto es posible únicamente desde finales de 2018. Existen dos publicaciones donde se describen algunas de sus características —Ahumada (2010) y Ahumada *et al.* (2011)—, en las que se hace una presentación y justificación de la nueva herramienta, por un lado, y una descripción técnica de algunos aspectos de lematización, por otro<sup>4</sup>, pero su reciente aparición y la falta de difusión ha hecho que todavía falte una descripción más completa; esto es lo que haremos aquí.

IBERIA surge como parte de un proyecto de más envergadura del profesor de investigación Ignacio Ahumada<sup>5</sup> en el marco de los llevados a cabo en el entonces Instituto de Estudios Documentales sobre Ciencia y Tecnología (IEDCYT) del CSIC. El IEDCYT era un centro derivado del antiguo Centro de Información y Documentación Científica (CINDOC), que se encargaba de suministrar la terminología española al *Eurodicautom* de la Unión Europea, de gestionar las bases de datos de vaciados de revistas ISOC y de realizar estudios bibliométricos<sup>6</sup>: estos aspectos son importantes para entender cómo surge el corpus y la trayectoria institucional en la que se inserta. Hoy el Instituto de Estudios Documentales sobre

4 Consideramos también la presentación del corpus en el X Congreso Internacional de Traducción, Textos e Interferencias (Augsburgo, 4 al 6 de septiembre de 2013).

5 Ficha del autor en la página del Centro: <http://cchs.csic.es/es/personal/ignacio.ahumada>.

6 Sobre el antiguo CINDOC (1992-2007) y los antiguos institutos que lo componían (ICYT, Instituto de Información y Documentación en Ciencia y Tecnología, e ISOC, Instituto de Información y Documentación en Ciencias Sociales y Humanidades), puede verse documentación en Digital CSIC: <https://digital.csic.es/handle/10261/24917>. También nuestro trabajo de 2010 (Pablo Núñez 2010), diapositivas de una presentación sobre la labor del CSIC, especialmente desde la número 38.

Ciencia y Tecnología ha desaparecido como tal, aunque sus investigadores se han integrado en algunos de los otros institutos que componen el Centro de Ciencias Humanas y Sociales del Consejo, en Madrid, como el Instituto de Lengua, Literatura y Antropología (ILLA).

El proyecto de Ahumada consistía en la creación de una estación de trabajo lexicográfica (ETL), entendida como un conjunto de recursos y herramientas tecnológicas usadas y/o creadas para el estudio de la lexicografía y la realización de diccionarios (o, como señala Cabré, una plataforma de “integración tecnológica de múltiples productos y recursos” para facilitar la tarea del lexicógrafo o el “diseño de productos lexicográficos en un entorno profesional”, Cabré *et al.* 2004: 278 y 286). La estación lexicográfica se inspira en los modos de trabajo habituales en las grandes editoriales del mundo anglosajón que realizan diccionarios. En el ámbito hispánico, han tratado también de estas estaciones Alvar Ezquerro (2003) y Cabré (2004).

Esta estación de trabajo lexicográfica, que persigue el “análisis y descripción del léxico científico-técnico del español con fines lexicográficos”, se completa, además de con el corpus IBERIA, con otros proyectos del grupo de investigación ELCI (“El español, lengua de la ciencia”) como son las bases de datos “Torres Quevedo”, sobre diccionarios de especialidad, y la Nebrija-Valdés, de la que han surgido los tres volúmenes del diccionario bibliográfico de la metalexigrafía del español (*DBME* 1-3)<sup>8</sup>. El observatorio de neonomía o del neologismo científico es otro de sus componentes<sup>9</sup>. Debemos indicar que, durante los años 2009-2010, colaboramos con el profesor Ahumada en la base de datos “Torres Quevedo” como investigador postdoctoral Juan de la Cierva, de ahí nuestro conocimiento interno del funcionamiento de este proyecto.

## 2.2. Cronología del lanzamiento

IBERIA se crea en el IEDCYT (CCHS) del CSIC en los años 2008-2009 como un corpus textual de carácter panhispánico y sincrónico sobre el discurso científico-técnico en español. Incluye textos desde 1985, si bien la mayor parte son posteriores al año 2000 (“los documentos anteriores a 2000 se han incorporado como referencia para los estudios de carácter lexicológico”). Tiene como principal propósito ayudar al estudio del léxico científico-técnico español y orientar sobre los procesos de obsolescencia, neología y consolidación terminológica.

El corpus —cito textualmente— “fue presentado oficialmente el 18 de junio de 2018 en la Universidad Internacional Menéndez Pelayo [UIMP], en el curso avanzado “La lengua española como factor de excelencia internacional” (Santander, 18-20 de junio de 2018)”,

7 Página electrónica del grupo de investigación (sin actualizar): <http://www.investigacion.cchs.csic.es/elci/node/8>. Esta línea es hoy en parte continuada por ES-CIENCIA, una plataforma interdisciplinar que promueve el español como lengua de comunicación científica. Cf. <https://pti-esciencia.csic.es/>

8 Ahumada Lara (dir.): *Diccionario bibliográfico de la metalexigrafía del español (orígenes-año 2000)*. Jaén: Universidad, Seminario de Lexicografía Hispánica, 2006 / *Diccionario bibliográfico de la metalexigrafía del español (años 2001-2005)*, ídem, 2009 / *Diccionario bibliográfico de la metalexigrafía del español (años 2006-2010)*, ídem, 2014.

9 Información extraída de conversación con el propio Dr. Ahumada y de la página electrónica de ELCI: <http://www.investigacion.cchs.csic.es/elci/>: “Para lograr sus objetivos ELCI cuenta con una Estación de Trabajo Lexicográfico (ETL) en la que se alojan herramientas como el corpus Iberia (el español, lengua de la ciencia en la sociedad del conocimiento), las bases de datos «Torres Quevedo» (diccionarios terminológicos del español) y Nebrija-Valdés (metalexigrafía del español), un tesoro terminológico a partir de una abundante producción terminográfica y, en fase de desarrollo, el observatorio de neonomía o del neologismo científico”.

donde participaron personas de instituciones como la RAE-ASALE, el Fondo de Cultura Económica, el Instituto Caro y Cuervo y la Red Iberoamericana de Innovación y Conocimiento Científico (REDIB) con el propósito de debatir sobre el uso del español en instituciones de investigación del mundo hispano<sup>10</sup>.

Es importante observar las fechas, pues en ese intervalo que media entre la realización en 2008-2009 y su presentación en línea en 2018, el corpus estuvo en funcionamiento, pero en estado latente y solo fue de acceso restringido para los integrantes del grupo ELCI. La publicación estaba prevista en un principio para la primavera de 2011 (Porta *et al.* 2011: 156), pero la crisis económica y las dificultades de financiación posteriores impidieron su lanzamiento.

A pesar del retraso en el lanzamiento, se siguió trabajando en el corpus. Durante esos años se promovieron acuerdos con instituciones como la citada REDIB (Red Iberoamericana de Innovación y Conocimiento Científico) y con Elsevier para la incorporación de materiales (revistas, fundamentalmente) al corpus, lo que ha permitido que creciera enormemente.

Siguiendo las publicaciones citadas de Ahumada (2010) y Porta *et al.* (2011), se puede conocer la progresión y aumento gradual del corpus: si en 2011 Iberia contenía unos 64~70 millones de palabras tomadas de 6600 documentos científicos en español (Porta, 2011, p. 146)<sup>11</sup>, en 2015 alcanzaba ya los 280 millones (66 584 documentos)<sup>12</sup> y en 2018 alcanzó los 318 millones (317 931 294 formas en 77 951 documentos con límite cronológico del 2016)<sup>13</sup>. Esto lo hacía similar en extensión al CORPES de la RAE<sup>14</sup>, y anticipaba una de las características principales ya mencionadas de los corpus: la rápida actualización y la marcación automática de sus materiales como corpus abierto.

La progresiva ampliación del corpus ha sido sin duda grande, pero otros cambios también han ocurrido en esos años, como la simplificación de la interfaz: entre las dos, la que como miembro del equipo de investigación pude usar internamente en 2010 y la presente en línea, el cambio más notable consiste en la supresión de la búsqueda avanzada, del filtrado del ancho de la concordancia y la eliminación de la pestaña de filtro “Lengua” (que es relevante, pues permitía buscar un término solo en los textos identificados en español o bien buscar por otras lenguas, pues IBERIA, al tomar como corpus documentos científicos ya publicados en revistas, como veremos, incluye también artículos en inglés y otras lenguas):

10 Se puede descargar el tríptico con la información sobre los participantes en la página electrónica de la UIMP: [http://www.uimp.es/agenda-link.html?id\\_actividad=63UW&anyaca=2018-19](http://www.uimp.es/agenda-link.html?id_actividad=63UW&anyaca=2018-19)

11 “IBERIA contains about sixty-four million tokens from over 6,600 scientific documents in Spanish» (Porta, 2011, p. 146); «En la actualidad contiene 6595 documentos y setenta millones de palabras aproximadamente referidos al español científico-técnico del español de España» (Ahumada, 2011, p. 12).

12 Según el texto presente en la “ayuda” del corpus: “En la actualidad contiene 66.584 documentos y 280 millones de palabras aproximadamente referidos al español científico-técnico del español de España entre 1985 y 2015”.

13 Nota de presentación del recurso de Infoling: <http://infoling.org/informacion/RecursoL223.html>

14 Corpus del Español del Siglo XXI (CORPES XXI): “Desde diciembre de 2018, el CORPES XXI (versión 0.91) cuenta con más de 285 000 documentos que suman alrededor de 286 millones de formas, procedentes de textos escritos y de transcripciones de textos orales. Con respecto a la versión anterior, publicada en julio de 2018, supone un incremento de 10 millones de formas” (<http://www.rae.es/recursos/banco-de-datos/CORPES-xxi>, 31/05/2019). No obstante, sucesivas actualizaciones (0.92, en mayo de 2020, 312 millones; 0.93 en febrero de 2021, 333 millones; 0.94, julio 2021, 350 millones) han superado este número (<https://www.rae.es/banco-de-datos/CORPES-xxi>, consulta: 01/04/22).

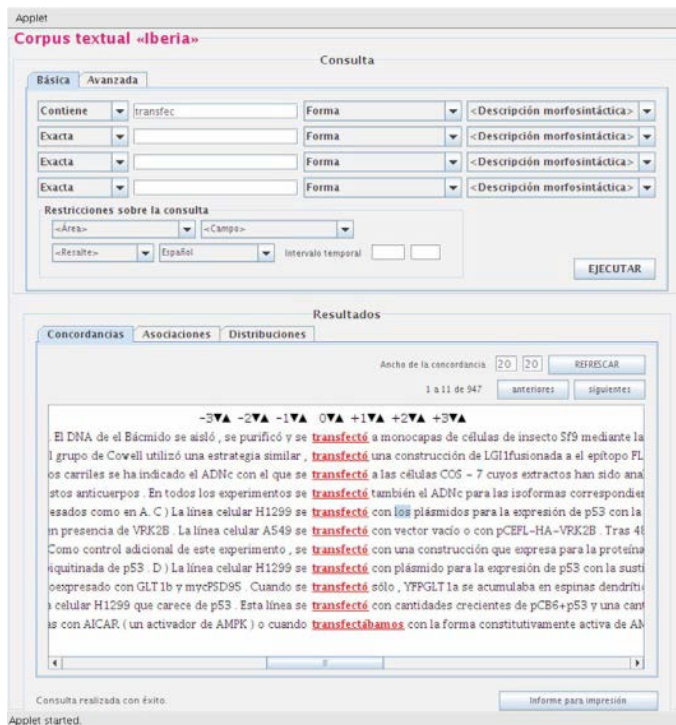


Figura 3. Corpus IBERIA, versión interna, año 2010, ejemplo de búsqueda.  
Fuente: Porta *et al.* (2011: 154); *cf.* asimismo Pablo (2010: 62-65).

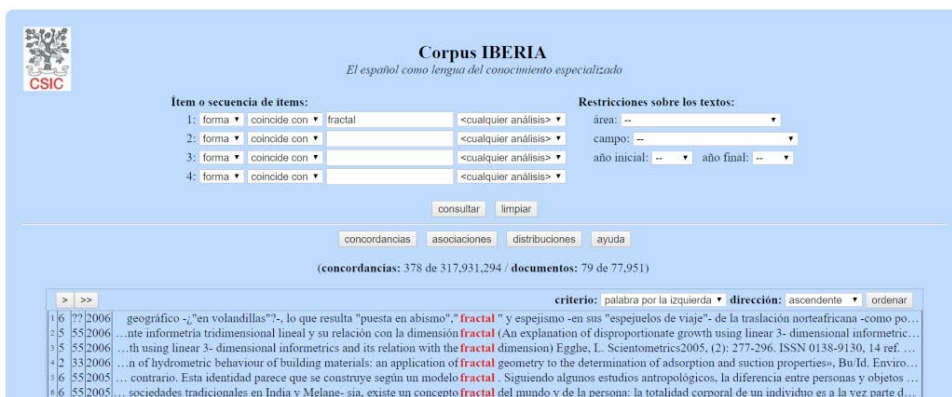


Figura 4. Corpus IBERIA, versión actual, interfaz en línea.

## 2.2. Documentos y representatividad

IBERIA recopila los textos aparecidos principalmente en revistas académicas de distintos ámbitos científicos. Así, se nutre fundamentalmente de las revistas y documentos del

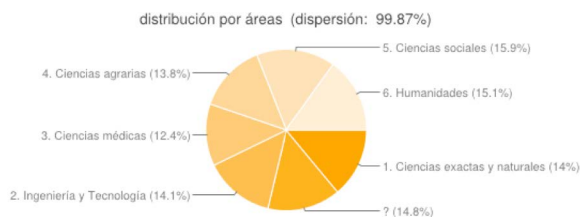


CSIC: incluye los textos digitalizados de las revistas que edita el Consejo y los textos del repositorio Digital CSIC (17 480 docs. en Digital CSIC en 2008-2009; 33 revistas en 2011 según Porta *et al.* (2011: 12); 36 en 2013, según la presentación del proyecto que se hizo en Augsburg (Ahumada 2013); en 2019, las revistas del CSIC eran 37, tras haberse creado una recientemente<sup>15</sup>). A estos se añaden otros documentos cedidos mediante acuerdos, como los volúmenes de la *Revista de Lexicografía* (1994-) o los de la *Flora ibérica* (vols. I-XV) de Castroviejo.

Ahora bien, las revistas del CSIC no cubren todos los ámbitos de la ciencia, ni los que cubre están representados con la misma proporcionalidad: así, el CSIC publica 21 revistas de arte y humanidades, sin duda la más representada, seguida por las 10 revistas sobre ciencia y tecnología (de áreas como biotecnología, botánica, bioquímica, geología, medicina, zoología, ingeniería de construcción); otras seis se engloban en las ciencias sociales (antropología, geografía, prehistoria, sociología, documentación, pero algunas de ellas también publican ocasionalmente estudios multidisciplinares o de lengua, literatura, como *Arbor*). A esto se debería también considerar la periodicidad: no todas las revistas publican el mismo número de volúmenes al año.

Otras revistas incorporadas son las de REDIB, la Red Iberoamericana de Innovación y Conocimiento Científico<sup>16</sup>, más otras revistas médicas del grupo Elsevier<sup>17</sup>, pero desconocemos el elenco de revistas incluidas con exactitud ni hasta qué momento, pues el número de revistas varía con los años, al ir incorporándose nuevas. Sabemos, por ejemplo, que REDIB agrupaba en 2019 2812 revistas, pero 4259 en 2022 (un incremento de unas 40 al mes): ¿se han incorporado todas, de forma automática? Del listado de 748 que se podían descargar en línea en 2019, 422 eran de España y el resto, de otros países, fundamentalmente hispanoamericanos (Argentina, Colombia, Chile, Ecuador, El Salvador, Venezuela, etc.); no obstante, si las revistas son de hispanística, se incluyen también de otros países, como Brasil, Japón, tres revistas estadounidenses y alguna italiana, más *Iberoamericana*, de Vervuert, clasificada como alemana. A fecha de 2022 es Brasil el país con más revistas incluidas en REDIB, con 1145, frente a las 963 de España, 546 de Colombia, 358 de Argentina, 241 de México, 135 de Chile, 15 de Bolivia u 8 de Puerto Rico, entre otros. No queda indicado cómo afecta esto a la representatividad del corpus actual.

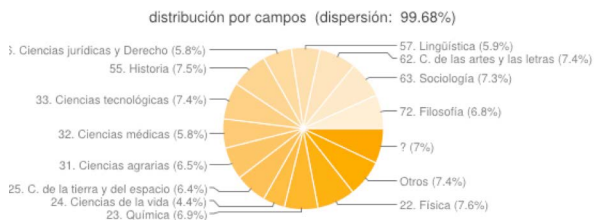
No obstante, en 2013 se respondía a la cuestión: la representatividad que aportaba entonces cada área científica al corpus no era superior a un 15 % del total:



15 Revistas CSIC: <http://revistas.csic.es/>.

16 Portal de la Red Iberoamericana REDIB: <https://revistas.redib.org/?lng=es>.

17 Revistas Elsevier del ámbito de Ciencias de la Salud: <https://www.elsevier.es/es-ambitos-ciencias-salud-kINPCC00091B-profesion-medicina-kINPBA000918>.



Figuras 5 y 6. Representatividad por áreas y campos de los documentos del corpus.  
Fuente: Ahumada (2013).



Figura 7. Distribución temporal de los documentos en el corpus. Fuente: Ahumada (2013).

Es importante notar que, en lo referido a la clasificación de áreas del corpus, IBERIA adopta para la clasificación temática de los textos la nomenclatura internacional de la UNESCO en áreas y campos por ser la seguida en la administración española: 1 Ciencias Exactas y Naturales, 2 Ingeniería y Tecnología, 3 Ciencias Médicas, 4 Ciencias Agrarias, 5 Ciencias Sociales, 6 Humanidades. Siguiendo estas, se realiza una subclasificación por campos:

- 11 Lógica
- 12 Matemáticas
- 21 Astronomía y Astrofísica
- 22 Física
- 23 Química
- 24 Ciencias de la vida
- 25 Ciencias de la Tierra y del espacio
- 31 Ciencias agrarias
- 32 Ciencias médicas
- 33 Ciencias tecnológicas
- 51 Antropología
- 52 Demografía
- 53 Ciencias económicas
- 54 Geografía
- 55 Historia
- 56 Ciencias jurídicas y Derecho
- 57 Lingüística
- 58 Pedagogía
- 59 Ciencia política

- 61 Psicología
- 62 Ciencias de las artes y las letras
- 63 Sociología
- 71 Ética
- 72 Filosofía

Estos campos se corresponden, en gran medida, con los de las revistas que nutren el corpus y sirven para restringir las consultas en caso de querer buscar algún término por alguno de esos ámbitos específicos. Los campos no cubiertos por estos se engloban en otro genérico denominado “Otros”.

En cuanto a la procedencia de los materiales por áreas geolingüísticas, el corpus busca un reparto equitativo de los materiales contenidos procedentes de las distintas regiones en donde se habla español: español de España / español centroamericano / español norteamericano / español septentrional / español ecuatorial / español austral, pero, al igual que en el caso de las revistas, no sabemos los porcentajes de cada uno ni de dónde se han tomado los textos para cada variedad (salvo en el caso del español centroamericano, mediante la cesión de las revistas de la Universidad de Costa Rica tras un acuerdo institucional).

### 2.3. Procesamiento

El procesamiento textual de los textos es automático: se hace una limpieza de formato de los textos de revista (normalmente, en PDF) y, mediante el reconocimiento óptico de caracteres —OCR Omnipage— y la posterior codificación a XML, se identifica la lengua y se realiza la marcación morfosintáctica automática y la indexación para integrar las voces en el corpus y facilitar su recuperación mediante el formulario de consultas de la interfaz (sobre estos aspectos técnicos, remitimos al artículo específico de Porta *et al.* (2011) y a otros recursos sobre corpus, como Pérez Hernández 2002, así como a la página de Grial sobre etiquetaje e interrogación en corpus textuales<sup>18</sup>).

En cuanto a sus funcionalidades, “todas las palabras están analizadas lingüísticamente (con el lema, la clase de palabras y los rasgos morfológicos) e indexadas. A través de la interfaz de consulta es posible recuperar no solo los contextos de aparición de estas palabras, sino información sobre otras palabras con la que coaparecen y sobre cómo distribuyen las apariciones en los distintos ámbitos científico-técnicos” (página electrónica del corpus, descripción).

## 3. EL ANÁLISIS DEL LÉXICO ESPECIALIZADO A TRAVÉS DE LOS CORPUS. BÚSQUEDAS

### 3.1. Cuestiones metodológicas

Tras la descripción de IBERIA, procedemos a su consulta. Hemos realizado búsquedas de un número cerrado de términos de los ámbitos astronómico y médico con mayor y menor grado de especialización; posteriormente hemos buscado esos mismos términos en otros corpus

18 El Grial (Pontificia Universidad Católica de Valparaíso), cf. <http://elgrial.cl/grial/>.

y comparamos sus resultados. De acuerdo con la comparación, mostramos la viabilidad, alcance y posibles limitaciones del corpus IBERIA para el estudio del léxico especializado.

Para este trabajo se han tenido en cuenta, además de IBERIA, los corpus de la Academia Española (CORDE, CREA, CORPES<sup>19</sup>), el corpus del español especializado del IULA<sup>20</sup>, que permite buscar en español, catalán e inglés, y los bancos de datos terminológicos IATE y Euroterm Bank de la Unión Europea, así como el *TermiumPlus* del gobierno de Canadá.

Para la búsqueda y selección de términos se han empleado también vocabularios terminológicos localizados a través de la página electrónica de la Universidad de Vaasa (Finlandia), un portal con enlaces a diccionarios, glosarios y bancos de datos en línea para traductores especialmente interesante porque verifican el funcionamiento y la actualización de los recursos<sup>21</sup>. Se han considerado también otros vocabularios electrónicos o impresos de instituciones fiables, como el *Recueil de terminologie spatiale = A Glossary of Space Terms* (ESA, 1982) o el *Vocabulaire des sciences et techniques spatiales* de la Commission générale de terminologie et de néologie de Francia (2013). La selección de términos empleados para las búsquedas abarca tanto léxico especializado como semiespecializado para poder comprobar mejor la difusión y dispersión o concentración de las ocurrencias y la funcionalidad de los corpus. No obstante, aquí solo tratamos unos pocos como muestra.

Una observación necesaria: al realizar estas búsquedas, hemos tenido en cuenta que las características de cada corpus no siempre los hacen comparables (por ser más/menos especializados, tener mayor o menor amplitud de palabras, abarcar mayor/menor amplitud cronológica). Del mismo modo ocurre con la diferente tipología de léxico —predominio de textos generales/especializados—, pero esta posibilidad es, precisamente, la que interesa mostrar para saber cuándo un corpus puede ser preferible o cuándo ofrece más o menos información para la búsqueda de información sobre términos del español. Nuestras conclusiones tendrán en cuenta estas posibles diferencias.

### 3.2. Búsquedas

Para comprobar las posibilidades que ofrece el corpus IBERIA, hemos seleccionado tres términos de la astronomía, pero también uno de medicina, de los que damos sus definiciones según los diccionarios académicos (*Diccionario de la lengua española* o *DLE* de la Academia Española-ASALE; *Diccionario de términos médicos* o *DTME* de la Real Academia Nacional de Medicina de España; *Diccionario español de ingeniería* o *DEI* de la Real Academia de Ingeniería; *Vocabulario científico-técnico* o *VCTRAC digital* de la Real Academia de Ciencias Exactas, Físicas y Naturales de España):

- *rayos gamma/rayos  $\gamma$* : “Ondas electromagnéticas extraordinariamente penetrantes, producidas en las transiciones nucleares o en la aniquilación de partículas” (*DRAE* 1992,

19 La Academia dispone también de un corpus de español científico-técnico, pero es de su uso interno y desconocemos su alcance. Según Ahumada (2013) incluiría textos pedagógicos hasta 2004. Asimismo, también tendrían corpus técnicos la Real Academia de Medicina y la Real Academia de Ingeniería, probablemente para la realización de sus respectivos diccionarios. Hoy también puede consultarse en línea el recurso Enclave de Ciencia, donde se unifica la consulta a varios diccionarios técnicos (cf. <https://enclavedeciencia.rae.es/>).

20 Institut Universitari de Lingüística Aplicada-Universitat Pompeu Fabra: *Corpus Tènic de l'IULA*, <http://bwananet.iula.upf.edu/>.

21 Vaasan yliopisto/Universidad de Vaasa, Terminology Forum (<https://sites.uwasa.fi/terminologyforum/>). Cf. Special Language Glossaries: <https://sites.uwasa.fi/terminologyforum/special-field-glossaries/>.

2001 y *DLE*). “Radiación electromagnética de longitud de onda muy corta (del orden de 10-6 a 10-4  $\mu\text{m}$ ) y de gran poder penetrante que es emitida durante la desintegración de muchos elementos radiactivos” (*TermiumPlus*, acepción 1, campos temáticos: Radiación electromagnética / Física radiológica (Teoría y aplicación) / Teledetección).

- *geosíncrono*, *-a*: referido a un satélite, aquel cuya órbita tiene el mismo periodo de rotación orbital que el de rotación sideral de la Tierra (DEIng, s. v. satélite geosíncrono; Terminesp: satélite síncrono de la Tierra. Cf. Enclave de Ciencia).

-  *cromosfera*: “Zona superior de la envoltura gaseosa del Sol, de color rojo y constituida principalmente por hidrógeno inflamado” (*DRAE* 1992, 2001 y *DLE*, corrigiendo la definición del diccionario usual que se daba ya en 1925, “Parte de la fotosfera que se supone compuesta de hidrógeno inflamado”. De forma complementaria, en la *Encyclopedia Britannica*: “Layer of the Sun’s atmosphere, several thousand miles thick, above the photosphere and below the corona” (*Britannica Concise Encyclopedia*, 2006: 402). En el *Trésor de la langue Française informatisé (TLFi)*: “Couche moyenne de l’atmosphère solaire située entre la photosphère et la couronne solaire et visible seulement lors des éclipses totales”.

- *mastodinia*: = mastalgia, Dolor en la mama que puede estar relacionado o no con el ciclo menstrual [...] (*DTME*); = mastopatía fibroquística (desus.), Cada una de las alteraciones clínicas e histológicas del normal desarrollo e involución de la mama, que abarca cambios diversos y lesiones benignas, tales como quistes mamarios, adenosis, fibrosis, papilomatosis y cambios apocrinos (*DTME*).

Debemos indicar que, además de estos cuatro términos, hemos realizado búsquedas de otros muchos más, pero la extensión de este artículo nos ha hecho suprimirlos en la exposición.

En algunos casos los términos no aparecen en los diccionarios, al ser específicos, así que hemos recurrido a otras obras. El *DLE* sí incluye varias de ellas, especialmente del ámbito de astronomía, debidamente marcadas, lo que muestra su preocupación por incluir voces especializadas.

### 3.3. Términos

#### 3.3.1. Rayos gamma

Buscando en IBERIA, aparecen 271 concordancias con este término complejo. Es preciso indicar la segunda palabra de la lexía compleja en el segundo campo del formulario, pues si los buscamos juntos no da resultados. Si buscamos solo “gamma”, salen 3876 concordancias, pero no se refieren necesariamente a los rayos gamma (pueden ser “peroxisomas gamma”, etc.), por lo que no es una búsqueda pertinente. No obstante, puede ser interesante para conocer otras combinaciones que tienen afinidad, como “radiación gamma”<sup>22</sup>.

Si hacemos la búsqueda restringiendo por área o campo, los resultados varían mucho: si seleccionamos el campo 21 (Astronomía y Astrofísica), no hay resultados; si buscamos solo en el 25, sobre Ciencias de la tierra y del espacio, da únicamente dos concordancias, y

22 Una observación respecto a las búsquedas: al buscar “gamma”, notamos que salen resultados procedentes de textos escritos en inglés junto con los escritos en español, lo que se debe a que, como parten de revistas especializadas, muchos artículos están en esta lengua: sería recomendable añadir un criterio de restricción por lengua (español/inglés) elegible entre los parámetros de búsqueda para evitar esto. Este criterio ya estaba en la interfaz de uso interno, pero se suprimió al poner el corpus en línea.

trece sobre el campo 22 (Física): la causa se debe a que la mayor parte de los resultados son del área de la Medicina (32, Ciencias médicas). La ordenación de los resultados por campos no es efectiva, pues aparecen los del campo 33 en medio del ámbito 22 o 25 y luego más concordancias de la 33... En cualquier caso, es un gran número de concordancias.

Si buscamos en singular, “rayo gamma”, aparecen otras 50 concordancias, pero aquí es llamativo que todas, salvo una del campo médico (32. Ciencias médicas), estén clasificadas en el campo de la tecnología (33. Ciencias tecnológicas).

The screenshot shows the 'Corpus IBERIA' search interface. At the top left is the CSIC logo. The title 'Corpus IBERIA' is centered, with the subtitle 'El español como lengua del conocimiento especializado' below it. The search criteria are set to 'coincide con' for six items, with 'rayo gamma' entered in the first item's text field. The 'Restricciones sobre los textos' section is empty. Below the search fields are buttons for 'consultar' and 'limpiar'. A navigation bar contains 'concordancias', 'asociaciones', 'distribuciones', and 'ayuda'. The search results are displayed in a table with columns for 'criterio', 'área', 'dirección', and 'ordenar'. The results list various scientific texts related to gamma rays, such as 'continuo, un fotodetector sensible a posición y una electrónica aso...', 'al campo de la física de detectores. Es un dispositivo que permite d...', 'que libera la energía emitiendo fotones en el rango de energía visibl...', 'tipo de éste. Por ejemplo, un fotomultiplicador normal sin resolucio...', 'partículas gamma interactúan con el cristal centelleador. Dicho es...', 'de entrada del foto-detector, porque estas dos direcciones no son su...', 'otras dimensiones, con lo cual siempre se introduce un error en la d...', 'posición del origen del rayo gamma. El error de paralaje es más im...', 'se obtiene una distribución de luz de centelleo, 20cuya anchura per...', 'continuo, un fotodetector sensible a posición y una 10electrónica as...', 'luz de centelleo. Dicho detector de rayos gamma proporciona infor...', 'PIN fotodiodo) sensible a posición y 30una electrónica asociada q...', 'cristal. El detector de rayos gamma con codificación de profundidad...', 'la posición bidimensional de los detectores de rayos gamma habitu...'. The interface indicates 50 concordances and 4 documents.

Figura 8. Corpus IBERIA, búsqueda de rayo gamma.

La búsqueda de rayos  $\gamma$  con la letra griega solo arroja un resultado:

Para las personas que desconocen las diferencias entre los dos tipos de energía del espectro electromagnético, la palabra “radiación” suele evocar “peligro” por su asociación a las radiaciones ionizantes (rayos X y rayos  $\gamma$ ) (“Telefonía móvil: ¿representa algún riesgo para la salud?”, *Revista Medicina clinica*, 132/14)

The screenshot shows the 'Corpus IBERIA' search interface with the search criteria changed to 'coincide con' for six items, with 'rayo gamma' entered in the first item's text field. The 'Restricciones sobre los textos' section is empty. Below the search fields are buttons for 'consultar' and 'limpiar'. A navigation bar contains 'concordancias', 'asociaciones', 'distribuciones', and 'ayuda'. The search results are displayed in a table with columns for 'criterio', 'área', 'dirección', and 'ordenar'. The results list various scientific texts related to gamma rays, such as 'palabra “radiación” suele evocar “peligro” por su asociación a las r... rayos y ). El objetivo de este artículo es revisar los resultados de los estudios publicados re...'. The interface indicates 1 concordance and 1 document.

Figura 9. Corpus IBERIA, búsqueda de rayo  $\gamma$ .

En cuanto a las asociaciones o términos con los que suele aparecer, vemos que, según IBERIA, “rayos gamma” aparece relacionada frecuentemente con voces como *sonda*, *crystal*, “*detector* (sust. o adj.), *detección*, *codificación*, *entrada*, *fuelle* (pero también incluye preposiciones, artículos e incluso signos ortográficos):

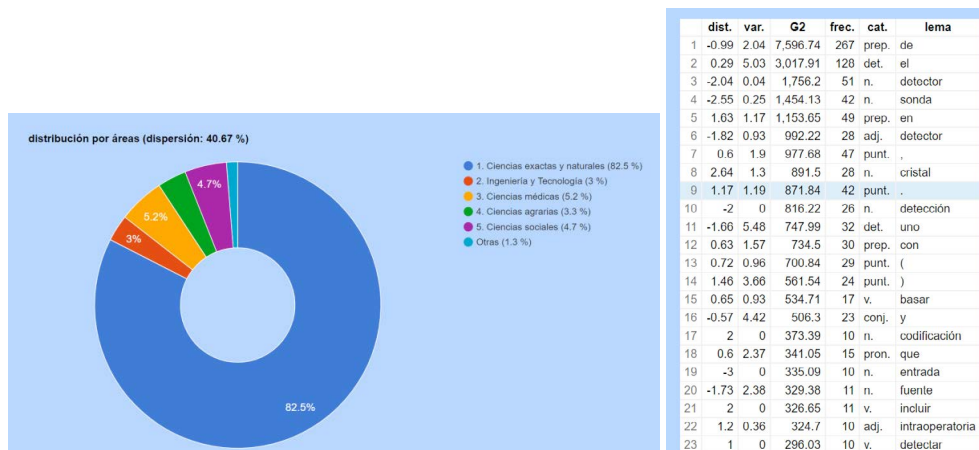


Figura 10. Corpus IBERIA, distribución por áreas y asociaciones de *rayos gamma*.

En cuanto a la búsqueda en los corpus académicos, “rayos gamma” aparece ya en CORDE con 36 ocurrencias en 7 documentos. Los primeros registros son de 1945, pero el número de concordancias fluctúa mucho y solo figuran para documentos de cuatro años concretos: 1945 (15 casos), 1961 (1 caso), 1963 (1 caso), 1968 (19 casos), y de dos países únicamente: España (casi el 93 % de los documentos) y Argentina (2,77 %). Los textos son prosa científica en un 41,66 % y prosa periodística en un 52,77; en narrativa aparece un 5,55 %.

Buscando por “rayo gamma” solo aparece un registro, y también es de 1968: “La intensidad del rayo gamma que va de la fuente de radiación al contador, al pasar por él el tren, disminuye en relación al peso de la carga que lleva el tren” (de la obra *Balanza de radiaciones gamma*, *El Correo de la Unesco*).

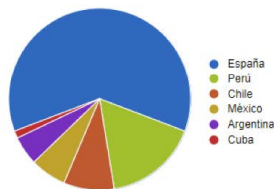
La documentación en CREA para “rayos gamma” es mayor: 88 veces en 55 documentos. Aquí sí es patente el uso técnico, por la preponderancia de los textos científicos (Ciencia y tecnología: 72,72 %); como “rayo gamma” solo hay cuatro concordancias, una de 1980, de la obra *Einstein o la armonía del mundo*, publicada en Argentina, y las tres restantes de un manual peruano de Ricardo Díaz Chávez, *Principios de Física Atómica*, 1997. La búsqueda en CREA anotado da el mismo resultado para “rayo gamma”, pero un número menor para “rayos gamma”: 78 casos en 40 documentos. En todo caso, podemos ver la distribución de resultados por zonas lingüísticas y países (con IBERIA no nos es posible saberla):

Distribución País

País	Freq	Fnorm
España	48	0,69
Perú	13	4,08
Chile	7	1,42
México	5	0,42
Argentina	4	0,33
Cuba	1	0,34

1 - 6 de 6 página: 1

Distribución País



Distribución Tema

Tema	Freq	Fnorm
Ciencias y tecnología	59	6,70
Artes, cultura y espectáculos	5	0,37
Ciencias sociales, creencias y pensamiento	4	0,23
Política, economía y justicia	4	0,11
Salud	4	0,55
Actualidad, ocio y vida cotidiana	1	0,06
Teatro	1	0,21

1 - 7 de 7 página: 1

Distribución Tema



Figuras 11 y 12. Distribución de *rayos gamma* en CREA anotado.

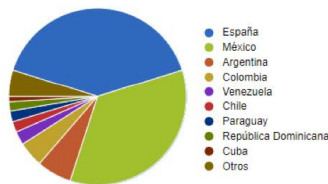
En cuanto a CORPES, si la versión beta 0.91 señalaba 356 casos en 132 documentos, la versión más reciente (0.94, julio 2021) recoge 410 en 156 documentos, lo que indica no solo un enriquecimiento del corpus, sino también un uso mucho más extendido de este término, con un predominio apabullante del ámbito de Ciencias y tecnología.

Distribución País

País	Freq	Fnorm
España	165	1,36
México	143	3,83
Argentina	26	0,84
Colombia	19	0,73
Venezuela	10	0,65
Chile	8	0,39
Paraguay	8	1,12
República Dominicana	7	1,02
Cuba	4	0,35
Perú	4	0,33

1 - 10 de 18 página: 1 2

Distribución País

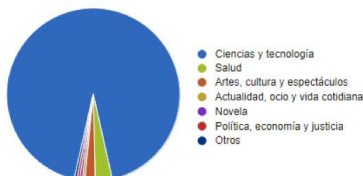


Distribución Tema

Tema	Freq	Fnorm
Ciencias y tecnología	381	10,69
Salud	13	0,57
Artes, cultura y espectáculos	8	0,23
Actualidad, ocio y vida cotidiana	2	0,04
Novela	2	0,02
Política, economía y justicia	2	0,03
Relato	1	0,07
Teatro	1	0,10

1 - 8 de 8 página: 1

Distribución Tema



Figuras 13 y 14. Distribución de *rayos gamma* en CORPES.

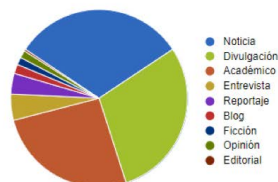


La tipología textual se reparte entre noticias en medios, divulgación y ámbito académico, pero otro gran porcentaje es para divulgación y noticias (científicas), lo que demuestra que es un término que ha ampliado la difusión más allá de los textos especializados.

Distribución Tipología

Tipología	Freq	Fnorm.
Noticia	92	2,58
Divulgación	86	2,77
Académico	76	2,54
Entrevista	14	1,96
Reportaje	11	0,74
Blog	5	1,00
Ficción	4	0,04
Opinión	4	0,81
Editorial	1	3,10
1 - 9 de 9		página: 1

Distribución Tipología

Figura 15. Tipología textual para la voz *rayos gamma* según CORPES.

Para “rayo gamma” solo se dan 8 ocurrencias, todas de México (2004, 2009, 2016), procedentes de revistas de ciencia; por ejemplo: “Cuando un rayo gamma colisiona con las partículas del aire en la parte superior de la atmósfera, inicia una reacción en cadena que provoca una cascada de partículas cargadas eléctricamente. Éstas pueden viajar a velocidades superiores a las de la luz en el aire” (Dultzin, D. (2016). “Nuevos ojos para ver la luz del cielo”. Ciencia. México D. F.: revistaciencia.amc.edu.mx).

### 3.3.2. Cromosfera

A través del corpus IBERIA obtenemos dos ocurrencias: una primera de 2006, de la revista *Física de la Tierra*, y una segunda de 2007, de un libro:

Corpus IBERIA  
El español como lengua del conocimiento especializado

Ítem o secuencia de ítems: 1. forma | coincide con | cromosfera

Restricciones sobre los textos: área: | campo: | año inicial: | año final: |

consultar limpiar

concordancias | asociaciones | distribuciones | ayuda

(concordancias: 2 de 317.931,294 | documentos: 2 de 77.931)

créditos: | palabra por la que | dirección: | concordante | ordenar

2006 comparan estas variaciones con las observadas en la actividad solar. cromosfera solar obtenidas también en el Observatorio. El interés por las VSMR ha estado t...  
2007 rias, de unos 4.000°C, que vemos oscuras en comparación con sus ... cromosfera se encuentra justo por encima de la fotosfera. Las imágenes que se han obteni...

Figura 16. Corpus IBERIA, voz *cromosfera*.

La segunda da un texto muy descriptivo:

Casi totalmente transparente, la cromosfera se encuentra justo por encima de la fotosfera. Las imágenes que se han obtenido de esta región han sido tomadas durante el principio y el final los eclipses del Sol totales, en los que aparece como un anillo rojizo, o con filtros

muy específicos. Dichas imágenes revelan una serie de fenómenos, como los filamentos, protuberancias y espículas (figura 3.1). Finalmente, la corona es la capa más externa de la atmósfera solar, formada por gas de muy baja densidad y con una extensión que supera los millones de kilómetros. [...] (Antxón Alberdi, Olga Muñoz (coords.): *Un viaje al Cosmos en 52 semanas*, Madrid, CSIC/Ministerio Educación y Ciencia, 2007: 27)

Como esta concordancia está extraída de un libro y este está digitalizado en abierto en Digital CSIC, lo hemos consultado y hemos tomado de él la figura 3.1 que indicaba la concordancia. Sorprendentemente, la limpieza del texto del libro a la hora de procesarlo en el corpus —que hace prescindir de tablas, ilustraciones, etc.— hace perder dos ocurrencias más en la misma obra (columna azul) que no se localizan en IBERIA:

estructuras muy características, como gránulos y manchas. En tanto que los primeros se deben a la "ebullición" del gas ya mencionada, las manchas son zonas más frías, de unos 4.000°C, que vemos oscuras en comparación con sus alrededores.

Casi totalmente transparente, la cromosfera se encuentra justo por encima de la fotosfera. Las imágenes que se han obtenido de esta región han sido tomadas durante el principio y el final los eclipses del Sol totales, en los que aparece como un anillo rojizo, o con filtros muy específicos. Dichas imágenes revelan una serie de fenómenos, como los filamentos, protuberancias y espículas (figura 3.1).

Finalmente, la corona es la capa más externa de la atmósfera solar, formada por gas de muy baja densidad y con una extensión que supera los millones de kilómetros. Podemos observarla durante los eclipses totales de Sol como un halo blanquecino y, por su fuerte emisión en rayos X debido a su elevada temperatura —cerca al millón de grados—, también con telescopios diseñados para esta longitud de onda. Con ellos se han obtenido imágenes que muestran "agujeros" en los polos de la corona, de donde se cree que procede el viento solar, un chorro de partículas eléctricamente cargadas que, con velocidades de unos 400 km/s, invade el espacio interplanetario.

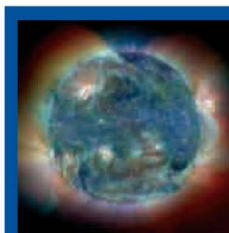


Figura 3.1. Filamentos y protuberancias en la cromosfera. Se trata de densas nubes de material más frío que quedan suspendidas sobre la superficie siguiendo los bucles del campo magnético. Como consecuencia de su menor temperatura, se muestran oscuras en el disco (filamentos) y brillantes en el limbo (protuberancias). En la cromosfera también se distinguen las espículas, pequeñas erupciones que ascienden y descienden a una velocidad del orden de 20 km/s, y cuyo aspecto puede compararse con el de una pradera en llamas.

Figura 17. *Voz cromosfera* en el documento original.  
Fuente: Digital CSIC, <http://hdl.handle.net/10261/3147>.

La búsqueda en los corpus académicos es productiva: el término figura ya en CORDE con quince casos en dos documentos (el primer testimonio se da en 1903 —Cuba— y los restantes en la obra española *Astronomía I y II* de José Comas Salas, 1919). A pesar de proceder de un relato literario, la primera documentación, la de 1903, da el significado de la voz, lo que indica que no era muy conocida aún (el término se acuñó en 1868 en inglés por propuesta del científico Dr. Sharpey tras observaciones espectroscópicas del sol):

—Y su constitución química —continuó Rosa— no será la misma que la de la tierra. —Por el espectroscopio sabemos —respondió Baranda— que en el sol hay hierro, calcio, níquel, cobalto, sodio, cobre, plomo, aluminio, oxígeno, etcétera. La cromosfera, por ejemplo, o sea la capa gaseosa rosada que se advierte alrededor de la superficie luminosa, se compone de hidrógeno (Bobadilla, E. (1903). *A fuego lento*. La Habana: Editorial Letras Cubanas).

En CREA encontramos 25 casos en 11 documentos, pero, salvo una documentación de 1987, significativa, pues se refiere a la historia de la óptica y astronomía, todas las demás son posteriores a 1992:

Dominique François Arago (1786-1853), astrónomo y físico francés, tiene contribuciones a la física además de las muy valiosas a la óptica a las que nos hemos referido. Con Prony midió la velocidad del sonido en el aire, publicó trabajos sobre electricidad y magnetismo y como astrónomo midió diámetros de los planetas y estudió la corona solar y la  cromosfera. Vino a España con Biot en 1806 [...] Durán Miranda, A. (1987). *La óptica en el siglo XIX [Historia de la Física en el siglo XIX]*, p. 142)

Consulta: cromosfera, en todos los medios, en CREA  
Resultado: 25 casos en 11 documentos.

OBTENCIÓN DE EJEMPLOS

Recuperar Concordancias Normal Clasificación Año

Agrupación Marcas

Cómo citar el CORPUS Concordancias. Pantalla: 1 de 1. Ver párrafos

Nº	CONCORDANCIA	AÑO	AUTOR
1	os de los planetas y estudió la corona solar y la cromosfera. Vino a España con Biot en 1806 para traba	** 1987	Durán Miranda, Armando
2	lómetros de altitud sobre la superficie. Tanto la cromosfera como las espículas y protuberancias pueden	** 1992	Oliver, José Mª
3	débil, que tiene a su alrededor. Por encima de la cromosfera está la corona, una especie de aureola de	** 1992	Oliver, José Mª
4	lo, pueden verse abundantes zonas eruptivas en la cromosfera y fotosfera, en las que juegan un importan	** 1992	Oliver, José Mª
5	ensidad luminosa. En cambio, la observación de la cromosfera o de las protuberancias debe realizarse ne	** 1992	Oliver, José Mª
6	fácilmente las protuberancias y el aspecto de la cromosfera. Por el contrario, la observación de la co	** 1992	Oliver, José Mª
7	esde el conocimiento de los helios galácticos a la cromosfera de las estrellas, desde los núcleos galáct	** 1996	PRENSA
8	o, que por su color ligeramente rosado es llamada cromosfera; es la atmósfera solar. Más allá de la cro	** 1996	Ayllón, Teresa
9	romosfera; es la atmósfera solar. Más allá de la cromosfera se extiende la corona solar, a varios mill	** 1996	Ayllón, Teresa
10	ha llamado el cuarto estado de la materia. De la cromosfera y la corona inferior brotan enormes géyser	** 1997	Fierro, Julieta
11	vista durante los eclipses totales de Sol, son la cromosfera y la corona. La cromosfera está más calien	** 1997	Fierro, Julieta
12	totales de Sol, son la cromosfera y la corona. La cromosfera está más caliente que la fotosfera y se ca	** 2000	PRENSA
13	las oscilaciones del gas ionizado o plasma en la cromosfera de las manchas solares, ambos fenómenos es	** 2000	PRENSA
14	ncialmente importante para el calentamiento de la cromosfera solar) está de hecho teniendo lugar en tod	** 2000	PRENSA
15	o el plasma magnetizado de las regiones externas (cromosfera) de la umbra de las manchas solares se est	** 2000	PRENSA
16	erfiles de polarización circular observados en la cromosfera de una mancha solar pasan de ser "normales	** 2000	PRENSA
17	do a la fotosfera y más caliente que ella está la cromosfera, que parece una pradera ardiente. En ella	** 2001	PRENSA
18	observa un resplandor rojizo que corresponde a la cromosfera, una capa de espesor irregular, más calien	** 2001	PRENSA
19	s proyectadas sobre el disco solar. Tampoco en la cromosfera termina el Sol: a medida que el eclipse co	** 2001	PRENSA
20	en las regiones externas de la atmósfera del Sol (cromosfera y corona). El plasma solar es un gas muy c	** 2001	PRENSA
21	nte zona más exterior, de aspecto rojizo, llamada cromosfera, y que se eleva por encima de la fotosfera	** 2002	Herrera Merino, Rosa María
22	nombre los chorros gaseosos que se producen en la cromosfera y se elevan a gran altura para acabar diso	** 2002	Herrera Merino, Rosa María
23	la superficie solar y deja sólo al descubierto la cromosfera; es anular cuando la Luna se superpone a l	** 2002	Herrera Merino, Rosa María
24	o de esta historia revelárenos las respuestas. La cromosfera solar, una atmósfera de gas ionizado a muy	** 2002	Altschuler, Daniel Roberto
25	que no pudo obtener unos ansados espectros de la cromosfera solar, porque el eclipse sólo duró 5 segun	** 2003	PRENSA

Ir arriba Pantalla: 1 de 1. Ver párrafos

Nueva consulta: CREA CORDE Nómina de autores y obras Ayuda.

Figura 18. Concordancias de la voz *cromosfera* en CREA.

En cuanto a la procedencia de los textos, la mayor parte son de España: solo cinco ocurrencias son de México y una del Uruguay. La tipología es fundamentalmente divulgativa: *Manual práctico del astrónomo aficionado* (editorial De Vecchi, 1992), *Guía para observar el firmamento* (LIBSA, 2002), aunque también se extraen de algunos textos especializados procedentes del boletín de noticias del Instituto Astrofísico de Canarias (La Laguna, 2000) o de un manual universitario de doscientas páginas sobre *Elementos de meteorología y climatología* (México D. F., editorial Trillas, 1996).

En cuanto a CREA anotado, este solo recoge 17 ocurrencias frente a las 25 del CREA, pero aporta una documentación unos años anterior de una obra divulgativa argentina: *Astronomía desde la terraza* (Buenos Aires, 1978): “Se extiende a su vez sobre la capa mencionada, hasta alcanzar los 10000 km de altura, la llamada  cromosfera, constituida principalmente por hidrógeno. En esta capa la temperatura desciende a los 5000 grados centígrados”. Todas las demás ocurrencias provienen de textos publicados en España o México.

La búsqueda en CORPES proporciona 34 ocurrencias en 14 documentos, de nuevo en su mayor parte de España y México, con solo una documentación de Uruguay y dos de Venezuela.

Corpus del Español del Siglo XXI  
(CORPES)

Versión beta (0.94) Cerrar sesión

Concordancias | Copariciones | Configuración | Ayuda | Estadística | Modo de cita | Sugerencias | Preguntas frecuentes

Lema:  Forma:  cromosfera Clase de palabra:   Grafía original  + Suborpus  + Posimitar

Concordancia | Estadística | Nueva consulta

34 casos en 14 documentos.

REF. (Clasificación, país) CONCORDANCIA Ordenar por: Año ascendente | sin criterio

1	2001 Esp.	trata de medir la relación entre la radiación ultravioleta y visible de Sol. la <b>cromosfera</b> emite preferentemente en ultravioleta y el disco solar, en visible", explicó Prieto
2	2001 Esp.	explicó Prieto: "Durante el eclipse total se oculta sólo el disco solar, no la <b>cromosfera</b> , por ello la radiación visible cae y casi sólo se registra la ultravioleta"
3	2001 Esp.	flotando a la fotosfera y más caliente que ésta está la <b>cromosfera</b> , que parece una granera ardiente. En ella destacan las protuberancias o filamentos
4	2001 Esp.	totalidad en un eclipse de Sol se observa un resplandor rojo que corresponde a la <b>cromosfera</b> , una capa de espesor irregular, más caliente que la fotosfera a la que rodea. En
5	2001 Esp.	se observan como cintas oscuras proyectadas sobre el disco solar. Tampoco en la <b>cromosfera</b> termina el Sol: a medida que el eclipse continúa resulta visible una majestuosa
6	2001 Esp.	más flabie los campos magnéticos en las regiones externas de la atmósfera del Sol ( <b>cromosfera</b> y corona). El plasma solar es un gas muy caliente que está parcialmente ionizado
7	2002 Ur.	La <b>cromosfera</b> solar, una atmósfera de gas ionizado a muy alta temperatura, se extiende millones
8	2003 Esp.	IAC), han descubierto un efecto físico de polarización de la luz del calcio en la <b>cromosfera</b> del Sol. Manso explica que la luz del calcio ionizado no debería estar polarizada
9	2003 Esp.	átomo no lo permita, informa el IAC. Sin embargo, observaciones polarimétricas de la <b>cromosfera</b> solar revelan la existencia de enigmáticos picos de polarización en la señal de
10	2003 Esp.	IAC), han descubierto un efecto físico de polarización de la luz del calcio en la <b>cromosfera</b> del Sol. Manso explica que la luz del calcio ionizado no debería estar polarizada
11	2003 Esp.	átomo no lo permita, informa el IAC. Sin embargo, observaciones polarimétricas de la <b>cromosfera</b> solar revelan la existencia de enigmáticos picos de polarización en la señal de
12	2003 Esp.	aficionados pidiendo disculpas porque no pudo obtener unos anidados espectros de la <b>cromosfera</b> solar, porque el eclipse sólo duró 5 segundos desde su lugar de observación. Conociendo
13	2004 Esp.	la fotosfera hay una «atmósfera» roja de unos 10.000 km de grosor, denominada <b>cromosfera</b> , en la que se proyectan gases a muy altas temperaturas y de la que sobresalen las
14	2004 Esp.	Tanto la <b>cromosfera</b> como las protuberancias pueden verse directamente en los momentos de la totalidad
15	2004 Esp.	Por encima de la <b>cromosfera</b> está la corona (figura 7) Capas del Sol. (Gráfico J.C. Casado). Una especie de
16	2004 Esp.	(i-otos y procesado J.C. Casado). En los primeros segundos se muestra parte de la <b>cromosfera</b> como un fino arco de intenso color rojo con brillantes protuberancias, que si
17	2004 Esp.	suficientemente grandes, desaparecen rápidamente tras el avance del disco lunar (figura 23) <b>Cromosfera</b> y protuberancias visibles en la totalidad del eclipse solar del 11 de agosto de
18	2004 Méx.	fotosfera es la capa más fría de nuestra estrella ya que las capas atmosféricas ( <b>cromosfera</b> , corona) tienen temperaturas mucho más altas. Muchos fenómenos muy interesantes
19	2009 Méx.	<b>Cromosfera</b>
20	2009 Méx.	La atmósfera solar está formada por dos capas llamadas <b>cromosfera</b> y corona que se encuentran encima de la fotosfera. En la <b>cromosfera</b> la temperatura

1 - 20 | Imprimir | Exportar | ExportarWord | Exportar TSV 1 de 2 | Ir a página: | Ir

Figura 19. Concordancias de la voz *cromosfera* en CORPES.

En cuanto a la tipología textual, abundan las obras académicas, pero también las divulgativas (como las relativas a la Semana de la Ciencia y la Tecnología 2003 de la Fundación Española para la Ciencia y la Tecnología, FECYT) y las noticias periodísticas, lo que mostraría su progresiva expansión más allá del ámbito académico:

Distribución Tipología

Tipología	Freq	Fnorm.
Académico	9	0,30
Divulgación	7	0,22
Noticia	7	0,19
Reportaje	2	0,13

1 - 4 de 4 página: 1

Distribución Tipología

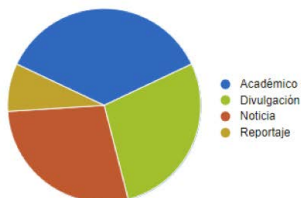


Figura 20. CORPES, voz *cromosfera*, distribución por tipología textual.

### 3.3.3. Geosíncrono, -a

La búsqueda en IBERIA nos proporciona solo una documentación tanto para su forma masculina como femenina, respectivamente. Ambas proceden de un único documento, un artículo de 2006 sobre satélites tomado de una revista mexicana:

Las investigaciones en el campo de los satélites de comunicación habían estado encaminadas, desde los años 40, a vencer un enorme desafío: aumentar la altura de los satélites aproximadamente a 36.000 kilómetros de la Tierra, donde el periodo de rotación alrededor del planeta fuera de 24 horas. De esta forma, cuando el satélite se ubicara sobre el plano del Ecuador, su rotación sería geosíncrona, es decir, a la par de la Tierra. El primer satélite geosíncrono o geoestacionario fue el Syncomm II, lanzado en 1963, que transmitió señales

de televisión durante los Juegos Olímpicos de Tokio, en 1964 (Herrera de la Rosa, 1989) (Sacristán Romero, F. (2006). “Retrospectiva histórica de los satélites de comunicación”, revista *Ra Ximhai*, 2/1).

En cuanto a los corpus académicos, CORDE no registra la voz *geosíncrono*, -a; CREA y CREA anotado solo la recogen en su forma femenina con cuatro documentaciones (una de 1991, México, y tres de 1996, España):

Corpus de Referencia del Español Actual  
(CREA. Versión anotada) Versión beta (0.0) [Cerrar sesión](#)

Concordancias | Copariciones | Configuración | Ayuda | Estadística | Modo de cita | Sugerencias | Preguntas frecuentes

Lema:  Forma:  geosíncrona Clase de palabra:  Grafía original:  Subcorpus:  Proximidad:

Concordancia Estadística Nueva consulta

REF. (Clasificación, país)	CONCORDANCIA	Ordenar por:
1 1991 Méx.	El perigeo de la órbita de transferencia <i>geosíncrona</i> está normalmente a una altura aproximada de 200 km sobre el nivel del mar y su	Año ascendente   sin criterio
2 1996 Esp.	para el empuje de satélites destinados a órbitas superiores (principalmente la <i>geosíncrona</i> ).	
3 1996 Esp.	Localización en puntos «estratégicos» del espacio exterior, que incluyen la órbita baja, la <i>geosíncrona</i> y la superficie de cualquier planeta o luna.	
4 1996 Esp.	45 centímetros de diámetro, fue lanzado al espacio en 1975 y colocado en órbita <i>geosíncrona</i> sobre el Atlántico (entre 42.00 y 25.000 kilómetros de altura). Hasta octubre de	

1-4 | [Imprimir](#) | [Exportar](#) | [Exportar HTML](#) | [Exportar TSV](#) | 1 de 1 | Ir a página:  | [Ir](#)

Figura 21. CREA anotado, concordancias para la voz *geosíncrona*.

En cuanto a CORPES, no da ningún resultado para la forma *geosíncrono*, y solo dos ocurrencias para la forma *geosíncrona*, ambas de 2019. Para este término concreto, IBERIA resulta un complemento notorio de CORPES.

Corpus del Español del Siglo XXI  
(CORPES) Versión beta (0.0) [Cerrar sesión](#)

Concordancias | Copariciones | Configuración | Ayuda | Estadística | Modo de cita | Sugerencias | Preguntas frecuentes

Lema:  Forma:  geosíncrona Clase de palabra:  Grafía original:  Subcorpus:  Proximidad:

Concordancia Estadística Nueva consulta

REF. (Clasificación, país)	CONCORDANCIA	Ordenar por:
1 2013 Chile	La agencia espacial ha informado que el asteroide ya se aleja del anillo <i>geosíncrona</i> de nuestro planeta, área en donde se encuentran los satélites meteorológicos y	Año ascendente   sin criterio
2 2014 Ur.	El asteroide pasará por el exterior de la órbita <i>geosíncrona</i> de los satélites de comunicaciones y meteorológicos que orbitan alrededor de 36.000	

1-2 | [Imprimir](#) | [Exportar](#) | [Exportar HTML](#) | [Exportar TSV](#) | 1 de 1 | Ir a página:  | [Ir](#)

Figura 22. CORPES, concordancias para la voz *geosíncrona*.

En el corpus del IULA encontramos una documentación, de la misma fecha que la de CORPES:

```
# Date: 28/5/2019      3:5Hs
# Corpus: CORPUSTECNIC_ES
# Name: [lemma="geosíncrono"] within txt;
# Size: 1 intervals/matches
#-----
```

```
1      <s>Otro tipo de órbita es la denominada ## geosíncrona ## , que implica una
      vuelta cada 24 horas pero situándose el satélite en un plano distinto al del ecuador. </s>
```

### 3.3.4. Mastodinia

Cerramos el apartado de búsquedas con la voz médica *mastodinia*. En el corpus IBERIA encontramos 36 ocurrencias desde 1996 hasta 2014, todas del ámbito de las Ciencias médicas (número 32).

**Corpus IBERIA**  
El español como lengua del conocimiento especializado

Item o secuencia de ítem:  
 1: forma > coincide con > mastodinia > <Quisque análisis>  
 2: forma > coincide con > <Quisque análisis>  
 3: forma > coincide con > <Quisque análisis>  
 4: forma > coincide con > <Quisque análisis>  
 5: forma > coincide con > <Quisque análisis>  
 6: forma > coincide con > <Quisque análisis>

Restricciones sobre los textos:  
 Área: < >  
 campo: < >  
 año inicial: < > año final: < >

concordancias asociaciones distribuciones ayuda

(concordancias: 36 de 317,931,294 / documento: 26 de 77,951)

crisis	2%	dirección	coincide	ordenar
91119994	el resultado de pasar la misma encuesta que han realizado a mujeres...		mastodinia y alteraciones vaginales a varones de más de 40 años. Si realmente existiese un...	
91119995	41% emblematas. Los 4 meses siguientes se ven patológicos. Se ver...		mastodinia o palpación de nódulo mamario. 92,33% eran patológicas (4 tumores mamario...	
91119998	de consulta más frecuente fue la revisión de rutina (44,8...		mastodinia (55,2%), siendo mucho menos frecuentes otros motivos. Un 72,3% de las mug...	
91119999	comparación en la edad del primer control, edad de la menopausia...		mastodinia, las modificaciones en la exploración física, recogidas en la historia clínica y l...	
91120000	manejo valorado por el informe manográfico. Resultados: De las...		mastodinia en el seguimiento, ni se detectaron modificaciones en la inspección ni en la p...	
91120001	diminución de la ansiedad y uno (16,67%) persistió. De las 37 pacie...		mastodinia en controles sucesivos, y en ninguna se detectó modificaciones en la inspecc...	
91120002	Se cree que el paso de progesterona hacia el torax circulato...		mastodinia, náuseas, edemas, etc. (2). El objetivo de este trabajo es revisar las seque...	
91120003	hamen es el sistema más frecuente (1, 2, 10), siendo raro otro sist...		mastodinia. A la exploración clínica se aprecia una tumescencia elástica y no adherida a p...	
91120004	atracción del pezco, úlcera o enfermedad de Paget, en 1/3 casos (...)		mastodinia es descrita por 118 pacientes (9,51 por 100) y la presencia de secreción a trav...	
91120005	de un complejo sintomático abigarrado con un gran número de sint...		mastodinia, etc.). Entamos, en realidad, una vez verificado síndrome causado por la hiper...	
91120006	Por un lado, como tejido hormono-dependiente que es, con frecuen...		mastodinia tras el inicio de la terapia. Estos síntomas suelen preocupar a la paciente y con...	
91120007	actual inicia los preparados del fide de agnóstico para el tratamie...		mastodinia. Todas estas indicaciones están reconocidas por el Comité E del Ministerio...	
91120008	en caso de vómitos profundos, puede administrarse la dosis por vía...		mastodinia (hasta el 47%), que suele durar pocos días. Otros efectos secundarios, más raras...	
91120009	La incidencia de náuseas es del 20% y la de vómitos del 5%. Mier...		mastodinia, cefalea o mareos son otros posibles efectos. La pauta de progesterona en 2...	
91120010	mayor acción o más afinidad sobre los alo (mamaria y tiro). Esta eng...		mastodinia su proliferación endometrial (no es necesario ablatir un postgrado). En el incid...	
91120011	antecedentes de tabaquismo, que ingresó en el hospital por deterio...		mastodinia. En la exploración destacaba una ginecomastia bilateral dolorosa. En la radiog...	
91120012	durante 2 semanas (1 mg/día) se reduce la concentración de estrad...		mastodinia (Tabla 1). No obstante, tuvo que suspenderse el fármaco por intolerancia digest...	
91120013	no se producen variaciones del peso corporal en las mujeres del A...		mastodinia y alteraciones cutáneas (dismeteos, amonreas, etc.) en las mujeres que c...	
91120014	determina la evolución con la combinación de sistemas ("ambo") de...		mastodinia, spotting) y de la temperatura basal ("termica"). Así se consigue una mayor e...	
91120015	mensual. Solo aparece amonrea en el 6,20% de los casos. Los a...		mastodinia, náuseas, insomnio, ganancia de peso) aparecen muy eventualmente.	
91120016	esta situación estaría más indicada la inserción de un DIU. Efec...		mastodinia. La pauta clásica de Yupra (200 µg de etinilestradiol) y 1 mg de N300 está...	
91120017	de acción antandrogénica, por lo que se eliminan los moléculas efec...		mastodinia, disminución de la libido, disfunción erectil) que el Altablex® ocasiona en l...	
91120018	alleviación el dolor, hay que hablar de mastalgia como la expresión...		mastodinia bilateral sin galactomas, difusión erectil o disminución de la libido. Se desca...	
91120019	hegros por vía parenteral. Coincidiendo con la detección del VHC...		mastodinia, 10 sujetos presentaron disfunción erectil, 2 sujetos presentaron estomatitis y ot...	
91120020	plazo, lo que obligó a su suspensión (como efecto adverso...		mastodinia, demostros) o mastectomías en los pliegues submamarios, además de retromam...	
91120021	este tipo de efectos adversos. En estos pacientes se ha descrito dolor...		mastodinia otros relacionados con la hiperfrotia de las mamas, deslizo de evolución. En e...	
91120022	manejada por el Servicio de Traumatología o por fisioterapia...		mastodinia al año actual. Fue estudiada en su país, pero no disponía de informes médicos...	
91120023	en 74 libros normal (7,76-19 mg/dl). La paciente refiere un episodio...		mastodinia y depresión. Se ha descrito en su uso prolongado por más de una década, que se...	
91120024	esta terapia ha sido tan efectiva como terapia y danzaci en ensay...		mastodinia en pacientes que realizan tratamiento de depresión hormonal para el cáncer...	
91120025	antecedentes. Síntesis de evidencia: No encontramos con 3 posibles...		mastodinia, dada su alta incidencia, basos que la exploración física sea un arma funcio...	
91120026	amoxicilina 20 mg/d que ocasiona una disminución importante de l...		mastodinia, mientras que en el caso de la ginecomastia establecida de larga evolución la c...	
91120027	20 mg/d constituye la mejor opción para el tratamiento y la preven...		mastodinia en pacientes que realizan tratamiento de depresión hormonal para el cáncer...	
91120028	antecedentes. Síntesis de evidencia: No encontramos con 3 posibles...		mastodinia, dada su alta incidencia, basos que la exploración física sea un arma funcio...	
91120029	amoxicilina 20 mg/d que ocasiona una disminución importante de l...		mastodinia, mientras que en el caso de la ginecomastia establecida de larga evolución la c...	
91120030	20 mg/d constituye la mejor opción para el tratamiento y la preven...		mastodinia, la dificultad diagnóstica en mamas densas y muchas veces, solido de vez de l...	
91120031	deprección de hematomas, seromas, la sacrosal de los tejidos, la e...		mastodinia...	

Figura 23. Corpus IBERIA, concordancias para la voz mastodinia.

En CORDE no se recuperan casos para la consulta. En CREA/CREA anotado se recuperan seis casos en tres documentos divulgativos, todos de España, el primero fechado en 1992 (*La menopausia: Cómo vivirla y superarla sin miedo*, 1992; *Nueva guía de ginecología: 100 respuestas para la mujer*, 1999):

**REAL ACADEMIA ESPAÑOLA**

Concordancias (RAE)

Consulta: mastodinia, en 1985-2004, en todos los medios, en CREA  
 Resultado: 6 casos en 3 documentos.

OBTENCIÓN DE EJEMPLOS

Recuperar Concordancias Normal Clasificación: Año  
 Agrupación Marcas

Cómo citar el CORPUS Concordancias.  
 Pantalla: 1 de 1. Ver párrafos

Nº	CONCORDANCIA	AÑO	AUTOR
1	a atención que debieran. Mastalgia la mastalgia o mastodinia es el dolor que acompaña a la congestión m	** 1992	Pérez López, Faustino R.
2	Enfermedades benignas de la mama Dolor de mama o mastodinia. Está generalmente asociado a la menstruac	** 1999	Rapado, Aurelio
3	estinales, sensación de letargo y mareo, sudores, mastodinia (dolor de mamas), hipotensión, taquicardia	** 1999	Rapado, Aurelio
4	la enfermedad. 1. Dolor de mamas o mastodinia. La mastodinia no significa que la mujer padezca un cánc	1999	García del Real, Eduardo
5	ce o disminuye tras la menstruación. Este tipo de mastodinia se relaciona más con una enfermedad muy co	** 1999	García del Real, Eduardo
6	que se padezca la enfermedad. 1. Dolor de mamas o mastodinia. La mastodinia no significa que la mujer p	** 1999	García del Real, Eduardo

Ir arriba Pantalla: 1 de 1. Ver párrafos

Figura 24. CREA, concordancias de la voz mastodinia.

En cuanto a CORPES, figuran solo dos ocurrencias: una de España, 2014, tomada de un blog en línea, y otra de Chile, 2016, de la revista *Buena Salud*, también en línea:

The screenshot shows the CORPES web interface. At the top, it says 'Corpus del Español del Siglo XXI (CORPES)' and 'Versión beta (0.94)'. Below the navigation menu, the search criteria are set to 'Lema: mastodinia' and 'Forma: mastodinia'. The results section shows '2 casos en 2 documentos.' and a table with the following data:

RF	Clasificación, país	CONCORDANCIA	Ordenar por
1	2014 Esp.	Cambios en la sed y el apetito, deseos intensos de comer ciertos alimentos.	Año ascendente
2	2014 Chile	síntomas como irritabilidad, edema (hinchazón provocada por la retención de líquido);	sin criterio

Additional concordance information for 'mastodinia' is shown: 'mastodinia o sensibilidad mamaria aumentada, hinchazón y aumento de peso, dolor de cabeza.' and 'mastodinia (tensión o sensibilidad mamaria); etc.'

Figura 25. CORPES, concordancias de la voz *mastodinia*.

En el análisis comparado de este término, es clara la superioridad cuantitativa de documentaciones del corpus IBERIA, si bien estas provienen de revistas académicas y las de CORPES, de textos semiespecializados y divulgativos.

#### 4. CONCLUSIONES

Lo mostrado a lo largo del artículo permite extraer las siguientes conclusiones:

—Si bien los corpus académicos no tienen como objetivo recoger léxico especializado, hemos comprobado que sí pueden contenerlo y pueden ser útiles para obtener concordancias cuando estos han sido incluidos en textos divulgativos.

—Los datos también muestran la tendencia de que, a medida que se avanza cronológicamente hacia nuestros días, las documentaciones de estos términos aumentan en número. Esto es patente en la voz *rayos gamma* y, en menor medida, en *chromosfera*. Los corpus permitirían ver, de este modo, el proceso de divulgación de ciertos términos técnicos.

—No obstante, las documentaciones obtenidas para voces técnicas son todavía pocas y presentan lagunas en ciertos años o periodos cronológicos, que aparecen vacíos o con muy pocas ocurrencias, pues no se ha volcado en los corpus académicos documentación técnica suficiente. Sería preciso reforzar este aspecto.

—Precisamente ante voces técnicas que no están representadas con suficiente cantidad en los corpus académicos, IBERIA puede ser un recurso complementario digno de tenerse en cuenta, pues en algunos casos da muchas más documentaciones que los corpus académicos. En el ámbito médico, como hemos visto con la voz *mastodinia*, IBERIA parece ser un corpus especialmente rico. En otros campos, IBERIA puede ofrecer concordancias complementarias a las de los corpus académicos, especialmente CORPES, con el que comparte más similitudes cronológicas: así lo hemos mostrado con las ocurrencias de los términos *chromosfera* y *geosíncrono*, *-a*, con testimonios correspondientes a años que no están documentados en CORPES.

Es evidente que sería preciso realizar más búsquedas y en más campos para poder generalizar estas conclusiones, pero, en todo caso, lo observable es válido para los términos analizados y probablemente extrapolable a otros muchos semejantes de sus mismas áreas y campos.

Las búsquedas también han mostrado que IBERIA presenta problemas puntuales en la clasificación automática de los documentos y en la recuperación de concordancias; sería esperable una actualización y enriquecimiento progresivo del corpus con documentación de los años recientes para que pudiera aprovecharse en el futuro todo su potencial, si bien esto quizá dependa no tanto del factor humano como de la financiación que reciba para que sea

posible. No obstante, es un recurso provechoso que merece conocer la comunidad académica para potenciar los estudios del léxico especializado español.

## REFERENCIAS BIBLIOGRÁFICAS

- Ahumada Lara, I. (2010). El corpus Iberia como recurso para la traducción especializada. *Hikma: Estudios de traducción / Translation Studies*, 9, 9-24.
- Ahumada Lara, I. (2011). El español de la ciencia: ¿la identidad en crisis? / *Scientific Spanish: Identity in crisis*. En *Word for Word. The social, economic and political impact of Spanish and English / Palabra por palabra. El impacto social, económico y político del español y del inglés* (pp. 309-328). Santillana, British Council, Instituto Cervantes.  
[https://www.britishcouncil.es/sites/default/files/book\\_word\\_for\\_word\\_completo.pdf](https://www.britishcouncil.es/sites/default/files/book_word_for_word_completo.pdf)
- Ahumada Lara, I. (2013). *Lingüística de corpus y traducción: el corpus Iberia*. Presentación en el X Congreso Internacional de Traducción, Textos e Interferencias (Augsburgo, 4 al 6 de septiembre de 2013).
- Alvar Ezquerro, M. (2003). *La estación lexicográfica*. En M.<sup>a</sup> A. Martín Zorraquino y J. L. Aliaga Jiménez (coords.). *La lexicografía hispánica ante el siglo XXI. Balances y perspectivas* (pp. 225-230). Gobierno de Aragón, Departamento de Educación, Cultura y Deporte, Institución Fernando El Católico.
- Cabré, M. T., L. de Yzaguirre, M. Lorente y A. Matamala. (2004). *La Estación de Trabajo Lexicográfico (ETL). La ingeniería lingüística al servicio del profesional de la lexicografía*. En M.<sup>a</sup> P. Battaner y J. DeCesaris (eds.). *De lexicografía: actes del I Symposium Internacional de Lexicografía* (Barcelona, 16-18 de maig de 2002) (pp. 277-286). Universitat Pompeu Fabra, Institut Universitari de Lingüística Aplicada (IULA).
- Castroviejo Bolibar, S. (coord.). 1986-2012. *Flora Iberica: Plantas vasculares de la Península Ibérica e islas Baleares* 1-8, 10-15, 17-18, 21. CSIC, Real Jardín Botánico.
- Halliday, M. A. K., W. Teubert, C. Yallop y A. Čermáková. (2004). *Lexicology and Corpus Linguistics: An Introduction*. Continuum.
- Martín Herrero, C. (2009). Aproximación a ciertas perspectivas en Lingüística de corpus. En P. Cantos Gómez y A. Sánchez Pérez (Eds.). *A Survey of Corpus-based Research* (pp. 1020-1032). Asociación Española de Lingüística del Corpus.
- Pablo Núñez, L. (2010). *Fuentes de información y otros recursos terminológicos para traductores: la labor del CSIC* (Aranjuez, CES Felipe II, 19/05/2010). <http://hdl.handle.net/10261/24626> (14-03-2022)
- Pérez Hernández, M. C. (2002). Explotación de los corpora textuales informatizados para la creación de bases de datos terminológicas basadas en el conocimiento. *Estudios de Lingüística del Español (ELiEs)*, 18. <http://elies.rediris.es/elies18/> (11-01-2022)
- Porta, J., E. Rosal y I. Ahumada. (2011). Design and Development of Iberia: A Corpus of Scientific Spanish. *Corpora*, 6(2), 145-158.

## Diccionarios, corpus en línea y bancos de datos electrónicos

- CORDE = RAE. Banco de datos *CORDE: Corpus diacrónico del español*. <https://corpus.rae.es/cordenet.html> (30-03-2022)
- CORPES = RAE: Banco de datos *CORPES XXI: Corpus del Español del Siglo XXI*. <http://www.rae.es/> (30-03-2022)
- CREA = RAE. Banco de datos *CREA: Corpus de referencia del español actual*. <https://corpus.rae.es/creanet.html> (30-03-2022)



- CREA anotado = RAE. Banco de datos *CREA* (versión anotada). *Corpus de referencia del español actual*. <https://apps2.rae.es/CREA/> (30-03-2022)
- DEI = Real Academia de Ingeniería. *Diccionario español de ingeniería* (DEI 1.0). <https://diccionario.raing.es/es> (30-03-2022)
- DLE = RAE/ASALE. *Diccionario de la lengua española*. 23.5 ed. <https://dle.rae.es/> (30-03-2022)
- DTME = Real Academia Nacional de Medicina de España. *Diccionario de términos médicos*. <http://dtme.ranm.es/index.aspx> (30-03-2022)
- Enclave de Ciencia = RAE/FECYT: <https://enclavedeciencia.rae.es/contenidos/inicio>
- IATE = Unión Europea. Banco de datos *IATE* (*Interactive Terminology for Europe*). En línea: <https://iate.europa.eu/home> (30-03-2022)
- IBERIA = Ahumada Lara, I. (dir.). *Corpus IBERIA: El español como lengua del conocimiento especializado*. En línea: <http://iberia.cchs.csic.es/cgi-bin/iberia.cgi> (30-03-2022)
- EuroTerm Bank = Unión Europea. *TermBank*. <https://www.eurotermbank.com/> (26-03-2022)
- TERMIUMPlus* = Canada. Banque de données terminologiques et linguistiques du Gouvernement du Canada. Bureau de la traduction. <https://www.btb.termiumplus.gc.ca/> (25-03-2022)
- VCTRAC = Real Academia de Ciencias Exactas, Físicas y Naturales de España. *Vocabulario científico-técnico*. <https://vctrac.es/> (30-03-2022)

## A corpus-based study of 4-grams in the research article genre

### Estudio basado en corpus de 4-gramas en el artículo científico

EVA LUCÍA JIMÉNEZ-NAVARRO  
Universidad de Córdoba, España

[lucia.jimenez@uco.es](mailto:lucia.jimenez@uco.es)

<https://orcid.org/0000-0001-9377-6921>

#### Abstract

The analysis of phraseology in the specialized discourse of science has sparked researchers' interest in the last few decades, probably because the use of word groupings in specific registers can provide information about certain typical features of the genre. For instance, Gledhill (2009) explores colligations of tenses in scientific articles and discovers that the present tense is used for qualitative and empirical expressions, while the past tense provides quantitative and research-oriented descriptions; Pérez-Llantada (2014) investigates 4-word lexical bundles in research articles, finding that these multiword combinations express referential meaning and organize the text; finally, Jiménez-Navarro (2019) analyzes adjective + noun collocations in a corpus of scientific papers and concludes that these phraseological units convey specific meanings when used in this genre, since they represent the contents of research articles.

The aim of the current study is to contribute to the analysis of 4-grams in the language of science. To this end, two specific objectives are defined: first,

#### Resumen

El análisis de la fraseología en el discurso especializado de la ciencia ha despertado el interés de los/as investigadores/as en las últimas décadas, probablemente porque el uso de grupos de palabras en registros específicos puede informar de algunas características típicas del género. Por ejemplo, Gledhill (2009) explora las coligaciones de tiempos verbales en artículos científicos y descubre que el tiempo presente se usa para expresiones cualitativas y empíricas, mientras que el tiempo pasado proporciona descripciones cuantitativas y orientadas a la investigación; Pérez-Llantada (2014) investiga grupos léxicos de cuatro palabras en artículos de investigación y descubre que estas combinaciones multilexémicas expresan significado referencial y organizan el texto; finalmente, Jiménez-Navarro (2019) analiza colocaciones de adjetivo + sustantivo en un corpus de artículos científicos y concluye que estas unidades fraseológicas aportan significados específicos cuando se usan en este género,

---

**To cite this article:** Jiménez-Navarro, E. L. (2022). A corpus-based study of 4-grams in the research article genre. *ELUA*, (38), 241-262. <https://doi.org/10.14198/ELUA.22267>

Recibido: 16/03/2022 Aceptado: 14/05/2022

© 2022 Eva Lucía Jiménez-Navarro



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

to ascertain the structure of 4-grams; second, to analyze the function they perform. The methodology was based on a corpus and entailed five major steps: (1) a specialized corpus of research articles was built, (2) a list of 4-grams was automatically extracted using the software Sketch Engine, (3) the resulting list was manually verified in order to suppress inaccurate candidates, (4) the selected units were classified depending on their structural framework, and (5) the selected units were categorized according to their function in the text. The findings show that, in terms of the first objective, the most typical 4-grams were noun phrases; and as for the second objective, the sequences examined mostly concerned the research conducted and the authorship of the texts. All in all, the 4-grams identified were structures that were specific to the genre under study but could also be used in other domains.

**KEYWORDS:** specialized corpus; 4-gram; function; research article; structure.

puesto que representan los contenidos del artículo de investigación.

El objetivo de este estudio es contribuir al análisis de 4-gramas en el lenguaje de la ciencia. Para ello, se han definido dos objetivos específicos: en primer lugar, establecer la estructura de estas secuencias de palabras; en segundo lugar, analizar su función. La metodología empleada se basó en corpus y conllevó cinco pasos principales: (1) la construcción de un corpus especializado de artículos científicos, (2) la extracción de una lista de 4-gramas de manera automática usando el software Sketch Engine, (3) la verificación manual de esa lista para eliminar candidatos inadecuados, (4) la clasificación de las unidades seleccionadas dependiendo de su estructura, y (5) la categorización de las unidades seleccionadas según su función en el texto. Los resultados muestran que, con respecto al primer objetivo, los 4-gramas más típicos fueron sintagmas nominales; en relación con el segundo objetivo, las secuencias examinadas trataban principalmente con la investigación llevada a cabo y la autoría de los textos. En conjunto, podemos decir que estas estructuras eran específicas del género estudiado, aunque también podrían ser usadas en otros dominios.

**PALABRAS CLAVE:** corpus especializado; 4-grama; función; artículo científico; estructura.

## 1. INTRODUCTION

According to Sinclair (2000: 197), 80% of the combinations of words are made through co-selections, as corpus evidence has demonstrated, which has relegated grammar to the background and may explain the exponentially growing interest in the field of phraseology in the last decades. Phraseology, deriving from Greek *phrasis* and *logia*, can be defined as the discipline which studies groups of words. It can be considered a dynamic one, for not only does it motivate the analysis of types of words that combine with each other, but also other aspects, such as the psychological factors behind the way we recall word combinations from memory, the steps involved in teaching and learning them in second language acquisition, their semantic status in specialized languages, and so forth.

The establishment of Phraseology as a discipline was supported by a group of Russian scholars conducting research in the 1940s. They became especially active in this field, their aim being to categorize all set expressions as well as to describe their varying degrees of fixedness. The group was led by Victor V. Vinogradov (he published his selected works in 1947) and Natalia N. Amosova (she studied the foundations of English phraseology in 1963), who prepared the ground for further work on the classification of word combinations. For

instance, Klappenbach (1968), Weinreich (1969), and Lipka (1974) followed Vinogradov's steps in the 1960s and 1970s, and other authors working in the 1980s and 1990s deserve special mention for stating more practical than theoretical purposes, namely Anthony P. Cowie (1981, 1998), Peter A. Howarth (1996), and Igor A. Mel'čuk (2006, 2012).

As to the phraseological units emerging from the categorizations of these groups of authors, these sequences were defined according to two main criteria: degree of fixedness and compositionality. As such, a continuum is conceived where free combinations (i.e., the most flexible and most transparent units) are located at one end and idiomatic expressions (i.e., the most fixed and most opaque structures) are at the other. In the middle are collocations, defined as typical combinations of two lexical items. They have normally been studied from two perspectives (Nesselhauf 2005; Granger & Paquot 2008; Gablasova *et al.* 2017): the phraseological approach, according to which they are partly stable and transparent, and the statistical approach, focused on the frequency of the co-occurrence of the two words.

Aside from that, other phraseological units have also been identified that are not based on the criteria of fixedness and compositionality. For example, *lexical bundles* are defined as "extended collocations: sequences of three or more words that show a statistical tendency to co-occur" (Biber & Conrad 1999: 183). Their importance in a language is highlighted by Biber and Barbieri (2007), who state that they "serve important discourse functions in both spoken and written texts" (p. 264). Another instance is the concept of *collostruction*, a term coined by Stefanowitsch and Gries in 2003 that refers to the combination of a lexeme and a construction, both of which being mutually attracted. Finally, a *collocational catena* is described as "an independently recurrent, multiword dependency subtree which is formally and/or contextually stereotyped" (Pęzik 2018: 96). To put it another way, several words which establish dependency relations combine to form longer structures, therefore, a collocational catena can be regarded as a cluster of several collocations.

Regarding the type of combination explored in this paper, an *n-gram* can be defined as "a recurrent string of uninterrupted word-forms" (Stubbs 2007: 90); the letter 'n' defines the number of elements involved. In this work, 4-grams are to be examined, which means that the focus will be on uninterrupted sequences of four words. Many scholars have acknowledged that n-gram is another name for lexical bundle (e.g., Allen 2010; Granger 2014; Kwary *et al.* 2017; Biel 2018). This type of multiword expression has usually been tackled through a frequency-based approach (Biber *et al.* 1999; Wray 2008) and has played a crucial role in linguistic production (Pawley & Syder 1983; Thomson 2017). For instance, n-grams are said to be a marker of proficient language use and native-like language competence, to contribute to natural language use and greater fluency, and to facilitate language production. These may be some of the reasons why n-grams have become a key issue in specialized languages, such as tourism (Fuster-Márquez 2014), academic writing (Shin *et al.* 2018), and economics (Aimenova *et al.* 2019).

This study seeks to contribute to the analysis of 4-grams in the specialized language of science, more specifically, in the research article genre. The underlying hypothesis being that these phraseological units can help to structure the text and convey specific nuances of meaning when used in a concrete text type. In order to test this hypothesis, two specific aims will be pursued: (1) to explore the structural frameworks of these sequences of words, and (2) to examine their function in the text. A compelling reason for focusing on 4-grams and not 3- or 5-grams is that 4-grams often subsume 3-grams (Pérez-Llantada 2014: 86).

Furthermore, 4-grams are far more frequent than 5-grams and present a clearer range of structures and functions (Hyland 2008: 44). The organization of the paper is as follows. I start with a broad overview of studies on word sequences in research articles (Section 2, Background). Then, I describe the methodology employed to achieve my objectives (Section 3, Methodology). After that, I present the main findings of the study (Section 4, Results) and compare them to those obtained in previous works (Section 5, Discussion). Finally, I summarize the conclusions drawn as well as provide some possible research lines for the future (Section 6, Conclusions and Further Research).

## 2. BACKGROUND

The phraseology that characterizes the scientific genre has attracted the attention of linguists for many years, and a wide variety of word combinations have been analyzed, such as collocations (Luzón Marco 2000; Jiménez-Navarro 2019), colligations (Gledhill 2009; Menon & Mukundan 2010), and lexical bundles (Pérez-Llantada 2014). These studies have supported Biber and Barbieri's (2007) idea that each register has a distinct set of phraseological units which are "associated with the typical communicative purposes of that register" (p. 265). To exemplify, Luzón Marco (2000) finds that the linguistic conventions of a genre determine the specific collocates chosen as well as their meaning in the text, the latter being related to measure, quantification, probability, or cause/result. On the other hand, Menon and Mukundan (2010) point out that some elements need to be interpreted beyond their literal sense as they convey a more specific and/or extended meaning in the scientific field. In this section, I will describe in more detail some of the investigations into the role of n-grams<sup>1</sup> that have been carried out in several disciplines.

First, Jalali *et al.*'s (2015) aim is threefold: (1) to retrieve the most frequent lexical bundles from COMRA, the Corpus of Medical Research Articles; (2) to explore the forms of these multiword units; and (3) to discover their function in the texts. Their corpus contains 790 articles published between 2009 and 2011 covering 33 subject areas (including oncology, psychiatry, surgery, public health, and health policy), resulting in around 2.4 million words. In their study, they only focus on 4-word sequences and extract 102 different items of this type, highlighting that just 0.3% of the whole corpus consists of lexical bundles. The results show that almost half (i.e., 44.5%) of these lexical bundles selected are prepositional phrases, and the most common function performed is related to the text itself. For example, more than 40% of these units are used to express transition and result, among others. The authors' conclusion is that "the context establishes the function of bundles" (p. 61).

Second, Jalilifar *et al.* (2017) try to identify common core lexical bundles across three disciplines, namely, arts and humanities, sciences, and social sciences. Their 6-million-word corpus contains over 1,300 research articles published between 2010 and 2015. Unlike Jalali *et al.* (2015), these authors consider 3-grams, 4-grams, and 5-grams, extracting separate lists from each discipline and then collapsing them into one single list which includes all types of n-grams retrieved from every discipline. They set a minimum frequency of three co-occurrences for a particular lexical bundle to be regarded as a core one. They extract a total of 11,670 bundles, 661 of them being core and 3-word sequences

---

<sup>1</sup> They are referred to as lexical bundles in the studies described here.

being the most frequent (85%). In their opinion, these results reveal that lexical bundles are highly frequent and significant in academic writing. As for their function in the text, they are mostly text-oriented (58.5%), which means that they are used to focus the reader on the text. In addition, these authors address the extent to which these core bundles overlap with the bundles from a corpus of applied linguistics and discover that 593 (out of 661) sequences are central to this discipline too.

And, finally, Jalilifar and Ghoreishi (2018) make use of the previous study to find the proportion of general and discipline-specific formulaic sequences in a corpus of research articles from the field of applied linguistics. They retrieve sequences of 3, 4, and 5 words, setting a minimum frequency threshold of 10 tokens per million words and a dispersion criterion of five texts (i.e., the lexical bundles should occur in at least five different texts from the corpus). Their results show that general sequences account for 53% of all the extractions, and discipline-specific bundles for 47%. In these authors' opinion, these findings demonstrate that multiword units have clear pedagogical implications for academic writing students, in that they should be aware of the existence and relevance of these word sequences in order to incorporate them into their written production. Furthermore, as to function, again text-oriented sequences are most common in both groups of the formulaic sequences under study, that is, 58.5% in the general context and 55.2% in the discipline-specific context.

In short, the main idea suggested by the previous studies is that the most recurrent n-grams found in research articles across different disciplines (i.e., medicine, arts and humanities, sciences, social sciences, and applied linguistics) are concerned with the organization of the text and its meaning. This is in consonance with the findings of other research into the language of science, such as Luzón Marco (2000) and Pérez-Llantada (2014). The current paper aims to learn more about the structure and the function performed by 4-grams in, specifically, scientific research articles. The next section addresses the methodological steps followed in order to achieve this goal.

### 3. METHODOLOGY

Without a doubt, the composition of the corpus is an essential element in corpus-based and corpus-driven studies, since the results obtained will depend solely on its characteristics and, of course, on the researchers' skills in obtaining and evaluating these results. For this reason, how it is compiled is of supreme importance. The construction of UCOSCIENCOR, the specialized corpus used in the current research, will comprise the first part of this section. After that, I will explain the extraction of 4-grams, their classification according to their structural framework, and their categorization based on their function in the text.

#### 3.1. Compilation of UCOSCIENCOR

UCOSCIENCOR is a specialized corpus of over three million words built for the linguistic analysis of the English language used by non-native speakers who are part of the Universidad de Córdoba (UCO, Spain). It contains more than 600 publicly available research articles dealing with the disciplines of medicine, veterinary science, and genetics, and published in high-impact international journals (such as *Nature Reviews Immunology*, *The Lancet*, *Nature Reviews Molecular Cell Biology*, *Vaccine*) from 1980 to 2020 by leading scholars. Given

the specificity of the corpus, the texts were manually selected first and then automatically annotated using the corpus management software Sketch Engine. The characteristics of the texts selected (e.g., authorship, place of publication, year of publication, length) and the large size of the corpus make UCOSCIENCOR representative of the genre under study.<sup>2</sup> Table 1 provides a summary of the main features of UCOSCIENCOR.

<b>Number of words</b>	3,009,296
<b>Type of corpus</b>	Specialized / electronic
<b>Size</b>	Medium
<b>Mode</b>	Written
<b>Language</b>	English
<b>Domain</b>	Science
<b>Subdomain</b>	Medicine / Veterinary Science / Genetics
<b>Genre</b>	Research article
<b>Text length</b>	Complete texts
<b>Purpose</b>	Linguistic analysis
<b>Communicative situation</b>	Specialized
<b>Publication date</b>	1980-2020
<b>Source of texts</b>	High-impact international journals
<b>Publishers</b>	Non-native UCO academic researchers

Table 1. Defining characteristics of UCOSCIENCOR.

### 3.2. Extraction of 4-grams

After having compiled my specialized corpus, the ‘N-grams’ function available at Sketch Engine was used to extract 4-word sequences typical of the corpus. Following Jalali *et al.* (2015), the frequency threshold set was 20 tokens per million words, which means that only those 4-grams occurring at least 60 times were retrieved. Using a minimum frequency criterion helps to identify groupings of words characteristic of the language under analysis and to avoid arbitrary uses of language. Apart from this, another criterion established was that no distinction was made between lowercase and uppercase versions of the same sequence.

This step produced a list of 251 items with a total frequency of 33,784 tokens in the corpus. However, manual work was needed to evaluate the results and 70 items were discarded for the following reasons:

- (1) They were linked to journal names, for instance, *j am soc nephrol*, *am j clin nutr*;
- (2) They were part of an institution’s name, such as *reina sofia university hospital*, *instituto maimónides de investigación*;
- (3) They were abbreviations of words, for example, *t i c l*, *e i n f*;

<sup>2</sup> According to Vargas Sierra (2006: 11) and Corpas Pastor and Seghiri Domínguez (2010: 124), specialized texts are denser than general language texts when it comes to terminology, consequently, a specialized corpus does not need to contain as many millions of words as it would be desirable for a reference corpus so as to be representative of a specific domain.

- (4) They were irrelevant to the focus of this study, such as *elsevier b.v. all rights, lists available at sciencedirect*.

The final list of 4-grams amounted to 181 items.

### 3.3. Structural categorization of 4-grams

The next step in this research was the categorization of the selected 4-grams on the basis of their structure following Biber *et al.* (1999: 96-106), who describe five types of phrase, namely: (1) noun phrases, that is, phrases with a noun as head (e.g., *her gold watch, any printed material discovered*); (2) verb phrases, which contain a lexical verb or primary verb as head or main verb, either alone or accompanied by one or more auxiliaries (e.g., *had been making, should have said*); (3) adjective phrases, with an adjective as head, optionally accompanied by modifiers (e.g., *slow to respond, guilty of a serious crime*); (4) adverb phrases, which contain an adverb as head, optionally accompanied by modifiers (e.g., *quite melodiously, much more quickly than envisaged*); and (5) prepositional phrases, consisting of a preposition and a complement, most typically in the form of a noun phrase (e.g., *in the morning, to Sue*).

### 3.4. Functional categorization of 4-grams

The final step of my methodology was the categorization of the 4-grams selected according to their function in the text. To do so, I followed Hyland (2008: 49), who suggests three broad categories loosely based on Halliday's (1994) linguistic macrofunctions: (1) research-oriented, which serves an ideational function; (2) text-oriented, that is, phrases concerned with textual functions; and (3) participant-oriented, expressing interpersonal meanings. More detail about this taxonomy is presented below:

- (1) Research-oriented sequences. They help writers to structure the information.
  - a. Location, which indicates time and place (e.g., *at the beginning of, at the same time, in the present study*).
  - b. Procedure, concerning methods and processes (e.g., *the use of the, the role of the, the purpose of the*).
  - c. Quantification, related to quantities (e.g., *the magnitude of the, a wide range of, one of the most*).
  - d. Description, used to present facts (e.g., *the structure of the, the size of the, were no significant differences*).
  - e. Topic, connected with the field of research (e.g., *in the Hong Kong, the currency board system, risk of cardiovascular disease*).
- (2) Text-oriented sequences. These are concerned with the organization of the text and the meaning of its elements as a message or argument.
  - a. Transition signals, which establish additive or contrastive links between elements (e.g., *on the other hand, in addition to the, in contrast to the*).
  - b. Resultative signals, which mark inferential or causative relations between elements (e.g., *as a result of, it was found that, these results suggest that*).
  - c. Structuring signals, defined as text-reflexive markers which organize stretches of



discourse or direct the reader elsewhere in text (e.g., *in the present study*, *in the next section*, *as shown in fig.*).

- d. Framing signals, which situate arguments through specifying limiting conditions (e.g., *with respect to the*, *on the basis of*, *with the exception of*).

(3) Participant-oriented sequences. These are focused on the writer or the reader of the text (Hyland 2005):

- a. Stance features, which convey the writer's attitudes and evaluations (e.g., *are likely to be*, *may be due to*, *it is possible that*).
- b. Engagement features, addressing readers directly (e.g., *it should be noted*, *as can be seen*, *it must be highlighted*).

Additionally, I decided to create a third subcategory of participant-oriented sequences that I called authorship-related, which concerns details linked to the authors of the publications and the publications themselves (e.g., *the author of this*, *profiles for this publication*, *would like to thank*), given the relevance of such phrases in UCOSCIENCOR, as I will explain in Section 4.2.

#### 4. RESULTS

In total, 181 items were manually selected and analyzed both structurally and functionally in this work. The full list of 4-grams is provided in Appendix 1, although Table 2 shows the top ten items. They are organized in terms of their frequency in the corpus and are accompanied by their frequency per million (i.e., their relative frequency), the number of texts in which they occur, and the relative document frequency (i.e., the percentage of texts that contain the item).

Rank	4-gram	Frequency	Frequency per million	Document frequency	Relative document frequency
1	<i>in the present study</i>	376	86.64	177	29.45
2	<i>the authors of this</i>	375	86.41	72	11.98
3	<i>of the authors of</i>	372	86.41	372	61.89
4	<i>of this publication are</i>	361	83.86	361	60.06
5	<i>some of the authors</i>	360	83.62	360	59.90
6	<i>the user has requested</i>	360	83.62	360	59.90
7	<i>page was uploaded by</i>	360	83.62	360	59.90
8	<i>and author profiles for</i>	360	83.62	360	59.90
9	<i>following this page was</i>	360	83.62	360	59.90
10	<i>user has requested enhancement</i>	360	83.62	360	59.90

Table 2. Top ten 4-grams extracted from UCOSCIENCOR.

As can be seen, the most frequent 4-gram is *in the present study*, with 376 tokens occurring in 177 different texts, which makes a corpus distribution of 29.45%. Other recurrent and more highly distributed sequences are *of the authors of* (372 tokens, 61.89%) and *of this publication are* (361 tokens, 60.06%), along with the rest of 4-grams included in the table,

which occur 360 times in 360 texts (59.9%). However, the case of *the authors of this* must be highlighted, given that it is highly recurrent (375 tokens) but unevenly distributed (11.98%).

At the other end of the list, there are nine 4-grams that occur only 60 times, which was the minimum frequency threshold set in the extraction step, namely: *was significantly higher in*, *that the presence of*, *study was to evaluate*, *seroprevalence of toxoplasma gondii*, *the area under the*, and *the number of*, *was supported by the*, *not included in the*, and *was found to be*. With respect to their distribution in the corpus, *was supported by the* and *not included in the* show the highest distribution (9.65% and 9.15%, respectively), followed by *study was to evaluate* and *and the number of* (8.48% and 8.15%, respectively). After that, the most widely distributed 4-grams are: *was significantly higher in* (7.15%), *that the presence of* (6.48%), *the area under the* (6.32%), and *was found to be* (6.32%). Finally, the least distributed sequence is *seroprevalence of toxoplasma gondii* (2.66%), which may be due to the fact that it is so much more specific than the others and therefore found in a lower number of texts (16 out of 601 texts).

#### 4.1. Structural features of 4-grams in UCOSCIENCOR

The first aim pursued in this paper was the structural classification of the 181 4-grams selected. As I have previously mentioned, Biber *et al.*'s (1999: 96-106) work inspired this step, thus, the 181 sequences were divided into five different groups: (1) noun phrases, (2) verb phrases, (3) adjective phrases, (4) adverb phrases, and (5) prepositional phrases. Table 3 shows this classification, which is organized in terms of the number of 4-grams found in each type of phrase.

Type of phrase	Number of types	Number of tokens	% in the corpus	Examples
Noun phrase	80	10,627	45.6	<i>some of the authors</i> , <i>the results of the</i> , <i>the present study was</i>
Prepositional phrase	52	6,938	29.8	<i>at the end of</i> , <i>in the presence of</i> , <i>for the treatment of</i>
Verb phrase	43	4,904	21.1	<i>also working on these</i> , <i>are shown in table</i> , <i>were included in the</i>
Adjective phrase	4	576	2.5	<i>it is possible that</i> , <i>was significantly higher in</i>
Adverb phrase	2	243	1	<i>as well as the</i> , <i>as well as in</i>
Total	<b>181</b>	<b>23,288</b>	<b>100</b>	

Table 3. Structural classification of the 4-grams selected.

Table 3 reveals that the largest structural category of 4-grams is noun phrases, with 80 distinct types (i.e., distinct phrases regardless of their repetition in the corpus), which accounts for almost half (45.6%) of the total number of sequences. The number of tokens found in the corpus is 10,627, and some of the types of phrases included in this group consist of a noun phrase plus a preposition. In terms of frequency, the combination noun phrase + *of* (e.g.,

*the end of the, the effect of the*) is the most recurrent (38.75%), followed at a considerable distance (7.5%) by the pattern noun phrase + *in* (e.g., *an important role in, a significant increase in*), noun phrase + *with* (6.25%, e.g., *patients with ankylosing spondylitis, patients with rheumatoid arthritis*), and noun phrase + *for* (5%, e.g., *anova for repeated measures, profiles for this publication*).

The second largest category is prepositional phrases (29.8%). In total, 52 4-grams are identified, with an overall frequency of 6,938 tokens in the corpus. The most typical preposition introducing a phrase is *in* (34.62%, e.g., *in advanced colorectal cancer, in the treatment of*), followed by *of* (19.23%, e.g., *of this publication are, of the mediterranean diet*) and *for* (11.54%, e.g., *for the management of, for the first time*). Some of the most recurrent phrases in UCOSCIENCOR are prepositional, such as *in the present study* (376 tokens, the highest frequency in the corpus), *of the authors of* (372 tokens), and *of this publication are* (361 tokens).

Verb phrase is the third more recurrent category in the corpus (21.1%). In total, 43 different 4-grams are found consisting of 4,904 tokens. Various subtypes are observed, the most common (37.2%) being the combination passive verb + prepositional phrase (e.g., *was approved by the, has been associated with*). After that, 16.28% of the recurrences belong to two types, namely, verb *to be* + adjective/noun (e.g., *be due to the, is one of the*) and main verb + prepositional phrase (e.g., *distributed under the terms, included in the study*). Next, in terms of quantity, 9.3% of the verb phrases is represented by the combination main verb + noun phrase (e.g., *plays an important role, has requested enhancement of*), and 6.97% belongs to two types, anticipatory *it* + passive (e.g., *it has been shown, it has been suggested*) and modal verb phrase (e.g., *could be due to, would like to thank*). The passive forms achieve the highest overall frequency with 1,870 tokens.

Finally, adjective phrases and adverb phrases are the least common types of sequence in UCOSCIENCOR. As for the former, only 4 distinct 4-grams (2.5%), with 576 tokens, are included, examples being *it is important to* and *it is possible that*. Regarding adverb phrases, only 2 different 4-grams (1%), with a total frequency of 243 tokens, are identified: *as well as the* and *as well as in*.

#### 4.2. Functions of 4-grams in UCOSCIENCOR

The second aim stated in this work was the functional classification of the 181 4-grams selected. To this end (as previously mentioned), I followed Hyland (2008: 49), who proposes three large categories: research-oriented, text-oriented, and participant-oriented sequences. Each of them presents several subcategories. Additionally, I decided to create a new subcategory in the participant-oriented group, that is, authorship-related. Table 4 shows the results of this classification and includes the number of 4-grams, their overall frequency in UCOSCIENCOR, and the percentage of the categories and subcategories in the corpus.

Category/Subcategory	Number of types	Number of tokens	% in the corpus
Research-oriented	109	10,715	59.9
Location	15	2,431	8.2
Procedure	29	2,777	16

Category/Subcategory	Number of types	Number of tokens	% in the corpus
Quantification	13	1,302	7.2
Description	27	2,146	14.8
Topic	25	2,059	13.7
Text-oriented	33	3,572	18
Transition signals	4	617	2.2
Resultative signals	14	1,191	7.6
Structuring signals	1	194	0.6
Framing signals	14	1,570	7.6
Participant-oriented	39	9,001	22.1
Stance features	11	805	6.7
Engagement features	0	0	0
Authorship-related	28	8,196	15.4
<b>Total</b>	<b>181</b>	<b>23,288</b>	<b>100</b>

Table 4. Functional classification of the 4-grams selected.

As the table makes clear, the largest amount of 4-grams (59.9%) belong to the research-oriented category, specifically, 109 items. The sequences included in this group describe various features of the research presented in the texts of UCOSCIENCOR: (1) 29 items (16%) address the methods and purposes of the research so are in the procedure subcategory (e.g., *aim of this study, play an important role, for the treatment of*); (2) 27 items (14.8%) represent qualities of the studies within the description subcategory (e.g., *an increased risk of, no significant differences in, the influence of the*); (3) 25 sequences (13.7%) are to be found in the topic subcategory, given that they are related to the field of research (e.g., *ankylosing spondylitis functional index, microbiology and infectious diseases, patients with rheumatoid arthritis*); (4) 15 sequences (8.2%) belong to the location subcategory and indicate time and place (e.g., *after consumption of the, at the same time, on these related projects*); and (5) the least recurrent subcategory is quantification with 13 items (7.2%), which describe amounts and numbers (e.g., *a large number of, a wide range of, the rest of the*). In spite of the fact that both the description and topic subcategories are represented by a much larger number of 4-grams than the location subcategory (27 and 25 items, respectively, vs. 15 items), it should be highlighted that, if overall frequency of the sequences is considered, the location subcategory is more recurrent than the others, having 2,431 tokens in total, while the description subcategory has 2,146 tokens and the topic subcategory, 2,059 tokens.

Secondly, 39 items (22.1%) are in the participant-oriented category. Despite having previously identified three subgroups, only two of them are in fact evoked by the 4-grams selected. On the one hand, the newly created authorship-related subcategory contains 28 sequences (15.4%) that concern details related to the authors of the publications and the publications themselves (e.g., *authors of this publication, for personal use only, this work was supported*). Needless to say, this new subcategory is essential, containing as it does 8,196 tokens in total. On the other hand, the stance features subcategory is less commonly found and includes 11 items (6.7%), which convey the writers' attitudes and evaluations (e.g., *has been associated with, it is important to, was found to be*). In this work, the engagement features subcategory is not evoked by any of the 4-grams selected.

Finally, 33 sequences (18%) are included in the text-oriented category, where four sub-categories are considered: (1) framing signals, which contain 14 sequences (7.6%) used to limit conditions (e.g., *in relation to the*, *in the absence of*, *in the case of*); (2) resultative signals, also 14 items (7.6%) although with a lower frequency in the corpus (1,191 vs. 1,570 for framing signals), which mark inferential or causative relations between elements (e.g., *as a consequence of*, *as a result of*, *studies have shown that*); (3) transition signals, with only 4 sequences (2.2%), are used to establish additive or contrastive links between elements (e.g., *in addition to the*, *on the other hand*, *as well as the*); and (4) structuring signals, with only 1 item (0.6%, i.e., *are shown in table*) which directs the reader elsewhere in the text.

### 4.3. Interconnection between structures and functions of 4-grams in UCOSCIENCOR

After having examined the most common structural frameworks found in UCOSCIENCOR and the functions they perform, the interconnection between these two central aspects cannot be disregarded. Thus, Table 5 illustrates the type of structure used to serve each of the functions; the number of 4-grams representing the distinct categories and their percentage in UCOSCIENCOR are shown.

	Research-oriented		Participant-oriented		Text-oriented	
Noun phrase	64	58.8%	14	35.8%	2	6.1%
Prepositional phrase	30	27.5%	4	10.3%	18	54.4%
Verb phrase	14	12.8%	18	46.2%	11	33.4%
Adjective phrase	1	0.9%	3	7.7%	0	0%
Adverb phrase	0	0%	0	0%	2	6.1%
Total	<b>109</b>	<b>100%</b>	<b>39</b>	<b>100%</b>	<b>33</b>	<b>100%</b>

Table 5. Structural frameworks used in the functional categorization.

Table 5 shows that each major functional category is mostly characterized by a different specific structural framework. Noun phrases are the most typical structure when it comes to research-oriented 4-grams (e.g., *statistically significant differences were*, *the influence of the*), with 64 structures of this type identified that account for 58.8% of the structures used in this function. Verb phrases are common to represent participant-oriented sequences (e.g., *be explained by the*, *be due to the*), 18 of them comprising 46.2% of the structures identified. A coherent explanation for this may be that it is specifically the use of certain verbs that makes possible the conveyance of the writer's attitude (e.g., *would like to thank*) or the carrying out of their work (e.g., *was supported by the*).

And finally, 18 prepositional phrases describe text-oriented 4-grams (e.g., *in the pathogenesis of*, *under the terms of*), making up 54.4% of the structures found in this function. Without a doubt, prepositions are very much needed to organize the text by introducing additions (e.g., *in addition to the*), contrasts (e.g., *on the other hand*), marking relations between elements (e.g., *as a consequence of*), or framing arguments (e.g., *in relation to the*), among other functions.

## 5. DISCUSSION

In the first place, in terms of the structural frameworks identified in this work, it must be pointed out that these results coincide with those obtained by Jalali *et al.* (2015) as regards the two most frequent types of phrases, that is, noun phrases and prepositional phrases, although the order is reversed in the two works. Noun phrases are the most typical combinations in UCOSCIENCOR (45.6%) with prepositional phrases in second place (29.8%), while prepositional phrases are the most typical sequences in COMRA (44.5%) followed by noun phrases (20.42%). In both studies verb phrase is in third place, and recurrence is similar in both works (21.1% in UCOSCIENCOR and 20.59% in COMRA).

Taking this into consideration, the results yielded in the current research are in consonance with the long-held assumption that nominal complexes have always been present in the language of science, and that complex noun phrases characterize the specialized language of science in English (Nagy 2014: 265). For example, Halliday (1993) performs a comprehensive analysis of works which support the idea that nominal structures have always been a distinctive feature of scientific texts (e.g., Chaucer 1391; Dalton 1827; Maxwell 1881). More recently, other authors have also focused on multiword units arranged around a noun, such as Luzón Marco (2000) and Pérez-Llantada (2014). What is more, noun phrases have also been found to be common in scientific texts written in Spanish (Soto Vergara & Zenteno Bustamante 2004), which is the mother tongue of the authors of the research articles included in UCOSCIENCOR, this being another plausible reason for the more varied repertoire of noun phrases compared to COMRA (i.e., the influence of the authors' first language).

On the other hand, Jalali *et al.* (2015) emphasize that prepositional phrases are used to identify a particular time period or location, although they do not specify why this is the case nor mention the most frequent prepositions. One explanation for the prevalence of prepositional phrases in COMRA may be the nature of the texts, since they may be more concerned with reporting these two aspects of the research conducted (i.e., time and location) or including time references to previous research (Swales 1990: 144, as cited in Gledhill 2000: 36). Additionally, Jalali *et al.* (2015) compare their work to previous studies and conclude that some prepositional phrases typical of their corpus were not extracted in other works, such as *in the present study* (this is in fact the most frequent 4-gram extracted from UCOSCIENCOR), *at the time of*, *of the present study*, *in the control group*, *in the current study*. Nevertheless, the results of both studies (i.e., Jalali *et al.* 2015 and the present research) are in line with Chen and Baker's (2010) findings that these two groups of phrases, nominal and prepositional, are frequently used in expert writing.

In second place, with respect to the functions of the 4-grams analyzed in this study, significant differences must be noted when they are compared to other works. As has been previously stated, research-oriented sequences are the most common (59.9%) in UCOSCIENCOR, followed by participant-oriented sequences (22.1%) and text-oriented sequences (18%). In contrast, text-oriented sequences are the most recurrent in the work of Salazar (2014), Jalali *et al.* (2015), Jalilifar *et al.* (2017), and Jalilifar and Ghoreishi (2018); to clarify, they account for 48%, 42.5%, 58.5%, and 58.51%, respectively, in the corpora analyzed by these authors. Jalali *et al.* (2015) justify their finding of the extensive use of text-oriented sequences by arguing that scholars employ them in order to show a sophisticated approach to language and demonstrate their mastery of it. Added to that, this type of sequence helps

to organize their discourse in the way that readers can better understand it, which is fully achieved thanks to bundles that make connections, present contrastive information, frame arguments, and so on.

However, the fact that research-oriented sequences are so prevalent in UCOSCIENCOR accords with Parvizi's findings in a study focusing on education (2011, as cited in Jalali *et al.* 2015: 65) that they outweigh all other functional types of bundles. Moreover, when it comes to the language of science, Jalilifar *et al.* (2017) suggest that this difference may also be due to the methods used in distinct types of science, that is, soft and hard, given that "writers in hard sciences [are required] to use less evaluative patterns of argument which are mostly realized by text-oriented bundles" (p. 11).

Finally, a common feature among all disciplines (i.e., science, applied linguistics, etc.) is that participant-oriented sequences have the lowest representation of all categories. It might be expected given that these bundles involve the reader or the writer of the texts, but in these fields the emotive or conative functions of language (Jakobson 1960) do not normally emerge, since the texts are referential and are more interested in transmitting accurate and precise information rather than focusing on the addresser or addressee of the message.

## 6. CONCLUSIONS AND FURTHER RESEARCH

The current study has delved into the use of 4-grams by academic researchers from the UCO in their scientific papers. After having compiled a specialized corpus, I automatically extracted a list of potential 4-grams and then manually discarded those which did not meet the eligibility criteria. In total, 181 items were selected and examined in order to achieve the two main objectives of this work.

First of all, and based on Biber *et al.*'s (1999: 96-106), I explored the structural frameworks of these sequences and discovered that noun phrases were the most typical structure (45.6%), followed by prepositional phrases (29.8%) and verb phrases (21.1%). The least common structures were adjective phrases (2.5%) and adverb phrases (1%). These results validate the widespread assumption that scientific language is characterized by complex nominals, albeit they are not in line with Jalali *et al.*'s (2015) findings, where prepositional phrases account for almost half (44.5%) of the multiword units analyzed. Nevertheless, it is important to realize that these two types of phrases are also said to be frequent in expert writing, which is the category where academic scholars may fit in.

After that, these sequences were classified according to the function they perform in the corpus, as also done by Hyland (2008: 49). I found that the majority of 4-grams (59.9%) were used to express information related to the research described in the scientific papers itself. Furthermore, these word combinations mostly concerned the authors of the articles, their publications, or their evaluations (22.1%), as well as the texts themselves (18%). At this point, it must be emphasized that these results do not conform to those obtained in Salazar (2014), Jalali *et al.* (2015), Jalilifar *et al.* (2017), or Jalilifar and Ghoreishi (2018), where text-oriented sequences were the most recurrent, which may be explained on the basis that scholars are likely to incorporate such bundles into their writing so as to make their discourse more elaborate.

As for the relation between these two aims, different structures were more frequently used depending on the type of function fulfilled. As such, noun phrases mainly represented

research-oriented 4-grams (58.8%), verb phrases were prominently identified with participant-oriented 4-grams (46.2%), and finally text-oriented 4-grams were typically described by prepositional phrases (54.4%). Without a doubt, the headword of these sequences contributed to the function served in the texts, as complex nominals are usually the most specialized multiword units (which explains their prevalence in the research-oriented category), while verbs are used to express evaluations and opinions (participant-oriented category), and prepositions frequently introduce phrases that are useful in organizing the text (text-oriented category).

My goals have therefore been successfully achieved and this paper has contributed to the description of multiword units (more specifically, 4-grams) used in the scientific research article genre. Moreover, I can confirm my hypothesis, that is, some of these phraseological units help to structure and organize the text (i.e., the text-oriented sequences) and acquire specific nuances of meaning in this particular context. To put it another way, despite knowing that they could also be employed in other domains, as some of the works mentioned above acknowledge (e.g., Pérez-Llantada 2014; Jalilifar & Ghoreishi 2018), it is important to realize that the research-oriented combinations analyzed, which were the most recurrent type, dealt with various specific aspects of the research presented in the corpus, such as detailed descriptions, procedures, and locations.

Future studies should focus on shorter and/or longer n-grams (e.g., 3-/5-grams) in the same corpus as well as in other languages, such as Spanish. Aside from that, these phraseological units may be explored separately in the different sections of the research article (e.g., abstract, methods, discussion) and the results compared afterwards. Furthermore, the same methodological steps could be employed to examine other specialized languages, such as academic writing, law, or the environment, and see whether the type and function of multiword structures coincide with those found here.

## REFERENCES

- Aimenova, M., A. Ospanova, A. Rakhimova, A. Sarsembayeva & Z. Mazhit. (2019). Phraseological terminology in the English economic discourse. *Xlinguae*, 12(1), 228-238. <https://doi.org/10.18355/XL.2019.12.01.18>
- Allen, D. (2010). Lexical bundles in learner writing: An analysis of formulaic language in the ALESS learner corpus. *Komaba Journal of English Education: KJEE*, 1, 105-127.
- Amosova, N. N. (1963). *Osnovy anglijskoj frazeologii*. Izdatelstvo Leningradskogo Universyteta.
- Biber, D. & F. Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, 26, 263-286. <https://doi.org/10.1016/j.esp.2006.08.003>
- Biber, D. & S. Conrad. (1999). Lexical bundles in conversation and academic prose. In H. Hasselgård & S. Oksefjell (Eds.), *Out of corpora: Studies in honour of Stig Johansson* (pp. 181-189). Rodopi.
- Biber, D., S. Johansson, G. Leech, S. Conrad & E. Finegan. (1999). *The Longman grammar of spoken and written English*. Longman.
- Biel, L. (2018). Lexical bundles in EU law: The impact of translation process on the patterning of legal language. In S. Goźdz-Roszkowski & G. Pontrandolfo (Eds.), *Phraseology in legal and institutional settings: A corpus-based interdisciplinary perspective* (pp. 11-26). Routledge. <https://doi.org/10.4324/9781315445724>
- Chaucer, G. (1391). *A treatise on the astrolabe*. Early English Text Society Edition.
- Chen, Y. & P. Baker. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology*, 14(2), 30-49.



- Corpas Pastor, G. & M. Seghiri Domínguez. (2010). Size matters: A quantitative approach to corpus representativeness. In R. Rabadán, M. Fernández López & T. Guzmán González (Eds.), *Lengua, traducción, recepción en honor de Julio César Santoyo* (pp. 111-145). Universidad de León, Área de Publicaciones.
- Cowie, A. P. (1981). The treatment of collocations and idioms in learners' dictionaries. *Applied Linguistics*, 2(3), 223-235. <https://doi.org/10.1093/applin/II.3.223>
- Cowie, A. P. (1998). Phraseological dictionaries: Some east-west comparisons. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis and applications* (pp. 209-228). Clarendon Press.
- Dalton, J. (1827). *A new system of chemical philosophy*. George Wilson.
- Fuster-Márquez, A. (2014). Lexical bundles and phrase frames in the language of hotel websites. *English Text Construction*, 7(1), 84-121. <https://doi.org/10.1075/etc.7.1.04fus>
- Gablasova, D., V. Brezina & T. McEnery. (2017). Collocations in corpus-based language learning research: Identifying, comparing, and interpreting the evidence. *Language Learning*, 67(S1), 155-179. <https://doi.org/10.1111/lang.12225>
- Gledhill, C. J. (2000). *Collocations in science writing*. Gunter Narr Verlag Tübingen.
- Gledhill, C. J. (2009). Colligation and the cohesive function of present and past tense in the scientific research article. In D. Banks (Ed.), *Les temps et les textes de spécialité* (pp. 65-84). L'Harmattan.
- Granger, S. (2014). A lexical bundle approach to comparing languages: Stems in English and French. *Languages in Contrast*, 14, 58-72. <https://doi.org/10.1075/bct.87.04gra>
- Granger, S. & M. Paquot. (2008). Disentangling the phraseological web. In S. Granger & F. Meunier (Eds.), *Phraseology. An interdisciplinary perspective* (pp. 27-49). John Benjamins Publishing Company. <https://doi.org/10.1075/z.139.07gra>
- Halliday, M. A. K. (1993). On the language of physical science. In M. A. K. Halliday & J. R. Martin (Eds.), *Writing science: Literacy and discursive power* (pp. 54-68). The Falmer Press.
- Halliday, M. A. K. (1994). *Functions of language. 2<sup>nd</sup> edition*. Arnold.
- Howarth, P. A. (1996). *Phraseology in English academic writing. Some implications for language learning and dictionary making*. Niemeyer. <https://doi.org/10.1515/9783110937923>
- Hyland, K. (2005). Stance and engagement: A model of interaction in academic discourse. *Discourse Studies*, 7(2), 173-191. <https://doi.org/10.1177/1461445605050365>
- Hyland, K. (2008). Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics*, 18(1), 41-62. <https://doi.org/10.1111/j.1473-4192.2008.00178.x>
- Jakobson, R. (1960). Linguistics and poetics. In T. A. Sebeok (Ed.), *Style in language* (pp. 350-377). The Massachusetts Institute of Technology.
- Jalali, Z. S., M. R. Moini & M. A. Arani. (2015). Structural and functional analysis of lexical bundles in medical research articles: A corpus-based study. *International Journal of Information Science and Management*, 13(1), 51-69.
- Jalilifar, A. & S. M. Ghoreishi. (2018). *From the perspective of*: Functional analysis of formulaic sequences in applied linguistics research articles. *International Journal of English Studies*, 18(2), 161-186. <https://doi.org/10.6018/ijes/2018/2/310351>
- Jalilifar, A., S. M. Ghoreishi & S. A. E. Roodband. (2017). Developing an inventory of core lexical bundles in English research articles: A cross-disciplinary corpus-based study. *Journal of World Languages*, 3(3), 182-203. <https://doi.org/10.1080/21698252.2017.1301279>
- Jiménez-Navarro, E. L. (2019). Nominal collocations in scientific English: A frame-semantic approach. In G. Corpas Pastor & R. Mitkov (Eds.), *Computational and corpus-based phraseology. Third International Conference, Europhras 2019. Malaga, Spain, September 25-27, 2019. Proceedings* (pp. 187-199). Springer. [https://doi.org/10.1007/978-3-030-30135-4\\_14](https://doi.org/10.1007/978-3-030-30135-4_14)
- Klappenbach, R. (1968). Probleme der Phraseologie. *Wissenschaftliche Zeitschrift der Karl-Marx-Universität Leipzig*, 17(2), 221-227.

- Kwary, D. A., D. Ratri & A. F. Artha. (2017). Lexical bundles in journal articles across academic disciplines. *Indonesian Journal of Applied Linguistics*, 7(1), 132-140. <https://doi.org/10.17509/ijal.v7i1.6866>
- Lipka, L. (1974). Probleme der Analyse englischer Idioms aus struktureller und generativer Sicht. *Linguistik und Didaktik*, 20, 274-285.
- Luzón Marco, M. J. (2000). Collocational frameworks in medical research papers: A genre-based study. *English for Specific Purposes*, 19, 63-86. [https://doi.org/10.1016/S0889-4906\(98\)00013-1](https://doi.org/10.1016/S0889-4906(98)00013-1)
- Maxwell, J. C. (1881). *An elementary treatise on electricity*. Clarendon.
- Mel'čuk, I. A. (2006). Explanatory combinatorial dictionary. In G. Sica (Ed.), *Open problems in linguistics and lexicography* (pp. 225-355). Polimetrica.
- Mel'čuk, I. A. (2012). Phraseology in the language, in the dictionary, and in the computer. *Yearbook of Phraseology*, 3(1), 31-56. <https://doi.org/10.1515/phras-2012-0003>
- Menon, S. & J. Mukundan. (2010). Analyzing collocational patterns of semi-technical words in science textbooks. *Pertanika Journal of Social Sciences & Humanities*, 18(2), 241-258.
- Nagy, I. K. (2014). English for special purposes: Specialized languages and problems of terminology. *Acta Universitatis Sapientiae, Philologica*, 6(2), 261-273. <https://doi.org/10.1515/ausp-2015-0018>
- Nesselhauf, N. (2005). *Collocations in a learner corpus*. John Benjamins Publishing Company. <https://doi.org/10.1075/scl.14>
- Parvizi, N. (2011). *Identification of discipline-specific lexical bundles in education*. Unpublished master's thesis, University of Kashan, Kashan, Iran.
- Pawley, A. & F. H. Syder. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication* (pp. 191-226). Longman.
- Pérez-Llantada, C. (2014). Formulaic language in L1 and L2 expert academic writing: Convergent and divergent usage. *Journal of English for Academic Purposes*, 14, 84-94. <https://doi.org/10.1016/j.jeap.2014.01.002>
- Pęzik, P. (2018). *Facets of prefabrication. Perspectives on modelling and detecting phraseological units*. Łódź University Press.
- Salazar, D. (2014). *Lexical bundles in native and non-native scientific writing: Applying a corpus-based study to language teaching*. John Benjamins Publishing Company.
- Shin, Y. K., V. Cortes & I. W. Yoo. (2018). Using lexical bundles as a tool to analyze definite article use in L2 academic writing: An exploratory study. *Journal of Second Language Writing*, 39, 29-41. <https://doi.org/10.1016/j.jslw.2017.09.004>
- Sinclair, J. (2000). Lexical grammar. *Naujoji Metodologija*, 24, 191-203.
- Sketch Engine. [Computer software]. Lexical Computing Limited.
- Soto Vergara, G. & C. Zenteno Bustamante. (2004). Los sintagmas nominales en textos científicos escritos en español. *ELUA (Estudios de Lingüística. Universidad de Alicante)*, 18, 275-292. <https://doi.org/10.14198/ELUA2004.18.14>
- Stefanowitsch, A. & S. T. Gries. (2003). Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics*, 8(2), 209-243. <https://doi.org/10.1075/ijcl.8.2.03ste>
- Stubbs, M. (2007). An example of frequent English phraseology: Distributions, structures and functions. In R. Facchinetti (Ed.), *Corpus linguistics 25 years on* (pp. 87-105). Rodopi. [https://doi.org/10.1163/9789401204347\\_007](https://doi.org/10.1163/9789401204347_007)
- Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge University Press.
- Thomson, H. (2017). Building speaking fluency with multiword expressions. *TESL Canada Journal*, 34(3), 26-53. <https://doi.org/10.18806/tesl.v34i3.1272>
- Vargas Sierra, C. (2006). Diseño de un corpus especializado con fines terminográficos: El corpus de la piedra natural. *Debate Terminológico*, 2(7), 1-20.

Vinogradov, V. V. (1947). *Izbrannye trudy. Leksikologija i leksikografija*. Nauka.

Weinreich, U. (1969). Problems in the analysis of idioms. In J. Puhvel (Ed.), *Substance and structure of language* (pp. 23-81). University of California Press. <https://doi.org/10.1525/9780520316218-003>

Wray, A. (2008). *Formulaic language: Pushing the boundaries*. Oxford University Press.

## APPENDIX 1

Rank	4-gram	Frequency	Frequency per million	Document frequency	Relative document frequency
1	<i>in the present study</i>	376	86.64	177	29.45
2	<i>the authors of this</i>	375	86.41	72	11.98
3	<i>of the authors of</i>	372	86.41	372	61.89
4	<i>of this publication are</i>	361	83.86	361	60.06
5	<i>some of the authors</i>	360	83.62	360	59.90
6	<i>the user has requested</i>	360	83.62	360	59.90
7	<i>page was uploaded by</i>	360	83.62	360	59.90
8	<i>and author profiles for</i>	360	83.62	360	59.90
9	<i>following this page was</i>	360	83.62	360	59.90
10	<i>user has requested enhancement</i>	360	83.62	360	59.90
11	<i>all content following this</i>	360	83.62	360	59.90
12	<i>this page was uploaded</i>	360	83.62	360	59.90
13	<i>content following this page</i>	360	83.62	360	59.90
14	<i>author profiles for this</i>	360	83.62	360	59.90
15	<i>authors of this publication</i>	360	83.62	360	59.90
16	<i>enhancement of the downloaded</i>	360	83.62	360	59.90
17	<i>profiles for this publication</i>	360	83.62	360	59.90
18	<i>on these related projects</i>	360	83.62	360	59.90
19	<i>of the downloaded file</i>	360	83.62	360	59.90
20	<i>for this publication at</i>	360	83.62	360	59.90
21	<i>working on these related</i>	360	83.62	360	59.90
22	<i>are also working on</i>	360	83.62	360	59.90
23	<i>has requested enhancement of</i>	360	83.62	360	59.90
24	<i>requested enhancement of the</i>	360	83.62	360	59.90
25	<i>also working on these</i>	360	83.62	360	59.90
26	<i>at the end of</i>	318	73.27	120	19.96
27	<i>on the other hand</i>	312	71.89	175	29.10
28	<i>the end of the</i>	239	63.41	120	19.96
29	<i>patients with ankylosing spondylitis</i>	201	46.31	34	5.65
30	<i>in the presence of</i>	200	46.08	89	14.80
31	<i>are shown in table</i>	194	45.06	142	23.61
32	<i>in the case of</i>	183	42.16	96	15.97
33	<i>as well as the</i>	170	39.48	127	21.13
34	<i>in the absence of</i>	155	35.71	94	15.63

Rank	4-gram	Frequency	Frequency per million	Document frequency	Relative document frequency
35	<i>the results of the</i>	151	35.07	92	15.30
36	<i>of the present study</i>	151	35.07	101	16.80
37	<i>an important role in</i>	149	34.33	100	16.63
38	<i>risk factors associated with</i>	147	33.87	52	8.65
39	<i>on the basis of</i>	147	33.87	93	15.47
40	<i>for the treatment of</i>	145	33.41	59	9.81
41	<i>of this study was</i>	143	32.95	122	20.29
42	<i>in the development of</i>	141	32.49	93	15.47
43	<i>in the treatment of</i>	137	31.82	54	8.98
44	<i>has been shown to</i>	136	31.59	101	16.80
45	<i>in patients with ankylosing</i>	134	31.12	31	5.15
46	<i>this study was to</i>	133	30.89	114	18.96
47	<i>this work was supported</i>	129	29.72	128	21.28
48	<i>an increase in the</i>	126	29.26	79	13.14
49	<i>work was supported by</i>	124	28.80	123	20.46
50	<i>the effect of the</i>	123	28.56	87	14.47
51	<i>was approved by the</i>	122	28.33	109	18.13
52	<i>the aim of this</i>	121	28.10	108	17.97
53	<i>ankylosing spondylitis disease activity</i>	119	27.63	27	4.49
54	<i>for the management of</i>	118	27.41	48	7.98
55	<i>as a result of</i>	117	27.17	78	12.97
56	<i>with respect to the</i>	117	27.17	71	11.81
57	<i>the presence of the</i>	113	26.24	65	10.81
58	<i>at the time of</i>	113	26.24	69	11.48
59	<i>the present study was</i>	112	26.01	89	14.80
60	<i>studies have shown that</i>	112	26.01	90	14.96
61	<i>were included in the</i>	112	26.01	86	14.30
62	<i>included in the study</i>	111	25.78	73	12.14
63	<i>the total number of</i>	105	24.38	70	11.64
64	<i>has been associated with</i>	104	24.15	66	10.98
65	<i>one of the most</i>	103	23.91	87	14.47
66	<i>aim of this study</i>	95	22.05	85	14.14
67	<i>is one of the</i>	95	22.05	79	13.14
68	<i>this is the first</i>	94	21.83	68	11.31
69	<i>area under the curve</i>	93	21.60	48	7.98
70	<i>there were no significant</i>	93	21.60	56	9.31
71	<i>for personal use only</i>	92	21.36	10	1.66
72	<i>patients with metastatic colorectal</i>	91	21.13	23	3.82
73	<i>with metastatic colorectal cancer</i>	91	21.13	24	3.99
74	<i>it is important to</i>	91	21.13	61	10.14
75	<i>these results suggest that</i>	90	20.90	59	9.81
76	<i>a systematic review and</i>	89	20.67	26	4.32

Rank	4-gram	Frequency	Frequency per million	Document frequency	Relative document frequency
77	<i>systematic review and meta-analysis</i>	89	20.67	26	4.32
78	<i>in accordance with the</i>	89	20.67	74	12.29
79	<i>was associated with a</i>	87	20.20	53	8.81
80	<i>be due to the</i>	87	20.20	68	11.31
81	<i>a significant increase in</i>	85	19.74	50	8.31
82	<i>bath ankylosing spondylitis disease</i>	85	19.74	25	4.15
83	<i>spondylitis disease activity index</i>	83	19.28	25	4.15
84	<i>the bath ankylosing spondylitis</i>	83	19.28	27	4.48
85	<i>in patients with chronic</i>	83	19.28	34	5.65
86	<i>peripheral blood mononuclear cells</i>	82	19.04	36	5.99
87	<i>in the pathogenesis of</i>	82	19.04	43	7.15
88	<i>were no significant differences</i>	82	19.04	52	8.65
89	<i>been shown to be</i>	81	18.81	64	10.64
90	<i>the aim of the</i>	80	18.57	70	11.63
91	<i>no significant differences were</i>	80	18.57	59	9.81
92	<i>of coronary heart disease</i>	80	18.57	43	7.15
93	<i>it has been shown</i>	79	18.35	59	9.81
94	<i>was carried out in</i>	79	18.35	58	9.65
95	<i>play an important role</i>	79	18.35	65	10.81
96	<i>in the number of</i>	78	18.11	60	9.98
97	<i>it has been suggested</i>	78	18.11	52	8.65
98	<i>study was carried out</i>	77	17.88	69	11.48
99	<i>anova for repeated measures</i>	76	17.65	32	5.32
100	<i>of the metabolic syndrome</i>	76	17.65	26	4.32
101	<i>after consumption of the</i>	76	17.65	18	2.99
102	<i>distributed under the terms</i>	76	17.65	75	12.47
103	<i>no significant differences in</i>	75	17.42	51	8.48
104	<i>has been suggested that</i>	75	17.42	51	8.48
105	<i>an increased risk of</i>	74	17.18	32	5.32
106	<i>and reproduction in any</i>	74	17.18	72	11.98
107	<i>important role in the</i>	74	17.18	64	10.64
108	<i>and the risk of</i>	74	17.18	39	6.48
109	<i>the end of each</i>	74	17.18	37	6.15
110	<i>to the presence of</i>	74	17.18	51	8.48
111	<i>reproduction in any medium</i>	73	16.95	72	11.98
112	<i>as well as in</i>	73	16.95	64	10.64
113	<i>test was used to</i>	73	16.95	65	10.81
114	<i>is associated with a</i>	73	16.95	52	8.65
115	<i>to be associated with</i>	73	16.95	48	7.98
116	<i>efficacy and safety of</i>	72	16.72	35	5.81
117	<i>a wide range of</i>	71	16.49	61	10.14
118	<i>microbiology and infectious diseases</i>	71	16.49	10	1.66

Rank	4-gram	Frequency	Frequency per million	Document frequency	Relative document frequency
119	<i>in advanced colorectal cancer</i>	70	16.26	18	2.99
120	<i>significant differences were found</i>	70	16.26	55	9.15
121	<i>to the development of</i>	70	16.26	48	7.98
122	<i>has been shown that</i>	70	16.26	52	8.65
123	<i>be related to the</i>	70	16.26	59	9.81
124	<i>it has been reported</i>	70	16.26	52	8.65
125	<i>the fact that the</i>	69	16.02	56	9.31
126	<i>the rest of the</i>	69	16.02	48	7.98
127	<i>the terms of the</i>	69	16.02	67	11.14
128	<i>with the use of</i>	69	16.02	37	6.15
129	<i>plays an important role</i>	68	15.79	48	7.98
130	<i>under the terms of</i>	68	15.79	67	11.14
131	<i>aim of the present</i>	68	15.79	59	9.81
132	<i>could be due to</i>	68	15.79	55	9.15
133	<i>patients with rheumatoid arthritis</i>	67	15.56	32	5.32
134	<i>and the presence of</i>	67	15.56	54	8.98
135	<i>for the first time</i>	67	15.56	55	9.15
136	<i>informed consent was obtained</i>	67	15.56	65	10.81
137	<i>to the fact that</i>	67	15.56	57	9.48
138	<i>mediterranean diet supplemented with</i>	66	15.33	16	2.66
139	<i>a large number of</i>	66	15.33	49	8.15
140	<i>of the mediterranean diet</i>	66	15.33	25	4.15
141	<i>of the university of</i>	66	15.33	46	7.65
142	<i>would like to thank</i>	66	15.33	63	10.48
143	<i>it is possible that</i>	65	15.09	52	8.65
144	<i>renal excretion of urates</i>	65	15.09	7	1.16
145	<i>in the regulation of</i>	65	15.09	50	8.31
146	<i>be explained by the</i>	65	15.09	56	9.31
147	<i>was used for the</i>	65	15.09	61	10.14
148	<i>the presence of a</i>	64	14.86	51	8.48
149	<i>of virgin olive oil</i>	64	14.86	18	2.99
150	<i>consent was obtained from</i>	64	14.86	62	10.31
151	<i>by the presence of</i>	64	14.86	52	8.65
152	<i>as a consequence of</i>	64	14.86	51	8.48
153	<i>has been reported that</i>	63	14.63	46	7.65
154	<i>the start of the</i>	63	14.63	37	6.15
155	<i>in chronic kidney disease</i>	63	14.63	24	3.99
156	<i>in any of the</i>	63	14.63	44	7.32
157	<i>has been reported in</i>	63	14.63	52	8.65
158	<i>the influence of the</i>	63	14.63	53	8.81
159	<i>in addition to the</i>	62	14.40	54	8.98
160	<i>for the diagnosis of</i>	62	14.40	34	5.65

Rank	4-gram	Frequency	Frequency per million	Document frequency	Relative document frequency
161	<i>received in revised form</i>	62	14.40	62	10.31
162	<i>statistically significant differences were</i>	62	14.40	42	6.98
163	<i>in relation to the</i>	62	14.40	45	7.48
164	<i>on behalf of the</i>	62	14.40	54	8.98
165	<i>in the expression of</i>	62	14.40	30	4.99
166	<i>used in this study</i>	62	14.40	52	8.65
167	<i>risk of cardiovascular disease</i>	61	14.16	31	5.15
168	<i>bath ankylosing spondylitis functional</i>	61	14.16	28	4.65
169	<i>ankylosing spondylitis functional index</i>	61	14.16	28	4.65
170	<i>were observed in the</i>	61	14.16	49	8.15
171	<i>at the same time</i>	61	14.16	45	7.48
172	<i>may be due to</i>	61	14.16	54	8.98
173	<i>was significantly higher in</i>	60	13.93	43	7.15
174	<i>that the presence of</i>	60	13.93	39	6.48
175	<i>study was to evaluate</i>	60	13.93	51	8.48
176	<i>seroprevalence of toxoplasma gondii</i>	60	13.93	16	2.66
177	<i>the area under the</i>	60	13.93	38	6.32
178	<i>and the number of</i>	60	13.93	49	8.15
179	<i>was supported by the</i>	60	13.93	58	9.65
180	<i>not included in the</i>	60	13.93	55	9.15
181	<i>was found to be</i>	60	13.93	38	6.32

## Verbs of cutting in the middle construction: Examining the compositional cospecification of patientive vs. Agent-instrument middles

Verbos de corte en la construcción media: examinando la co-especificación  
composicional de las medias pacientes y agentivo-instrumentales

MACARENA PALMA GUTIÉRREZ  
Universidad de Córdoba, España  
l82pagum@uco.es  
<https://orcid.org/0000-0002-0558-9358>

### Abstract

The purpose of this paper is to examine both prototypical Ergative-like middles (e.g. *The cake cuts well*) and their metonymically-motivated extensions, called Agent-Instrument middles (e.g. *The saw cuts like a dream*). The former incorporate as subjects patientive entities and the latter instruments. Both structures allow the incorporation of verbs of cutting, but they represent divergent portions of the action chain, and thus, they differ in the nature of their subject entities and their processes of compositional cospecification in terms of *qualia* structure. On the whole, on the basis of the prototype effects of the middle construction analysed here, they should be considered proper middles. This paper is based on a corpus study of contextualised examples (1700+) to examine their collostructional schema in the English middle construction (cf. Stefanowitsch and Gries 2003). Particularly, we explore the

### Resumen

El propósito de este trabajo es examinar tanto las medias prototípicas de tipo ergativo (e.g.: *El pastel se corta bien*) como sus extensiones de motivación metonímica denominadas medias agentivo-instrumentales (e.g.: *La sierra corta de ensueño*). Las primeras incorporan entidades pacientes como sujeto, y las segundas, instrumentos. Ambas estructuras permiten la incorporación de verbos de ‘cortar’, pero representan porciones divergentes de la cadena de acción y, por tanto, difieren en la naturaleza de sus entidades sujeto y en sus procesos de cospecificación composicional en términos de estructura de *qualia*. En general, ambas deben considerarse medias propias en virtud de los efectos prototípicos de la construcción media que aquí se analiza. Este trabajo se basa en un estudio de corpus de ejemplos contextualizados (más de 1700) para examinar el esquema colostruccional en la construcción media inglesa (cf. Stefanowitsch y Gries 2003). En particular,

**To cite this article:** Palma Gutiérrez, M. (2022). Verbs of cutting in the middle construction: Examining the compositional cospecification of patientive vs. agent-instrument middles. *ELUA*, (38), 263-279. <https://doi.org/10.14198/ELUA.21752>

Recibido: 21/01/2022 Aceptado: 11/05/2022

© 2022 Macarena Palma Gutiérrez



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



Verb + Adverb collocation found in middles by focusing on 29 predicates belonging to Levin's (1993: 156) class of verbs of cutting, in combination with facility-/quality-oriented adjuncts (cf. Davidse and Heyvaert, 2007). Following a cognitive perspective, this paper examines the semantic roles of the subject referents analysed (patients and instruments), as well as their distinct processes of compositional cospecification as factors that contribute to the prototype effects of the middle construction. On the basis of corpus data, despite the syntactic, semantic, and pragmatic common schemas found in prototypical Ergative-like middles and Agent-Instrument middles, the prototype effects of the middle construction highlight the divergent processes of compositional cospecification that these two types of middles follow. More specifically, a shift of semantic weight from a telic to a constitutive value occurs in Agent-Instrument middles, while there is lack of such a semantic shift in terms of *qualia* structure in prototypical Ergative-like middles with patientive subjects. Thus the middle construction cannot be considered as a discrete category of its own, but rather a prototype category (cf. Taylor, 1995; Hundt, 2007) that permits the accommodation of central and peripheral members. Consequently, in addition to the traditionally accepted structures in which the subject referent is a patientive entity, other less archetypal nominals can occur as middle subjects, particularly, instruments whose action affects implied and patientive entities.

**KEYWORDS:** middle construction; compositional cospecification; Ergative-like middle; Agent-Instrument middle; prototype effects; verbs of cutting

exploramos la colocación Verbo + Adverbio que se encuentra en las medias, centrándonos en 29 predicados pertenecientes a la clase de verbos de 'cortar' de Levin (1993: 156), en combinación con adjuntos orientados a la facilidad/calidad (cf. Davidse y Heyvaert 2007). Siguiendo una perspectiva cognitiva, este trabajo examina los roles semánticos de los referentes temáticos analizados (pacientes e instrumentos), así como sus distintos procesos de cospecificación composicional como factores que contribuyen a los efectos prototípicos de la construcción media. Sobre la base de los datos del corpus, a pesar de los esquemas sintácticos, semánticos y pragmáticos comunes que se encuentran en las medias prototípicas de tipo ergativo y en las medias agentivo-instrumentales, los efectos prototípicos de la construcción media ponen de manifiesto los procesos divergentes de cospecificación composicional que siguen estos dos tipos de estructuras. Por lo tanto, se produce un cambio de peso semántico de un valor télico a uno constitutivo en las medias agentivo-instrumentales, mientras que no hay tal cambio semántico en términos de estructura de *qualia* en las medias prototípicas de tipo ergativo con sujetos pacientes. Así, la construcción del medio no puede considerarse como una categoría discreta propia, sino como una categoría prototípica (cf. Taylor 1995; Hundt 2007), que permite la acomodación de miembros centrales y periféricos. En consecuencia, además de las estructuras tradicionalmente aceptadas en las que el sujeto referente es una entidad paciente, pueden darse otros nominales menos arquetípicos como sujetos medios, en particular, instrumentos cuya acción afecta a entidades implícitas y pacientes.

**PALABRAS CLAVE:** construcción media; cospecificación composicional; media de tipo ergativo; media agentivo-instrumental; efectos prototípicos; verbos de 'cortar'

## 1. INTRODUCTION

Middle structures have been largely examined in the traditional literature (Van Oosten, 1977 & 1984; Keyser & Roeper 1984; Fellbaum 1986; Fagan 1992; Levin 1993; Kemmer 1993; Ackema & Schoorlemmer 1994; among many others), always with the aim of trying to narrow down the essence of middlehood by restricting the conditions of well-formedness and grammaticality of this construction. Middability has generally been associated with the following

features: (i) transitive verbs in their one-argument intransitive use (*cf.* Fagan 1992; Levin 1993); (ii) non-agentive and affected participants occurring in Subject position, thus, fulfilling the role of Patient (*cf.* Van Oosten 1977 and 1986; Keyser and Roeper, 1984; Fagan, 1992); (iii) implicit agentive participants with an arbitrary and generic nature (*cf.* Ackema & Schoorlemmer 1994; Fagan 1992; Keyser & Roeper 1984); (iv) the need to incorporate Subject-oriented adjuncts whose Experiencer arguments are necessarily identified with the implied Agent (*cf.* Dixon 1982; Hoekstra & Roberts 1993); (v) non-eventive situations which lack a specific time reference (*cf.* Ackema & Schoorlemmer 1994); and (vi) Subject referents with certain letting/facilitating (or hindering) properties leading to the conduciveness of the event denoted by the predicate (*cf.* Fawcett 1980; Kemmer 1993; Davidse & Heyvaert 2007). In accordance with these core features, a traditionally accepted middle structure is illustrated in (1) below:

- (1) *Crystal vases break easily.* (Marín Arrese, 2011: 37).

In this example, the Subject referent ‘crystal vases’ has a patientive nature since it is affected by the ‘breaking’ event. The implied Agent (either a person acting deliberately or any spontaneous event such as a gust of wind or a ball hitting the vases in question by accident) is syntactically absent, but is semantically and/or contextually recoverable. The vases in question are characterised by being prone to breaking ‘easily’. In fact, it is the speaker who evaluates the breaking event easily occurring, basing their evaluation on the inherent properties of the Subject referent. Therefore, the natural disposition or breakability of the vases in question is seen by the speaker as being conducive to the action denoted by the predicate in the way indicated by the adjunct.

Formalist/projectionist verb-centred approaches to the process of middle formation (Levin 1993; Hale & Keyser 2002; Pinker 1989, and others) merely focus on structural information in order to restrict the conditions for acceptability, but such a rigid view does not always capture the essence of middlehood. This is so because these approaches advocate for the idea that the syntactic behaviour of verbs is determined by their meaning, and they therefore assume that verbs with common semantic features need to participate in the same syntactic alternations. However, these verb-centred approaches to the process of middle formation do not successfully take into account the process of lexical-constructional interaction, and thus they fail to recognise that “the constructions in which verbs occur are meaningful in and by themselves” (Lemmens 1998: 38). In fact, in line with Goldberg’s (1995, 2006, 2019) constructionist model, a verb and its arguments are necessarily determined by the construction itself, because it is the construction (understood as a conventional pairing of form and meaning) which determines the combinatory possibilities of a verb and its arguments.

A cognitive/constructionist perspective on the process of middle formation might, therefore, shed light on this issue, particularly, through applying the notions of ‘family-resemblance’ (*cf.* Wittgenstein 1985) and ‘prototype effects’ (*cf.* Taylor 1995; Lakoff, 1987; Langacker 1987, 2008; Goldberg 1995, 2006 and 2019). Thus, in considering the middle construction as a radial network or category (rather than a discrete category of its own), in this paper we can examine it as enabling the accommodation of central and peripheral members (see also Hundt 2007 and Palma Gutiérrez 2022).<sup>1</sup>

<sup>1</sup> Along the lines of Hundt’s (2007: 63) ideas, and also by extending the corpus sample from previous research projects (*cf.* Palma Gutiérrez, 2022), this paper argues that instrumental middles involve the occurrence of subject

In this paper, we focus on the nature of verbal predicates and the notion of ‘affectedness’ in non-agentive Subjects as a restricting factor for middlehood as proposed by the traditional literature (Van Oosten 1977, 1986; Keyser & Roeper 1984; Fagan 1992), and we provide a reanalysis of it. To do so, we compare two types of middle structures incorporating Levin’s (1993) class of verbs of cutting and we examine them in terms of *qualia* structure (cf. Pustejovsky 1991, 1995) and their processes of compositional cospecification (cf. Yoshimura 1998; Yoshimura & Taylor 2004). These two types of structures are the so-called prototypical Ergative-like middles (like *This cake cuts easily*) and the metonymically-motivated extensions of the prototype, known here as Agent-Instrument middles (like *The saw cuts like a dream*).

The theory of *qualia* structure postulates the idea that *qualia* roles are idiosyncratic, lexical-semantic features that “structure our basic knowledge” about an entity (Pustejovsky 1991: 427). The *qualia* structure of an entity lets us conceptualise it by means of its cognitively-activated inherent properties, namely, its *formal* features (Qf), its *constitutive* features, or internal parts, (Qc), its *agentive* features, which depend on the factors involved in its origin or bringing about (Qa), and its *telic* features, or purpose and function (Qt).<sup>2</sup> The *qualia* structure, in this sense, refers to semantic constraints “based on the idea that there is a system of relations that characterises the semantics of nominals”; and notably, “serves to specify the reading of a verb” (Yoshimura 1998: 115). This is fundamental in cospecification phenomena, since “middles will be licensed only insofar as there is a proper matching between the verb’s meaning and one of the *qualia* of its subject” (Cortés-Rodríguez & Mairal Usón 2013: 234).

The process of cospecification implies that “semantic information of the complement (of a verb) contributes to the specification of a unique and appropriate meaning of the verb” (Yoshimura 1998: 114). This idea is based on the fact that “just as a verb can select for an argument-type, an argument itself is able to select the predicate that governs it” (Yoshimura 1998: 114). In compositional analysis, the *qualia* structure of middles might lead to a shift in semantic importance when the value of the adjunct is added to that of the match between the meaning of the verb and the semantics of the Subject (cf. Yoshimura, 1998; Yoshimura & Taylor 2004).

In order to clarify this notion of compositional cospecification, consider a middle structure like *This book reads easily*.<sup>3</sup> Any arbitrary person can read a book due to their own skills (*i.e.*, literacy). Nevertheless, if the speaker specifies that the reading event was ‘easily’ performed, the abilities of the Agent (the reader) are being backgrounded and the responsibility of the inherent properties of the book in question (like its clarity and liveliness of

---

entities fulfilling the semantic role of Instrument, in combination with omitted patientive entities in Object position. The latter, despite their lack of syntactic projection, rely on an absolute interpretation at the semantic level of analysis, exhibiting more agent-like features associated with the Instrumental role of subject entities than prototypical middles (which characterised by a higher level of affectedness). These ideas are further elaborated in this section, as well as in Section 2.

2 According to Yoshimura (1998: 119-120) and Yoshimura and Taylor (2004: 308), telic and constitutive *qualia* modes (Qt and Qc) are the most relevant types regarding the process of compositional cospecification in middles. These two types of *qualia* roles are the ones explored in this paper, although formal and agentive *qualia* (Qf and Qa) are also found in middles (cf. Cortés-Rodríguez & Mairal Usón 2013: 234).

3 This is an Action-oriented middle, as explained in section 3. In this paper, this type of middle is not examined but it is mentioned here as an illustrative instance to convey the notion of ‘compositional cospecification’.

style) are being foregrounded and assessed as being conducive to the action denoted by the predicate in the way specified by the adjunct. In terms of *qualia* structure and compositional cospecification, then, we can affirm the following: the meaning of the nominal ‘book’ is cospecified by virtue of the semantics of the predicate ‘read’, reflecting a telic value (Qt). That is, books are meant to be read (by an implied Agent, a reader), as this is their purpose or function. However, when the semantic value of the adjunct ‘easily’, in this case, is added in compositional analysis, we understand that the book in question can be read ‘easily’ because it possesses certain inherent or contingent properties that enable any implied reader (independently of their literary abilities) to carry out the reading event with ease. Therefore, in the process of compositional cospecification of the middle structure ‘*This book reads easily*’, a shift in semantic importance from telic to constitutive *qualia* occurs (from Qt to Qc).

In this paper we explore the extent to which the process of compositional cospecification is different in the two types of middle structures examined here. Usage-based data is used to support the idea that a shift of semantic weight from a telic to a constitutive value occurs in Agent-Instrument middles (QtàQc), whereas there is a lack of any such shift in prototypical Ergative-like middles, since the constitutive value is maintained throughout the process of compositional cospecification (QcàQc).

This paper is organised as follows: section 2 focuses on the distinction between prototypical Ergative-like middles and extensions of the prototype called Agent-Instrument middles, based on the prototype effects displayed in the construction; section 3 shows the usage-based methodology employed in this work; section 4 presents the data and results obtained; and section 5 offers some final remarks.

## 2. PROTOTYPE EFFECTS IN THE MIDDLE CONSTRUCTION: PROTOTYPICAL ERGATIVE-LIKE MIDDLES AND AGENT-INSTRUMENT MIDDLES

According to the traditional approaches to middle formation, it is generally assumed that the Subject entities found in middle constructions fulfill a patientive role, which is provided by their affectedness or change of state and their non-agentive value (cf. Van Oosten 1977, 1986; Keyser & Roeper 1984; Fagan 1992). However, this phenomenon, called ‘Affectedness constraint on middle formation’ fails to account for canonical examples that are frequently cited in the literature as proper middles and which contain other than a Patient entity as Subject. Consider the following instances in this regard:

- (2) *I am more than satisfied with this card stock. The paper cuts easily and cleanly, scores and folds well too. I am using this in conjunction with the Taupe brown card stock for wedding invitations.* (enTenTen13 corpus, Concordance section, Sketch Engine)
- (3) *The installation looks better than if the factory had done it and the car drives like a dream. It will break the tires loose easily in third gear.* (enTenTen13 corpus, Concordance section, Sketch Engine)

In example (2) above, the ‘paper’ in question is patientive as it suffers from a change of state due to the cutting action in the way indicated by the adjunct (*i.e.*, ‘easily’). However, no change of state occurs in the case of the ‘car’ in (3) above. That is, the ‘car’ in question is not affected at all by the driving activity. These two instances have something in common,

though: the profiling of the non-agentive Subject and the defocusing of the Agent. In both cases, the speakers assess the inherent properties of the Subject referents as being conducive to the actions denoted by the predicates in the way indicated by their adverbs, independently of the skills of any implied agentive participant. Therefore, the cutting of the ‘paper’ in (2) does not depend on the abilities of any implied Agent, but rather, it relies on the internal constitution of the ‘paper’ in question: the materials that compose it and other internal features like its flexibility, rigidity, etc. Similarly, the fact that the driving event in (3) is easily carried out does not depend on the driving abilities of any implied driver, but rather, it has to do with the inherent properties of the ‘car’ in question: the car body, the braking system, the steering wheel, the tyres, etc.

Apart from this pragmatic commonality between examples (2) and (3) by which there is a process of non-agentive Subject-profiling and Agent-defocusing, these two examples also share syntactic and semantic underlying schemas. The syntactic schema that both examples undergo is that of intransitive structures: [N – V – (Adv)]. In terms of semantic analysis, both examples follow the pattern [X (by virtue of some property P) IS SUBJECTIVELY ASSESSED BY THE SPEAKER AS BEING CONDUCTIVE TO ACT].

Based on the theory of prototypes (*cf.* Taylor, 1995) and the contrast between middles with affected participants (as in example (2) above) and middles with non-affected entities in Subject position (as in example (3) above), Sakamoto (2001) proposes the distinction between Ergative-like middles (like (2)) and Action-oriented middles (like (3)).<sup>4</sup> As the author explains it, Ergative-like middles incorporate ergative verbs (like ‘open’, ‘close’, ‘break’, and ‘cut’) since they “specify how the change of state proceeds”; whereas Action-oriented middles are constructed with unergative verbs (like ‘read’, ‘translate’, ‘drive’, and ‘handle’), since they “specify the manner of action” (Sakamoto, 2001: 101).

As detailed in Palma Gutiérrez (2021b), prototypical Ergative-like middles and Action-oriented middles can also be distinguished by the divergent processes of compositional cospecification that the two structures undergo. In terms of *qualia* structure, the semantic relation that specifies the meaning of the predicate ‘drive’ in example (3) above by virtue of the semantics of the entity ‘car’ has a telic nature (*i.e.*, Qt). In other words, the final purpose or function of a ‘car’ is to be ‘driven’. However, when the semantic charge of the adverb ‘like a dream’ is considered, the inherent properties of the ‘car’ (its Qc) are foregrounded as being responsible for the conduciveness of the driving activity, backgrounding the role of the implicit Agent, the driver. Thus, a shift in semantic importance from a telic to a constitutive value occurs in compositional analysis in the case of Action-oriented middles. This phenomenon is captured in the formula [QtàQc].

On the other hand, in the process of compositional cospecification in prototypical Ergative-like middles, there is a lack of an initial telic relation between the nominal entity occurring as Subject and the verb of the middle construction. In the case of (2) above, for example, there is no telic relation between ‘paper’ and ‘cut’. The fact that the ‘paper’ in question cuts at all relies on its inherent properties (its Qc), and the addition of the adjunct ‘easily’ does not suppose a shift of semantic importance. In fact, the adverb intensifies the responsibility of the inherent properties of the ‘paper’ in question (its Qc) as letting the

4 In this paper, the term ‘Ergative-like middle’ is used instead of the actual term found in Sakamoto’s (2001) work, ‘unaccusative-based middle’. See also Palma Gutiérrez (2021b).

cutting action occur in an easy way. Thus, prototypical Ergative-like middles follow the formula [QcàQc] in compositional analysis.

Once argued that there are middle structures that do not follow the traditional restriction of affectedness (as occurs in Action-oriented middles), it is possible to think of the middle construction as a prototype category, rather than as a discrete category. That is, the middle construction can be understood as a radial category or network in which some exemplars are more prototypical and others are more marginal or peripheral, but they have in common the fact that they instantiate the same syntactic, semantic, and pragmatic underlying schemas. In this way, the middle prototype category can be analysed as a family of related but diverse structures. Besides, other less prototypical middles can be accommodated into this family of structures.

Following Sakamoto (2001) as well as the theory of prototypes (*cf.* Taylor 1995; Lakoff 1987; Langacker 1987, 2008; Goldberg 1995, 2006), in this paper we argue that a restricted set of middles, referred to here as Agent-Instrument middles, are identified as a hybrid combination of certain features of prototypical Ergative-like middles and Action-oriented middles. Consider example (4) below in this regard:

- (4) *A lot of people call this a sharpening steel. It doesn't actually sharpen your blade. Its purpose is to hone the blade. When everything is lined up straight, your knife cuts smoothly. When it is not, it feels like there is something dragging while making your cut.* (enTenTen13 corpus, Concordance section, Sketch Engine)

According to the core aspects traditionally accepted as middle-forming, example (4) displays most of the fundamental features of middles: (i) the predicate 'cut' occurs in its one-argument intransitive use; (ii) the Agent is implicit and has an arbitrary and generic nature; (iii) the adverb 'smoothly' is Subject-oriented, as its Experiencer argument is necessarily identified with the implied Agent; (iv) a non-eventive situation lacking a specific time reference is denoted; and (v) the Subject referent 'knife' is subjectively assessed by the speaker as being conducive to the action denoted by the predicate due to its letting/facilitating properties.

However, there is one feature generally associated with middles which is not present in example (4): the affectedness of the Patientive Subject referent. In this case, the 'knife' in question does not suffer from any change of state due to the cutting event. Instead, the 'knife' has an Instrumental role in semantic analysis, as it is seen as an artifact used by an implicit Agent to perform the cutting activity, and thus, as affecting an implied Patient (for instance, vegetables). Therefore, a different portion of the action chain is described in middles like (4) in contrast to prototypical Ergative-like middles like (2). Consider the following instances with regards to this contrast:

- (5) *Use a type of wood that does not require the use of power tools. Basswood carves easily and is lightweight, but it should be painted or stained because it's rather plain.* (enTenTen13 corpus, Concordance section, Sketch Engine)
- (6) *Certain tapping processes require that the Sawing Molybdenum be heated to above 325 degrees Fahrenheit, however it should not be heated above 500 degrees unless in a protective, non-oxidizing atmosphere. The Sawing Molybdenum saws easily with power bend saws and hacksaws.* (enTenTen13 corpus, Concordance section, Sketch Engine)

Example (5) instantiates a prototypical Ergative-like middle containing a Patientive Subject ('basswood'), a Subject-oriented adverb whose Experiencer participant coincides with the implied Agent ('easily'), and a transitive verb used in its one-argument intransitive form ('carve'), the predicate 'carve' belonging to Levin's (1993) class of verbs of cutting, within her set of 'carve verbs'. In addition, example (5) involves an implied Agent who carries out the carving event, producing a change of state on the 'basswood' with ease.

In contrast, instance (6) is an example of an Agent-Instrument middle, as the Subject entity (the 'Sawing Molybdenum') is not affected at all by the sawing activity. In fact, it is the Instrument used by the implied Agent to perform the action denoted by the predicate in the way indicated by the adjunct ('easily') in order to affect an implied Patient ('metal'). The predicate 'saw' in (6) does also belong to Levin's (1993) class of verbs of cutting, specifically, to her set of 'cut verbs'. Example (6), therefore, focuses on a different portion of the action chain than do middles like (5) above. This can be illustrated in the following instance:

- (7) *This type of metal saws easily with the Sawing Molybdenum.* (enTenTen13 corpus, Concordance section, Sketch Engine)

Example (7) is a paraphrase of (6) that results from the altering of the order of participants in the action chain being described. On this occasion, the Subject entity, the 'type of metal' in question, is Patientive, that is, it is affected by the sawing activity and its change of state occurs 'easily' due to its inherent properties. The implied Agent carries out the sawing event by using the 'Sawing Molybdenum'. Thus, example (7) constitutes an instance of a prototypical Ergative-like middle. However, when the argument focus is moved from the Patient to the Subject position (in this case, the Instrument used by the Agent), another portion of the action chain emerges, resulting into a less prototypical type of middle: an Agent-Instrument middle, like in (6) above.

Regarding the type of predicates allowed in Agent-Instrument middles, only a restrictive set of Ergative-like verbs (*i.e.*, those focused on "how the change of state proceeds" (Sakamoto, 2001: 101)) can occur. This is the case of Levin's (1993) class of 'verbs of cutting' (including 'carve verbs' and 'cut verbs'). In contrast, prototypical Ergative-like middles allow a more ample range of predicates besides these, like the class of 'break verbs', for example (see Palma Gutiérrez (2021b) for a detailed explanation in this regard).

So far, we have shown the similarities and divergences between prototypical Ergative-like middles and Agent-Instrument middles. In addition, Agent-Instrument middles do also possess some of the features found in Action-oriented middles, and that is why they are considered here to be hybrid structures.<sup>5</sup> Consider the following instances in this regard:

- (8) *On the track, where it is legal to push a car like this to the limits, the car handles well. It has a solid sports suspension that keeps all four wheels on the ground even in the hardest cornering, and it breaks as smoothly and as well as it accelerates.* (enTenTen13 corpus, Concordance section, Sketch Engine)

<sup>5</sup> In a constructionist framework, constructions that have partial similarities with others are connected via subpart inheritance links, referred to as multiple inheritance constructions (*cf.* Goldberg 1995: 78; see also Hilpert 2019).

- (9) *The saw has been sharpened and cuts like a dream.* (enTenTen13 corpus, Concordance section, Sketch Engine)

Example (8) above instantiates an Action-oriented middle. The Subject entity (the ‘car’ in question) is not affected at all by the handling activity, and the predicate “specifies the manner of action”, rather than focusing on “how the change of state proceeds”, as happens in prototypical Ergative-like middles (Sakamoto 2001: 101). In addition, as proposed in Palma Gutiérrez (2021b), Action-oriented middles follow the pattern [QtàQc] in compositional cospecification, which differs from the pattern found in prototypical Ergative-like middles, [QcàQc]. Therefore, Action-oriented middles are characterised by a telic relation that cospecifies the reading of the verb by virtue of the semantics of the nominal, and this telic value is lacking in prototypical Ergative-like middles. In Agent-Instrument middles like (9) above, we observe the same pattern occurring in Action-oriented middles: there is a telic relation between the predicate ‘cut’ and the nominal ‘saw’. Besides, in both Action-oriented middles and Agent-Instrument middles, the addition of the semantic value of the adjunct provokes a shift in semantic weight towards a constitutive value (Qc) which highlights the fact that the inherent properties of the Subject entity are responsible for carrying out the action, irrespective of the abilities of any implied Agent.

In addition, as previously mentioned, Agent-Instrument middles incorporate other than Patientive Subjects, the same as happens with Action-oriented middles. In the case of Agent-Instrument middles, the particular type of non-Patientive entity they employ is an Instrument.<sup>6</sup> Finally, in contrast with Action-oriented middles, Agent-Instrument middles cannot be construed with verbs that ‘specify the manner of action’. Instead, they are construed with a restrictive set of verbs that show ‘how the change of state proceeds’; *i.e.*, those belonging to Levin’s (1993) class of ‘verbs of cutting’.

Basically, the hybridity of Agent-Instrument middles is due to the fact that they coincide with the collostructional schema of prototypical Ergative-like middles (Inanimate Subject + Verb of cutting + Adverb), but they project non-Patientive Subjects and have a telic value that cospecifies the meaning of the predicate by virtue of the semantics of the nominal, as happens in Action-oriented middles, following the pattern [QtàQc] in compositional cospecification. Therefore, Agent-Instrument middles can be considered as metonymically-motivated extensions from prototypical Ergative-like middles which profile an Instrument instead of a Patient. This divergence in the order of the elements of the action chain triggers a metonymic interpretation of the Instrument as an extension of the implied Agent: the Agent physically manipulates an artifact and produces a change of state in an implied Patient (see also Hundt, 2007: 63). Notably, the semantic relation between the Instrumental Subjects and the set of verbs allowed in Agent-Instrument middles has a telic value, as occurs in Action-oriented middles. In both cases, though, the addition of the semantics of the adjunct provokes a shift in semantic importance towards a constitutive value (Qc). Thus, both types of peripheral middles are characterised by the pattern [QtàQc] in compositional cospecification.

<sup>6</sup> In the case of Action-oriented middles, as stated in Palma Gutiérrez (2021b), the type of entity occurring in Subject position fulfils the semantic role of ‘Enabler’.



### 3. METHODOLOGY

This paper is based on a corpus study of contextualised examples which is used to analyse the collocational schema of the English middle construction.<sup>7</sup> The corpus consulted was the enTenTen13 within the *Sketch Engine* tool (last accessed 02/05/2022). By using the Concordance section of this corpus, we explore the combinations of Inanimate Subject referents (Patientive and Instrumental entities) when in conjunction with a Verb + Adverb collocation, specifically focusing on: (i) affirmative and negative occurrences of 29 predicates belonging to Levin's (1993: 156) class of verbs of cutting, as listed in Table 1;<sup>8</sup> and (ii) adjuncts belonging to the facility- and quality-oriented subtype (Heyvaert 2003, 2001; and Davidse & Heyvaert 2007), as also detailed in Table 1. The total sample of instances analysed is 1703 examples, distributed as shown in Figures 1, 2 and 3 in section 4.

As mentioned in previous sections, Levin's (1993: 156) class of verbs of cutting is made of two sets of verbs: *cut* and *carve* verbs. There are 10 predicates belonging to Levin's class of *cut* verbs, and 33 predicates in the set of *carve* verbs. In this paper, the 29 predicates appearing in Table 1 were selected (6 predicates within the class of *cut* verbs and 21 predicates within the set of *carve* verbs). The rest of the predicates from Levin's classification had no entries in the corpus consulted which adjusted to the collocational schemas examined here.

According to Heyvaert (2003: 136), facility-oriented middles specify how easy (or difficult) it is to perform the action denoted by the predicate through the incorporation of adjuncts like 'easily', whereas quality-oriented middles comment on the way in which the process can be carried out through the addition of adverbials that express a quality judgement (like 'smoothly') or a comparison of quality (such as 'like a dream').

Verbs of cutting ( <i>cut</i> and <i>carve</i> verbs)	Facility-/Quality-oriented adjuncts
(i) <i>Cut</i> verbs: 'clip', 'cut', 'saw', 'scrape', 'scratch', and 'slash'.	(i) <u>Facility-oriented adjuncts</u> : 'easily', 'with ease', and 'effortlessly'.
(ii) <i>Carve</i> verbs: 'bore', 'bruise', 'carve', 'chip', 'crush', 'dent', 'drill', 'file', 'fillet', 'grate', 'grind', 'mash', 'mince', 'mow', 'nick', 'perforate', 'pulverise', 'punch', 'prune', 'shred', 'slice', 'squash', and 'squish'.	(ii) <u>Quality-oriented adjuncts</u> : 'well', 'smoothly', 'quickly', 'fast', 'rapidly', <sup>9</sup> 'like a dream', 'like butter', and 'like mad'. <sup>10</sup>

Table 1. Collocational schemas examined: Verbs of cutting + Facility-/Quality-oriented adjuncts.

7 Even though in present-day corpus linguistics there are other more complex and reliable techniques that can be used to measure the association or attraction between lexical items and constructions, as well as the combinatory patterns of co-occurring lexemes and constructions throughout large amounts of data (*cf.* Stefanowitsch & Gries 2003, 2005), this paper uses a raw frequency data analysis due to the difficulties of implementing a lexicogrammatical attraction technique in the specific corpus used and the low frequency of occurrence of the middle construction in general terms.

8 This paper extends Palma Gutiérrez's (2002) corpus by incorporating not only affirmative but also negative structures containing both Present Simple 3<sup>rd</sup> person (singular and plural) occurrences of these 29 predicates, as well as verbal phrases with auxiliary verbs (will/would/may/might) followed by these 29 verbs.

9 Middle constructions with the adverbials 'quickly', 'fast', and 'rapidly' are considered time-oriented, a subtype within the quality-oriented group in Davidse and Heyvaert's (2007: 68) work. The facet of the interaction between the non-agentive Subject and the process that these middles highlight refers to the time it takes to carry out a certain process as influenced by the (inferable) properties of the Subject entity.

10 The adjuncts 'like a dream', 'like butter', and 'like mad' are referred to as like-phrase adverbials in the tables in section 4.

The collostructional units examined in this paper, therefore, contain an Inanimate Subject referent<sup>11</sup> (manually identified, as *Sketch Engine* does not have a model to identify semantic notions like ±Animate, ±Patientive, or ±Instrumental), namely one of the 29 verbs of cutting that appear in Table 1, and one of the facility- or quality-oriented adverbials specified in the same table.

#### 4. RESULTS

In this section we present the results of the corpus analysed, taking into account the combinatory possibilities of the collocated items and the frequency of occurrence of the elements conforming the collostructional schemas examined here. Figures 1 and 2 focus on the Verb + Adverb combinations and frequency of occurrence of prototypical Ergative-like middles that include Levin’s (1993) classes of ‘cut’ verbs and ‘carve’ verbs, respectively. Figure 3 shows the Verb + Adverb combinations and frequency of occurrence of Agent-Instrument middles which incorporate Levin’s (1993) class of verbs of cutting (which involves both, the set of ‘cut’ verbs and the set of ‘carve’ verbs).

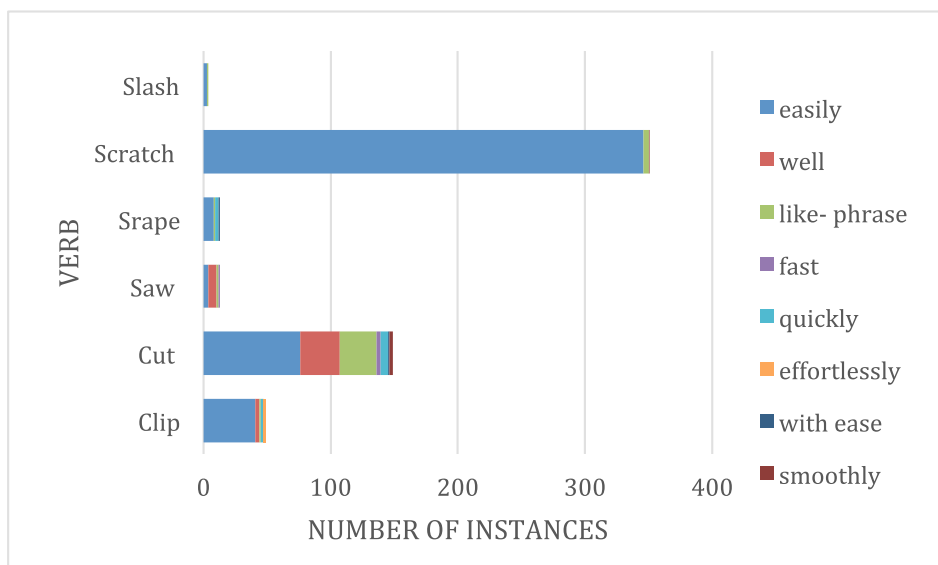


Figure 1. Frequency of occurrence of prototypical Ergative-like middles with ‘cut’ verbs.

The total sample of prototypical Ergative-like middles which incorporate ‘cut’ verbs is 579. The collostructional schema examined in this paper with regards to this type of middles is illustrated in the following [FORM] ↔ [MEANING] pattern: [SUBJ<sub>inanimate patient</sub> V<sub>enabling event</sub>

11 The Inanimate entities examined have either a Patientive or an Instrumental nature. No +Animate entities were considered in this paper, as they are a less archetypal type of subject entity in middles (cf. Yoshimura, 1998: 123). See Palma Gutiérrez (2021a) on non-prototypical middles containing +Animate Subject entities in combination with Levin’s (1993) class of ‘amuse verbs’ and facility-oriented adjuncts.

(cut) ADV<sub>facility/quality-oriented</sub>] ↔ [X (by virtue of some property P) IS SUBJECTIVELY ASSESSED BY THE SPEAKER AS BEING CONDUCTIVE TO ACT]. As shown in Figure 1, the set of ‘cut’ verbs found in the corpus examined (in combination with the adjuncts included in the same table)<sup>12</sup> are the following: ‘clip’, ‘cut’, ‘saw’, ‘scrape’, ‘scratch’, and ‘slash’. With regards to the results shown in Figure 1, the most productive middle structures are those which incorporate the verbs ‘scratch’, ‘cut’, and ‘clip’ in combination with the facility-oriented adjunct ‘easily’, followed by those which combine the predicate ‘cut’ with quality-oriented adjuncts (particularly, ‘well’ and the *like*-phrase ‘like butter’). Consider the following instances in this regard:

- (10) *[About food containers] Some people prefer to avoid plastic containers as plastic scratches easily and then it becomes difficult to keep clean.* (enTenTen13 corpus, Concordance section, Sketch Engine)
- (11) *Mat board is very tough to cut – a large paper cutter, mat board cutter, or even a scalpel will be necessary to work with it, but the foam core board cuts easily with a hobby knife.* (enTenTen13 corpus, Concordance section, Sketch Engine)
- (12) *Ideal for areas of heavy pedestrian and vehicle traffic or when it’s dark outside, the LED taillight clips easily to your belt or backpack.* (enTenTen13 corpus, Concordance section, Sketch Engine)
- (13) *The vanilla sugar cookie recipe from Cookies and Brownies is a cookie standard – and heavy on the vanilla flavor. This is a fun, flexible cookie that cuts well, maintains its shape, and is great to decorate with all sorts of things.* (enTenTen13 corpus, Concordance section, Sketch Engine)
- (14) *It is open for lunch, dinner and Sunday brunch. I would strongly remember the grilled swordfish, which cuts like butter. It comes with vegetables, but you can switch this for the mashed potatoes.* (enTenTen13 corpus, Concordance section, Sketch Engine)

In examples (10) – (14), the Inanimate entities occurring as Subjects are Patientive: the plastic container in (10), the foam core board in (11), the LED taillight in (12), the cookie in (13), and the grilled swordfish in (14) experience a change of state due to the actions denoted by their respective predicates in the way indicated by their adjuncts. The affectedness of these entities is not due to the skills of any implied Agent, but rather, it is triggered by the inherent properties of these entities. In fact, the role of the Agent is backgrounded and the responsibility of the Patient is foregrounded. For instance, the grilled swordfish of example (14) is not cut because the Agent (the eater) has the ability to do so; instead, this is due to the tenderness of the fish in question, its Qc. Besides, the addition of the semantic value of the adjuncts does not suppose a shift of semantic importance. In fact, the Qc mode is maintained, as it intensifies the responsibility of the inherent properties of the Subject entity in the carrying out of the action. Therefore, the grilled swordfish of example (14), for instance, ‘cuts like butter’ due to its inherent properties, its tenderness. Hence, in terms of the process of compositional cospecification, this type of middle illustrates the schema [QcàQc].

12 The adverb ‘rapidly’ is not included within the set of prototypical Ergative-like middles which incorporate ‘cut’ verbs due to the lack of corpus entries which fix the collostructional schema examined here.

As shown in Figure 2, the corpus data reveal that similar results are obtained with regards to the prototypical Ergative-like middles which incorporate the ‘carve’ verbs examined in this paper.

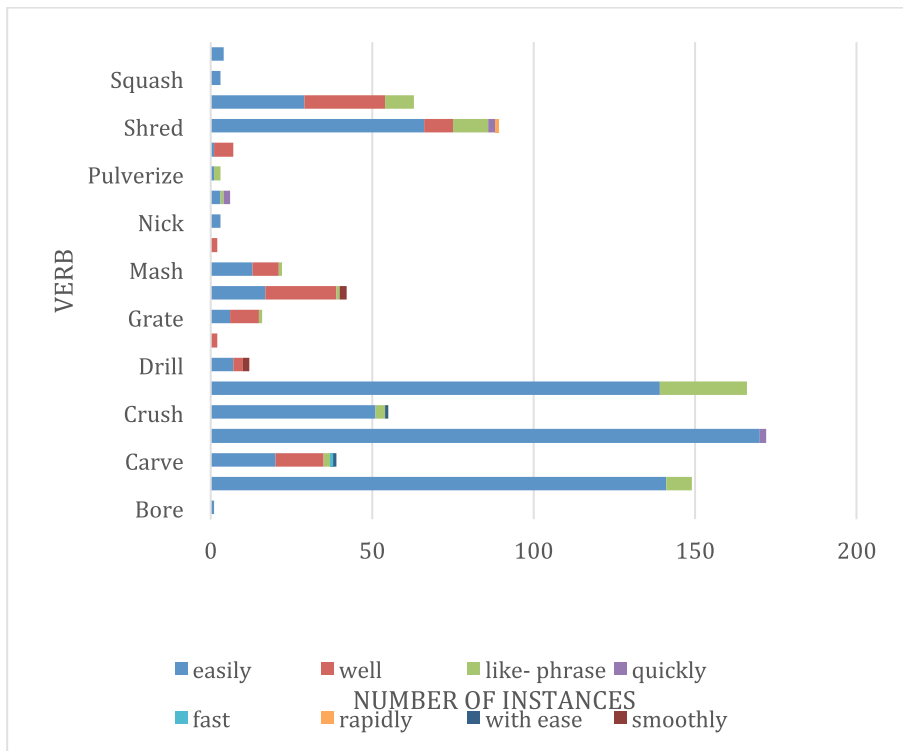


Figure 2. Frequency of occurrence of prototypical Ergative-like middles with ‘carve’ verbs.

The total sample of prototypical Ergative-like middles which incorporate ‘carve’ verbs is 856. The collostructional schema examined in this paper with regards to this type of middles is illustrated in the following [FORM] ↔ [MEANING] pattern: [SUBJ<sub>inanimate patient</sub> V<sub>enabling event</sub> (carve) ADV<sub>facility/quality-oriented</sub>] ↔ [X (by virtue of some property P) Is SUBJECTIVELY ASSESSED BY THE SPEAKER AS BEING CONDUCIVE TO ACT].<sup>13</sup> With regards to the results shown in Figure 2, the most frequent middle structures are those which incorporate the verbs ‘bruise’, ‘chip’, and ‘dent’ in combination with the facility-oriented adjunct ‘easily’. Other frequent structures are those which combine the predicates ‘crush’, ‘shred’, and ‘slice’ also with the adverb ‘easily’. Consider the following instances in this regard:

- (15) *A simple way of dealing with excess basil is to chop it fine with a very sharp knife (basil bruises easily) and whirl it in a food processor with a few tablespoons of olive oil.* (enTenTen13 corpus, Concordance section, Sketch Engine)

13 The adverb ‘effortlessly’ is not included within the set of prototypical Ergative-like middles which incorporate ‘carve’ verbs due to the lack of corpus entries which adapt to the collostructional schema examined here.

- (16) *Paint and peel nail polish is not recommended for prolonged wear – the nail polish chips easily.* (enTenTen13 corpus, Concordance section, Sketch Engine)
- (17) *Vinyl siding requires little maintenance and can be matched to the siding of your home, but it dents easily. Cedar siding is a more expensive choice, but cedar wood is resistant to rot and insects.* (enTenTen13 corpus, Concordance section, Sketch Engine)

In examples (15) – (17), the Inanimate entities occurring as Subjects are Patientive: basil in (15), the nail polish in (16), and the vinyl siding in (17) are affected by the actions denoted by their respective predicates in the way indicated by their adverbial modifiers. The change of state of these entities is not caused by the abilities of the implied Agent, rather, it is due to the inherent properties of these entities. For instance, the vinyl siding of example (17) tends to get dented due to its inherent properties (its Qc), not because of the action of the implied Agent, and the semantic value of the adjunct intensifies this condition. As previously explained in the case of prototypical Ergative-like middles with ‘cut’ verbs, those which incorporate ‘carve’ verbs follow the same pattern in the process of compositional cospecification: [QcàQc].

Figure 3 shows the combinatory possibilities, distribution, and frequency of occurrence of the Agent-Instrument middles with verbs of cutting examined in this paper. On this occasion, both ‘cut’ verbs and ‘carve’ verbs are in the same table because only two predicates within the set of ‘cut’ verbs (‘cut’ and ‘saw’) adjusted to the constructional schema analysed at this point.

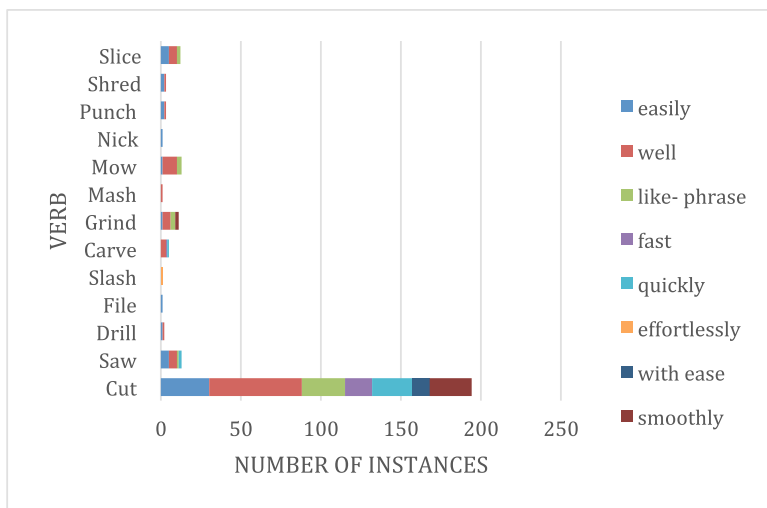


Figure 3. Frequency of occurrence of Agent-Instrument middles with verbs of cutting.

The total sample of Agent-Instrument middles which incorporate verbs of cutting is 265, of which 207 instances incorporate a ‘cut’ verb and 58 instances include ‘carve’ verbs. As shown in Figure 3, the most productive middle structures are those which incorporate the verb ‘cut’ in combination with the adjuncts ‘well’ and ‘easily’. Other salient structures in terms of frequency are those which combine the predicate ‘mow’ with the quality-oriented adjunct ‘well’. Consider the following instances in this regard:

- (18) *This saw cuts well and is simple to use. It is also light weight. Women enjoy doing home projects, and finding the right tool can be hard. Black and Decker offers the Project Mate. It is a three in one tool featuring a scraper, detail sander, and a screwdriver.* (enTenTen13 corpus, Concordance section, Sketch Engine)
- (19) *One of the coolest new tools we've had in the last couple of years is the Cheese Knife! This cleverly shaped hard plastic blade cuts easily thru hard or soft cheeses WITHOUT the cheese sticking to the knife.* (enTenTen13 corpus, Concordance section, Sketch Engine)
- (20) *[About a lawn mower] It cuts great, bag doesn't let much dust out, light weight and easy to handle, I love this mower, stated first pull, mows well, clean cut, and I love the bagger; it mulches beautifully and the job was done in record time. This lawn mower (almost) makes me want to go out and mow the lawn.* (enTenTen13 corpus, Concordance section, Sketch Engine)

In examples (18) – (20), the Inanimate entities occurring as Subjects are non-Patientive, indeed, they fulfill an Instrumental role: the ‘saw’ in (18), the ‘knife’ in (19), and the lawn mower in (20) do not experience a change of state due to the actions denoted by their respective predicates. In fact, following Hundt’s (2007: 63) analysis, it is the implied Agent that physically manipulates the Instrument, and consequently, the implied Patient is affected by the action denoted by the verb in each case. Because the argument that occupies the Subject position in this type of middles has an Instrumental role and the Agent is still implicit, the same pragmatic schema is represented in Agent-Instrument middles like (18) – (20) as in other types of middles such as prototypical Ergative-like middles. Thus, the Agent is backgrounded and another argument, in this case the Instrument, is profiled and occupies the Subject position. In terms of semantic analysis, this implies that the Subject entities have certain inherent properties which are subjectively assessed by the speaker as being conducive to the actions denoted by the respective predicates in the way indicated by their adjuncts. The main difference between Agent-Instrument middles like (18) – (20) and prototypical Ergative-like middles is that in the former there is a telic value that cospecifies the reading of the Subject by virtue of the semantics of the predicate, as happens in Action-oriented middles: a ‘saw’ is made to ‘cut’, the purpose of a ‘knife’ is to ‘cut’, and the function of a ‘lawn mower’ is to ‘mow’.

However, as argued in section 2, this Qt mode is backgrounded in the process of compositional cospecification, and the Qc mode referred to the inherent properties of the Instrument is foregrounded due to the addition of the semantic value of the adjunct. The ‘saw’ in (18) ‘cuts well’ because of its lightweight and versatile nature; the ‘knife’ in (19) ‘cuts easily’ due to its ‘cleverly shaped hard plastic blade’; and the ‘lawn mower’ in (20) ‘mows well’ because its blades produce clean cuts. Thus, Agent-Instrument middles illustrate the schema [QtàQc] in compositional cospecification, just as Action-oriented middles do.

## 5. CONCLUSIONS

In this paper, we have examined both prototypical Ergative-like middles and their metonymically-motivated extensions called Agent-Instrument middles. Both structures allow the incorporation of Levin’s (1993) class of verbs of cutting, but they represent divergent portions of the action chain, and thus, the semantic role of their Subject referents and their processes of compositional cospecification are different. However, these two types of

middles have common syntactic, semantic, and pragmatic underlying schemas, as shown by the prototype effects displayed by the middle construction and the family-resemblance analysis provided here (cf. Palma Gutiérrez 2022). It has thus been demonstrated that Agent-instrument middles are hybrid structures which combine features of both prototypical Ergative-like middles and also properties of Action-oriented middles. The common property that Agent-Instrument middles share with prototypical Ergative-like middles refers to the fact that both can be formed with verbs of cutting. On the other hand, the features that Agent-Instrument middles share with Action-oriented middles refer to the fact that both incorporate non-patientive subject referents, and their process of compositional cospecification is reflected in the formula [QtàQc].

## REFERENCES

- Ackema, P. & M. Schoorlemmer. (1994). The middle construction and the syntax semantics interface. *Lingua*, 93, 59-90. [https://doi.org/10.1016/0024-3841\(94\)90353-0](https://doi.org/10.1016/0024-3841(94)90353-0)
- Cortés-Rodríguez, F. J. & R. Mairal Usón. (2013). Constraints on English middle structures: A Lexical-Constructional analysis. *Onomázein*, 27, 220-240.
- Davidse, K. & L. Heyvaert. (2007). On the middle voice: An interpersonal analysis of the English middle. *Linguistics*, 45(1), 37-83. <https://doi.org/10.1515/LING.2007.002>
- Dixon, R. M. W. (1982). *Where have all the adjectives gone? And other essays in Semantics and Syntax*. Berlin: Mouton de Gruyter. <https://doi.org/10.1515/9783110822939>
- Fagan, S. (1992). *The syntax and semantics of middle constructions: A study with special reference to German*. Cambridge: Cambridge University Press.
- Fawcett, R.P. (1980). *Cognitive Linguistics and Social Interaction: Towards an Integrated Model of a Systemic Functional Grammar and the Other Components of a Communicating Mind*. Exeter and Heidelberg: Julius Groos Verlag and Exeter University Press.
- Fellbaum, C. (1986). *On the middle constructions in English*. Bloomington: Indiana University Linguistics Club.
- Goldberg, A. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.
- Goldberg, A. (2006). *Constructions at work: The nature of Generalization in Language*. Oxford: Oxford University Press.
- Goldberg, A. (2019). *Explain me this: Creativity, competition, and the partial productivity of constructions*. Princeton, NJ: Princeton University Press. <https://doi.org/10.1515/9780691183954>
- Hale, K. & S. J. Keyser. (2002). *Prolegomenon to a Theory of Argument Structure*. Cambridge, Massachusetts: The MIT Press. <https://doi.org/10.7551/mitpress/5634.001.0001>
- Heyvaert, L. (2003). *A cognitive-functional approach to nominalization in English*. Berlin: Mouton de Gruyter. <https://doi.org/10.1515/9783110903706>
- Heyvaert, L. (2001). Nominalization as an interpersonally-driven system. *Functions of language*, 8(2), 283-324. <https://doi.org/10.1075/fol.8.2.06hey>
- Hilpert, M. (2019). *Construction Grammar and its application to English* (2<sup>nd</sup> ed.). Edinburgh: Edinburgh University Press.
- Hoekstra, T. & I. Roberts. (1993). Middle Constructions in Dutch and English. In E. Reuland and W. Abraham. (Eds.), *Knowledge and Language (vol. 2): Lexical and Conceptual Structure* (pp. 83-220). Dordrecht: Kluwer. [https://doi.org/10.1007/978-94-011-1842-2\\_9](https://doi.org/10.1007/978-94-011-1842-2_9)
- Hundt, M. (2007). *English mediopassive constructions*. Amsterdam/New York: Rodopi.
- Kemmer, S. 1993. *The Middle Voice*. Amsterdam: John Benjamins. <https://doi.org/10.1075/tsl.23>
- Keyser, S. J. & T. Roeper. (1984). On the Middle and Ergative constructions in English. *Linguistic Inquiry*, 15(3), 381-416.

- Lakoff, G. (1987). *Women, Fire and Dangerous things: What categories tells us about the mind*. Chicago: Chicago University Press. <https://doi.org/10.7208/chicago/9780226471013.001.0001>
- Langacker, R.W. (1987). *Foundations of Cognitive Grammar (Vol. 1), Theoretical prerequisites*. Stanford, California: Stanford University Press.
- Langacker, R.W. (2008). *Cognitive Grammar. A basic introduction*. Oxford/New York: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195331967.001.0001>
- Levin, B. (1993). *English Verb Classes and Alternations. A Preliminary Investigation*. Chicago: University Press.
- Marín Arrese, J. (2011). Spontaneous and facilitative events revisited. In P. Guerrero Medina (Ed.), *Morphosyntactic Alternations in English: Functional and Cognitive Perspectives* (pp. 137-160). London: Equinox.
- Palma Gutiérrez, M. (2021a). Un enfoque contrastivo del Sujeto +Animado en la construcción media inglesa y española. In G. Corpas Pastor, M.R. Bautista Zambrana, & C.M. Hidalgo-Ternero. (Eds.), *Sistemas fraseológicos en contraste: enfoques computacionales y de corpus* (pp. 101-120). Albolote, Granada: Comares.
- Palma Gutiérrez, M. (2021b). Prototype effects in the process of Compositional Cospecification in the middle construction. *Estudios Interlingüísticos*, 9.
- Palma Gutiérrez, M. (2022). *A family-resemblance analysis of the middle construction: A functional-cognitive approach*. (PhD Thesis) Córdoba: UCO Press.
- Pinker, S. (1989). *Learnability and Cognition. The acquisition of argument structure*. Cambridge, Massachusetts/London, England: The MIT Press.
- Pustejovsky, J. (1991). The Generative Lexicon. *Computational Linguistics*, 17(4), 409-441.
- Pustejovsky, J. (1995). *The Generative Lexicon*. Cambridge: The MIT Press.
- Sakamoto, M. (2001). The middle and related constructions in English: A Cognitive network analysis. *English Linguistics*, 18(1): 86-110. <https://doi.org/10.9793/elsj1984.18.86>
- Stefanowitsch, A. & S. T. Gries. (2003). Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics*, 8(2), 209-243. <https://doi.org/10.1075/ijcl.8.2.03ste>
- Taylor, J. R. (1995). *Linguistic Categorization. Prototypes in Linguistic Theory* (2<sup>nd</sup> Ed.). Oxford: Clarendon Press.
- Van Oosten, J. (1977). Subjects and Agenthood in English. *Papers from the 13<sup>th</sup> Regional Meeting of the Chicago Linguistic Society*, 459-471.
- Van Oosten, J. (1984). *The nature of Subjects, Topics and Agents: A cognitive explanation*. PhD Thesis. Berkeley, California: University of California.
- Van Oosten, J. (1986). *The nature of subjects, topics and agents: A cognitive explanation*. Bloomington, Indiana: Indiana University Linguistics Club.
- Wittgenstein, L. (1958). *Philosophical Investigations* (2<sup>nd</sup> Edition) (Translated by G.E.M. Anscombe). Oxford: Basil Blackwell.
- Yoshimura, K. (1998). Encyclopedic structure of nominals and middle expressions in English. *Kobe Papers in Linguistics*, 1, 112-140. <https://doi.org/10.1017/S136067430400139X>
- Yoshimura, K. & J. R. Taylor. (2004). What makes a good middle? The role of qualia in the interpretation and acceptability of middle expressions in English. *English Language and Linguistics*, 8(2), 293-321.

### Online Resources

Sketch Engine: <https://www.sketchengine.eu/>





## Extracción automática de unidades terminológicas guiada por Framenet: una aplicación al corpus electrónico CORD-19

Automatic framenet-guided extraction of terminology: an application to the cord-19 electronic corpus

MARIO CRESPO MIGUEL

Universidad de Cádiz, España

[mario.crespo@uca.es](mailto:mario.crespo@uca.es)

<https://orcid.org/0000-0002-8570-8678>

### Resumen

La evolución de la terminología va unida a las nuevas tecnologías y al desarrollo de plataformas de trabajo o interfaces, que permitan crear un glosario técnico de forma semiautomática o incluso automática. Las unidades terminológicas y sus propiedades están conectadas a la expresión del conocimiento de los campos especializados en los que aparecen, por lo que estos enfoques automáticos no solo se enfrentan a la tarea de determinar cuáles son las unidades terminológicas de un campo determinado, sino a expresar cómo se estructura dicha información en esa disciplina. Muchas de las unidades terminológicas que se pueden encontrar en un ámbito científico determinado aparecen en otros campos e, incluso, en el uso general. Los términos suelen estar presentes en el acervo léxico de las lenguas y comparten con las unidades léxicas un complejo conjunto de relaciones. La semántica de marcos es un modelo particularmente

### Abstract

The evolution of Terminology is joined to new technologies and the development of work platforms or interfaces that allow creating a technical glossary semi-automatically or even automatically. Terms and their properties are connected formally to the expression of knowledge of specialized fields in which they occur, so automatic approaches are not only faced with the task of determining which are the terminological units of a given field, but to express how such information is structured in their technical field. Most of the terms occurring in a scientific domain are also found in other disciplines and even in everyday language. Terms are often present on the lexical stock of languages and share with lexical units a complex set of relationships. Frame Semantics is a particularly attractive model for the terminological work, interested in accounting for the connection between the conceptual structure of a specialized field and the

---

**Para citar este artículo:** Crespo Miguel, M. (2022). Extracción automática de unidades terminológicas guiada por Framenet: una aplicación al corpus electrónico CORD-19. *ELUA*, (38), 281-300. <https://doi.org/10.14198/ELUA.22357>

Recibido: 29/03/2022 Aceptado: 14/05/2022

© 2022 Mario Crespo Miguel



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

atractivo para el trabajo terminológico, interesado en dar cuenta de la conexión entre la estructura conceptual de un campo de conocimiento especializado y los elementos utilizados para transmitir este conocimiento. Esto ha llevado a muchos investigadores a utilizar FrameNet como forma de representar la terminología. FrameNet es un recurso en línea para el inglés basado en la semántica de marcos y respaldado por pruebas de corpus. Un marco se fundamenta en el hecho de que ciertas palabras evocan determinadas situaciones en las que tienen lugar determinados participantes. Estas situaciones o marcos son estructuras estereotipadas que representan áreas de experiencia y conocimiento sociocultural. Presentamos un enfoque estadístico basado en corpus que es capaz de seleccionar los marcos de FrameNet que mejor representan un conjunto de textos electrónicos sobre COVID-19 e indicar cuáles de sus unidades léxicas funcionan como unidades terminológicas de ese corpus. Los resultados muestran cómo esta metodología puede ser un buen apoyo al trabajo terminográfico, ya que no solo permite la extracción de unidades terminológicas, sino el uso del esqueleto formal de FrameNet como medio para estructurar este conocimiento.

**PALABRAS CLAVE:** Terminología, FrameNet, Lenguaje especializado, Extracción de términos, Análisis de corpus, Frecuencia de aparición del término - Frecuencia inversa del documento (*tf-idf*)

## 1. INTRODUCCIÓN

La terminología es una rama multidisciplinar de la lingüística aplicada que analiza cómo se estructura el conocimiento en los diferentes dominios especializados y cuáles son las características de las unidades usadas para expresar ese conocimiento (L'Homme 2018). Existe un esfuerzo continuo por aumentar el número de recursos terminológicos que recojan las unidades terminológicas del ámbito científico, analicen su interpretación y permitan su inclusión en glosarios o léxicos especializados para la comunidad científica (Ramírez Salado 2019). Sin embargo, la construcción de este tipo de recursos es bastante difícil de realizar. En primer lugar, la continua investigación hace que la lista de nuevos términos esté en constante crecimiento. Además, las unidades terminológicas se relacionan entre sí según diferentes relaciones semánticas y conceptuales que se deben determinar (Casas Gómez 2014). Es por ello que esta disciplina se está ligando a las nuevas tecnologías y al desarrollo de plataformas de trabajo o interfaces que permitan crear un glosario técnico

elements used to transmit this knowledge. This has led to many researchers to use FrameNet as a way of representing terminology. FrameNet is an online resource for English based on Frame Semantics and supported by corpus evidence. A frame is founded on the basis that certain words evoke certain situations in which particular participants take place. These situations or frames are stereotyped structures representing areas of sociocultural experience and knowledge. We present a statistical approach based on corpus able to select most representative FrameNet frames that best represent a set of electronic texts on COVID-19 and show which of their lexical units work as terminological units. Results confirm that this methodology can be a good support for terminographic work, since it not only allows the extraction of terminological units, but also the use of the FrameNet framework to structure this knowledge.

**KEYWORDS:** Terminology, FrameNet, Specialised language, Term extraction, Corpus analysis, Term frequency – Inverse document frequency (*tf-idf*).

de forma semiautomática o incluso automática como parte de la tendencia general en Lingüística (Crespo 2020b). Los términos y sus propiedades están conectados formalmente a la expresión del conocimiento de los campos especializados en los que aparecen, por lo que los enfoques automáticos no solo se enfrentan a la tarea de determinar cuáles son las unidades terminológicas de un campo determinado, sino indicar cómo se estructura dicha información en el conocimiento general de esa disciplina.

La semántica de marcos (Fillmore 1977) es un modelo particularmente atractivo para el trabajo terminológico, interesado en dar cuenta de la conexión que existe entre la estructura conceptual de los campos de conocimiento especializados y las unidades lingüísticas utilizadas para transmitir este conocimiento (L'Homme, Subirats Rüggeberg y Robichaud 2016; Azoulay 2017). Esta corriente se basa en la suposición de que los significados de las unidades léxicas se construyen sobre los conocimientos previos de los individuos (experiencia, creencias, convenciones, etc.). Esta conexión se establece mediante marcos semánticos: estructuras estereotipadas que representan áreas de experiencia sociocultural que describen un tipo de evento o situación, y los participantes que aparecen en él. Se trata de un modelo que hace explícita la relación entre las unidades léxicas y este conocimiento. Los hablantes entienden los significados de las palabras en virtud de los marcos que evocan (Potęga 2017), sin importar que pertenezca a dominios generales o especializados. Dentro de esta forma de entender las relaciones léxicas, FrameNet (Baker *et al.* 1998; Ruppenhofer *et al.* 2006) es un recurso online basado en la semántica de marcos que pretende agrupar el vocabulario del inglés en diferentes marcos y establecer sus relaciones conceptuales. Se trata de un proyecto en desarrollo que, hasta ahora, ha definido 1.224 marcos para 13.640 unidades léxicas diferentes del inglés.

El trabajo que aquí presentamos introduce un enfoque automático para reutilizar el modelo e información codificada en FrameNet con la finalidad de estructurar el conocimiento y las unidades terminológicas que aparecen en textos de especialidad. Este trabajo parte del corpus en línea COVID-19 Open Research Dataset (CORD-19), de 100 millones de palabras para uso científico. En primer lugar, se determinan cuáles de los 1.224 marcos codificados actualmente en FrameNet son los más representativos de este ámbito médico. Para ello, seguimos un método estadístico que selecciona los marcos de FrameNet más relevantes a partir de la comparación de la frecuencia de las unidades codificadas en este proyecto tanto en el corpus online COVID-19 Open Research Dataset como en uno de temática general. De la lista de marcos resultante, elegimos las unidades léxicas más relevantes de acuerdo con un análisis *tf-idf* (del inglés *Term frequency – Inverse document frequency*) usando el mismo corpus especializado. Los resultados muestran cómo FrameNet puede utilizarse para estructurar el conocimiento de un dominio especializado, determinar los marcos más representativos del mismo, establecer relaciones entre ellos y analizar cuáles de las unidades léxicas de estos marcos funcionan como unidades terminológicas de este corpus técnico. Por último, observamos que los resultados se ven afectados por la falta de cobertura de unidades léxicas de los marcos de FrameNet, ya que este se encuentra en proceso de creación. Para resolver este problema, ampliamos el número de palabras mediante su conexión de esta base de datos con WordNet (Miller *et al.* 1993), una base de datos léxica que organiza el vocabulario inglés en grupos según conceptos y relaciones semánticas. Tras la conexión de ambos recursos, los resultados generales mejoran.

## 2. TERMINOLOGÍA BASADA EN LA SEMÁNTICA DE MARCOS

La semántica de marcos (Fillmore 1985; Fillmore y Baker 2010) es un desarrollo dentro de la lingüística cognitiva que ha llamado la atención de muchos investigadores interesados en dar cuenta de las asociaciones entre el léxico y el conocimiento que se supone que comparten los hablantes de una lengua (L'Homme, Subirats Rüggeberg y Robichaud 2016). Se parte de la base de que los significados de las unidades léxicas se construyen en relación con el conocimiento representado en los marcos semánticos de los que estas forman parte. Los marcos semánticos son estructuras que representan situaciones específicas (por ejemplo, una situación médica) (Bernier-Colborne y L'Homme 2015) y participantes implicados en ellas. Los modelos cognitivos (Lakoff 1987, cit. en Durán-Muñoz 2016) dan la posibilidad de definir las unidades léxicas de una forma sistemática y flexible según la situación comunicativa en la que se producen.

Como indicábamos en el apartado anterior, FrameNet es un recurso en línea para el inglés basado en la semántica de marcos con el objetivo de identificar y definir todas las posibles situaciones evocadas por las unidades léxicas de una lengua, identificar los participantes de tales situaciones (lo que en la práctica supone la determinación de los roles semánticos evocados) y dar cuenta de las relaciones sintáctico-semánticas de todos ellos mediante pruebas de corpus (Ruppenhofer *et al.* 2006). Para ejemplificar cómo debe concebirse cada una de estas estructuras, podemos partir del marco 'cure'. En él se describe una situación estereotipada en la que un *Healer* ('sanador') trata y cura una *Affliction* ('afección') de un *Patient* ('paciente'). Otros participantes o roles semánticos centrales de este marco serían: *Body part* ('parte del cuerpo'), *Medication* ('medicación') y *Treatment* ('tratamiento'). De esta manera, unidades léxicas como *alleviation.n*, *curable.a*, *curative.a* o *cure.v* evocarían este conocimiento al aparecer en el discurso (Ruppenhofer *et al.* (2016)<sup>1</sup>. A partir de aquí el objetivo de FrameNet es analizar y anotar diferentes enunciados de un corpus lingüístico para mostrar las estructuras sintácticas en las que aparecen estas unidades y cómo se estructuran tales roles semánticos sobre ellas (Verdaguer 2020). De esta manera, las anotaciones de FrameNet son formalmente conjuntos de triples que representan las realizaciones de los elementos del marco para cada frase anotada, cada una de las cuales consiste en rol semántico (por ejemplo, *Affliction*), una función gramatical (e.g., objeto) y un tipo de frase (e.g., sintagma nominal (NP)):

<sup>1</sup> <https://framenet.icsi.berkeley.edu/fndrupal/>

Roles semánticos	Estructuras en las que aparece y número	Ejemplo en el corpus
Affliction	2nd.-- (2) DNI.-- (4) INI.-- (1) NP.Ext (6) NP.Obj (12) PP[of].Dep (6) PPing[of].Dep (1)	On the way <b>Jesus CURED</b> the woman with the haemorrhage with the haemorrhage . If the procedure succeeds , Carly will be able to lead a normal life , offering hope to thousands of sufferers with diseases like cystic fibrosis and muscular dystrophy who could be <b>CURED</b> by similar techniques . I could never <b>CURE</b> him .DNI Rosemary apparently has to cope with her husband 's anxiety that she could <b>only</b> be <b>CURED</b> by becoming more ` assertive " .DNI
Degree	AVP.Dep (2)	I 'm very positive that <b>Natalie</b> will be <b>CURED</b> .CNIDNI
Healer	NP.Ext (13) CNI.-- (4)	<b>Isis</b> then <b>CURED</b> Re by reciting a spell and with her new power became one of the mightiest goddesses .DNI
Manner	AVP.Dep (3)	<b>You</b> can not <b>CURE</b> yourself .INI
Patient	NP.Obj (9) NP.Ext (4) PP[re].Dep (1) INI.-- (12) DNI.-- (1) 2nd.-- (3) PP[in].Dep (2)	I do n't want to suggest that <b>social problems</b> can be <b>CURED</b> by the application of wealth .INI [It was almost as if she was suffering from <b>some dreadful disease</b> that could only be <b>CURED</b> by his physical removal .DNI By a visit to a doctor , a psychiatrist , or even through a programme of counselling and guidance , <b>mental illnesses</b> can be <b>CURED</b> , and almost always are .INI It occurred mainly in communities with poor diets and was eventually identified as a <b>deficiency disease</b> which could be <b>promptly and completely</b> <b>CURED</b> by supplying adequate food .INI <b>All forms of early syphilis</b> can be <b>completely</b> <b>CURED</b> , and some of the later manifestations can be markedly improved , or , at least , the progression of the disease can be arrested .CNIINI It could point to plenty of ailments that the Spanish economic <b>rejuvenation</b> so far has failed to <b>CURE</b> .INI The others are dozing while I chatter , attempting to <b>CURE</b> their <b>somnambulism</b> with my words .
Treatment	PP[by].Dep (4) PPing[by].Dep (4) NP.Ext (9) PP[with].Dep (2) PP[without].Dep (1)	

Figura 1. Anotación en FrameNet para la unidad ‘Cure.v’. Fuente: Ruppenhofer *et al.* (2016). <https://framenet.icsi.berkeley.edu/fndrupal/>.

La anotación se realiza para cada uno de los sentidos que poseen las palabras analizadas. Así, una palabra polisémica debe pertenecer a marcos semánticos diferentes ya que cada uno de sus sentidos evoca un marco o estructura conceptual particular. La Figura 1 ilustra los diferentes significados asignados a la unidad léxica ‘accept.v’, cada uno de los cuales lleva parejo unos roles y una estructura sintáctica diferente:

Marcos asociados a la unidad léxica ‘accept.v’	
<i>Respond to proposal</i>	<i>Receiving</i>
A Speaker addresses a Proposal made by an Interlocutor, either agreeing to it or rejecting it.	A Recipient comes into possession of the Theme as a result of the joint action of the Donor and the Recipient.
accept.v, acceptance.n, rebuff.n, rebuff.v, refuse.v, reject.v, rejection.n, turn down.v	accept.v, receipt.n, receive.v

Figura 2. Unidad léxicas ‘accept.n’ en dos marcos diferentes de FrameNet.

Actualmente FrameNet está en proceso de creación y hasta ahora se han completado 202.232 anotaciones. La versión utilizada en este trabajo es la R1.7. El estado actual de FrameNet puede verse en la Tabla 1. Esta versión contiene 1.224 marcos que cubren 13.640 unidades léxicas diferentes y 10.542 roles semánticos:

	FrameNet versión: R1.7
Marcos	1.224
Unidades léxicas ('Triggers')	13.640
Promedio de unidades léxicas por marco	12,5
Roles semánticos	10.542
Promedio de roles semánticos por marco	9,7

Tabla 1. Estado de FrameNet en su versión R1.7.

Como señalan Johnson y Fillmore (2000), las situaciones representadas en FrameNet se pueden agrupar en torno a dominios del conocimiento y la experiencia humana, tales como "Body, Chance, Cognition, Communication, Emotion, Health, Life Stages, Motion, Perception, Society, Space, Time, Transaction, or a General domain". Además, los marcos se conectan entre sí según diferentes relaciones conceptuales e interrelacionan el léxico de múltiples maneras. La Figura 3 (Gildea y Jurafsky 2002) muestra los marcos vinculados al dominio de la cognición ('cognition'):

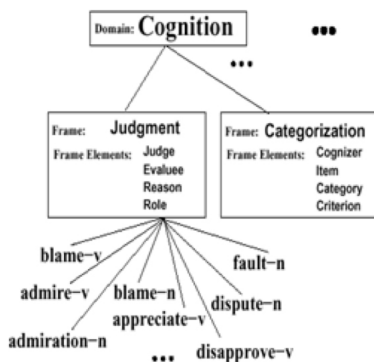


Figura 3. Dominio de la cognición. Fuente: FrameNet.

Como ya indicábamos en el apartado anterior, la ligazón que existe en FrameNet entre estructura conceptual y unidades léxicas ha llevado a muchos investigadores a utilizar FrameNet como forma de representar la terminología. A este respecto, Malm *et al.* (2018) indican que la mayoría de los términos que se encuentran en un dominio científico también se encuentran en otros campos (por ejemplo, tense.n en Lingüística). Aparte de su terminología específica, un dominio científico hace uso de las mismas unidades léxicas y patrones sintácticos que la lengua general. Según Verdaguer (2020), el lenguaje del dominio de la biomedicina no es muy diferente del lenguaje ordinario. Asimismo, Guerrero Ramos y Pérez Lagos (2003), y Cabré (2005) señalan que muchas unidades terminológicas que se

encuentran en el léxico del hablante general pueden llegar a adquirir valores terminológicos en determinadas situaciones. Muchas unidades terminológicas están presentes en el acervo léxico de las lenguas y comparten con las unidades léxicas un complejo conjunto de relaciones (L’Homme, Robichaud y Subirats Rüggeberg 2020). Entre los estudios que han aplicado FrameNet a registros especializados se encuentran Dolbey et al. (2006) para textos biomédicos, Venturi (2013) para el lenguaje jurídico, Malm et al. (2018) para la lingüística o L’Homme et al. (2020) para la ecología.

Estos proyectos suelen utilizar una metodología *bottom-up* que parte del análisis de las unidades terminológicas que aparecen en los corpus científicos para asignarles el marco que mejor representa su significado en ese contexto. La gran desventaja de esta forma de trabajar es que analiza secuencialmente las unidades que van apareciendo, por lo que no se discrimina de ninguna manera entre marcos que realmente sean específicos y representen ese campo del conocimiento. Por el contrario, una metodología *top-down* partiría del proceso contrario: define las unidades de conocimiento (o marcos) y sus participantes (roles semánticos), y a partir de ahí enumera qué elementos son los que se asocian a esos marcos conceptuales. Sin embargo, esta última metodología se enfrenta al reto de identificar por adelantado la gama de marcos que mejor representa un determinado conjunto de textos científicos. Entre las posibles soluciones a este problema se encuentra el recurrir a la opinión del experto (que deberá familiarizarse previamente con la semántica de marcos), o recurrir a un análisis macroscópico de las unidades del corpus, que permita identificar patrones y los contenidos que se desarrollan en un corpus técnico de manera general.

Este trabajo se sitúa en esta segunda opción y presenta un método estadístico para hacer esa selección de una manera fiable a partir del análisis automático del corpus. La siguiente sección presenta la metodología necesaria para determinar los marcos más significativos de FrameNet para un corpus determinado. Para ello se compara cómo se comportan las unidades léxicas de cada marco desde el punto de vista de su frecuencia en un corpus específico respecto a uno general. En este estudio se ha trabajado con COVID-19 Open Research Dataset (CORD-19) de 100 millones de palabras y el Corpus of Contemporary American English (COCA) de 560 millones de palabras. Una vez realizada la selección de marcos, vinculamos sus unidades léxicas con los términos *tf-idf* más representativos del corpus CORD-19. Los resultados se presentan al final de este artículo.

### 3. METODOLOGÍA

#### 1.1. Selección automática de marcos de FrameNet

Yarowsky (1995) afirma que el tema global de un determinado documento o corpus se refleja en el vocabulario que se puede encontrar en el mismo (por ejemplo, en un documento médico no es raro encontrar el término ‘cardiología’ o ‘auscultación’ frente a otros tipos de texto donde tales unidades no aparecerán). Esto no solo tiene implicación respecto a la aparición o no de determinadas unidades en el texto en función de la temática de los mismos, sino que existen diferencias significativas respecto a las frecuencias observadas del léxico que comparten todo estos textos según el corpus donde aparezcan. De esta manera, habrá elementos que tengan una mayor frecuencia en los textos científicos que en los generales (por ejemplo, es el caso de términos como “pulmón” o “cura” en un ámbito médico). En la



Tabla 2 se comparan la frecuencias por millón de palabras de algunas unidades de los marcos FrameNet *Medical\_conditions* y *Medical\_professional* en un corpus médico y otro general según los datos proporcionados por el corpus COCA, del que hablaremos más adelante.

	Médico	General
cancer.n	407	111
asthma.n	165	8
doctor.n	162	19
surgeon.n	198	24

Tabla 2. Frecuencias por millón de palabras en un corpus médico y otro general.

Como se puede observar, estas unidades léxicas ‘cancer.n’, ‘asthma.n’, ‘doctor.n’ y ‘surgeon.n’ tienen una mayor aparición en el dominio médico que en el general. Esto indica que la distribución de frecuencias de las unidades del corpus podría ser un indicador de la pertenencia a un determinado campo temático. Siguiendo esta premisa, es posible analizar la frecuencia de las unidades léxicas de cada marco de FrameNet en un corpus científico y compararlas respecto a cómo aparecen en un corpus general. La idea es determinar si existen diferencias estadísticas entre ambas. Las unidades de los marcos más relacionados con la temática del corpus estarán sobrerrepresentadas en su frecuencia respecto al otro corpus.

Como ya se ha indicado en la Tabla 1, el estado actual de FrameNet clasifica 13.640 unidades léxicas en 1.224 marcos diferentes. Evidentemente las unidades del inglés exceden las dimensiones actuales de FrameNet. El desarrollo de recursos léxicos requiere de un gran esfuerzo, lo que justifica que sea un proceso lento. Es por ello por lo que poder reutilizar o transferir la información de otros repositorios léxicos a FrameNet se considera como una de las posibles soluciones al problema (Cristea y Pistol 2012). Entre las diferentes posibilidades se ha planteado su alineación con WordNet, de tal manera que la información contenida en este recurso pueda integrarse en otras bases de datos léxicas (Crespo 2021).

WordNet (Miller et al. 1993) es una base de datos léxica que organiza el vocabulario del inglés a partir de relaciones conceptuales y semánticas. Estos grupos, llamados synsets, reflejan la disponibilidad del léxico inglés para reflejar una idea o concepto. Los synsets son conjuntos de “sinónimos cognitivos”. WordNet ha sido desarrollado en el Cognitive Laboratory de la Universidad de Princeton bajo la dirección de George A. Miller y contiene unas 150.000 unidades distribuidas entre las diferentes categorías verbales de sustantivos, verbos, adjetivos y adverbios. En la siguiente figura se aprecia los synsets asociados a la unidad ‘pain.n’ en inglés y las unidades asociadas a cada uno de sus sentidos:

Noun: “pain.n”

S: (n) **pain, hurting** (a symptom of some physical hurt or disorder) “the patient developed severe pain and distension”

S: (n) **pain, painfulness** (emotional distress; a fundamental feeling that people try to avoid) “the pain of loneliness”

S: (n) **pain, pain sensation, painful sensation** (a somatic sensation of acute discomfort) “as the intensity increased the sensation changed from tickle to pain”

S: (n) **pain, pain in the neck, nuisance** (a bothersome annoying person) “that kid is a terrible pain”

S: (n) **annoyance, bother, botheration, pain, infliction, pain in the neck, pain in the ass** (something or someone that causes trouble; a source of unhappiness) “washing dishes was a nuisance before we got a dish washer”; “a bit of a bother”; “he’s not a friend, he’s an infliction”

Figura 4. Synsets asociados a la unidad léxica “pain.n”.<sup>2</sup>

A partir del repositorio de WordNet, Crespo (2021) amplía la cobertura de FrameNet conectando los marcos de FrameNet con los synsets que los representan en este recurso. Este enfoque alcanza una precisión del 88% y amplía la cobertura de FrameNet de 13.631 unidades a 25.305. Sobre este segundo número de unidades aplicamos un procedimiento estadístico basado en un test de hipótesis para la determinación de los marcos más representativos.

Como es ampliamente sabido, una prueba de hipótesis se utiliza para comparar dos conjuntos de datos y poder afirmar si una propiedad que es constante para una determinada población bajo análisis, es compatible con otra muestra de la misma población. Si no es compatible, se rechaza la hipótesis nula y se considera que las dos muestras proceden de poblaciones diferentes. Este trabajo sigue el procedimiento metodológico propuesto por el Crespo (2020a), que utiliza la prueba de rangos con signo de Wilcoxon para comparar si existen diferencias estadísticamente significativas entre las frecuencias observadas en uno u otro corpus. La prueba de rangos con signo de Wilcoxon se trata de una prueba de hipótesis estadística no paramétrica que compara dos muestras relacionadas, muestras emparejadas o mediciones repetidas sobre una única muestra para evaluar si la población de diferencias tiene una mediana de cero, por lo que las hipótesis nula y alternativa son las siguientes (Triola, 2007):

H0: Los pares emparejados tienen diferencias que provienen de una población con una mediana igual a cero.

H1: Los pares emparejados tienen diferencias que provienen de una población con una mediana distinta de cero.

Si  $n \leq 30$ , el estadístico de prueba es T.

Si  $n > 30$ , el estadístico de prueba es

$$\text{Si } n > 30, \text{ el test estadístico es } z = \frac{T - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$$

<sup>2</sup> Ejemplo extraído de <https://wordnet.princeton.edu>.

Esta es la fórmula de la prueba de rangos con signo de Wilcoxon, donde  $T$  es la menor de las dos sumas siguientes:

1. El valor absoluto de la suma de los rangos negativos de las diferencias ( $d$ ) que no sean cero.  $U$
2. La suma de los rangos positivos de las diferencias ( $d$ ) que no sean cero.

Sobre los valores de  $T$  o  $z$  se comprueban sus valores críticos. En nuestros experimentos, utilizamos las frecuencias relativas de las unidades y adoptamos una hipótesis de prueba de cola a la derecha con una significación del 95% y del 99%. Esto significa que la región crítica a considerar estará en la región extrema derecha (cola) bajo la curva. Esto es debido a que solo tomamos aquellas diferencias en las que las frecuencias observadas en el corpus CORD-19 sea estadísticamente superior a las observadas en el ámbito general. Una prueba de cola derecha se realiza cuando la hipótesis alternativa ( $H_1$ ) contiene la condición  $H_1 > x$  (mayor que una cantidad determinada).

### 3.2. El corpus CORD-19

El COVID-19 Open Research Dataset (CORD-19)<sup>3</sup> es un repositorio en constante actualización con artículos científicos y preprints sobre el COVID-19, y toda la investigación relacionada con otros coronavirus como el SARS y el MERS (Wang et al. 2020). Este corpus ha sido diseñado para facilitar todo tipo de estudios entre los que se encuentra el desarrollo de sistemas capaces de extraer información relevante de los mismos como la minería de textos y la recuperación de información. Los artículos de CORD-19 proceden de PubMed Central (PMC), PubMed, la base de datos COVID-19 de la Organización Mundial de la Salud y los servidores de bioRxiv, medRxiv y arXiv.

De este corpus se descargaron aleatoriamente 40.000 artículos que contenían 102.716.238 palabras. Este corpus fue tokenizado, etiquetado con partes de la oración y lematizado con TreeTagger. El resultado de este etiquetador contiene palabras asignadas a partes de la oración (por ejemplo, verbos, nombres, adjetivos) y lematizadas a sus formas canónicas (por ejemplo, formas adjetivas singulares y masculinas, formas verbales de infinitivo). Las etiquetas de TreeTagger se simplificaron a las observadas en FrameNet, de modo que compartiesen el mismo tipo de anotaciones. La frecuencia de los lemas se obtuvo a partir de este corpus.

Para disponer de una lista fiable de ocurrencias de palabras que representaran el ámbito de general, se optó por utilizar el corpus COCA (Davies, 2019). El Corpus of Contemporary American English (COCA) es uno de los mayores corpus de inglés de libre acceso. Contiene más de 560 millones de palabras (20 millones de palabras por cada año del periodo 1990-2017) y está dividido por igual entre textos hablados, de ficción, revistas populares, periódicos y textos académicos. Las unidades han sido lematizadas y etiquetadas con su clase de palabra (POS-tagged). El corpus COCA contiene unos 60.000 lemas. La Tabla 3 muestra el aspecto de este corpus:

3 <https://www.kaggle.com/datasets/allen-institute-for-ai/CORD-19-research-challenge>.

rank	lemma/word	POS	disp	totFreq
25083	piglet	n	0,88	239
25088	woodsman	n	0,70	300
25090	candied	j	0,87	242
25093	metacognitive	j	0,69	306
25107	industry-wide	j	0,89	236
25108	health-food	j	0,85	246
25110	posterior	n	0,88	240

Tabla 3. Ejemplo de unidades en el corpus COCA. Fuente: COCA corpus.

### 3.3. Selección de las unidades léxicas de FrameNet como posibles unidades terminológicas

Como afirma Witschel (2005), los enfoques automáticos para la extracción de terminología se basan en las frecuencias de las unidades del corpus. Las palabras que se repiten con más frecuencia en un documento son buenos descriptores de su contenido, siempre y cuando estos elementos no aparezcan en muchos otros documentos (como “el”, “sobre” o “cree”). La medida *tf-idf* se basa en este supuesto. Se trata de uno de los enfoques más comunes para la extracción automática de unidades terminológicas o palabras clave a partir del análisis de corpus (Lossio-Ventura, Jonquet, Roche y Teisseire, 2014). En lugar de representar a las diferentes unidades del corpus por su frecuencia absoluta o relativa, con *tf-idf* cada elemento se pondera dividiendo su frecuencia de aparición por el número de documentos del corpus que contienen tal unidad. Las unidades léxicas o lemas con una alta clasificación *tf-idf* se calculan de la siguiente manera:

$$W_{x,y} = tf_{x,y} * \log \frac{N}{df_x}$$

$tf_{x,y}$  = frecuencia de  $x$  en  $y$

$df_x$  = número de documentos o partes que contienen  $x$

$N$  = número total de documentos o partes del corpus

El resultado de esta operación es un valor que indica el grado de relevancia de la unidad analizada en ese corpus. A partir del CORD-19 lematizado y etiquetado con POS, se extrajo la lista de términos *tf-idf* ordenados según su puntuación:

Unidad	Valor
cell.n	0,037
virus.n	0,027
use.v	0,025
protein.n	0,021
infection.n	0,019
study.n	0,017

Unidad	Valor
figure.n	0,015
show.v	0,013
patient.n	0,012
gene.n	0,011

Tabla 4. Primeras 10 unidades en el ranking de *tf-idf*.

Una vez que se obtuvo esta lista se cruzó con las unidades léxicas presentes en FrameNet. Partimos de la hipótesis de que la selección de unidades terminológicas que realiza este último algoritmo debería coincidir con los principales elementos que encontramos en los marcos más representativos del CORD-19. La sección de resultados presenta los resultados de todo el proceso metodológico: selección automática de marcos, extracción de las unidades léxicas de cada uno de ellos y cruce de tales unidades con las procedentes de la selección de la herramienta *tf-idf*. Los resultados muestran hasta qué punto la selección de marcos y la lista *tf-idf* coinciden, así como el nivel de precisión al asignar estos términos a un determinado marco.

## 4. RESULTADOS

### 4.1. Selección de marcos del CORD-19

Para la selección de los marcos más representativos, el sistema itera a través de todos los marcos de FrameNet tomados uno por uno, y calcula la prueba de rango con signo de Wilcoxon utilizando las frecuencias relativas de sus unidades léxicas en el corpus CORD-19 y COCA. Sin embargo, en este análisis no se incluyeron todas las unidades léxicas de cada marco. Si buscamos en FrameNet, hay unidades que pertenecen a verbos frasales como ‘break up’, ‘look for’, etc., expresiones idiomáticas ‘hold (one’s) tongue’, ‘give thought’, etc. o compuestos ‘family room’, ‘screw-up’, etc. Como ya se indicó anteriormente, el análisis de frecuencias se ha realizado a partir del corpus COCA y este tipo de combinaciones no se incluyen; sólo se recogen elementos léxicos simples. Por eso se decidió descartar estas unidades para calcular la significación estadística del grupo, ya que, en caso contrario, se les asignaría la frecuencia “cero”. Este mismo procedimiento se aplicó en el caso de que la unidad léxica de un marco determinado no tuviera representación ni en el corpus general ni en el médico. Igualmente fueron descartados los marcos con un solo elemento (no hay mediana posible). De los 1.222 marcos de FrameNet, 149 no tienen ninguna unidad y 53 sólo tenían un elemento. Por último, realizamos la prueba de hipótesis con signo de Wilcoxon para comprobar si la media de la muestra procedente de CORD-19 es superior a la obtenida con los mismos elementos en el corpus general. Los resultados a un nivel de significación del 95% y del 99% son los siguientes:

95%	99%
<ul style="list-style-type: none"> <li>• Biological mechanisms</li> <li>• Building subparts</li> <li>• Condition symptom relation</li> <li>• Containers</li> <li>• Control</li> <li>• Dispersal</li> <li>• Distinctiveness</li> <li>• Diversity</li> <li>• Extreme value</li> <li>• Health response</li> <li>• Identity</li> <li>• Measure mass</li> <li>• Measure volume</li> <li>• Medical conditions</li> <li>• Medical specialties</li> <li>• Ordinal numbers</li> <li>• Origin</li> <li>• Progression</li> <li>• Quantity</li> <li>• Similarity</li> <li>• Undergo transformation</li> </ul>	<ul style="list-style-type: none"> <li>• Condition symptom relation</li> <li>• Dispersal</li> <li>• Diversity</li> <li>• Health response</li> <li>• Medical conditions</li> <li>• Medical specialties</li> <li>• Ordinal numbers</li> <li>• Origin</li> <li>• Similarity</li> </ul>

Figura 5. Resultados de la selección de marcos al 95% y 99% de nivel de significatividad.

El número de marcos seleccionados difiere entre el nivel de significación del 95% y el del 99%. Este último es más restrictivo. Pasa de una lista de 21 elementos a sólo 9 situaciones de FrameNet. Como se puede observar, en general los resultados son coherentes respecto a la temática del CORD-19.

#### 4.2. Integración de elementos de Tf-idf y FrameNet

Este análisis parte exclusivamente de los verbos y sustantivos del corpus CORD-19. A partir de esta selección analizamos cuántas de estas unidades de FrameNet coincidían con el ranking obtenido de *tf-idf*. Como se ha señalado anteriormente, utilizamos la cobertura de FrameNet propuesta por el Crespo (2021) con 25.305 términos. La siguiente figura muestra la proporción de la lista *tf-idf* encontrada en todo FrameNet desde las primeras 100 unidades hasta las 15.000:

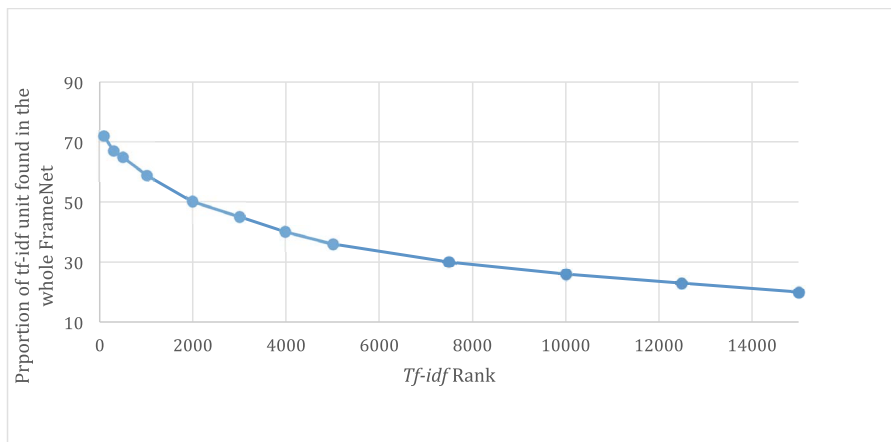


Figura 6. Proporción de coincidencia entre FrameNet y la lista *tf-idf* desde sus 100 primeros elementos hasta 15.000.

A medida que aumentamos la lista, la proporción de coincidencia entre las unidades léxicas de FrameNet y el *tf-idf* disminuye. Debemos recordar que el ámbito de aplicación de FrameNet es general, por lo que las unidades del CORD-19 coincidirán parcialmente con las de FrameNet. Para la experimentación, sólo nos centramos en los primeros 100, 300, 500 y 1000 sustantivos y verbos del ranking *tf-idf*, ya que superan el 50% de coincidencias de acuerdo con la Tabla 5. La proporción de coincidencia para tales valores son las siguientes:

100 tf-idf	300 tf-idf	500 tf-idf	1000 tf-idf
72 de 100: 72%	201 de 300: 67%	326 de 500: 65%	589 de 1000: 59%

Tabla 5. Proporción de coincidencia entre FrameNet y *tf-idf* para los primeros 100, 300, 500 y 1000.

Para los marcos seleccionados, enumeramos sus unidades léxicas. Si una de estas unidades coincidía con la lista *tf-idf* en alguno de sus grados (100, 300, 500 y 1000), se enumeraba. Las coincidencias fueron revisadas manualmente para determinar si los elementos asociados eran coherentes con los marcos en los que aparecían. Los resultados de la selección de marcos a un nivel de significación del 95% y del 99% son los siguientes:

Marcos al 95%			
100 tf-idf	300 tf-idf	500 tf-idf	1000 tf-idf
Número de unidades de nuestra selección de FrameNet que aparecen en la lista <i>tf-idf</i>			
27 de 72: 38%	54 de 201: 27%	68 de 326: 21%	104 de 589: 18%
Número de marcos con alguna de sus unidades en la lista <i>tf-idf</i>			
52%	67%	76%	90%
11 de 21	15 de 21	16 de 21	19 e 21

Marcos al 95%			
Precisión del cruce FrameNet-tf-idf			
85.1% correcto - 23 de 27 posibles-	93% correcto - 50 de 54 posibles-	93% correcto - 63 de 68 posibles-	94% correcto - 98 de 104 posibles-

Tabla 6. Resultados de correcta asociación entre FrameNet y *tf-idf* al 95%.

Marcos al 99%			
100 tf-idf	300 tf-idf	500 tf-idf	1000 tf-idf
Número de unidades de nuestra selección de FrameNet que aparecen en la lista <i>tf-idf</i>			
18 de 72 posibles: 25%	38 de 201 posibles: 19%	46 de 326 posibles: 14%	70 de 589 posibles: 12%
Número de marcos con alguna de sus unidades en la lista <i>tf-idf</i>			
44.4% 4 de 9	66.6% 6 de 9	77.7% 7 de 9	89% 8 de 9
Precisión del cruce FrameNet-tf-idf			
100% correcto - 18 de 18 posibles -	100% correcto - 38 de 38 posibles -	96% correcto - 44 de 46 posibles -	94% correcto - 66 de 70 posibles -

Tabla 7. Resultados de correcta asociación entre FrameNet y *tf-idf* al 99%.

Como se observa, la proporción de términos de FrameNet asignados disminuye a medida que aumenta la lista de *tf-idf*. Los mejores resultados proporcionales se encuentran al 95% y al 100 *tf-idf*. En todos los casos, los marcos con términos en *tf-idf* aumentan a medida que se incrementa esta lista. Los mejores resultados de precisión se encuentran en 100 y 300 *tf-idf* y en el 99%. En estos casos el 100% de las unidades terminológicas encontradas son correctas.

## 5. DISCUSIÓN

Observamos en los resultados algunos problemas señalados por L'Homme, Subirats Rüggeberg y Robichaud (2016):

- Muchos elementos registrados como términos en la lista *tf-idf* no aparecen en FrameNet. A medida que aumentamos esta lista de 100 a 1000, la proporción de términos que aparecen en FrameNet disminuye. Podemos atribuir esto a la falta de unidades léxicas en FrameNet y al hecho de que la lista *tf-idf* también se basa en un corpus que recoge la investigación en curso sobre COVID-19, por lo que contiene mucha terminología nueva.
- Por otra parte, nos encontramos con que un elemento léxico puede estar registrado en FrameNet, pero el significado contabilizado no es el requerido. Este es el caso del marco seleccionado *Building\_subparts* y de las unidades léxicas 'room.n', 'study.n', 'cell.n' y 'level.n' que refieren a algo diferente a lo expresado por el marco. Se ha tomado como un error.
- Es necesario crear nuevos marcos para dar cuenta de la información del corpus. Este es el caso de la unidad 'cell.n' que solo se encuentra en el marco *Building\_subparts*. Debería incluirse un nuevo marco que la represente.



Como se ha visto, la mejor coincidencia entre las unidades léxicas de FrameNet y la lista *tf-idf* es del 38%, pero en la mayoría de los casos la vinculación entre ambas ronda el 20-25%. Esto se explica porque FrameNet incluye términos generales y un corpus científico incluye unidades más específicas.

Podemos intentar paliar esta carencia añadiendo nuevos términos no considerados anteriormente. Como señala Pérez Hernández (2002), los términos científicos se relacionan entre sí mediante distintas relaciones conceptuales. Entre ellas están la inclusión (“hiperonimia-hiponimia”), la equivalencia (“sinonimia”), la semiequivalencia (“parasinonimia”) o la oposición conceptual (“antonimia”) (Casas Gómez 2006). A continuación, podríamos ampliar la cobertura de FrameNet añadiendo la gama de relaciones conceptuales descritas en WordNet. Entre ellas se encuentran (Vossen 1998; 2002): ‘*hypernym*’, ‘*hyponym*’, ‘*holonym*’, ‘*meronym*’, ‘*near\_synonym*’, ‘*near\_antonym*’ or ‘*has\_derived*’. Sin embargo, WordNet está estructurada jerárquicamente de elementos generales a más específicos, en una relación de tipo is-a, por lo que la hiperonimia y la hiponimia será la relación más numerosa. De estas dos interesan las segundas tal como indica Carrió Pastor (2010) sobre el lenguaje de especialidad, ya que las primeras son elementos más generales.

De esta manera ampliamos los marcos seleccionados con unidades más específicas de acuerdo con el procedimiento que se muestra en la siguiente figura:

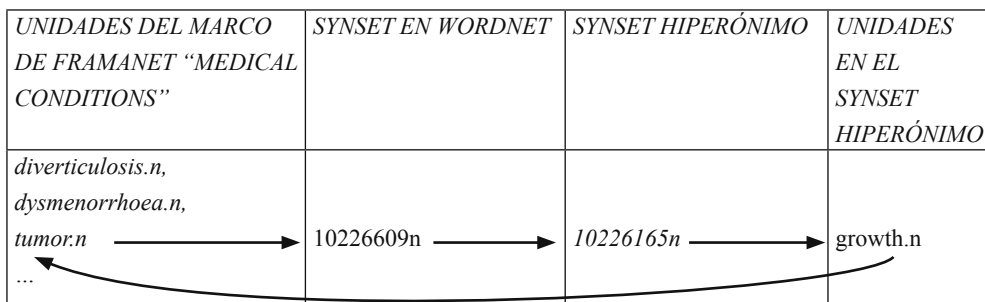


Figura 7. Método de adición de nuevos elementos desde WordNet

Utilizamos el enlace FrameNet-WordNet propuesto por el Crespo (2021) que desarrolla un algoritmo para de alinear las diferentes unidades del marco con los synsets de WordNet que mejor los representan, llegando a una precisión del 88%. Los resultados tras la adición de hipónimos son los siguientes:

Marcos al 95%			
100 tf-idf	300 tf-idf	500 tf-idf	1000 tf-idf
Número de unidades de nuestra selección de FrameNet que aparecen en la lista <i>tf-idf</i>			
49 de 72 posibles: 68%	127 de 201 posibles: 63%	185 de 326 posibles: 57%	299 de 589 posibles: 51%
Número de marcos con alguna de sus unidades en la lista <i>tf-idf</i>			
67% 14 de 21	86% 18 de 21	86% 18 de 21	95% 20 de 21

Marcos al 95%			
Precisión del cruce FrameNet-tf-idf			
92% correcto - 58 de 73 -	88% correct - 140 de 159-	88% correcto - 204 de 231-	85% correcto - 255 de 299-

Tabla 8. Resultados de correcta asociación entre FrameNet y *tf-idf* al 95%.

Marcos al 99%			
100 tf-idf	300 tf-idf	500 tf-idf	1000 tf-idf
Número de unidades de nuestra selección de FrameNet que aparecen en la lista <i>tf-idf</i>			
23 de 72 posibles: 31.9%	61 de 201 posibles: 30%	83 de 326 posibles: 25.4%	133 de 589 posibles: 23%
Número de marcos con alguna de sus unidades en la lista <i>tf-idf</i>			
67% 6 de 9	89% 8 de 9	89% 8 de 9	89% 8 de 9
Precisión del cruce FrameNet-tf-idf			
96% correcto - 26 de 27 -	93% correcto - 62 de 67-	94% correcto - 87 de 93-	90% correcto - 120 de 133-

Tabla 9. Resultados de correcta asociación entre FrameNet y *tf-idf* al 99%.

Observamos cómo los resultados mejoran al incluir hipónimos. Para la selección de marcos al 95% de significación, el número de términos de FrameNet encontrados en la lista *tf-idf* duplica los resultados anteriores y, en algunos casos, se triplica. El número de marcos con representación en la lista *tf-idf* aumentó en todos los casos. El nivel de precisión disminuye ligeramente, pero sigue siendo muy alto.

A pesar de que el nivel de significación del 99% es más restrictivo, también mejora los resultados. La coincidencia de FrameNet y la lista *tf-idf* duplica los resultados anteriores, pero sólo para los primeros 500 y 1000. En consecuencia, el número de marcos seleccionados con representación en la lista *tf-idf* también aumenta. La precisión de los términos sólo disminuye en torno al 5%. Esta reducción se explica en parte por el hecho de que el enlace FrameNet-WordNet en el Crespo (2021) presenta una precisión del 88%, por lo que deben haberse incluido algunas coincidencias erróneas.

## 6. CONCLUSIONES

Hemos propuesto una metodología basada en corpus para gestionar la terminología. Este enfoque permite reutilizar la información proporcionada por FrameNet para estructurar la información terminológica de un corpus técnico. Aparte de su terminología específica, un campo científico hace uso de las mismas unidades léxicas y patrones sintácticos que el lenguaje general. Este enfoque permite reutilizar el marco FrameNet para representar la terminología presentada en un corpus y determinar cómo se relacionan los términos conceptualmente. Entre sus ventajas se encuentra el hecho de tratarse de una metodología ‘top-down’ que define, en primer lugar, las unidades de conocimiento (o marcos), y a partir de ahí enumera qué elementos son los que se asocian a esos marcos conceptuales. Los resultados muestran su utilidad para apoyar el trabajo terminográfico.

Para la selección de marcos representativos, este estudio parte de la base de que los marcos representativos pueden determinarse observando las frecuencias de las unidades que aparecen en nuestro corpus CORD-19 frente a las mismas palabras en un dominio general. Nuestro enfoque para la selección de marcos hace uso de la prueba de rangos con signo de Wilcoxon, un método no paramétrico, que no requiere una suposición de normalidad para la distribución de los datos. Para las primeras 1000 unidades tf-idf, el sistema es capaz de asignar 299 términos a sus marcos correctos con una precisión del 85%.

Observamos que este método ofrece buenos resultados en general, por lo que puede aplicarse para la determinación de los marcos que más representan un dominio científico determinado. Sin embargo, FrameNet está orientado a dominios generales, por lo que no cubre todos los usos científicos. Además, al estar aún en proceso de creación, la lista de activadores de marcos aún no está completa. Intentamos resolver la escasez de cobertura léxica de FrameNet enriqueciéndola con nuevos términos de WordNet. Esta base de datos léxica organiza el vocabulario de diferentes idiomas según conceptos y relaciones semánticas. Estos grupos, llamados synsets, tratan de reflejar la disponibilidad léxica para un determinado concepto. Para la adición de términos se han utilizado las relaciones de hipónimos. Los resultados mejoran notablemente con la inclusión de estos nuevos términos, lo que indica que los lenguajes específicos hacen uso de unidades léxicas más específicas.

Hemos observado igualmente que muchos términos no se incluyeron o el sistema los asignó a un marco equivocado. Esto significa que es necesario construir nuevos marcos e identificar las unidades léxicas a las que se asocian. El trabajo futuro en la terminología basada en FrameNet incluye la ampliación del número de marcos para dominios específicos y la mejora de las conexiones entre WordNet y FrameNet. Pensamos que es interesante pasar a otros campos especializados diferentes y observar cómo trabaja este modelo. Esto nos permitiría establecer equivalencias y diferencias entre diferentes dominios desde el punto de vista de la semántica de los marcos.

Por último, la conexión entre FrameNet y los términos de un corpus nos permite el desarrollo de diferentes aplicaciones semánticas que incluyen la recuperación de información multilingüe, el resumen automático, la extracción de información, la traducción automática, la clasificación de documentos multilingües, los sistemas pregunta-respuesta, el procesamiento de diálogo o la organización del conocimiento.

## AGRADECIMIENTOS

Esta investigación se enmarca dentro del proyecto *Lingüística y Humanidades Digitales: base de datos relacional de documentación lingüística* (PY18-FR-2511) Entidad financiadora: Convocatoria 2018 de Ayudas a proyectos I+D+i (Modalidad «Frontera Consolidado») del Plan Andaluz de Investigación, Desarrollo e Innovación Duración del proyecto: 01/01/2020 -31/03/2023. Cuantía de la subvención: 71.800 €. Investigador responsable: Miguel Casas Gómez.

## REFERENCIAS BIBLIOGRÁFICAS

- Azoulay, D. (2017). Frame-based knowledge representation using large specialized corpora. En L. Steels y J. Feldman (Eds.), *2017 AAAI Spring Symposium on computational construction grammar and natural language understanding* (pp. 119-126). Palo Alto, California: AAAI Press. <https://www.aaai.org/ocs/index.php/SSS/SSS17/paper/view/15324>
- Baker, C., C. J. Fillmore y J.B. Lowe. (1998). The Berkeley FrameNet project. En C. Boitet y P. Whitelock (Eds.), *Proceedings of the Thirty-Sixth Annual Meeting of the Association for Computational Linguistics and Seventeenth International Conference on Computational* (pp. 86-90). San Francisco, California: Morgan Kaufmann Publishers.
- Cabré, M. T. (2005). La Terminología, una disciplina en evolución: pasado, presente y algunos elementos de futuro. *Debate Terminológico*, 1. <http://riterm.net/revista/ojs/index.php/debateterminologico/article/view/23/45>
- Carrió Pastor, M. L. (2010). La variación del lenguaje de especialidad en artículos científicos. *Pragmalingüística*, 15-16, 71-83. <https://doi.org/10.25267/Pragmalinguistica.2017.i25>
- Casas Gómez, M. (2006). Modelos representativos de documentación terminográfica y su aplicación a la terminología lingüística. *Revista de Lingüística y Lenguas Aplicadas*, 1(1), 25-36.
- Casas Gómez, M. (2014). A Typology of Relationships in Semantics. *Quaderni di semantica: Rivista Internazionale di Semantica Teorica e Applicata*, 35 (2), 45-74.
- Crespo, M. (2020a). *Automatic Corpus-based translation of a Spanish FrameNet medical Glossary*. Colección Lingüística. Sevilla: Universidad de Sevilla.
- Crespo, M. (2020b). Lingüística digital: revisión de su estado actual y retos en el Instituto Universitario de Investigación en Lingüística Aplicada de la Universidad de Cádiz. *Pragmalingüística*, 28, 148-165. <https://doi.org/10.25267/Pragmalinguistica.2020.i28.08>
- Crespo, M. (2021). Aproximación al trasvase automático de predicados de Framenet al español mediante Wordnet. *Revista de Lingüística y Lenguas Aplicadas*, 16, 49-62. <https://doi.org/10.4995/rlyla.2021.14408>
- Cristea, D. y I. C. Pistol. (2012). Multilingual linguistic workflows. *Multilingual Processing in Eastern and Southern EU Languages. Low-resourced Technologies and Translation* (pp. 228-246). Cambridge, Reino Unido: Cambridge Scholars Publishing.
- Davies, M. (2019). *The Corpus of Contemporary American English (COCA): 560 million words, 1990-present*. Disponible online en <https://www.english-corpora.org/coca/>
- Dolbey, A., M. Ellsworth y J. Scheffczyk. (2006). BioFrameNet: A Domain-specific FrameNet Extension with Links to Biomedical Ontologies. En O. Bodenreider (ed.), *Proceedings of the "Biomedical Ontology in Action" Workshop at KR-MED* (pp. 87-94). Baltimore, Maryland: National Library of Medicine.
- Durán-Muñoz, I. (2016). Producing frame-based definitions: A case study. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication*, 22(2), 23-249. <https://doi.org/10.1075/term.22.2.04mun>
- Fillmore, C. J. (1977). Scenes and Frames Semantics. En A. Zampolli (Ed.), *Linguistic Structures Processing* (pp. 55-82). Amsterdam: North Holland.
- Fillmore, C. J. (1985). Frames and the semantics of understanding. *Quaderni di Semantica*, 6(2): 222-254.
- Fillmore, C. J. y C. F. Baker. (2010). A Frames Approach to Semantic Analysis. En B. Heine y H. Narrog (Eds.), *The Oxford Handbook of Linguistic Analysis* (pp. 313-339). Oxford: Oxford University Press.
- Gildea, D y D. Jurafsky. (2002). Automatic Labelling of Semantic Roles. *Computational Linguistics*, 28, 245-288. <https://doi.org/10.1162/089120102760275983>
- Guerrero Ramos, G. y M. F. Pérez Lagos. (2003). Lexicografía, terminología y diccionario. En E. Ortega Arjonilla, A. B. Martínez López y E. Echeverría Pereda (Eds.), *Panorama actual de la investigación en traducción e interpretación* (pp. 541-563). Granada: Atrio.

- Johnson, Christopher y C. J. Fillmore. (2000). The FrameNet tagset for frame-semantic and syntactic coding of predicate-argument structure. En J. Wiebe (Ed.), *Proceedings of the 1st Meeting of the North American Chapter of the Association for Computational Linguistics (ANLP-NAACL 2000)* (pp. 56–62). Seattle WA: ACM.
- L’Homme, M. C. (2018). Maintaining the balance between knowledge and the lexicon in terminology: a methodology based on Frame Semantics. *Lexicography*, 4(1), 3-21. <https://doi.org/10.1007/s40607-018-0034-1>
- L’Homme, M. C., B. Robichaud y C. Subirats Rüggeberg. (2020). Building Multilingual Specialized Resources Based on FrameNet: Application to the Field of the Environment. En T. Timponi Torrent, C. F. Baker, O. Czulo, K. Ohara, y M. R. L. Petruck (Eds.), *Proceedings of the International FrameNet Workshop 2020: Towards a Global, Multilingual FrameNet* (pp. 85-92). Marseille: ELRA.
- L’Homme, M. C., C. Subirats Rüggeberg y B. Robichaud. (2016). A Proposal for combining ‘general’ and specialized frames. En M. Zock, A. Lenci y S. Evert (Eds.), *Proceedings of the 5th Workshop on Cognitive Aspects of the Lexicon (CogALex-V)* (pp. 156-165). Osaka, Japan: ACL.
- Lossier- Ventura, J. A., C. Jonquet, M. Roche y M. Teisseire. (2014). Towards a mixed approach to extract biomedical terms from text corpus. *International Journal of Knowledge Discovery in Bioinformatics (IJKDB)*, 4(1), 1-15. <https://doi.org/10.4018/ijkdb.2014010101>
- Malm, P., V. Mumtaz, B. Shafqat, Lars y A. Saxena. (2018). LingFN: Towards a framenet for the linguistics domain. En T. Timponi Torrent, L. Borin y C. F. Baker, *11th edition of the Language Resources and Evaluation Conference* (pp. 37-43). Paris, France: ELRA.
- Miller, G. A., R. Beckwith, C. Fellbaum, D. Gross y K. Miller (Eds.). (1993). *Five Papers on WordNet, cls report 43. Technical report*. New Jersey: Cognitive Science Laboratory. Princeton University.
- Pérez Hernández, C. (2002). Explotación de los corpórea textuales informatizados para la creación de bases de datos terminológicas basadas en el conocimiento. *Estudios de lingüística del español*, 18. Disponible online en <http://elies.rediris.es/elies18/>
- Potęga, P. (2017). Frame Based Modelling of Specialist Languages. *Studia Anglica Resoviensia 14*: 121–31. <http://dx.doi.org/10.15584/sar.2017.14.10>.
- Ramírez Salado, M. (2019). *Terminología y lingüística forense: usos terminológicos relacionados con los ámbitos de actuación de la lingüística forense y su interfaz con otras disciplinas*. (Tesis doctoral, Universidad de Cádiz).
- Ruppenhofer, J., M. Ellsworth, M. R. L. Petruck, C. Johnson y J. Scheffczyk. (2016). *FrameNet II: Extended Theory and Practice*. <https://framenet2.icsi.berkeley.edu/docs/r1.7/book.pdf>
- Triola, M. (2007). *Elementary statistics*. 10th ed. Boston: Addison-Wesley.
- Venturi, G. (2013). A semantic annotation of Italian legal texts. A FrameNet-based approach. En M. Fried y K. Nikiforidou (Eds.), *Advances in Frame Semantics* (pp. 51-84). Amsterdam / Philadelphia: John Benjamins Publishing Company. <https://doi.org/10.1075/bct.58.02ven>
- Verdaguer, I. (2020). Semantic frames and semantic networks in the Health Science Corpus. *Estudios de lingüística del español*, (Anejo 1), 117-155.
- Vossen, P. (Ed.). (1998). *EuroWordNet: A Multilingual Database with Lexical Semantic Networks*. Dordrecht: Kluwer Academic Publishers. <https://doi.org/10.1007/978-94-017-1491-4>
- Vossen, P. (Ed.). (2002). *EuroWordNet: general document*. URL: <http://vossen.info/docs/2002/EWNGeneral.pdf>
- Wang, L. L., k. Lo, Y. Chandrasekhar, R. Reas, J. Yang, D. Burdick y S. Kohlmeier. (2020). *CORD-19: The Covid-19 Open Research Dataset*. ArXiv. <https://aclanthology.org/2020.nlp-covid19-acl.1.pdf>
- Witschel, H. F. (2005). Terminology extraction and automatic indexing - comparison and qualitative evaluation of methods. En B. Nistrup Madsen y H. Erdman Thomsen (Eds.), *Proceedings of 7th International Conference on Terminology and Knowledge Engineering*. (pp. 363-374). Copenhagen: Association for Terminology and Knowledge Transfer.
- Yarowsky, D. (1995). Unsupervised word sense disambiguation rivaling supervised methods. En H. Uszkoreit, *Proceedings of the 33rd annual meeting on Association for Computational Linguistics* (pp. 189-196). Cambridge, Massachusetts: ACL.

# On empathy in online tourism experience. High-frequency adjective patterns in English and Spanish promotional language

La empatía en la experiencia turística online. Adjetivos y patrones de uso más frecuentes del lenguaje promocional en inglés y en español

EVA MARÍA LLORENS SIMÓN  
Universidad de Alicante, España  
[eva.llorens@ua.es](mailto:eva.llorens@ua.es)  
<https://orcid.org/0000-0002-0872-4789>

## Abstract

This paper focuses on the identification and analysis of recurring adjective patterns in Spanish and English tourism promotional content available online. The study discusses how prototypical utterances may trigger empathy in each language, with high-frequency adjectives and emotional relevance being key in the addressees' expectations as predictors of genuineness. This research offers a corpus-based approach to linguistic tendencies and presents empathy as a step beyond persuasion in translated or copywritten materials. To that end, a bilingual, *ad hoc* corpus recently compiled from institutional websites was examined in order to collect significant samples of persuasive elements in online tourism discourse. The results obtained show both common and contrastive aspects in the two languages, with empathy being preferably boosted under the assumption that positiveness and a specific proportion of frequent adjectives

## Resumen

El presente trabajo tiene por objeto la identificación y el análisis de patrones con adjetivos recurrentes en contenidos online sobre promoción turística en inglés y en español. En particular, el estudio plantea el potencial que presentan los segmentos prototípicos de cada lengua como disparadores de la empatía, a la vez que se centra en la frecuencia y la relevancia emocional de los adjetivos más habituales por su capacidad predictiva respecto a las expectativas de los destinatarios y el carácter genuino del lenguaje. Así mismo, esta investigación ofrece un enfoque basado en el análisis de corpus sobre tendencias lingüísticas en las que la empatía se concibe como un estadio adicional a la persuasión susceptible de debate en materiales traducidos o géneros publicitarios. A este respecto, se ha partido de un corpus bilingüe *ad hoc* elaborado a partir de páginas web institucionales con idea de reunir muestras significativas sobre

---

**To cite this article:** Llorens Simón, E. M. (2022). On empathy in online tourism experience. High-frequency adjective patterns in English and Spanish promotional language. *ELUA*, (38), 301-323. <https://doi.org/10.14198/ELUA.22381>

Recibido: 31/03/2022 Aceptado: 16/05/2022

© 2022 Eva María Llorens Simón



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

yield desirable effects among speakers of Spanish and English.

**KEYWORDS:** tourism promotional language, linguistic empathy triggers, adjective positiveness, specialised translation, translation in tourism.

tales elementos persuasivos en el discurso turístico online. Como resultado, se ha concluido que existen puntos en común y aspectos contrastivos entre ambas lenguas que ocupan una posición preferente como desencadenantes de la empatía al asumir que la carga positiva en determinadas proporciones de adjetivos frecuentes produce efectos deseables entre los hablantes de cada idioma.

**PALABRAS CLAVE:** lenguaje turístico promocional, disparadores lingüísticos de la empatía, adjetivos con carga positiva, traducción especializada, traducción en turismo.

## 1. INTRODUCTION

In recent decades, most companies seem to have confined their promotional efforts to publishing content with the simple aim of having an online presence or instantly offering a corporate image with a massive impact. The priority for many publications seemed to be SEO (Search Engine Optimization) or the automatic repetition of keywords to achieve top positions on Google Search and other web engines rather than achieving a notable quality or persuasive effect through their commercial messages. This entailed a sort of overgeneralisation (Coupland 2003) and a sometimes vague benefit for companies. On this respect, tourism promotion was no exception given the production of online texts and publications clearly diverging from natural language.

However, in our current and globally developed era of mass communication (Crystal 1997; Mateo Martínez 2007) a new conception appears to be increasingly accepted in terms of optimising website content and achieving real results with the said content, which includes the need for attention to language as a tailor-made tool for real efficiency and strategic communication.

While concepts such as empathy and prototypicality might seem to be easily understood at first sight, they are not always so easily achieved, since the extent to which they are visible, evident, or effective in a language is still little understood. That said, their potential is a key for translators and copywriters.

Research in Applied Linguistics and Corpora Analysis could be considered an ideal starting point to assess language in use and quantify to at least some extent the impact of such cognitive and pragmatic effects (Fogg 2003; Lakoff 1987; Langacker 2008; Wilson & Sperber 2002; Yus Ramos 2006). In this sense, a specific corpus of tourism-based persuasive messages in English and Spanish provides the basis for this study and offers an opportunity to show how the language of tourism is preferentially used.

This research seeks, therefore, to evaluate adjective recurrence in frequent utterances of the corpus to provide some preliminary insight into empathy triggers in English and Spanish. By exploring our specialised corpus on the language of tourism, we try to determine how empathy activation can be achieved with the high-frequency adjectives in these two languages, assuming that empathy triggering is a step beyond persuasion and efficiency.

Moreover, this article is intended to raise awareness of the importance of using prototypical language to meet customers' expectations. In this regard, predictability is linked to implied user knowledge and the probable emotions aroused in readers' minds when exposed to regular discourse. Specifically, our initial hypothesis is that prototypical items in tourism promotional language are better candidates to trigger empathy than random structures or personally biased terms. Furthermore, reproducing the common tendencies and proportions of a specific language is considered to boost empathy as well.

On the contrary, it is assumed that shocking or unnatural expressions are likely to produce distrust and distance, so identifying high-frequency items also helps us recognize less-recurring elements and decide how convenient they are in a specific context. In fact, distinguishing something as infrequent, uncommon, or atypical may lead writers and translators to avoid mistakes or eventually create more natural and effective content by choosing more frequent patterns instead. This may help them to become familiar with the most significant tendencies found in corpora as far as the adjective function is concerned.

This article starts with a short introduction about the cognitive effects related to regular discourse and expectations in tourism promotional language. The following section, Empathy and prototypical patterns in language for specific purposes, includes simple and representative clarifications about some key concepts: empathy, recurring patterns, and language for specific purposes (LSP). In these preliminary sections, tautology and positive connotations are also explored, with references to them as possible boosters of or candidates for empathy. The third section, Methodology, provides details on the selection of corpus analysis as the specific method with which recurrent adjectives are identified, listed, and analysed in both languages, focusing particularly on the contrastive points and high-frequency items. The corpus characteristics are visually presented and its features are described to make clear the extent to which the corpus-based study is appropriate for achieving the research goals of this study. In the fourth section, Results and discussion, the findings obtained after examining the corpus are graphically exhibited along with their interpretation. A general discussion is also provided about which items or segments may prompt a positive response in terms of empathy and, indirectly, which ones may disable its effects. Ultimately, in the final section, Conclusions, some general reflections on the presumed empathy-moving patterns found are presented, and ideas for future research projects are suggested.

## **2. EMPATHY AND PROTOTYPICAL PATTERNS IN LANGUAGE FOR SPECIFIC PURPOSES**

Tourism discourse has been traditionally associated with professional and academic languages considering its condition as jargon for a wide group of specialists (Alcaraz 2000-2004; Alcaraz *et al.* 2007). It is also considered that tourism language has a specific organizational structure and covers specific communicative intentions expressed in a formal register for a common goal. This combination of elements is assumed to be different from those normally linked to general language, which is an additional reason to consider tourism language, and particularly promotional materials in this genre, as an LSP (Cabré 1999; Calvi 2016; Mapelli 2008; Pickett 1989).

Considering promotional discourse in tourism as an LSP is particularly important to validate the representativeness of the prototypical patterns and repetitive aspects found in



the corpora (Leech 2007), since frequency in use expressly reflects their significance and genuine character. Frequent terms may indeed be considered as a sort of visible trace that reveals how the specific purpose of language in this area is often preferentially materialised.

On the other hand, empathy activation is regarded as a desirable effect, given the impact it may have on readers in terms of customer loyalty. In this sense, it is conceived as the product of a cognitive process that is assumed to be intentionally fostered by copywriters and translators. This means, therefore, that empathy can be what establishes the distinction between conscious, repetitive consumption of a service and the occasional choice of customers who were not moved by the promotional message from the very beginning. The promotional message may result in no loyalty if the service quality is not above average, but a positive inclination linguistically associated with the persuasive content may be the initial guarantee for the positive features to be remembered and internalized towards further consumption (Cialdini 2001 & 2005). With similar levels of quality, attention, and satisfaction, it is presumed that an effective and empathy-moving message is of key importance to gaining customers' devotion and fidelity. But what exactly is empathy and how is it achieved in tourism language?

### 2.1. The concept of empathy

Empathy is perceived as a sort of positive and desirable effect aroused in readers' minds after they process reasoning or emotions related to the message. Most experts consider empathy to be the result of several cognitive processes including logical reasoning, motivation and feelings of affinity, affective experiences, and emotional responses among others. But not all scholars agree on the degree to which each of these aspects is important, moving, or inspiring to customers or addressees whose desirable reaction is their selection of a specific destination.

For instance, Eisenberg identifies empathy as a predictable reaction that may influence people's behaviours or emotions depending on their mood (Eisenberg 2000). For some experts, empathy is a cognitive skill to be used for the evaluation of and proper response to thoughts and viewpoints that take into consideration the recipient's motivations (Deutsch & Madle 1975; Regan & Totten 1975). Similarly, Hoffman bases his definition of empathy on affective moves produced after assessing the potential emotions the others may have (1982), whereas Gallo (1989) finds in empathy a cognitive effect arising from the appreciation of and capacity to understand someone else's feelings.

Consequently, we may think that the extent to which empathy is achieved or activated may also depend on the speakers' level of motivation, reasoning, or emotional movements. It is, however, certainly true that the measurement of such complex processes does not depend exclusively on linguistic items. What can be assumed and linguistically studied here is, in any case, the set of possible combinations where language items and functions can contribute positively to empathy activation and, therefore, constitute a powerful but not exclusive strategy in tourism promotion.

Copywriters or translators can use empathy triggers to guess recipients' needs or preferences and develop the capacity to predict their reactions even without having any real awareness of the reasons behind them. Indeed, predictability in each language is conceived as a strategic tool with a very significant application in tourism campaigns, but the intentional use of language may also contribute to expected reactions and, therefore, predictable responses.

## 2.2. Cognitive effects: recurrence and positiveness

Cognitive effects are especially relevant in persuasive discourse bearing in mind readers' involvement. In this sense, as far as persuasion is concerned, it is not surprising to find explicit appeals to recipients or clarifying comments about ideal reactions to promotional messages. If we particularly focus on language for tourism promotion, this reader-oriented dimension or interpersonal function (Halliday 1978) is even reinforced by products being marketed as worthwhile experiences or life-long memories among real or potential addressees whose expected reactions are assumed to be preferentially positive (Cialdini 2005; Dann 1996).

As human beings, a positive response to a specific proposal is always more than welcome, so we may consider our cognitive systems also accept this premise logically and automatically. The pursuit of pleasure and the avoidance of pain is nothing new, from Epicurus' hedonism to the most cosmopolitan theories in neuro-linguistic programming (NLP) or self-help books. Consequently, positive connotations are integrated into our minds and language as desirable features and, therefore, as preferential aspects with a sort of emotional sense to be exploited for persuasion. In Dann's words (1996: 65), 'positive connotations or euphoria constitute one of the distinctive aspects of tourism discourse' to support the conative or appellative function.

Given all the previous considerations, positive adjectives are assumed to be of key importance in tourism language. But, obviously, it does not mean the indiscriminate use of these items absolutely and unconditionally guarantees a persuasive effect or, even, the next step, empathy activation. One of the hypotheses tacitly contained in this article points to the fact that the community of speakers and experts of each language strategically and implicitly decides or influences how frequently some adjectives or adjective patterns are used in the area of tourism. What is more, we contend that a detailed analysis is needed to determine the lexical variety of adjective patterns involved and the extent to which lexical repetition can be used for prototypical effects in English and Spanish.

However, the way readers are persuaded through cognitive processes needs to be discussed from a multifaceted viewpoint, since hundreds of simultaneous actions are likely to be activated at the same time. As an example, 'motivation, attention, memory, and audio-visual perception' (Fajardo 2007: 68), among other aspects, are conceived as cognitive precursors for empathy triggers with a special impact on language.

On the other hand, the frequency in key terms is an indicator of speakers' and professionals' preferences as well. Nevertheless, recurrence, or tautology according to Dann (1996), and its acceptance as a strategy, if so, can be regarded as a linguistic, cultural, or stylistic convention. Repetitions may be associated with poor style or, on the contrary, with clarity, precision and, smartness. This study is, therefore, also designed to determine frequency tendencies and conventions for each language (Spanish and English) in tourism as well as the variety or proportion of items that are repeated in the two languages for this genre, since the combination of these two factors in adjective patterns may offer guidance about linguistic conventions and uses.

## 2.3. Adjective functions and effects

On the whole, adjectives are commonly identified with their capacity to combine with nouns and provide extra details or limitations which are not intrinsically implied by the nouns on

their own. Moreover, a starting consideration for this study is that descriptive adjectives (Halliday 1994; Payne *et al.* 2010; Pustet 2006; Quirk *et al.* 1985) are especially relevant for tourism discourse given their connotative sense.

Proper adjectives, —demonyms or place adjectives (e.g. Spanish, English...) (Demonte 1999; Quirk *et al.* 1985) — also constitute an important class for tourism persuasive messages. They are considered to be regular and easily identifiable, but what is evident from a desirable position is their presence will be expected in a destination promotion with empathy triggers to be activated whenever the place under the campaign is somewhat referred to.

Indeed, a basic conception is that adjectives are essentially modifiers (normally in noun phrases), while nouns and verbs are, respectively, the heads of noun phrases and predicates (Bhat 1994). However, the case is that adjectives and adjective patterns are also very often used on their own (unaccompanied and not directly accompanying a noun but referring to it as a subject) as predicates. Adjectives modifying a noun are attributive, whereas those used as predicates are predicative (Quirk *et al.* 1985), making them interesting items to analyse in corpora of tourism language. In this article, adjectives in general have been analysed, no matter whether they are attributive or predicative in our corpus.

Some scholars, such as Englebretson (1997), have used corpora studies to quantify the proportion of attributive or predicative adjectives used in interactions, concluding that what defines adjective choice is not the grammar and norms of the language itself or the medium through which messages are transmitted (communicative channel), but rather the speech act (Gordon & Lakoff 1971; Searle 1985), the amount of shared knowledge among the participants in communication and the level of formality ascribed to the discourse. Considering the thousands of possible contexts that exist for tourism commercial language, we might understand that tourism persuasive content, identified as one-way communication, requires extra efforts to be put in place to establish what knowledge is shared between the recipients in a particular campaign and those who published or ordered the promotional messages (Hyland 2010).

Furthermore, the speech act implications are quite clear in tourism promotion, but the level of formality can be another source of discussion after studying the register and the level of colloquiality established as ideal in English and Spanish. Thus, locating and quantifying the specific high-frequency adjectives with a positive sense, along with the patterns they are integrated into, is a significant strategy for finding language-based tendencies and to thus conduct a qualitative analysis of pragmatical and contextual issues.

#### **2.4. Candidates for empathy: prototypical patterns in tourism discourse**

Some considerations may be needed to resolve in what ways frequency levels and positive charge in adjectives are materialised and interpreted in terms of predictability and preferences for speakers as far as tourism promotional discourse is concerned. Indeed, it may be of great help to find proper candidates for empathy triggering to be used as more appropriate, genuine, and prototypical items by translators, copywriters, and experts in communication.

Under the theoretical framework, it is generally presumed that some segments, combinations, or patterns including adjectives are better candidates to activate empathy in English or Spanish in proportion to how frequently they are used and how visibly the positive connotations are evidenced in each language corpus.

Thus, considering empathy a significant mechanism, sometimes automatised, for tourism promotional purposes, some specific assumptions can be made as a starting point to identify what is supposed to prompt it for confirmation purposes on the corpus. Firstly, we may consider that adjective prototypes positively contribute to empathy, with speakers expecting to find them in tourism discourse as representative elements of this LSP which enable users to identify it as a genre (Llorens Simón 2021; Rosch 1975-1983; Suau Jiménez 2006; Swales 1990).

Additionally, positiveness in recurring adjectives is also assumed to be characteristic and frequent in the language of tourism (Dann 1996: 65; Durán Muñoz 2012: 108-109; Kelly Holmes 2005: 8), and that it makes a special contribution to activating empathy, thus moving a step beyond persuasion. Tourism language being an LSP, as many scholars state, we may also consider that recurrently using positive adjectives yields a natural sense while identifying frequent positive qualities lets us find naturalness. Hence the interest in studying a corpus to identify patterns and understand their underlying cognitive reasoning. In this regard, recurrence shows relevance and users' expectations, which means empathy is probably activated with predictably typical and expected adjectives. For this reason, an English speaker may be shocked or puzzled by an unexpected formula that is probably a direct reproduction of Spanish style (e.g. *a proposal of very wide plans of guides\**) in the same way that a Spanish user may find some expressions quite unexpected and questionable, probably because they were derived from an English collocate (e.g. *esta representación es justo tan vibrante como la anterior\**).

Indeed, if customers have expectations towards the promotional message they read, using common adjectives will probably have a positive impact and a desirable empathy move. On the contrary, a translator or copywriter selecting uncommon elements to use in their text will probably result in mistrust, scepticism, or even disloyalty in terms of its consumption among recipients. Corpus analysis is thus the key to evaluating genuineness, prototypicality, and, therefore, appropriateness.

### 3. METHODOLOGY

Discourse addressees play a very relevant role in the tourism experience, not just because they are supposed to identify the reaction they are expected to show after interpreting a promotional message, but also because their expectations are also essential to designing the persuasive content necessary for marketing campaigns and quality promotional content (Agarwal 2004; García & Saad 2008; Scarpa 2020). This means specialised translation and copywriting should be based on common messages the speakers, whether predisposed or not, identifies with. Furthermore, empathy as a desirable effect adds extra value to the message, which makes the identification of empathy triggers more than welcome.

Indeed, first-sight impressions on how prototypical and empathy-moving some patterns may be cannot get accepted as evidence unless a significant amount of materials has been completely examined for confirmation purposes (Kenning 2010; Laviosa 1997; McEnery 2019; McEnery & Xiao 2007). As a consequence, corpus analysis is considered to be a suitable methodology to go beyond speculations and suppositions and overcome the limitations of manually processed samples. For this research, a comparable, bilingual, *ad hoc* corpus has been compiled online in order to carry out both a quantitative and qualitative analysis

of the tourism promotional language used on institutional websites. The technological tool *Sketch Engine* was chosen for corpus creation and treatment, which comprises the compilation of two subcorpora, the identification of frequently used adjectives, the listing of common expressions with positive, recurring adjectives; and the context displays. The most prominent advantage of using a corpus-based methodology and a computer-aided procedure includes the option of conducting a rigorous study based on empirical data and a significant sample of real language in use.

In summary, we have confirmed that corpus analysis is an appropriate method for recurrent adjectives to be identified, quantified, and examined in context. Specifically, a specialised corpus has been expressly compiled to find frequency levels, conventional expressions in context, and common word combinations as support to evaluate representative sense, or persuasive effect predictably and additionally resulting in empathy activation. After determining these significant and empathy-oriented elements, specialised translation or copywriting decisions can be better explained, while the introduction of the corpus approach is presented as a core strategy to do so (Leech 1991-2007; McEney 2019; Olohan 2004; Parodi 2004).

### 3.1. Corpus dimension

The corpus size is defined by the number of words it is composed of, this aspect being relevant in terms of representativeness. The general data for the corpus and the specific details on each subcorpus are shown in the table below.

	Global corpus	English subcorpus	Spanish subcorpus
Number of words	657,618	328,377	329,241

Table 1. Corpus size data.

This corpus is representative, being medium-sized according to the generally accepted classification (Berber Sardinha 2002; Leech *et al.* 1994; Vargas-Sierra 2012). Specialised corpora are assumed to be relevant enough even when their size is below 1 million words, since their use is limited to the specific LSP they were designed for (Baker 2010; Fuster Márquez & Pennock-Speck 2015; Leech, 2007; Ornia 2015, Parodi 2004; Vargas-Sierra 2006). Additionally, its representativeness is reinforced by the content institutional sense and the significant value of currently online available texts.

### 3.2. Corpus design

Our *ad hoc* corpus integrates both real materials as a reference and recurrence indicators as factual data in order for a contrastive study to be conducted, with technical support provided by the *Sketch Engine* platform. The online software has been used for corpus creation, and content analysis including frequency and context assessment functions. The corpus creation is, therefore, the basis for the research under the criteria below, whose application responds to general considerations for corpus arrangement (Bowker & Pearson, 2002; Vargas Sierra, 2006):

Criteria	Type	Corpus description/component
Language production channel & origin	Written, materials available online	Online content posted on the websites of the 10 most visited cities in the UK and Spain. The content authors are not identified, but what is significant is the institutional sense of the aforesaid websites, which eventually implies a normative function
Corpus size – number of words	Medium-sized (250.000-1m)	657,618 words
Text distribution	Balanced	Web content is divided into a similar number of sections, most of them with similar names/references (Home/About us, Things to do, Attractions, etc.), and normally linked to each other. The number of texts is, therefore, not clearly defined; although the content included in both subcorpora has this original characteristic.
Content	Specialised	Specialised content in the tourism domain.
Size in samples	Complete samples	All the materials available on the above websites are included in the corpus.
Coding/tags	Tagged	The corpus is tagged and the analysis of the basic research aspects is partially centred on the tags (especially those related to adjectives). It includes part-of-speech tagging and lemmatization.
Documented/non-documented	Non-documented	No specific elements, relations, links, or files can be identified with texts; considering web content is offered as a unit rather than as a collection of well-delimited texts.
Period	Recent	All the materials used for the corpus compilation currently appear on their original websites and have recently been used for promotional purposes.
Updating options	Monitor	The content integrated into the corpus can be both increased and updated.
Purpose	Specialised, <i>ad hoc</i>	Promotional purposes in a specialised area. Tourism promotion is the key purpose, assuming the promotional initiative was officially and institutionally started. The expository-argumentative genre is predominant, but not exclusive in this domain.
Selection of the samples	Proportional sampling	All the materials available on the aforesaid websites are mass compiled and can be proportionally analysed.
Languages	Bilingual, comparable	Spanish & English. The corpus comparable sense is especially important for a contrastive study to be applied in specialised translation, copywriting, or communicative tasks.

Table 2. Corpus design criteria.

Given that tourism discourse is conceived as an LSP, a comparable, bilingual corpus was created with a subcorpus in English and a subcorpus in Spanish designed in order to analyse language in tourism contrastively. In this regard, the comparable trait involves both subcorpora having similar traits as regards size, genre, source, period, and domain (Kenning 2010; Laviosa 1997; McEnery & Xiao 2007), which empirically validates their comparison for a contrastive study.

On the other hand, the features attributed to the corpus, to the extent they are described before, are also relevant to conceive it as a representative source for the specialised language and purpose indicated above. Thus, studying how the promotional language of tourism behaves online is assumed to be particularly accurate with the arrangement of the corpus under such criteria.

### 3.3. Corpus analysis

Different functions and features are offered by *Sketch Engine* to base a corpus analysis on, namely Word Sketch, Word Sketch Difference, Thesaurus, Concordance, Collocation, Word lists, N-grams, Keywords/Terminology, Trends, and Parallel corpus. However, not all of them are activated for a particular type of corpus. For example, the Parallel analysis tab is hidden in a comparable corpus display. In this particular case, examining the corpus has been possible by using three main functions available on the software: Word List, Concordance, and Word Sketch. As a result, data processing has been arranged in three steps. With the Word List tab, adjectives as a category can be found, so a list of high-frequency adjectives has been obtained along with recurrence indicators in each subcorpus. This is of great help to assess both the frequency levels and the number or variety of common adjectives as a contrastive feature to be compared between English and Spanish. While the Sketch Engine interface also includes Keywords as a function, it is designed for terminology extraction. The results obtained from a search initiated with this option would contain several grammar categories, so the frequency list and order resulting from the process would not be exclusive to adjectives.

Secondly, the Concordance function lets us find all the corpus applied uses of a high-frequency adjective, providing a significant collection of actual samples in context. This is valuable to prove how appropriate a specific adjective may be when comparing corpus contexts with our target or communicative background in specialised translation or copywriting. Finally, the Word Sketch section directly leads to possible collocates or word combinations that include each frequently used adjective identified with Word List. This helps us in evaluating how prototypical segments are or can be, with special attention to common patterns that integrate nouns and adjectives, prepositional groups, and conventional objects for specific verbs.

Both quantitative and qualitative details have, therefore, been explored. This means that first, recurrence indicators could be established, then positiveness could be identified and finally, typical patterns could be confirmed. On this point and because of its prototypicality, we can consider a recurring adjective with a positive charge as a probable empathy trigger, especially when included in a corpus-based collocate which displays the propensity proven by the corpus review. As a result, the findings obtained with our corpus analysis have representatively revealed how often certain adjectives are used in each language's tourism discourse, how varied they are in terms of their global recurrence, and how context-based their positive sense is; which is relevant to theorizing on degrees in empathy triggers as a first insight.

On the basis that any recurring adjective may boost empathy as a sample of prototypical persuasive language, a gradation can be theoretically established to show empathy activation and its possible levels under frequency ranges. The table below presents some preliminary formulae to provisionally categorize empathy moves, although an adjustment will be added to these initial patterns on the basis of the global results obtained from the corpus analysis.

Positions in order of frequency	Approximate frequency range	Empathy triggering expectation	Provisional formula/expression
1-75	850 to 100	High-empathy expectation	FE1
76-200	99 to 30	Medium-empathy expectation	FE2
201-300	29 to 15	Pro-empathy attitude	FE3

Table 3. Preliminary formulae on empathy triggering expectations.  
FE: Frequency as empathy trigger.

The aforementioned parameters consider frequency as one of the precursors for empathy triggering, but the findings from the analysis also demonstrate why recurrence values cannot be regarded as absolute or exclusive figures. As anticipated in a previous section, the research is not only dedicated to finding frequency levels but also to determining proportions and tendencies along with positiveness, so results have been specifically evaluated in each subcorpus for empathy prediction purposes. The specialised character of some high-frequency adjectives has also been examined taking into account frequency ranges and common contexts, which is especially important to determine the extent to which certain recurrence patterns or tendencies respond to a professional strategy or a dominium-based convention with empathy impact. Conversely, a medium-low frequency in relatively common adjectives can also be a source of debate, since frequency results may entail a preference for lexical variety to be discussed as well in terms of empathy. As a consequence, the data in the next section are displayed in such a way as to find the logical interpretation for the above-indicated aspects which extends the preliminary formulae.

#### 4. RESULTS AND DISCUSSION

Under the previous reflections, the interpretation of the results is centred on the recurrence and positiveness related to adjectives, with the data below showing proportions, tendencies, and speakers' expectations in each language. Thus, analysing the information in the tables lets us understand what speakers are used to finding in English and Spanish tourism promotion messages and, therefore, what may trigger empathy in their minds when feeling identified with the language they are familiar with.

Firstly, the most frequent adjectives are presented in order of frequency along with their recurrence indicators.



English subcorpus Wordlist (2,418 items/31,921 total frequency)		Spanish subcorpus Wordlist (2,554 items/28,959 total frequency)	
Item	Frequency	Item	Frequency
good	731	grande	841
new	487	santo	606
great	445	turístico	572
itinerary	419	nuevo	472
beautiful	365	mejor	397
unique	347	cultural	362
large	333	antiguo	342
perfect	326	histórico	304
exciting	312	natural	302
local	293	mayor	299
free	288	principal	274
welcome	284	único	272
famous	257	español	228
old	231	internacional	208
modern	229	real	208
available	228	importante	206
historic	226	bueno	204
independent	219	nacional	203
top	201	pequeño	195

Table 4. List of frequent adjectives in English and Spanish subcorpora – positions 1-20.

If the data in table 3 are considered, all the previous adjectives can be associated with a high-empathy expectation and can be categorized as FE1. But, although these adjectives are in the top positions in the corpus, their specificity level is not very high, because they are sometimes systematically repeated due to their obvious positive sense. However, the subsequent positions (21-40) give more specific samples.

English subcorpus Wordlist (2,418 items/31,921 total frequency)		Spanish subcorpus Wordlist (2,554 items/28,959 total frequency)	
Item	Frequency	Item	Frequency
contemporary	198	especial	166
open	196	alto	161
stunning	189	gastronómico	161
iconic	177	romano	160
high	175	municipal	158
full	173	amplio	155
special	156	largo	154
fantastic	155	viejo	151
British	152	tradicional	140
easy	150	mundial	139
fine	149	público	136
small	148	posible	135
short	144	urbano	131
stylish	141	máximo	121
friendly	138	comercial	120
award-winning	136	moderno	117
international	135	diferente	113
big	134	necesario	113
private	133	emblemático	110
vibrant	133	mediterráneo	110

Table 5. List of frequent adjectives in English and Spanish subcorpora – positions 21-40.

Combining more specificity with empathy, we can assume that the positions after the first 20-25 high-frequency adjectives are more and more specialised or specific, with lower figures in terms of frequency but a still significant recurrence. This condition can be especially empathy-moving for experienced users (those who are more familiar with this LSP, whether as experts or travellers) as more conscious speakers, for which empathy may be labelled as EE (empathy related to experience in this domain). The following table provides a logical grading:

Positions in order of frequency	Approximate frequency range vs. specialisation	Empathy triggering expectation	Logical formula/expression
20-75	200 to 100 Specialisation/Specificity C	Pro-empathy attitude for experienced speakers	EE3
76-150	99 to 30 Specialisation/Specificity B	Medium-empathy expectation for experienced speakers	EE2
151-250	29 to 15 Specialisation/Specificity A	High-empathy expectation for experienced speakers	EE1

Table 6. Logical formulae on empathy triggering with experienced speakers.  
EE: Extra empathy arising from experience in promotional tourism discourse.

These categories can be illustrated with some examples along with their frequency indicators, which entail relevant implications in any context related to tourism.

**EE3** → In English: *contemporary* (198), *stunning*, (189), *iconic* (177), *special* (156), *stylish* (141), *award-winning* (136), *vibrant* (133) / In Spanish: *medio* (175), *especial* (166), *gastronómico* (161), *amplio* (155), *tradicional* (140), *urbano* (131), *comercial* (120), *moderno* (117), *emblemático* (119).

**EE2** → In English: *magnificent* (84), *impressive* (78), *picturesque* (63), *exotic* (52), *colourful* (46), *authentic* (36), *legendary* (36) / In Spanish: *espectacular* (99), *contemporáneo* (84), *auténtico* (77), *clásico* (65), *interesante* (60), *profesional* (55), *sostenible* (50), *prestigioso* (35).

**EE1**→In English: *prestigious* (27), *immersive* (26), *renowned* (25), *peaceful* (23), *inspiring* (23), *trendy* (22), *cosmopolitan* (22), *organic* (16) / In Spanish: *inmaterial* (28), *creativo* (25), *permanente* (25), *autéctono* (25), *innovador* (24), *reciente* (22), *imponente* (18), *señorial* (15).

In lower frequency positions, many adjectives share the same recurrence level, since their use is more precise and the number of repetitions is both similar and limited to a specific context (e.g. *self-guided*, *brilliant*, and *eclectic* (33) in English and *artesanal*, *visual*, and *alternative* (23) in Spanish).

For contexts and positiveness to be exemplified, two adjectives from the EE3 category have been chosen, *vibrant* & *modern*, for several reasons. Firstly, they are representative of a pro-empathy attitude and very close to a medium-empathy expectation, which means their specificity level is representative of the genre (tourism promotional websites), as well as reasonable and understandable at first sight. Furthermore, their meaning and relevance in the tourism experience are quite evident for any speaker of the language they are expressed

in, not to forget the positive charge they involve on their own from a pragmatic and cultural viewpoint. These conditions are common in good candidates for empathy triggering and proper indicators of this language behaviour considering both frequency levels and positive connotations.

The signs of the specialised sense in terms of empathy are shown to a certain extent below, displaying some examples of the above-mentioned pro-empathy adjectives.

CQL [lempos_lc=="vibrant-j"] · 133			
Details	Left context	KWIC	Right context
1	visitmanchester.com	running street food event GRUB is located in the <b>vibrant</b> Green Quarter just	
2	visitliverpool.com	national bank, it's now restored back to life as a <b>vibrant</b> drinking establishme	
3	visitbrighton.com	Hove's iconic beach huts with their colourful & <b>vibrant</b> facades & enjoy the	
4	visitbristol.com	this incredible city. </s><s> Bristol is diverse and <b>vibrant</b> and there is always	
5	visitoxford.org	a common phenomenon in Oxfordshire, lies a <b>vibrant</b> restaurant scene. </s><	
6	visitcambridge.org	the region, you can choose to dine at some of the <b>vibrant</b> and bustling inner c	

Table 7. Examples of use in context for *vibrant*.  
KWIC: Key word in context

**Ex. 1 – Extended context:** Manchester's longest-running street food event GRUB is located in the **vibrant** Green Quarter just 5 minutes' walk from Manchester Victoria rail station </s><s> From fine dining and afternoon tea to pop-up markets and humble street food, Greater Manchester is packed with options to suit any mood offering any and every cuisine you can think of </s><s> Revolution: Food and cocktails - See 670 traveller reviews, 155 candid photos, and great deals for Manchester, UK, at Tripadvisor </s><s> Cocktails are shaken up late into the night here so you'll be sure to get your fix of liquid lusting </s><s> From the delicious food to the thirst-quenching cocktails and classy ... </s><s>: stop & space.

CQL [lempos=="moderno-j"] · 117			
Details	Left context	KWIC	Right context
1	es madrid.com	zona sur de la ciudad </s><> Se trata de un <b>moderno</b> espacio en el que la	
2	visitvalencia.com	y XIX, sin olvidar propuestas más actuales y <b>modernas</b> </s><> Platos decora	
3	malagaturismo.com	amplias, confortables, luminosas y muy <b>modernas</b> equipadas con todo luj	
4	zaragoza.es	vintage se unen en un lugar en el que lo <b>moderno</b> y lo tradicional se mezclan	
5	turismodecordoba.com	edificio ofrece unos espacios museísticos <b>modernos</b> atractivos y funcio	
6	bilbaoturismo.net	nuevas funciones combinando lo histórico y <b>moderno</b> sin olvidar la funciona	

Table 8. Examples of use in context for *moderno*.  
KWIC: Key word in context

**Ex. 2 – Extended context:** Desde el siglo XIX, establecimientos como Colla Monleó, Chez Ramón y la Casa de los Botijos **llevan hasta Valencia las creaciones de**

**maestros ceramistas de toda el área metropolitana** </s><s> Arturo Mora, Vicent Monlló y Rafael Mora son algunos de los nombres propios que firman sus colecciones, que recrean estilos tradicionales diversos: de los motivos religiosos medievales, al azul del gótico (siglo XVI), a los reflejos metálicos (siglo XV) y la cerámica colorida y ornamental de los siglos XVII y XIX, sin olvidar propuestas actuales más **modernas** </s><s> Platos decorativos, vajillas, vasos y recipientes pueden hallarse en sus estanterías, al igual que piezas icónicas y singulares de la historia valenciana </s><s>

The contexts of the 1-50 high-frequency adjectives have been examined, all of them being found to have positive connotations and an evident positive approach. From these positions onwards, some contexts have also been randomly reviewed, with positive charges found in all of them. No negative sense has been identified at all, which leads to the assumption that on tourism promotional websites positiveness is exclusive, typical, and expected. This circumstance can be evidenced with the symbol P+. Additionally, and surprisingly, in the process of context evaluation, opposite adjectives have been found very close to each other in the same line or paragraph, both with a positive sense or connotation. This involves the addressees' expectations are presented as positive and desirable concerning similar referents, as exemplified below:

E.g. There are lots of **new** attractions and events planned for families around Birmingham in the **ancient** woodland walk...

A walk is here an ancient and desirable attraction, while new attractions are also conceived positively.

E.g. Situada en pleno centro comercial de Córdoba, su **novedoso** diseño camina de la mano con el respeto por lo **tradicional**...

The design is presented here as innovative and respectful of traditions at the same time, and both qualities are shown as positive.

On the other hand, some of the above-listed adjectives are to be shown in combination with other segments to confirm what typical collocates and expressions are like. For this purpose, two pro-empathy adjectives, *contemporary* and *emblemático*, have been selected to show their Word Sketch layout, above all after considering their reasonably specialised sense, their evocative charge as high-frequency adjectives, and their collection of positive connotations which contribute to making them good candidates for empathy activation.

Along with their recurrence in tourism discourse, what can be observed in these adjectives and their use in real samples is not only their context-based positive sense but also their natural character when contained in common segments speakers are assumed to expect. After analysing the items surrounding the adjectives in question and the ideas they may evoke, we find both a persuasive sense and a pro-empathy move which seems to be identifiable with feelings, history, tradition, society, and culture-based experiences particularly offered as desirable for addressees. Thus, logically speaking, these adjectives are closer to empathy activation than any qualification item that is randomly chosen or simply based on someone's intuition.

<b>contemporary</b> as an adjective · 198x	<b>emblemático</b> as an adjective · 110x
<b>Nouns modified by ‘contemporary’</b>	<b>Nouns modified by ‘emblemático’</b>
<b>art</b> modern and contemporary art	<b>lugar</b> los lugares más emblemáticos
<b>design</b> of traditional and contemporary design set in	<b>rincón</b> algunos de los rincones más emblemáticos de la ciudad
<b>apartment</b> of 67 stylish contemporary apartments consisting of	<b>edificio</b> los edificios más emblemáticos
<b>culture</b> of traditional & contemporary Scottish culture with Burns	<b>establecimiento</b> algunos de los establecimientos chocolateros más emblemáticos y degustando sus
<b>twist</b> takes on a contemporary twist and offers	<b>monumento</b> de los monumentos más emblemático
<b>comfort</b> town house – contemporary comfort with gorgeous rooms	<b>casa-patio</b> una casa-patio emblemática de la tradición
<b>guesthouse</b> a stylish, contemporary family-run guesthouse, nestled in	<b>espacio</b> espacios verdes emblemáticos
<b>work</b> broad range of contemporary work from both emerging	<b>torre</b> la emblemática Torre del Oro.
<b>gallery</b> is Birmingham’s leading contemporary art gallery located in	<b>tienda</b> tiendas más emblemáticas
<b>bedroom</b> contemporary bedrooms	<b>comercio</b> comercios emblemáticos
<b>hotel</b> contemporary hotel	<b>calle</b> las calles más emblemáticas de

Table 9. Examples of ‘Word Sketch’ results with *contemporary* and *emblemático*.

On the basis of the results listed above, some word combinations are shown to be very prototypical if we pay attention to how frequently they are used as collocates. If so, the categorisation of empathy can be marked as NE+ (standing for naturalness empathy in formulae). Specifically, *contemporary* seems to be closely related to ideas such as style, art, culture, and top facilities, whereas *emblemático* is very commonly associated with superlative expressions containing place nouns. As an example, this information is of great help in making decisions in specialised translation, copywriting or communication tasks, to the extent that we can understand appropriateness, quality and persuasion may depend on a corpus-based search rather than just on a dictionary entry or a personal inclination. Likewise, empathy moves are likely to be better followed or tracked with corpus monitoring rather than with individual preferences, since the selection of items and terms by tossing a coin is probably not an ideal strategy for specific purposes to be achieved.

So far, in this article, empathy triggering modalities have been defined through various

options to achieve prototypicality; such as via top-frequency positions (FE), secondly with specificity to be identified by experienced language users or speakers (EE), and thirdly with natural collocates (NE+). An additional modality can also be arranged by rooting genuineness in the degree of recurrence which is preferential for each language and its reproduction as a tendency. This property can be explored by comparing the number of adjectives and frequency levels in English and Spanish to find out the specific conventions in each language as regards tourism promotional content.

	English subcorpus	Spanish subcorpus
Nr items/adjectives	2,418	2,554
Total frequency	31,921	28,959

Table 10. Comparative figures for recurring adjectives and frequency indicators.

Observing the figures above, some evidence is available and observable. Firstly, we can see that there is a more reduced collection of adjectives, which are more frequently used in English; whereas we find a higher number of adjectives, which are less frequently used in Spanish. Additionally, after examining each list of words (Table 4), we can confirm that the order of recurrence does not coincide in both languages, to the extent that *good* is the most frequently used adjective in English and *grande* is the most recurrently used in Spanish.

The same propensity can be detected with other adjectives and their list positions, with some examples of surprising and relevant recurrence indicators for such adjectives as *modern* and *independent* in English or *santo* and *real* in Spanish. Undoubtedly, extra attention is needed in the future for research dedicated to the pragmatic and cognitive reasons for clarifying why these adjectives are preferential in each language and culture.

As an extension of the aforementioned tendencies, interestingly, the most recurring adjectives in English are condensed into the first 50 results, whereas in Spanish the most repetitive adjectives are more expansively distributed all throughout the first 75 positions, which reinforces the conclusion about a higher propensity to recurrence in English.

In terms of empathy triggering, the tendency of each language, either recurrence or more lexical variety and less recurrence, can be accordingly reproduced to achieve it. Therefore, if the aforesaid tendencies are properly followed, for example with express repetitions in English and with more synonyms in Spanish, the tendency variable will also be considered empathy moving (TE+ or empathy activated by following the suitable tendency). Nevertheless, this being a first approach to empathy, the specific proportions for each convention to be reproduced still need to be thoroughly studied and debated. Nevertheless, if we divide the total frequency by the number of adjectives used in each language, the result can act as a guide to the point at which we can consider that each tendency is being followed, although the achievement limits are still to be defined. The result of the division is 13.20 in English and 11.34 in Spanish. Therefore, we can interpret that the prototypical recurrence in English is followed with a result of 13.20 or a superior figure, whereas in Spanish variety (which is the opposite strategy to recurrence) is followed with a result of 11.34 or an inferior figure, establishing the reasonable variation in a  $\pm 10\text{-}15\%$ .

In sum, the combination of all the empathy triggering parameters can be evaluated in order to determine what empathy-moving language is like (EM), to identify it, reproduce it, and correct it, if so in translated or copywritten content.

$$EM = P+ \pm FE \pm EE \text{ s/A/B/C+} \pm NE+ \pm TE+$$

This means empathy moves or triggering are based on the use of high-frequency adjectives (FE) with positive connotations (P+) whose specific character (A, B, or C) is to be assessed along with their naturalness and their attention to the linguistic tendency related to each language.

## 5. CONCLUSIONS AND PROSPECTS

Posting promotional content on the Internet is an unavoidable and more than relevant strategy for companies offering tourism-related products or services. Therefore, efforts to optimise the effects of their message are both worthwhile in terms of profit-increase policies and advantageous in terms of brand equity or customer loyalty. This consideration is directly opposed to the massive sense consumers feel with millions of texts available on thousands of websites, which were specially designed to promote services and destinations as desirable experiences.

So where is the key to making tourism language particularly effective, distinctive, and empathy-moving in each language? Indeed, corpus analysis is a very powerful tool, not only to answer this question, but also to measure the extent to which each term is appropriate in a specific context and combination of words. Additionally, a specific collection of adjectives, among other categories, may be conceived as a fusion of proper candidates for empathy triggering, given their frequency of use and their positive charge in a language tourism background after their proven value as a part of speakers' expectations to recognize the genre in question. Consequently, corpus-based research is a convenient instrument to delve into specialised languages and, more particularly, into tourism language traits.

On the other hand, our corpus study has revealed some findings as regards frequency, naturalness, and propensities in the language of tourism. Specifically, frequency indicators can offer some guidance on empathy predictions, with high-frequency adjectives being good candidates as empathy triggers on the basis of their recurrence in each corpus (under a preliminary ranking FE1, FE2, or FE3). Considering the specificity of the adjectives, some extra aspects related to empathy activation can be seen; especially those identified as proper strategies to catch the attention of experienced speakers who may be captivated by more colourful and specific language (with the ranking EE1, EE2, or EE3). Additionally, adapting segments to typical collocates is also a source of interest in terms of empathy triggering and prototypicality, with naturalness being attributed to common expressions and collocations located in each corpus (ranking NE+ if all the collocates have been found in the corpus). Finally, recurrence tendencies have also been covered by the study as an indicator of empathy move (ranking TE+ if the natural tendency is followed). Recurrence as a strategy for prototypicality was more specifically found in English, with a smaller number of adjectives being more frequently used than in Spanish. In contrast, a larger collection of adjectives that are less frequently used was identified in the Spanish subcorpus, which probably entails a higher variety of adjectives. As a consequence, we may conclude that the recurrence of



some adjectives is more likely to trigger empathy in English than in Spanish, this strategy being more familiar to and predictable for English speakers.

However, not only contrasts but also common points, can be found after corpus analysis. In both languages, positive charge is of key importance after checking the context for the adjectives. No matter the collocate they are integrated into, adjectives are all positively used, even when two opposite qualities are mentioned in the same section or line, both considered as positive and desirable for similar referents. This makes adjectives in tourism promotional language completely context-dependent in such a way that whenever an adjective is used in tourism promotional messages, either in English or Spanish, it is to be interpreted positively.

Thus, in tourism promotional content as a multifaceted LSP, speakers' expectations and social conventions may entail, for instance, that spaces can be similarly ideal whether they are either small or large under the preferences or inclinations of consumers, tourists, or users may have towards size in each context. Small areas can be positively evaluated to the same extent that large spaces can also be idealised depending on how highly appreciated traits under promotion are exhibited for each place or destination.

In terms of empathy, positiveness – or Dann's euphoria (1996)- is a key point to be considered and reproduced so that tourism discourse can be recognised as genuine. This means positive adjectives are good candidates or prototypes to trigger empathy or, at least, they contribute to not restraining it.

Furthermore, it is clearly worth mentioning the cultural impact of history, society, and even traditional religions on linguistic tendencies. For example, the most frequent adjectives in Spanish are *grande* and *santo*, whilst the top positions in English are filled by *good* and *new*. On this basis, it also comes as no surprise that we found *soleado* around 150 times in Spanish compared to *sunny* on just 36 occasions in English, which is obviously related to the typical weather in each country. Indeed, thousands of cognitive and pragmatical reflections can be referred to explain recurrence levels for each particular adjective, so future research on this topic includes a vast area.

There is absolutely no doubt that much more still needs to be studied, thoroughly analysed, and discussed in relation to adjectives in order to evaluate the degree of empathy a text or collection of contents may trigger. In future research projects, adjectives and other categories should also be examined to find out if their influence prompts or improves empathy activation.

Boosting empathy is a complex phenomenon with language playing a key —although not exclusive— role and offering an essential contribution. The next steps in this research will, hopefully, shed light on the significance of cognitive processes for empathy triggering.

## REFERENCES

- Agarwal, S. (2014). Neuromarketing in action: How to talk and sell to the brain. *Journal of Consumer Marketing*, 31(5), 404-405. <https://doi.org/10.1108/JCM-02-2014-0879>
- Alcaraz Varó, E. (2000). *El inglés profesional y académico*. Alianza Editorial.
- Alcaraz Varó, E. (2004). El lenguaje profesional y académico. In *Nuevas formas de aprendizaje en lenguas extranjeras* (pp. 91-114). Subdirección General de Información y Publicaciones.
- Alcaraz Varó, E., J. Mateo Martínez & F. Yus Ramos. (Eds.). (2007). *Las lenguas profesionales y académicas*. Ariel.
- Alesón, M. (2011). An overview of the Evolution of English Lexicography in the Industries of Leisure and Tourism. In I. Balteiro (Ed.), *New Approaches to Specialised Lexicology and Lexicography* (pp. 15-46). Cambridge Scholars Publishing.

- Baker, M. (2003). *Lexical Categories*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511615047>
- Baker, P. (2010). *Sociolinguistics and corpus linguistics*. Edinburgh University Press
- Berber Sardinha, T. (2002). Corpora eletrônicos na pesquisa em tradução. *Cadernos de Tradução*, 1(9), 15-59.
- Beale, W. H. (1978) Rhetorical Performative Discourse: A New Theory of Epideictic. *Philosophy and Rhetoric*, (11), 221-46.
- Bhat, D.N.S. (1994). *The adjectival category. Criteria for differentiation and identification*. John Benjamins. <https://doi.org/10.1075/slcs.24>
- Bosque, I. (1993). Sobre las diferencias entre los adjetivos relacionales y los calificativos. *Revista Argentina de Lingüística*, (9), 10-48.
- Bowker, L. & Pearson, J. (2002). *Working with specialised language: A practical guide to using corpora*. Routledge. <https://doi.org/10.4324/9780203469255>
- Cabré, M.T. (1999). *La terminología: Representación y comunicación. Elementos para una teoría de base comunicativa y otros artículos*. Universitat Pompeu Fabra. Institut Universitari de Lingüística Aplicada.
- Calvi, M. V. (2010). Los géneros discursivos en la lengua del turismo: una propuesta de clasificación. *Ibérica 19*, 9-32
- Calvi, M. V. (2016). Léxico de especialidad y lengua del turismo. In A. L. Duffé Montalván (Ed.), *Estudios sobre el léxico* (pp. 65-105). Peter Lang.
- Cialdini, R. B. (2001). Systematic opportunism: An approach to the study of tactical social influence. In J. P. Forgas & K. D. Williams (Eds.), *Social influence. Direct and indirect processes* (pp. 25-39). Psychology Press.
- Cialdini, R. B. & Sagarin, B. J. (2005). Principles of interpersonal influence. In T. C. Brock & M. C. Green (Eds.), *Persuasion. Psychological insights and perspectives* (pp. 143-169). Sage.
- Coupland, N. (2003). Sociolinguistic Authenticities. *Journal of Sociolinguistics*, 7(3), 417-431. <https://doi.org/10.1111/1467-9481.00233>
- Croft, W. (1991). *Syntactic Categories and Grammatical Relations: The Cognitive Organization of Information*. University of Chicago Press.
- Crystal, D. (1997). *English as a Global Language*. Cambridge University Press
- Dann, G.M.S. (1996). *The Language of Tourism: A Sociolinguistic Perspective*. CAB International.
- Deutsch, F. & Madle, R. A. (1975). Empathy: Historic and current conceptualizations, measurement, and a cognitive theoretical perspective. *Human Development*, (18), 267- 287. <https://doi.org/10.1159/000271488>
- Demonte, V. (1999a). El adjetivo: clases y usos. La posición del adjetivo en el sintagma nominal. In I. Bosque & V. Demonte (Eds.), *Gramática Descriptiva de la Lengua Española* (pp.129-215). Espasa Calpe.
- Demonte, V. (1999b). A minimal account of the Spanish adjective position Interpretation. In J. Franco, A. Landa & J. Martín (Eds.), *Grammatical Analyses in Basque and Romance Linguistics. Papers in Honor of Mario Saltarelli* (pp. 45-76). John Benjamins. <https://doi.org/10.1075/cilt.187.05dem>
- Durán Muñoz, I. (2012). Caracterización de la traducción turística: problemas, dificultades y posibles soluciones. *Revista de Lingüística y Lenguas Aplicadas*, 7, 103-113. <https://doi.org/10.4995/rlyla.2012.1127>
- Durán Muñoz, I. (2014). Aspectos pragmático-lingüísticos del discurso del turismo de aventura: Estudio de un caso. *Normas (Revista De Estudios Lingüísticos Hispánicos)*, 4, 44-69. <https://doi.org/10.7203/Normas.4.4687>
- Eisenberg, N. (2000). Emotion, regulation, and moral development. *Annual Review of Psychology*, 51, 665-697. <https://doi.org/10.1146/annurev.psych.51.1.665>
- Englebretson, R. (1997). Genre and grammar: Predicative and attributive adjectives in spoken English. *Berkeley Linguistics Society*, 23, 411-421. <https://doi.org/10.3765/bls.v23i1.1272>

- Fajardo Uribe, L.A. (2007). La lingüística cognitiva: Principios fundamentales. *Cuadernos de Lingüística Hispánica*, 9, 63-82. <https://www.redalyc.org/articulo.oa?id=3222/322230194007>
- Fogg, B. J. (2003). Prominence-Interpretation Theory: Explaining how people assess credibility online. In *CHI'03 extended abstracts on human factors in computing systems* (pp. 722-723). Stanford University. <http://credibility.stanford.edu/mostcredible.html>
- Gallo, D. (1989). Educating for empathy, reason, and imagination. *The Journal of Creative Behavior*, 23(2), 98-115. <https://doi.org/10.1002/j.2162-6057.1989.tb00680.x>
- Garcia, J. R., & Saad, G. (2008). Evolutionary neuromarketing: Darwinizing the neuroimaging paradigm for consumer behavior. *Journal of Consumer Behaviour: An International Research Review*, 7(4-5), 397-414. <https://doi.org/10.1002/cb.259>
- Gordon, D., & Lakoff, G. (1971). Conversational postulates. *Papers from the 7th Regional Meeting, Chicago Linguistic Society*, 63-84.
- Gotti, M. (2006). The Language of Tourism as Specialised Discourse. In O. Palusci & S. Francesconi (Eds.), *Translating Tourism. Linguistic/cultural representations* (pp. 15-34). Università degli Studi di Trento.
- Halliday, M.A.K. (1978). *Language as Social Semiotic*. Edward Arnold.
- Halliday, M. A. K. (1994). *An Introduction to Functional Grammar*. Edward Arnold.
- Hoffman, M.L. (1982). Development of prosocial motivation: Empathy and guilt. In Eisenberg (Ed.), *The development of prosocial behaviour* (281-313). Academic Press. <https://doi.org/10.1016/B978-0-12-234980-5.50016-X>
- Hyland, K. (2010). Constructing proximity: Relating to readers in popular and professional science. *Journal of English for Academic Purposes*, 9, 116-127. <http://www.sciencedirect.com/science/article/pii/S1475158510000159>
- Kelly, D. (2000). Text Selection for Developing Translator Competence: Why texts from the tourist sector constitute suitable material. In C. Schäffner & B. Adab (Eds.), *Developing Translation Competence* (pp. 157-170). John Benjamins. <https://doi.org/10.1075/btl.38.15kel>
- Kelly Holmes, H. (2005). *Advertising as Multilingual Communication*. Palgrave Macmillan. <https://doi.org/10.1057/9780230503014>
- Kenning, M. (2010). What are parallel and comparable corpora and how can we use them? In *The Routledge handbook of corpus linguistics* (pp. 487-500). Routledge.
- Lakoff, G. (1987). *Women Fire and Dangerous Things: What Categories Reveal about the Mind*. University of Chicago Press. <https://doi.org/10.7208/chicago/9780226471013.001.0001>
- Langacker R. (2008). *Cognitive Grammar: A Basic Introduction*. Oxford University Press.
- Laviosa, S. (1997). How comparable can 'comparable corpora' be? Target. *International Journal of Translation Studies*, 9(2), 289-319. <https://doi.org/10.1075/target.9.2.05lav>
- Leech, G., Francis, B. & Xu, X. (1994). The use of computer corpora in the textual demonstrability of gradience in linguistic categories. In C. Fuchs & B. Victorri (Eds.), *Continuity in linguistic semantics* (pp. 57-76). John Benjamins.
- Leech, G. (2007). New resources, or just better old ones? The holy grail of representativeness. In M. Hundt, N. Nesselhauf & C. Biewer (Eds.), *Corpus linguistics and the web* (pp. 133-150). Rodopi
- Llorens Simón, E. M. (2017). ¿Empatía, persuasión y/o comunicación efectiva? Traducción y multimodalidad en el sector turístico. Mensajes corporativos en el ámbito del ocio y el tiempo libre. *E-Aesla*, (3), 290-310.
- Llorens Simón, E. M. (2021). Patrones recurrentes en promoción turística. Disparadores de la empatía en inglés y en español. In T. Barceló Martínez, I. Delgado Pugés & F. García Luque (Eds.), *Tendencias Actuales en Traducción Especializada, Traducción Audiovisual y Accesibilidad* (pp. 459-480). Tirant Humanidades.
- Mapelli, G. (2008). Las marcas de metadiscursos interpersonal de la sección turismo de los sitios web de los ayuntamientos. In M.V Calvi, G. Mapelli & J. Santos López (Eds.), *Lingue, culture, economia: Comunicazione e pratiche discorsive* (pp. 173-190). FrancoAngeli.
- McEnery, T. (2019). *Corpus linguistics*. Edinburgh University Press. <https://doi.org/10.1515/9781474470865>

- McEnery, T. & Xiao, R. (2007). Parallel and comparable corpora: What is happening? In *Incorporating corpora* (pp. 18-31). Multilingual Matters. <https://doi.org/10.21832/9781853599873-005>
- Mateo Martínez, J. (2007). El lenguaje de las ciencias económicas. En E. Alcaraz Varó, J. Mateo Martínez y R. Yus Ramos (Eds.), *Las lenguas profesionales y académicas* (pp. 191-204). Ariel.
- Olohan, M. (2004). *Introducing Corpora in Translation Studies*. Routledge. <https://doi.org/10.4324/9780203640005>
- Parodi, G. (2004). Textos de especialidad y comunidades discursivas técnico-profesionales: Una aproximación basada en corpus computerizado. *Estudios filológicos*, 39, 7-36. <https://doi.org/10.4067/S0071-17132004003900001>
- Payne, J., Huddleston, R., & Pullum, G. K. (2010). The distribution and category status of adjectives and adverbs. *Word Structure*, 3(1), 31-81. <http://www.lel.ed.ac.uk/~gpullum/E1750124510000486.pdf>
- Pickett, D. 1989. The sleeping giant: investigations in business English. *Language International*, 1(1), 5-11.
- Pustet, R. (2006). *Adjectives*. Encyclopedia Elsevier
- Quirk, R., S. Greenbaum, G. Leech & J. Svartvik. (1985). *A Comprehensive Grammar of the English Language*. Longman
- Regan, D.T., & Totten, J. (1975). Empathy and attribution: Turning observers into actors. *Journal of Personality and Social Psychology*, 32, 850-856. <https://doi.org/10.1037/0022-3514.32.5.850>
- Rosch, E. (1975). Cognitive representation of Semantic Categories. *Journal of Experimental Psychology, General*, 104, 192-232. <https://doi.org/10.1037/0096-3445.104.3.192>
- Rosch, E. (1983). Prototype classification and logical classification. In E. Scholnik (Ed.), *New Trends in Cognitive Representation. Challenges to Piaget's Theory* (pp. 73-86). Erlbaum
- Rogers, M. (2015). *Specialised Translation: Shedding 'the non-literary tag'*. Springer. <https://doi.org/10.1057/9781137478412>
- Scarpa, F. (2020). Quality in Specialised Translation. In *Research and Professional Practice in Specialised Translation* (pp. 291-365). Palgrave Macmillan
- Searle, J. R. (1985). *Expression and meaning: Studies in the theory of speech acts*. Cambridge University Press.
- Stubbs, M. (1996). *Text and Corpus Analysis (Computer-Assisted Studies of Language and Culture)*. Blackwell
- Suau Jiménez, F. (2006). *El metadiscursos en el género "Servicios y productos turísticos" en inglés y español: Importancia de su traducción como recurso para la persuasión del cliente*. [Ponencia]. I Congreso Internacional de Traducción Especializada organizado por el Colegio de Traductores Públicos de la ciudad de Buenos Aires. Universidad de Buenos Aires, BA, Argentina.
- Suau Jiménez, F. & Dolón Herrero, R. (2007). The importance of metadiscourse in the genre 'Promotion of Touristic Services and/or Products': Differences in English and Spanish. In D. Galová (Ed.), *Languages for Specific Purposes: Searching for Common Solutions* (pp. 71-79). Cambridge Scholars Publishings.
- Suau Jiménez, F. & Dolón Herrero, R. (2008). El metadiscursos en la traducción de textos performativos. Ponencia. *Congreso Internacional de Aesla*. Almería, abril 2008.
- Swales, J. M. (1990). *Genre Analysis: English in Academic and Research Settings*. Cambridge University Press.
- Vargas-Sierra, C. (2006). Diseño de un corpus especializado con fines terminográficos: El Corpus de la Piedra Natural, *Debate Terminológico*, 2 (7/2006), París: RITERM (Red Iberoamericana de Terminología). [http://rua.ua.es/dspace/bitstream/10045/9426/1/chelo\\_vargas\\_sierra.pdf](http://rua.ua.es/dspace/bitstream/10045/9426/1/chelo_vargas_sierra.pdf) 10-01-14
- Vargas Sierra, C. (2012). La tecnología de corpus en el contexto profesional y académico de la traducción y la terminología: Panorama actual, recursos y perspectivas. *Tecnología, Traducción y Cultura*, 67-99.
- Wilson, D. & Sperber, D. (2002). Relevance Theory, *UCL Working Papers in Linguistics*, 14, 249-290.
- Yus Ramos, F. (2006), Relevance theory, In K. Brown (Ed.), *Encyclopedia of Language and Linguistics* (2nd ed., vol. 10, pp. 512-519). Elsevier. <https://doi.org/10.1016/B0-08-044854-2/00313-8>

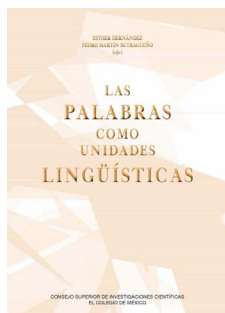


# RESEÑAS



**Hernández, E. y P. Martín Butragueño (eds.) (2020).**  
***Las palabras como unidades lingüísticas.***  
**Madrid/ Ciudad de México: CSIC. 467 Páginas**

RAQUEL RODRÍGUEZ FERNÁNDEZ  
National University of Ireland, Galway  
[raquel.rodriguez-fernandez@nuigalway.ie](mailto:raquel.rodriguez-fernandez@nuigalway.ie)  
<https://orcid.org/0000-0002-0466-2170>



La definición del concepto de palabra y de sus límites como unidad de trabajo ha constituido uno de los mayores retos para la lingüística. Como parte esencial del lenguaje y máximo exponente de nuestra capacidad comunicativa, la palabra se ha caracterizado tradicionalmente desde múltiples ángulos. La mayoría de las descripciones de este elemento a menudo se han basado bien en sus propiedades gramaticales –morfológicas o sintácticas–, bien fonológicas, llegando incluso a proponerse la existencia de dos tipos diferenciados de unidades: las palabras gramaticales, por un lado, y las fonológicas, por otro (Dixon y Aikhenvald, 2002, p. 13-25). También encontramos quienes, ante la dificultad de hallar unos criterios comunes y aplicables a todas las lenguas, rechazan la universalidad de este concepto (Haspelmath, 2011; Hippisley, 2015). Todo ello contribuye a demostrar que, a pesar de nuestra habilidad como hablantes para distinguir esta entidad de otras a las que no puede atribuírseles este estatus, el establecimiento a nivel teórico de unos parámetros de estudio unánimes continúa siendo una cuenta pendiente para los especialistas del lenguaje.

Como unidad de análisis, la palabra es objeto de interés no solo para la morfología, la sintaxis o la fonología, sino también para la lingüística histórica, la sociología, la psicolingüística o la adquisición de lenguas (Taylor, 2015). De hecho, los diversos cambios, particularidades y variaciones de esta entidad se abordan desde una multiplicidad de perspectivas y planteamientos teóricos que buscan analizar en mayor medida su naturaleza y sus límites. Este debate sobre la palabra como unidad de trabajo es, precisamente, el tema principal de

---

**Para citar esta reseña:** Rodríguez Fernández, R. (2022). Hernández, E. y P. Martín Butragueño (eds.) (2020). *Las palabras como unidades lingüísticas*. Madrid/Ciudad de México: CSIC. 467 Páginas. *ELUA*, (38), 327-332. <https://doi.org/10.14198/ELUA.21321>

Recibido: 14/11/2021 Aceptado: 14/12/2021

© 2022 Raquel Rodríguez Fernández



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



la obra que aquí reseñamos, que recoge una serie de estudios caracterizados por compartir un conjunto de intereses comunes, como los procesos de variación y cambio lingüísticos o la diversidad tipológica. El análisis de la palabra como unidad de trabajo se aborda aquí desde distintas ramas de la lingüística e incluye trabajos tipológicos, funcionales, cognitivos y sociales, que evalúan la dimensión sincrónica y diacrónica de este fenómeno, recurriendo, para ello, a datos empíricos. Además de esta variedad de perspectivas, otro de los aspectos más destacados de este manual es su carácter inclusivo, ya que no solo se han incorporado investigaciones sobre el español, sino también sobre otras lenguas románicas y europeas, el euskera, el tagalo, el tepehuano y el tlapaneco.

Los distintos estudios que constituyen este libro se dividen en tres secciones: i) “Palabras prosódicas y gráficas”, centrada en la dimensión sonora y escrita de estas unidades; ii) “La dimensión morfosintáctica y las clases de palabras”, dedicada a la dimensión gramatical de la palabra; y iii) “El papel del léxico”, donde se revisan varios aspectos lexicológicos. La distribución de los trabajos incluidos en cada uno de estos bloques es bastante desigual, ya que, si bien el primero y el tercero cuentan con cuatro estudios cada uno, el segundo reúne un total de nueve colaboraciones. La inclusión en este último caso de varios capítulos que analizan cuestiones más pragmáticas que gramaticales sugiere que quizá hubiera sido conveniente establecer un cuarto apartado que permitiera reorganizar la información de una manera más equitativa a partir de unos criterios clasificatorios más claros. Además, la división de los trabajos agrupados en este bloque en dos secciones distintas ayudaría a homogeneizar la extensión de las secciones y, con ello, a dinamizar la lectura.

Tras una introducción en la que los editores presentan y delimitan los objetivos del volumen, el primer bloque de este manual se abre con la investigación de Concepción Company Company, titulada “El concepto de palabra gramatical. La evidencia de los *usus scribendi* en español antiguo”. En ella, Company examina los sandhis externos –esto es, la unión gráfica de dos formas– de documentos novohispanos redactados en español para determinar si las palabras gramaticales forman una subclase de palabra. Esta hipótesis se confirma tras el análisis de la unión y separación de palabras presentes en estas prácticas escriturarias, que muestra, por un lado, que la primera palabra involucrada en la creación de un sandhi suele ser una palabra gramatical y, por otro, que gramática y fonología juegan un papel complementario en la creación de estas formas, ya que tanto su primer como su segundo miembro están sujetos a fuertes restricciones fonéticas.

El segundo estudio, “La palabra fonológica como dominio de la culminatividad en tepehuano del norte”, realizado por Carlos Ivanhoe Gil Burgoin, aborda cómo la culminatividad del tono puede ayudar a explicar el comportamiento de los pronombres clíticos y su adjunción al verbo en tepehuano del norte, una lengua mexicana de la familia yutoazteca. En él, se establece una distinción entre clíticos sufijales que se fusionan prosódicamente con la base verbal y clíticos o morfemas que, debido a su especificación tonal, pasan a constituir palabras fonológicas independientes con un tono propio. Este último tipo infringe la tendencia universal por la que afijos y radicales forman parte de una misma palabra fonológica, mostrando la falta de emparejamiento entre palabra fonológica y palabra morfológica en tepehuano.

“Polaridad, elisión vocálica y minimidad en la palabra fonológica del tlapaneco”, de Esther Herreras Zendejas, es el siguiente trabajo incluido en este primer bloque. Centrado en el ámbito de la palabra fonológica del tlapaneco, este trabajo revisa la elisión vocálica de la base al sufijar el morfema nominal de 2ª persona y la polaridad que da lugar a la inversión de

los puntos de articulación [coronal] y [dorsal] de la vocal de la base en la 1ª y 3ª persona. La palabra fonológica se concibe aquí como el dominio a partir del cual pueden analizarse, desde la teoría de la optimidad, los distintos procesos segmentales y morfofonológicos presentes en la flexión nominal. Volviendo al ámbito del español, la palabra fonológica o prosódica es, asimismo, objeto de la investigación realizada por Pedro Martín Butragueño, titulada “Tamaño y métrica en la construcción enunciativa de las palabras prosódicas en español”. En ella, el autor examina los problemas que plantean el tamaño y la métrica de las palabras prosódicas cuando se encuentran dentro de un enunciado específico y la relación de estas con respecto a la frase entonativa. A partir de una serie de datos empíricos, se establece que: i) el tamaño prototípico de la palabra enunciada en español es de dos sílabas; ii) las sílabas clíticas suelen localizarse más en lado izquierdo que en el derecho; y iii) las palabras enunciadas constan en español de un solo pie que, en función de la relación entre sílabas que lo conforman, dan lugar a dos clases rítmicas: la monoico-yámbica y la trocaico-dactílica. En lo que respecta a la relación de las palabras prosódicas con la frase entonativa, los resultados reflejan que esta última unidad suele estar formada por entre uno y cuatro niveles prosódicos superiores ( $\omega'$ ).

La segunda sección de este manual, centrada en varias cuestiones gramaticales, se inicia con el trabajo de Álvaro Arias, “Morfología y sintaxis del neutro de materia en castellano”, que profundiza en el estudio del índice morfológico conocido como *neutro de materia*, un rasgo dialectal de carácter morfológico. Esta particularidad permite distinguir entre sustantivos medibles concretos –o de materia– y sustantivos contables, facilitando la diferenciación entre lo continuo, lo discreto y lo plural. Así, en su análisis, además de estudiar la relación de distintas clases de palabras con este fenómeno y los posibles condicionantes sintácticos que pueden darse, este autor también examina el valor semántico de lo continuo y las condiciones en que se produce su adquisición. Asimismo, investiga la relación que facilita la distinción de lo medible en cada caso y revisa tanto los paradigmas que se proyectan sobre el sustantivo como las diferencias entre ellos.

En el segundo trabajo de este bloque, “Estudio comparativo de la modalidad habilitativa en predicados no verbales en lenguas romances, germánicas y euskera”, Elena Castroviejo e Isabel Oltra-Massuet analizan cómo la palabra *capaz* genera diferentes interpretaciones lingüísticas según su entorno lingüístico y la forma en que el desarrollo de estas interpretaciones puede concebirse como el resultado de un proceso de cambio lingüístico en algunas lenguas. De esta manera, se pretende demostrar que las dos posibles lecturas del adjetivo *capaz* en castellano –una puramente habilitativa y otra de carácter epistémico– y de sus equivalentes en otras lenguas se vinculan a consideraciones sintácticas.

Dentro de esta segunda sección se incluye también la investigación de Violeta Demonte y Olga Fernández-Soriano, que reexaminan semántica y sintácticamente el operador evidencial *dizque*, sosteniendo que la explicación de esta unidad reside en la interfaz entre la sintaxis y la semántica/pragmática. Se sostiene que este elemento –resultante de un proceso de gramaticalización– se trata de un operador evidencial reportativo en aquellos casos en los que introduce oraciones y de un modal epistémico en los que actúa como modificador de constituyentes. En lo referido su posición sintáctica, la libertad de la que goza *dizque* se debe a que precede al foco informativo de la oración, indicando cuál es la información remática. Además, el hecho de que este elemento señale la condición de posible de la proposición y presuponga que el hablante basa dicha posibilidad en una evidencia externa, ayuda a explicar que *dizque* se conciba como un evidencial modal con una presuposición de reportatividad.

Dejando a un lado este ámbito de los operadores, el trabajo de Silvia Gumiel-Molina, Norberto Moreno-Quibén e Isabel Pérez- Jiménez, “Variación dialectal en las oraciones copulativas en español: los adjetivos de edad y tamaño”, estudia cómo la combinación de los adjetivos de *edad* y *tamaño* con la cópula *estar* muestran, en ciertas variedades dialectales del español, un uso semejante al de los adjetivos evaluativos. Este fenómeno de variación sintáctico-semántica se debe a que, en este tipo de variedades, los adjetivos de edad y tamaño poseen un argumento experimentante/perceptor que fomenta una lectura perspectivizada/evidencial de la construcción *«estar + adjetivo»*. Por tanto, la extensión de *estar* en estas variedades del español no se debe a cambios en la semántica o en las propiedades de selección de la cópula, sino a un proceso de cambio lingüístico consistente en la ampliación de las clases de adjetivos compatibles con un experimentante.

La quinta contribución de esta sección es la perteneciente al estudio de Leonor Orozco, “Marcadores discursivos como elementos de variación dialectal: *según* y *saabe* en el español mexicano”, en la que la autora revisa el uso de dos marcadores discursivos propios del discurso oral en el español mexicano y en la variedad regional del occidente de este país. El análisis de datos de varios corpus revela que el uso de *según* como marcador discursivo es más frecuente entre los hablantes jóvenes y se genera en estratos sociales bajos, extendiéndose posteriormente a niveles superiores, con una escolaridad mayor. La limitada información sobre *saabe* no permite comprobar ni su estratificación sociolingüística ni su distribución dialectal. No obstante, el empleo de este marcador en redes sociales y en internet muestra que suele utilizarse para expresar un valor de desconocimiento o reafirmación y que se considera un elemento identificativo de la varias de las variedades dialectales del oeste de México. Situado también en el ámbito mexicano, la investigación de Asela Reig Alamillo, “La variación en los adverbios del foco exclusivo *nomás*, *nada más*, *solo* y *solamente* en el español de México”, analiza la alternancia de los adverbios de foco exclusivo en el español de México desde un punto de vista variacionista. La revisión de un corpus de datos orales muestra que la edad, el nivel educativo de los hablantes son los factores externos que influyen sobre la selección de estas formas, en cuyo uso también influyen el tipo de constituyente en el foco y la posición del adverbio con respecto al elemento crucial.

Desarrollado desde una perspectiva diacrónica, el trabajo de Javier Rodríguez Molina, “La difícil sucesión del latín *vix* en las lenguas romance: a propósito de los adverbios medievales *abés* y *malavez*”, estudia las dificultades etimológicas tras el origen de estas formas. Tras revisar las hipótesis sobre la etimología de estos dos términos y analizar los datos del autor propone la familia del latín *vertere* y su posible cruce con *adversum* como el origen de *abés* y vincula el adverbio *malavez* con la gramaticalización de la locución *mala + vez*. De nuevo en el plano de la evidencialidad, en “(Estrategias) evidenciales del tagalo”, Jennifer Tan analiza empíricamente las categorías gramaticales y los marcadores empleados en tagalo para expresar la fuente de información de la que el hablante adquiere la información que transmite. Se examinan, en concreto, los clíticos adverbiales *daw* y *yata*, que indican que la información se ha obtenido por testimonio de otro hablante y por inferencia, respectivamente.

Finalmente, la última contribución en este bloque, realizada por Violeta Vázquez Rojas Maldonado y titulada “Los límites de la palabra nominal en purépecha: marcas de número y caso”, rebate la caracterización morfológica de los sustantivos de esta lengua. En contra de lo establecido en la investigación previa, se sostiene que la flexión de número y caso no se marca en la palabra en sí, sino en la frase. Se argumenta, por tanto, que estas dos categorías

están constituidas por morfemas de naturaleza enclítica y que las palabras nominales del purépecha, al carecer de morfología flexiva, son unidades simples.

La tercera y última sección de este manual, en la que se revisan varias cuestiones relacionadas con el léxico, comienza con la aportación de Pilar García Mouton, “Anglicismos urgentes”. En su trabajo, García recurre a las recomendaciones publicadas por la Fundación del Español Urgente (Fundéu) en su página web para analizar el proceso de adaptación de los anglicismos al español de los medios de comunicación. La rápida respuesta de esta entidad ante la entrada de neologismos procedentes del inglés ha permitido dar cuenta de cuándo y cómo se introducen los préstamos léxicos, así como resolver algunas de las dudas que genera su uso hasta que se acomodan en el sistema. Toda esta información, por consiguiente, ayuda a identificar las distintas pautas seguidas por estas nuevas voces en su incorporación al español y a determinar el tipo de intervención planteado por la Fundéu, que suele seguir varios pasos: i) recomendación de términos alternativos; ii) adaptaciones fonéticas, ortográficas y morfológicas del anglicismo al español; iii) uso de comillas o cursiva para las palabras sin adaptar.

A continuación, el trabajo de Esther Hernández, “Americanismos formados con el sufijo *-illo*, *-illa* en el siglo XVI”, se centra en los americanismos léxicos surgidos entre 1495 y 1600 mediante la derivación con el sufijo apreciativo *-illo*, *-illa*. Para llevar a cabo este análisis, la autora examina los distintos indicadores lingüísticos y discursivos que revelan el origen de una nueva palabra a partir de una voz patrimonial. De acuerdo con los datos recopilados, la lexicalización de las palabras formadas por estos sufijos puede producirse tras una fase previa en la que se registra una variación semántica y morfológica. Así, tras un primer estadio en el que tiene lugar un cambio morfológico, estos afijos adquieren una connotación de “americanidad”, surgiendo de todo este proceso nuevas palabras, nuevos significados y nuevas unidades pluriverbales.

Isabel Molina Martos es la responsable del siguiente estudio de este bloque, “Mecanismos de gradualidad léxica en el cambio fónico”, que explora la variación gradual de la dental sonora /d/ en final de palabra en el léxico y examina el papel que los procesos cognitivos juegan en el cambio fónico de esta consonante. El análisis se centra en la comunidad de habla de Madrid y revela que las palabras más frecuentes son aquellas que tienen una mayor tendencia a experimentar una reducción fonética. Este cambio reductivo, propenso a la elisión, es habitual de las generaciones mayores, mientras que, en el caso de los hablantes más jóvenes, se observa un cambio fónico por refuerzo, causado por el mayor contacto con el estándar.

Por último, la investigación de Niktelol Palacios, titulada “Compuestos sintagmáticos y locuciones nominales en el español de México: criterios léxico-semánticos para su distinción”, tiene como objeto de estudio los compuestos sintagmáticos y las locuciones nominales del español de México, cuyas propiedades formales, semánticas y funcionales examina dentro de una serie de unas parcelas léxicas que implican un conocimiento especializado. Para el estudio de estas unidades establece, además, una diferenciación entre los procesos de denominación y designación que permite distinguir, en los campos léxicos analizados, entre compuestos sintagmáticos designativos y denominativos, por un lado, y locuciones denominativas, por otro. A partir de esta revisión, se proponen, asimismo, una serie de pautas para facilitar la inclusión y el tratamiento de estas dos unidades en la nomenclatura de los diccionarios de lengua general y fraseológicos.

En suma, las distintas aportaciones que configuran esta obra contribuyen a ofrecer una visión actualizada de los múltiples retos asociados a la noción de palabra y a su caracte-

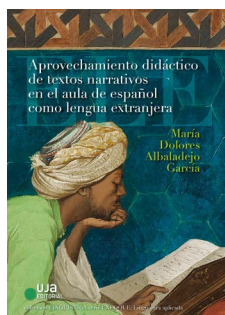
rización como unidad de trabajo. Además, la amplia gama de perspectivas desde las que se abordan las distintas cuestiones relacionadas con este elemento ayuda al lector a tomar conciencia de los difusos límites de esta entidad no solo en español, sino también en otras lenguas como el euskera, el tepehuano o el tagalo, por ejemplo. En este sentido, la inclusión de trabajos centrados no solo en la dimensión sincrónica de esta unidad, sino también en su dimensión diacrónica y en las divergencias que surgen en su análisis desde un punto de vista contemporáneo permite profundizar en la complejidad intrínseca que reviste el estudio de la palabra. En definitiva, nos hallamos ante un manual imprescindible para aquellos especialistas interesados en las últimas aportaciones de varias ramas de la lingüística al estudio de esta unidad que tan debatida ha sido en la historia de esta disciplina.

### REFERENCIAS BIBLIOGRÁFICAS

- Dixon, R. M. W. y A. Y. Aikhenvald (eds.) (2002). *Word: A cross-linguistic typology*. Cambridge University Press.
- Haspelmath, M. (2011). The indeterminacy of word segmentation and the nature of morphology and syntax. *Folia Linguistica*, 45, 1, 31-80. <https://doi.org/10.1515/flin-2017-1005>
- Hippisley, A. (2015). “The Word as a Universal Category”. En J. Taylor (ed.) *The Oxford Handbook of the Word* (pp.246-269). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199641604.001.0001>
- Taylor, J. R. (2015). *The Oxford Handbook of the Word*. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199641604.001.0001>

**Albadejo García, M. D. (2020).**  
***Aprovechamiento didáctico de textos narrativos en el aula de español como lengua extranjera.* Editorial Universidad de Jaén**

NEUS GIMENO GIMENO  
Universidad Jawaharlal Nehru, Nueva Delhi (India)  
[neusgigi@gmail.com](mailto:neusgigi@gmail.com)



El libro que se reseña fue publicado por la Universidad de Jaén, dentro de la colección de Lingüística, de la serie *Enfoque: Lingüística aplicada*, 2, bajo la dirección de Ventura Salazar García, escrito por M<sup>a</sup> Dolores Albadejo García, profesora titular del Instituto Cervantes con más de dos décadas de experiencia profesional en el campo de la docencia del español como lengua extranjera.

La autora analiza el uso de los textos literarios en el proceso de enseñanza y aprendizaje del español como lengua extranjera a través de 272 páginas. Gracias a su extensa experiencia en el campo en diferentes contextos, ha podido constatar que la enseñanza de esta lengua con textos literarios, así como los materiales y recursos puestos a disposición del docente para este fin son deficientes. De hecho, el papel jugado por el texto literario en este contexto, con una mayor presencia en las últimas décadas, no dejar de tener una presencia exigua. Por ello, la autora, en la primera parte del libro, realiza un análisis teórico al respecto y, en segundo lugar, presenta unas propuestas didácticas para el aprovechamiento didáctico de los textos literarios en el aula del español como lengua extranjera.

Dentro de la primera parte llamada «Marco de trabajo», para contextualizar al lector, la autora realiza un breve recorrido histórico a través de los diferentes enfoques metodológicos que han brotado a lo largo del siglo XX hasta nuestros días en el capítulo llamado «Antecedentes». En este, quiere mostrar cómo la noción de funcionalidad de los textos literarios con respecto al aprendizaje de una lengua extranjera ha ido variando notablemente con el tiempo: desde mediados de siglo con el *método tradicional*, pasando por la segunda mitad de siglo con la corriente llamada *método directo*, el enfoque *nocio-funcional* hasta llegar

---

**Para citar esta reseña:** Gimeno Gimeno, N. (2022). Albadejo García, M. D. (2020). *Aprovechamiento didáctico de textos narrativos en el aula de español como lengua extranjera*. Editorial Universidad de Jaén. *ELUA*, (38), 333-337. <https://doi.org/10.14198/ELUA.22421>

Recibido: 13/10/2021 Aceptado: 12/01/2022

© 2022 Neus Gimeno Gimeno



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

al paradigma de enseñanza *comunicativo* y a los *programas procesuales o de tareas*. La percepción de la utilidad del texto literario para fines didácticos en la enseñanza del español como lengua extranjera ha ido evolucionando a lo largo de estos periodos: desde el uso un tanto abusivo del mismo debido a su centralidad en la propuesta de enseñanza, su caída en desuso por la falta de practicidad para la comunicación oral (el principal objetivo de métodos como el nociofuncional o comunicativo) y el rechazo que sufre por el tipo de lenguaje —escrito, estático, alejado de la comunicación diaria— hasta su apreciación por tratarse de un «asunto de máxima actualidad» (Naranjo Pita, 1999a, p.8).

No obstante, a día de hoy, a pesar de que se vislumbra un intento por incorporar los textos literarios en los manuales actuales, aparecen «siempre como mero trazo ornamental, cultural y casi exótico, del que no surgen apenas propuestas metodológicas entroncadas en la unidad didáctica» (Naranjo Pita, 1999a, p.9). Como indica Sanz (2006, p.21), es necesario empezar a «perder el miedo al texto literario con todos los aspectos... y normalizar su utilización en el contexto de las clases de lengua» para poder democratizar su uso, haciéndolo así accesible y útil para la clase de lengua a través de su «didactización».

A continuación, Albadejo García detalla los argumentos que se han esgrimido en contra del uso de textos literarios para la enseñanza de español LE/L2 para poder demostrar los beneficios que se pueden obtener frente a los obstáculos que de su uso se puedan derivar. Entre ellos está el hecho de que los estudiantes de lenguas extranjeras no se sientan atraídos a leer en una lengua que no sea la materna por la falta de utilidad inmediata, tengan miedo hacia el texto literario por las connotaciones históricas atribuidas al mismo, presenten dificultades de comprensión no solo a nivel de significado, sino también conceptual y, por último, en relación a esto que conciben el texto con único significado.

Estos argumentos le sirven a la autora de premisa para justificar el valor que tiene el texto literario como material didácticamente rentable para la enseñanza de una lengua extranjera: la universalidad de los temas, la autenticidad de los materiales, el valor cultural de la literatura, así como la riqueza lingüística, la capacidad de involucrar al lector personalmente y, por tanto, potenciar el desarrollo de su conciencia emocional. Como dice la autora citando a Tomlinson (1986, p.34), «ser profesores de lengua nos convierte en educadores a la vez que instructores, y es por tanto nuestro deber contribuir al crecimiento de nuestros estudiantes a todos los niveles».

En el siguiente capítulo, se presentan las características de los textos literarios junto con los rasgos distintivos del lenguaje literario. Tras definir qué es un texto literario como acto de comunicación en consonancia con la pedagogía, es inevitable replantear el papel del estudiante frente a este tipo de textos, que deja de ser pasivo a ser activo y, por tanto, es quien da sentido al texto, ya que «el proceso de lectura es un proceso (...) de interacción creativa entre el lector y el texto» (Gwin, 1990, p.11). Otro aspecto en el que la autora profundiza es en el lenguaje literario y las características de este: uso estético de la lengua (predominio de la función poética o estética), literalidad (inalterabilidad de la lengua), perdurabilidad (impacto en el intelecto de las actuaciones lingüísticas), plurisignificación (interpretación particular de cada lector) y ficcionalidad (representación de la realidad). Estas especificaciones son esenciales para demostrar la viabilidad del uso de estos textos en el aula de lenguas extranjeras.

Tras demostrar que es posible el empleo de textos literarios en el aula de lenguas extranjeras y de sus ventajas, la autora analiza cuál debería ser el objeto de enseñanza-aprendizaje: la lengua en sí misma con el objetivo de desarrollarla; la competencia cultural como compo-

nente de la competencia comunicativa —pues “lengua y cultura no son más que el texto y el contexto en el que dicho texto se interpreta y/o produce” (Sanz, 2010, p.130)—; el texto literario no solo como fin, sino también como medio para poder desarrollar la competencia literaria; la lectora a través de las microhabilidades lectoras necesarias para leer estos textos; y, por último, la competencia metafórica como recurso de comunicación, ya que, a su vez, revierte en el desarrollo de la competencia literaria, cultural y lingüística de los estudiantes.

Para finalizar este capítulo, se define el modelo de enseñanza de la lengua y la literatura basado en la conciliación bajo la consigna *reflexión* (actividad cognitiva del estudiante), *cambio* (revisión de los preceptos pedagógicos consolidados hasta ahora), *acción* —o *interacción*— *creativa* (característica propia del texto literario) e *integración* [para poder «desarrollar objetivos tanto comunicativos, así como culturales y propiamente literarios» (Sanz, s.f.)]. No obstante, no es la intención de la autora dar a conocer un nuevo enfoque metodológico; sino que más bien, aboga por el *paradigma comunicativo*, y ajustado al *enfoque por tareas*, definir cuatro niveles de planificación siguiendo la terminología de Ventura Salazar (1999): *paradigma, enfoque, método y técnica*.

En el capítulo número cinco, se detallan los criterios para llevar a cabo la selección de textos literarios. Pese a que las lecturas adaptadas han estado presentes en el aula ELE, aquí se alienta a que los textos originales sean *accesibles* o *lecturables* (adaptados al nivel del aprendiente), *significativos* y *motivadores* (según los intereses del estudiante) y que *posibiliten la integración de destrezas* (rentables didácticamente) y que *contengan implicaciones socio-culturales* (alcance cultural), quedando estos dos últimos criterios a discreción del profesor.

Otra cuestión que tener en cuenta es el momento en que se puede empezar a introducir este tipo de textos, por ello se reitera la relevancia de la nivelación de los *inputs* literarios según el nivel de interlengua actual de los alumnos de una manera progresiva, como indica Pérez Parejo (2009, p.21). Esto, por ejemplo, en los niveles iniciales implica emplear, fábulas breves, microrrelatos, entre otro tipo de textos de corta extensión. A medida que se va progresando en las etapas de aprendizaje, la extensión de los textos también tiene que ir aumentando. La autora asimismo menciona su inclinación hacia el uso de textos íntegros frente a fragmentos textuales de una obra que puedan desvirtuarla y, como muchos autores del campo, prioriza los textos contemporáneos frente a los clásicos por la cercanía y facilidad de asimilación que los primeros presentan.

En cuanto a la elección del texto, se señalan tres criterios que deben ser tenidos en cuenta: el tipo de curso y los objetivos propuestos, los estudiantes y sus particularidades, y la adecuación del texto a los criterios de selección fijados. Teniendo esto en mente, se plantean tres fases sucesivas para este objetivo. La primera fase está centrada en los alumnos. En ella, el docente debe conocer con más detalle las inclinaciones lectoras de los alumnos, ya sea a través de un cuestionario o a través de la lectura del comienzo o resumen de una obra. En la segunda fase, el docente tiene que elegir el texto literario siguiendo unos criterios establecidos en un listado (adaptado de Lazar, 1993, p. 56) según el tipo de curso, de estudiantes y pertinencia del texto. Por último, en la tercera fase se debe medir la adecuación real del texto mediante la siguiente estrategia: cálculo del índice de *lecturabilidad* a través de un *cloze test* con un resultado mayor al 60%. No obstante, esta herramienta nos da una información sesgada y, por eso, la autora insiste en la relevancia de poner en práctica los tres filtros planteados (estudiantes, profesor y texto) para poder alcanzar un mayor grado de adecuación texto-alumnos.



Los mismos criterios se deberían aplicar para las lecturas de textos extensivos, además de la implementación de ciertos recursos para favorecerla: fichas de lectura, recomendaciones entre alumnos, archivos personales referidos a la experiencia lectora y, como hizo la autora con sus estudiantes, la creación de un *Banco de Textos Literarios* creado por los propios estudiantes, que, además, es extensible a otros centros y contextos. Atendiendo las necesidades actuales, este banco podría tener un formato digital en vídeo o de blog literario, inspirándose en los llamados *booktubers*, sin dejar de lado el uso de las redes sociales.

En la segunda parte del libro, se presentan diez propuestas didácticas que plasman ese intento de recuperar el texto literario y sirven como ejemplo de aprovechamiento didáctico de estos textos en el aula ELE. En concreto, están diseñadas para los aprendientes que se encuentren entre los niveles B2 y C2, pese a que a que el docente podría modificarlas para que fueran accesibles para otros perfiles, tanto para estudiantes adolescentes como quienes tengan un dominio inferior de la lengua.

Para llevar a cabo la selección de textos, la autora se ha servido del cuestionario propuesto por Lazar y explicado anteriormente, en el que se incluyen las tres áreas a tener en cuenta: curso, alumnos y texto. Estas propuestas didácticas incluyen tanto microrrelatos, como cuentos, novelas y fragmentos de novelas de autores mayoritariamente contemporáneos y algunos clásicos españoles, hispanoamericanos, sufi y andalusí. No obstante, son los docentes quienes deben comprobar la idoneidad de los textos según el perfil del estudiantado en contextos educativos particulares (educación reglada frente a la no reglada, por ejemplo), pues estos han sido seleccionados sin la participación de los propios alumnos. Ciertamente es que quizás se echen en falta otros géneros como la poesía o el teatro. El motivo de ello es que se ha querido favorecer un área de la literatura que, por su extensión y complejidad estructural y temática, ha quedado en desventaja frente a las otras.

Las etapas para llevar a cabo la explotación didáctica hay que pautar del siguiente modo. En primer lugar, se deben desarrollar actividades de prelectura, actividades durante la lectura (para mantener el interés de la lectura y la explotación de puntos cruciales), actividades de poslectura y, por último, la tarea final. Este modo de enfocar el diseño de las explotaciones didácticas está basado, en gran parte, en un enfoque por tareas en gran parte enmarcada dentro del enfoque por tareas y, a su vez, del paradigma comunicativo de enseñanza de la lengua, si bien no en su totalidad toda vez que las tareas resultan de las características del texto y no de una tarea final programada.

Respecto a los agrupamientos, varían de una actividad a otra con el objetivo de atender a la diversidad de los estudiantes participantes, a la vez que se crea cohesión de grupo y un ambiente óptimo para el trabajo. Este también es el motivo por el que la explotación de textos y la trayectoria a seguir en cada secuencia didáctica es abierta: la explotación didáctica debe ajustarse al grupo-clase y no a la inversa.

En cuanto a las actividades que se proponen en estas explotaciones didácticas que se ofrecen como modelo, cabe destacar dos cuestiones. La primera es que no hay un predominio particular de una destreza concreta. Esta elección dependerá de las necesidades que el profesor detecte. La segunda es el carácter flexible y no restrictivo de la secuencia, pues la didactización del texto debe partir de la naturaleza del texto para evitar la sensación de encorsetamiento de unas actividades que tienen que encajar con el texto.

Es necesario, además, señalar que estas explotaciones didácticas abogan porque el trabajo se lleve a cabo tanto en horas lectivas como fuera del aula y será el docente quien

flexibilizará los tiempos hasta lograr el equilibrio deseado. No solo esta será su función, sino también exponer los pasos a seguir, organizar el espacio y los agrupamientos, supervisar el funcionamiento de la actividad, resolver o aclarar dudas, guiar a los alumnos en el proceso lector, etc. para que el proceso didáctico sea considerado vivo, cambiante y, al mismo tiempo, emocionante. Otro elemento que juega un papel relevante es la actitud y la formación literaria del docente para que la experiencia didáctica tenga éxito. De ahí que el profesor desarrolle permanente su propia competencia literaria.

En cuanto a la evolución formativa, se recomienda prescindir de un examen o evaluación sumativa, pues esta herramienta no plasma la evolución global en el aprendizaje del español de los estudiantes. No obstante, se recomienda llevar a cabo una evaluación formativa de carácter cualitativo a través de cuestionarios, la autoevaluación y co-evaluación, el diario del profesor, el cuaderno de trabajo del estudiantado, las tareas finales, el comentario magistral, la asistencia y la participación. El objetivo no es cuantificar los conocimientos adquiridos, sino conseguir el progreso en el aprendizaje a través del modelo presentado desde una perspectiva holística de la lengua.

En el último apartado del libro, se proponen unas actividades para introducir la secuencia didáctica número 1 para que los alumnos y el docente se conozcan mejor (actividad de rompehielo) y otra para ayudar a resolver las dudas léxicas que planteen los alumnos (actividad con efecto bola de nieve). A continuación, se hace una descripción detallada de las actividades creadas para cada una de las diez secuencias didácticas proyectadas junto con las soluciones.

En suma, esta obra ofrece un recorrido histórico por la importancia de la literatura en la enseñanza del español como lengua extranjera y, de una manera práctica a través de ejemplos, muestra como llevarlo a cabo en el aula.

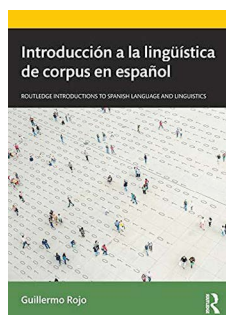
## BIBLIOGRAFÍA

- Gwin, T. (1990). Language skills through literature. *English Teaching Forum*, 28(3), p.10–17.
- Lazar, G. (1993). *Literature and language teaching* (Vol. 2002). Cambridge University Press.
- Naranjo Pita, M. (1999). *La Poesía como Instrumento Didáctico en el Aula de Español como Lengua Extranjera*. Edinumen.
- Pérez Parejo, R. (2009). *Modelos de Mundo socioculturales en la historia de la literatura española (Automatización y descodificación para alumnos E/LE)*. Junta de Andalucía. <http://iesgballester.juntaextremadura.net/web/profesores/tejuelo/index.htm>
- Salazar García, V. (1999). El enfoque por tareas como búsqueda de un currículum centrado en el alumno. *Lingüística aplicada a la enseñanza de lenguas extranjeras*, 307–332.
- Pastor Sanz, M. (2006). Didáctica de la literatura: el contexto en el texto y el texto en el contexto. *Carabela*, 59, 5–23.
- Tomlinson, B. (1986). Using poetry with mixed ability language classes. *ELT Journal*, 40(1), 33–41.



**Rojo, Guillermo (2021): *Introducción a la lingüística de corpus*.  
Londres: Routledge. 380 páginas**

MARÍA EUGENIA CONDE NOGUEROL  
GI *Gramática del español*. Universidade de Santiago de Compostela  
[meugenia.conde@usc.es](mailto:meugenia.conde@usc.es)  
<https://orcid.org/0000-0002-5568-7423>



La lingüística de corpus (LC, en adelante) —entendida como enfoque metodológico o ámbito de la lingüística que basa sus investigaciones en datos obtenidos a partir de corpus (Parodi, 2010: 14-15)— ha ido adquiriendo en las últimas décadas una presencia considerable en el campo de los estudios lingüísticos hispánicos; hasta el punto de que ha ido modificando la concepción misma del objeto de estudio de la lingüística descriptiva tradicional. Son señales de su auge el creciente interés en la construcción y ampliación de numerosos y variados corpus en español, la cada vez más abundante bibliografía sobre la explotación de sus datos y la publicación de manuales y compendios. A ello han contribuido autores como

Guillermo Rojo, cuyas aportaciones al desarrollo de la LC en español han tenido enorme repercusión en la lingüística hispánica.

Una buena prueba de ello es la publicación de este libro que, aunque concebido originariamente como una introducción práctica con una marcada orientación didáctica —como así lo expresa el autor en el prólogo y así lo demuestra la propia concepción estructural del texto—, trasciende ese propósito inicial para convertirse en una obra que eleva el estatus epistemológico de la LC hispánica y contribuye a seguir dotando la investigación producida en español de una orientación científica de profunda solidez.

El autor se propone dos objetivos: enseñar LC a los estudiantes (de los últimos cursos de grado, de máster o doctorandos) que deseen profundizar o adquirir formación en LC, y publicar la primera introducción a esta corriente en español, pues no existen introducciones

---

**Para citar esta reseña:** Conde Noguero, M. E. (2022). Rojo, Guillermo (2021): *Introducción a la lingüística de corpus*. Londres: Routledge. 380 páginas. DOI: <https://doi.org/10.4324/9781003119760> ELUA, (38), 339-341. <https://doi.org/10.14198/ELUA.21356>

Recibido: 18/11/2021 Aceptado: 14/12/2021

© 2022 María Eugenia Conde Noguero



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

escritas en español que muestren de forma práctica cómo se pueden resolver problemas de lingüística española con datos obtenidos de corpus de español.

Con este doble eje vertebrador —el de ayudar a adquirir conocimientos y el de dotar de una perspectiva global e integradora la investigación en LC en español— Guillermo Rojo ha creado una herramienta que responde a la necesidad actual de dotar de modelos y referentes a futuros investigadores en el manejo de datos lingüísticos en español. Para el desempeño de tal empresa asume la responsabilidad de atender tanto a los problemas sobre los que se puede trabajar, como a la forma en que hay que manejar las aplicaciones de consulta de diferentes corpus, aportando a la investigación en LC en español un soporte material similar al que ya posee la tradición anglosajona con los manuales *Routledge Corpus Linguistics Guides* (2015-2021), en los que la investigación en LC se aplica a una extensa variedad de áreas temáticas: pragmática, vocabulario, gramática, estudios contrastivos, enseñanza de la lengua inglesa, etc.

Es justo aquí donde reside otro de los intereses de esta obra: aportar un carácter globalizador e integrador a esta disciplina lingüística, que se desvela en el afán del autor por aprehender y actualizar con exhaustividad todo el saber relacionado con la LC, para lo que despliega todo un conjunto de conocimientos teórico-prácticos —fruto de una dilatada experiencia— acerca de los avances y desafíos que afronta la LC en el marco de la lingüística actual en español.

Desde sus primeras páginas se subraya el carácter empírico de la LC como ciencia cultural y en ella se destacan, no solo la complejidad y variabilidad como elementos consustanciales de esa concepción empírica y funcionalista del lenguaje, sino también, su estatus epistemológico, que se demuestra en la posibilidad de observación de datos objetivos a gran escala en favor de la *explicabilidad* total (*total accountability*).

El esfuerzo del autor por estructurar y ordenar la densidad de este campo de conocimiento se refleja en la vertebración de la obra en siete capítulos de complejidad progresiva, cuyos contenidos van de lo particular y teórico —qué es un corpus, para qué sirve, cuáles son sus tipos, qué es la LC—, a lo práctico y complejo —cómo se diseña y construye un corpus para su explotación y cómo se recupera e interpreta la información contenida en los corpus (en relación al léxico y los fenómenos gramaticales)—; para finalizar con un recorrido histórico por la LC y sus líneas futuras, y aportar, asimismo, una ampliación y resumen de los procedimientos de recuperación de datos.

Así, el capítulo 1, “La explotación básica de los corpus”, tiene la función introductoria de presentar cuáles son las características de un corpus textual, definir qué es la lingüística de corpus y describir sus aplicaciones a la investigación sobre elementos léxicos y categorías gramaticales, cuestiones diacrónicas, aspectos semánticos y sociolingüísticos y sobre enseñanza y aprendizaje de lenguas, entre otros.

El capítulo 2, “La lingüística de corpus y la metodología de la investigación lingüística”, se ocupa de cuestiones metodológicas y se centra en explicar el carácter empírico de la LC en el marco de la lingüística actual, frente a la lingüística racionalista y a la descriptiva tradicional. Sitúa el marco del debate teórico acerca de la concepción de la LC y desarrolla argumentalmente que se trata de una aproximación al estudio de los hechos lingüísticos de orientación empírica.

De la construcción de corpus se ocupa el capítulo 3, “Diseño, construcción y explotación de corpus”, que consiste en la descripción de las tareas y los aspectos más importantes del trabajo con estos recursos, y en el que se detallan los rasgos de diseño requeridos

para lograr una explotación adecuada de los datos, como son la representatividad, el formato, el volumen, la codificación y la anotación.

La aplicación de las técnicas ya explicadas a fenómenos léxicos y gramaticales son materia de estudio de los capítulos 4 y 5, respectivamente: “Recuperación de información contenida en corpus textuales: el léxico” y “Recuperación de información contenida en corpus textuales: fenómenos gramaticales”. En ellos se han seleccionado ejemplos prácticos reales que atienden a los diferentes ejes variacionales (diacrónico, sincrónico, diatópico y diastrático).

El capítulo 6, “Otras cuestiones centrales en lingüística de corpus”, trata de cuestiones generales relativas a la estructura estadística y al tamaño de los corpus, así como de las ventajas de su uso.

Por último, el capítulo 7, “Herramientas de recuperación de datos: resumen y ampliación”, es una introducción a la ciencia de los datos y se centra en cuestiones avanzadas para aplicaciones de consulta y herramientas informáticas de utilidad no tratadas en capítulos anteriores.

Cada capítulo concluye, asimismo, con un apartado dedicado a lecturas complementarias recomendadas y a cuestiones, problemas y temas de investigación que contienen actividades que retan y motivan a poner en práctica lo aprendido. El libro se cierra con un glosario que sirve de consulta rápida, un índice temático y las referencias bibliográficas a recursos y obras citadas en el texto.

El estudiante o investigador, tanto iniciado como experto, que se acerque a esta obra encontrará un recorrido teórico ordenado, pero también curiosidades, ejemplos de extracción de datos con resultados empíricos, ideas, líneas y temas de investigación que inspiran y estimulan a nuevas investigaciones. A su vez, en el contexto universitario, el docente tiene a su alcance un manual de uso pedagógico cuyos capítulos pueden reforzar y complementar la exposición en el aula.

En la obra se promueve, por tanto, una actitud activa de intercambio con un lector interesado por problemas lingüísticos, al que se dota del conocimiento y las habilidades necesarios para llevar a cabo su propia investigación basada en corpus ya existentes —a la vez que ofrece los pasos para diseñar el suyo propio—, y al que se le capacita para tomar el relevo del maestro/autor.

A lo largo de la lectura se desvela un entusiasmo contagioso por compartir con la comunidad científica, para inspirarla e impulsarla a utilizar la metodología, los recursos y las herramientas de que dispone la LC, gracias a los avances, entre otros factores, de la lingüística computacional. En definitiva, un libro que eleva el estatus de la investigación con datos de corpus y que da fe del compromiso de Guillermo Rojo con la comunidad científica a que va destinado.

## REFERENCIAS BIBLIOGRÁFICAS

- Parodi, Giovanni (2010). *Lingüística de Corpus: de la teoría a la empiria*. Iberoamericana Vervuert.  
McCarthy, Mike & Anne O’Keeffe (2015-2021). *Routledge Corpus Linguistics Guides*. Oxon, UK: Taylor & Francis.



**Jorge Fernández Jaén (2019).**  
***El abecé de la lingüística cognitiva.***  
**Arco Libros. 96 páginas. ISBN: 978-84-7635-996-9**

IRENE ARANCE RUSILLO  
Universidad de Córdoba  
i92arrui@uco.es  
<https://orcid.org/0000-0003-4177-4531>



Siguiendo la orientación de obras como *Introducción a la lingüística cognitiva* (Josep Cuenca y Hilferty Longanecker, 1999), *Cognitive Linguistics* (Croft y Cruse, 2004) y *Lingüística cognitiva* (Ibarretxe-Antuñano y Valenzuela, 2012), Fernández Jaén realiza una breve introducción a la lingüística cognitiva en *El abecé de la lingüística cognitiva* (2019), promoviendo así la difusión de la disciplina. La obra, junto con otras publicaciones como *La metáfora* (Masid Blanco, 2019) o *La meronimia* (Regueiro Rodríguez, 2018), se enmarca en un contexto actual en el que prima la divulgación.

La finalidad de la obra es, tal y como se indica en su prólogo (pp. 9-10), mostrar los conceptos y principios fundamentales de esta ciencia en un análisis claro y conciso; por ende, está orientada especialmente a un público que no sea experto en la materia como, por ejemplo, estudiantes de grado o de máster de carreras de filologías o traducción.

A propósito de su estructura, esta obra se divide en cuatro capítulos diferentes, organizados según su enfoque o las teorías que estos incluyen. Así, tras el prólogo, en el que se introducen algunas nociones generales de la disciplina para enmarcar la obra y su justificación, Fernández Jaén realiza una división de los capítulos basándose en una clasificación de los conceptos que se proporciona en el primer capítulo. Los capítulos del dos al cuatro se basan en las habilidades cognoscitivas humanas de Croft y Cruse: el capítulo 2 se centra en las teorías y conceptos relacionados con la atención, el capítulo 3 en aquellos relacionados con la comparación y el 4, con la perspectiva. Finalmente, se incluye una serie de ejercicios con sus correspondientes soluciones.

**Para citar esta reseña:** Arance Rusillo, I. (2022). Jorge Fernández Jaén (2019). *El abecé de la lingüística cognitiva*. Arco Libros. 96 páginas. ISBN: 978-84-7635-996-9. *ELUA*, (38), 343-345. <https://doi.org/10.14198/ELUA.22728>

Recibido: 16/05/2022 Aceptado: 09/06/2022

© 2022 Irene Arance Rusillo



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



En detalle podemos comentar que el contenido del primer capítulo es introductorio, pues en él se explican las bases del enfoque cognitivo y se nombran los lingüistas más influyentes en el campo, como Lakoff, Fillmore o Langacker. Se parte de una distinción entre la lingüística cognitiva y otros enfoques lingüísticos como el estructuralista y el generativista, y, a partir de esta, se definen nociones más concretas (el *compromiso cognitivo* de la disciplina, las *hipótesis básicas*, p. 12, en las que se basa la división de la obra, o la *corporeización lingüística*, entre otras). Tipográficamente, los conceptos se introducen utilizando la cursiva, lo que permite identificarlos rápidamente. Además, se incluye una breve definición de cada término para después explicarlo de forma más detallada, beneficiando a aquellos estudiantes que todavía no conocen la materia y necesitan investigar sobre algún término específico. Esto supone una diferencia con respecto a otras obras como la de Croft y Cruse, citada por el mismo autor, pues, aunque los conceptos se aborden en ambas, adoptan una formulación más didáctica en la obra de Fernández Jaén y, consecuentemente, más superficial.

Siguiendo la estructura mencionada anteriormente, el segundo capítulo versa sobre la atención y el significado verbal. Dentro de este capítulo se hace una división entre los marcos semánticos y el nivel de especificidad y la metonimia; esta última se presenta como resultado de las teorías del dominio cognitivo o del marco semántico. En este capítulo se incluyen figuras como, por ejemplo, tablas para explicar la teoría del nivel básico (p. 16) o la del dominio cognitivo (p. 21), facilitando la comprensión de estas teorías tan abstractas. Algunos de los ejemplos que se aportan en esta sección son clásicos, como “animal-perro” o “planta-árbol” en la teoría del nivel básico de concreción semántica. Otros permiten conectar los diferentes conceptos y comprenderlos mejor como la “playa”, que se utiliza para relacionar los marcos semánticos y extrínsecos de Fillmore, o “nudillo-dedo-mano”, que atañe al dominio cognitivo, ámbito y prominencia de Langacker.

Por su parte, el tercer capítulo trata sobre la comparación, la categorización y la ordenación y, análogamente al segundo capítulo, se realiza una división entre el proceso cognitivo (en este caso, la categorización) y su resultado en las expresiones lingüísticas, como la metáfora. Así, el autor entrelaza teorías de diversos lingüistas como el Modelo Cognitivo Idealizado (MCI) de Lakoff o la categorización continua de Rosch; además, pone de manifiesto las diferencias entre ellas y las ejemplariza con el objetivo de facilitar su comprensión. Por ejemplo, la metáfora “feliz es arriba” (p. 40) establece las diferencias entre los conceptos de la *metáfora cognitiva* y los *esquemas de imagen*, pues son teorías que, para aquellas personas que no conozcan las nociones más relevantes de la disciplina, pueden ser confusas al estar relacionadas. Por ende, el autor emplea de manera conveniente varios ejemplos para dejar patente las principales diferencias y características de estas teorías.

No obstante, el cuarto capítulo cuenta con una distribución diferente, ya que se presentan tres teorías: la gramática cognitiva de Langacker, la teoría de Talmy y la iconicidad lingüística. En algunos subapartados de este capítulo se parte de una distinción con la lingüística generativista, tal y como se hizo en el primer capítulo, pero con una mayor profundidad e ilustrando las diferencias que hay entre ambos enfoques con conceptos concretos como la *gramática universal* de Chomsky y la *gramática cognitiva* de Langacker. Asimismo, el autor introduce las teorías de varios autores en diferentes secciones, como es el caso de la teoría de Langacker y la de Talmy. De esta manera, explica ambas por separado para, en última instancia, relacionar algunos de los conceptos como *movimiento ficticio* y *orientación deíctica*. Al tratarse de conceptos relacionados con la gramática, tener un mayor grado de

abstracción y existir una mayor cantidad de nociones, explicar de una manera abstracta las teorías podría dificultar la comprensión a aquellos lectores que no poseen ningún conocimiento con respecto a estas, sin embargo, la ejemplificación de los conceptos después de las definiciones, al igual que en el resto de la obra, contribuye a una mejor asimilación. Del mismo modo, el autor incluye figuras como la tabla de los esquemas sintácticos y, para cada noción, incluye varios ejemplos ilustrativos para ofrecer diferentes puntos de vista para un mismo concepto (por ejemplo, en la jerarquía de empatía, p. 68).

Finalmente, se incluye una serie de ejercicios con sus correspondientes soluciones en las que prima una breve explicación, recapitulando así los conceptos. En ellos se combinan los términos y nociones de diferentes teorías, como en el segundo ejercicio, en el que se incluyen conceptos de las teorías de Langacker, Haiman, Fillmore y Talmy. Además, en algunos casos se incluye la solución comentada como en “este ingeniero se ha comprado un traje nuevo [figura] porque tiene una entrevista de trabajo [fondo]” (p. 87).

Como se aprecia en todas las secciones, el objetivo de una obra como *El abecé de la lingüística cognitiva* es didáctico, pues prima la ilustración de las teorías más relevantes para el enfoque y los diferentes conceptos relacionados con ellas. La constante ejemplificación permite que hasta un público lego pueda comprender las nociones más abstractas y el uso de la cursiva en los conceptos que, justo a continuación, incluyen una definición facilita la lectura y su posterior búsqueda y localización en caso de que sea necesario. El carácter pedagógico y el enfoque de la obra son los aspectos más destacables, ya que, gracias a estos, se consigue su propósito.

## REFERENCIAS

- Croft, W. y Cruse, A. (2004). *Cognitive Linguistics*. Cambridge University Press.
- Fernández Jaén, J. (2019). *El abecé de la lingüística cognitiva*. Arco/Libros, S. L.
- Ibarretxe-Antuñano, I. y Valenzuela, J. (2012). *Lingüística cognitiva*. Anthropos.
- Josep Cuenca, M. y Hilferty Longanecker, J. C. (1999). *Introducción a la lingüística cognitiva*. Editorial Ariel, S. A.
- Masid Blanco, O. (2019). *La metáfora*. Arco/Libros, S. L.
- Regueiro Rodríguez, M. L. (2018). *La meronimia*. Arco/Libros, S. L.



**La investigación en marcha. Reseña a Santos Gargallo, Isabel y Pastor Cesteros, Susana (dirs.) (2022). *Metodología de la investigación en la enseñanza-aprendizaje del español como segunda lengua (L2)/ lengua extranjera (LE)*. Arco Libros, 391 pp., ISBN 978-84-71353-864-8**

MIGUEL ÁNGEL MORA SÁNCHEZ  
Universidad de Alicante/ Escuela Oficial de Idiomas de Alicante  
[ma.mora@ua.es](mailto:ma.mora@ua.es)  
<https://orcid.org/0000-0002-3497-2006>



En los últimos años, y especialmente desde la implementación en el sistema de enseñanza superior del modelo de grados y másteres, los trabajos de investigación en el ámbito de la enseñanza del español como lengua extranjera (ELE) o lengua dos (EL2) se han visto incrementados de forma considerable, es decir, de forma cuantitativa. Harina de otro costal será si ese incremento numérico ha ido acompañado de una mejora en la calidad de los mismos y, a su vez, de un deseado progreso en la investigación de la lingüística aplicada a dicho campo de estudio. Tampoco resultan ajenas las condiciones en las que se han venido realizando dichas investigaciones, tanto en lo que se refiere a posibilidades de financiación, como a los referentes intelectuales y a las herramientas epistemológicas necesarias en todo proceso investigador. Es por eso que las editoras de este libro han marcado, de forma muy acertada, los condicionantes que la sociedad actual impone a la investigación, en general, y por tanto a la que se aplica en ELE, en particular. El interés creciente por el conocimiento de lenguas y, por ende, su estudio y acreditación, va acompañado de la preocupación por la didáctica, donde conviven viejas y nuevas formas de enseñar, en una constante transformación y con una diversidad de métodos que no hacen sino enriquecer las múltiples perspectivas de análisis (p. 12 y ss.). Todos estos condicionantes han llevado a plantear la organización de una guía de investigación el ELE que sirva de punto de apoyo imprescindible para llevar a buen puerto la labor investigadora

**Para citar esta reseña:** Mora Sánchez, M. Á. (2022). La investigación en marcha. Reseña a Santos Gargallo, Isabel y Pastor Cesteros, Susana (dirs.) (2022). *Metodología de la investigación en la enseñanza-aprendizaje del español como segunda lengua (L2)/ lengua extranjera (LE)*. Arco Libros, 391 pp., ISBN 978-84-71353-864-8. *ELUA*, (38), 347-352. <https://doi.org/10.14198/ELUA.22809>

Recibido: 25/05/2022 Aceptado: 26/05/2022

© 2022 Miguel Ángel Mora Sánchez



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

futura en el campo de nuestra especialidad. La ausencia de herramientas de apoyo de este tipo estaba muy poco justificada, llegados al punto de producción científica en el que se encuentra la disciplina, de ahí que la aparición de este libro sea celebrada tanto por quienes se dedican a la labor investigadora desde hace tiempo, como por aquellos que tienen la intención de iniciarse en ella. Su función puede cumplir una labor de imbricación y optimización de estas dos vertientes de los recursos humanos con que cuenta la disciplina.

La obra, organizada en tres bloques, se abre con una primera parte a la que las directoras han denominado “El diseño de la investigación”, que aparece articulada en cuatro capítulos: 1. Fundamentos de metodología, de Susana Martín y Ana Doquin; 2. El proceso de investigación, de Ana Blanco; 3. La documentación, de Iria da Cunha; 4. El diseño de investigación y análisis de datos, de Irini Mavrou. El bloque en su conjunto aborda diversos aspectos referentes a los diferentes enfoques y perspectivas que se pueden afrontar en las investigaciones de ELE, pasando por el necesario proceso de revisión de la documentación, sobre todo, en un momento como el actual de creciente producción de trabajos, donde la repetición de algunos temas, ideas y métodos de investigación resultan, cuando menos, sorprendentes. En el capítulo dedicado a los fundamentos de metodología, las autoras realizan una distinción epistemológica sobre las diferentes formas de la investigación. Destaca la vinculación de la clasificación de los tipos de estudios con las perspectivas de las preguntas de investigación, lo que da lugar a una distinción básica entre estudios con perspectiva hipotético-deductiva y estudios con perspectiva exploratoria o heurística. Sin duda, la necesidad de trabajar con una metodología experimental es algo que se viene demandando desde el propio ámbito de la investigación, donde también se denuncia el desequilibrio entre la elección de modelos cuantitativos y cualitativos, en favor del segundo sobre el primero<sup>1</sup>. Este hecho resulta palmario en el repaso de las líneas de investigación en ELE/EL2 que hacen las autoras al final de este capítulo, donde se observa una clara prevalencia de las áreas de carácter teórico. Si nos metemos ya de lleno en el segundo capítulo, vemos que aparece el tema del proceso de investigación, que se ocupa de identificar los pasos que forman parte de este proceso y su secuenciación. El esquema básico se centra en cuatro pasos: planteamiento, exploración, diseño y ejecución. Está claro que en cualquiera de ellos la persona investigadora puede encontrar obstáculos que den pie al replanteamiento de la investigación, que en ningún caso ha de ser traumático, sino que, muy al contrario, ha de suponer un enriquecimiento de los planteamientos iniciales. El tercer capítulo de este apartado gira en torno a la documentación necesaria en todo trabajo de investigación y sobre el no deseado plagio de trabajos o ideas anteriores. La importancia de la documentación pasa por el hecho de que una investigación científica que se precie no surge de la nada y ha de contar con una documentación exhaustiva, esto es, completa: no pueden quedar fuera trabajos anteriores sobre aspectos relacionados con dicha investigación, ya que se corre con uno de los peligros de los adolecen algunos de los trabajos ya realizados en nuestro campo, donde se produce una no deseada repetición de ciertos aspectos de la investigación en ELE/L2, a modo de vueltas de tuerca<sup>2</sup>, que claramente impiden el avance de las propuestas en nuestra disciplina. Y llega-

1 Llamativas son las relativamente recientes palabras de una reconocida profesora y formadora de profesores: “los estudios cualitativos son abrumadoramente mayoritarios y la teoría que defiende sus virtudes frente a los estudios cuantitativos, cada vez más abundante”. Ruiz Fajardo, G. (2016). El enfoque comunicativo en los libros de texto de español como nueva lengua, en Ruiz Fajardo, G. y Ríos Rojas, A. (eds.). *Inmigración: nuevas lenguas, nuevos hablantes, nuevas perspectivas*. Universidad Internacional de Andalucía, p.119.

2 Esta expresión la tomamos de Guadalupe Ruiz Fajardo, en el trabajo citado (2016): 117

mos al último capítulo de este conjunto de trabajos, centrado en el diseño de investigación y análisis de datos. Dentro del diseño de una investigación en particular encontramos elementos imprescindibles para el desarrollo de esta como son: las preguntas de investigación, la selección y muestreo de los participantes, las unidades de análisis, las variables y los análisis de datos, así como sus consecuencias. Todos estos elementos se postulan en torno a dos grandes paradigmas de investigación: el cuantitativo y el cualitativo. Aunque desde el punto de vista teórico, ambos tienen validez en el ámbito de la investigación, la producción científica en el campo de ELE se ha venido decantando más por el cualitativo que por el cuantitativo, dejando un poco de lado la necesidad que tiene la introspección de salir fuera y ofrecer estudios de orden cuantitativo que alivien la impresión de que muchas veces trabajamos con intuiciones, en ocasiones alejadas de la evidencia científica. No somos los primeros -y, por desgracia, tampoco creemos que los últimos- en denunciar la distancia entre la investigación y su aplicación práctica en el aula<sup>3</sup> que aqueja a nuestra disciplina.

El segundo bloque de la obra lo ocupan “los métodos e instrumentos de investigación”. A partir de los dos capítulos que abordan los dos grandes paradigmas de la investigación lingüística en el aula de lenguas extranjeras, a saber, el cualitativo y el cuantitativo (Mercedes Pérez y Jon Andoni Duñabeitia, por un lado, e Isabel García, por el otro), se pasa al estudio de caso, a cargo de Macarena Ortiz, centrado en un análisis teórico y una presentación de una tipología de estudios de este tipo. El siguiente paso va en la línea de la investigación en el aula, que sin duda despierta gran interés entre los profesores en activo, con inquietudes investigadoras y de mejora docente, y sobre el que se centra el capítulo de Olga Esteve Ruescas. La autora hace hincapié en la necesaria repercusión de este tipo de investigación sobre el aula y, por ende, sobre sus protagonistas, tanto el alumnado como el profesorado.

Sobre el enfoque de la investigación en acción versa el siguiente capítulo del bloque dos del libro, de la mano de Alicia Hernando Velasco, donde tras un recorrido a la historia de esta orientación, se centra en el proceso de investigación como tal y sus instrumentos (preguntas y cuestionarios), que van a ser el objeto de análisis del que le sigue, de la mano de M.<sup>a</sup> Cecilia Ainciburu. En este se intenta dejar clara la distinción entre pregunta, cuestionarios y entrevistas, y se articula una provechosa reflexión sobre la insoslayable validación de estos instrumentos. En el capítulo siguiente se avanza hacia los trabajos de investigación referidos a los materiales didácticos, por parte de Ernesto Martín Peris, entre los que destaca el libro de texto, aunque no solo se centra en este recurso didáctico. Destacan algunas propuestas muy tenidas en cuenta en trabajos que menciona en su repaso histórico de las investigaciones de esta índole en las últimas tres décadas. Pero, sin duda, un recurso frecuente, de los más recientes trabajos de investigación en ELE, es el de los corpus lingüísticos, de los que se ocupan Mar Cruz Piñol y Krys Bruyse. Los autores muestran la aplicación de los recursos tecnológicos a la investigación didáctica. Su trabajo – al igual que otros anteriores – ilustra cómo se trabaja con estos corpus y clasifica las herramientas existentes con la intención de facilitar la elección

---

3 En la colaboración de una reciente publicación Pedro Gras expresa esa queja con mucha claridad: “Por lo general no existe una comunicación fluida entre los lingüistas teóricos y los profesores de lenguas”. En Gramática de construcciones para profesores de ELE, Herrera, F. y Sans, N. (coords.) (2018). *Enseñar gramática en el aula de español. Nuevas perspectivas y propuestas*. Difusión, p. 65. También Teresa Cadierno se pregunta “si estos dos mundos [la lingüística cognitiva y la enseñanza de ELE] pueden aprender el uno del otro y así beneficiarse mutuamente”, en *Tendiendo puentes entre la lingüística cognitiva y el español LE/L2*, Ibarretxe-Antuñano, I., Cadierno, T. y Castañeda Castro, A. (eds.) (2019). *Lingüística cognitiva y español LE/L2*. Routledge, p. 3.

de la más adecuada a los intereses de la persona investigadora, de manera que puedan alcanzar los objetivos de su investigación con éxito. Pero, unido al conocimiento de estas herramientas de corpus, resulta también imprescindible cierta habilidad en el tratamiento estadístico de los resultados obtenidos. De esta estadística para lingüistas se ocupa M.<sup>a</sup> del Rosario Martínez Arias en el siguiente capítulo. Este conocimiento es insoslayable en las investigaciones que se acogen al paradigma cuantitativo. A partir de uno de los programas estadísticos más utilizados en la investigación en ELE, la autora reflexiona sobre la necesidad de saber interpretar con el suficiente rigor metodológico y científico los datos resultantes de investigar con aplicaciones estadísticas. El cierre de este segundo bloque de trabajos, se hace con un tema, que no por novedoso – en el tratamiento teórico, sobre todo – deja de ser relevante: la ética de la investigación científica. Este tema lo centra Alberto Rodríguez Lifante en un repaso de los conceptos clave en relación a la ética de la investigación, por un lado, y a la ética en la investigación, por otro. Resulta evidente que este es un aspecto que afecta tanto al diseño de la investigación como al método o a los instrumentos en su aplicación, de ahí la importancia de enmarcarlo dentro de la investigación en ELE.

Llegados al tercer bloque que conforma este libro, las directoras del mismo han decidido agrupar cuatro trabajos bajo el epígrafe de “transferencia y difusión del conocimiento”, que se va a materializar en tres artículos centrados en géneros textuales académicos muy vinculados entre sí: el trabajo fin de grado (TFG), el trabajo fin de máster (TFM), la tesis doctoral y el artículo científico. El bloque lo cierra un cuarto artículo que aborda la cuestión de la difusión de todo este trabajo científico. No resulta casual que esta obra se cierre con un bloque centrado en las formas discursivas en las que acaba materializándose el trabajo de la persona investigadora, puesto que estas son manifestaciones propias de la cultura académica y científica, que son las encargadas de tender puentes entre la teoría y la práctica, entre la investigación y la aplicación al aula, entre los logros del mundo académico y los avances en la sociedad de la que parten y que las nutre.

De esta manera, en el capítulo quince, la profesora Mar Galindo Merino ofrece una práctica visión de dos tipos de trabajos académicos supervisados de amplia repercusión como son el TFG y el TFM. Resulta evidente el creciente interés que suscita el ámbito de ELE entre los recién graduados y la necesidad de una especialización, que, por otra parte, ha ido acompañada de una abundante y abrumadora oferta de estudios de máster. De igual forma, la tesis doctoral, como forma prototípica de investigación universitaria, entendida como la culminación de la formación del profesional de ELE, queda abordada en el capítulo dieciséis por Javier de Santiago-Guervós y Jesús Fernández González, capítulo en el que los autores ofrecen una visión abarcadora y global de todo el proceso, que supone el mayor grado de especialización en el ámbito académico para los egresados universitarios. Y está claro que la elaboración de un trabajo académico necesita, para que tenga una realización plena, de una proyección en la sociedad (y, ¿por qué no?, también dentro del mismo mundo académico), que se materializa de forma discursiva en el artículo científico que es el tema que aborda, en el capítulo diecisiete, Antoni Nomdedeu Rull, en el que se plantean los principios generales sobre los que se asienta la confección del artículo científico. Destacan en este trabajo las orientaciones prácticas para la elaboración, como género discursivo particular que es y suponen una perfecta guía de redacción. Y no hay duda sobre la importancia que tiene la transferencia para la investigación. Si la investigación no da pie a que la sociedad cambie, evolucione y mejore es difícil que se mantenga como tal investigación, de ahí que tan im-

portante como ella misma lo sea también su difusión y su capacidad de acercar estructuras nuevas de conocimiento a las prácticas profesionales. A ello contribuyen, sin duda, las revistas científicas y didácticas, objeto de análisis de esta última colaboración, de la mano de Paz Villar-Hernández y Santiago Mengual-Andrés que destaca la importante función como agentes de transferencia de conocimiento que tiene dichas publicaciones.

Quizá un tema que queda abierto sea el de cuál es el impacto real de esta transferencia, si los mecanismos de transferencia son efectivos, cuál es la realidad del alcance de las investigaciones realizadas, la capacidad de implementación de los hallazgos realizados. Esto nos conduce a nuevas preguntas, ¿quiénes son los agentes de la investigación y quiénes los beneficiarios? ¿Son los investigadores formadores de profesores de ELE? ¿Son los profesores de ELE también agentes de la investigación o solo los receptores primarios de ella? ¿Cuándo, cómo y en qué condiciones llegan al aula los resultados de la investigación? Lo que parece claro es que la investigación científica tiene que adaptarse a las nuevas formas sociales y a sus necesidades, por eso habría que entender que, junto a proyectos de investigación como tesis doctorales -que constituyen lo que nosotros denominamos *macroinvestigación*-, conviven pequeñas “píldoras” de investigación -que podríamos bautizar como *microinvestigación*-, que presentan las mismas exigencias de rigurosidad metodológica, pero que resultan más accesibles y prolongables a lo largo de la vida profesional del profesor-investigador de nuestra especialidad. Y es que no debemos contentarnos con que la investigación constituya un elemento exclusivo de la formación inicial de estos profesionales, sino que a ellos los debemos enfocar hacia una investigación a lo largo de toda la vida, como mecanismo para poder avanzar en la formación, el enriquecimiento y el progreso. Estas perspectivas de futuro son las que hacen tan necesario un libro como el que estamos comentando, porque en él se establecen las bases que garantizan el rigor y la calidad que deben formar parte de la labor docente e investigadora. Por eso, y para finalizar, nos quedaría rescatar una de las conclusiones optimistas, en torno a este tema, de la profesora Ruiz Fajardo -ya citada-, acerca del estado de la investigación en ELE, en la que nos estamos acostumbrando a “por un lado, apreciar las virtudes de la investigación y por otro, a dejar de inventar la rueda”.<sup>4</sup>

Lo único que nos queda, pues, es agradecer a las profesoras Susana Pastor Cesteros e Isabel Santos Gargallo haber realizado este encomiable trabajo de coordinar a un amplio elenco de especialistas en la investigación en ELE y sus contribuciones, en aras de facilitar la imprescindible labor de investigación tanto de neófitos como de reincidentes -y recalci-trantes- en estas lides, que aporten su granito de arena al progreso, la mejora y el buen hacer del trabajo diario de los que nos dedicamos a esta hermosa tarea de guiar por el español a aquellas personas que quieren/necesitan desenvolverse en una nueva lengua y en una nueva forma de vida. Esperamos que este no sea su único trabajo de colaboración puesto que el resultado ha sido esta imponente obra de referencia para la investigación en ELE.

## REFERENCIAS BIBLIOGRÁFICAS

- Brown, J. D (1988). *Understanding Research in Second Language Learning*. Cambridge-University Press.
- Centellas Rodrigo, A. (2021). *Lingüística aplicada. Adquisición del español como Lengua Extranjera*. En clave ELE-Udima.

4 Ruiz Fajardo 2016: 154.



- De Santiago Guervós, J. y Fernández González, J. (2017). *Fundamentos para la enseñanza del español 2/L*. Arco Libros.
- De Santiago-Guervós, J. y Díaz Rodríguez, L. (2021). *Lingüística textual y enseñanza del español LE/L2*. Routledge.
- Jiménez Calderón, F. y Rufat Sánchez, A. (eds.) (2019). *Manual de formación para profesores de ELE*. SGEL.
- Lacorte, M. (coord.) (2007) *Lingüística aplicada del español*. Arco Libros.
- Lacorte, M. y Reyes Torres, A. (2021). *Didáctica del español como 2/L en el siglo XXI*. Arco Libros.
- Larsen- Freeman, D. y Long, M. H (1991). *An Introduction to Second Language Research*. Logman.
- Muñoz-Bassols, J., Gironzetti, E. y Lacorte, M. (2019) *The Routledge Handbook of Spanish Language Teaching. Metodologías, contextos y recursos para la enseñanza del español L2*. Routledge.
- Pastor Cesteros, S. (2017). *Lingüística aplicada a la enseñanza de segundas lenguas. Panorama actual y líneas de investigación*, en M. E. Olimpio de Oliveira Silva e I. Penadés Martínez (eds.). *Investigaciones actuales en Lingüística. Vol. I: Sobre la Lingüística y sus disciplinas*. Universidad de Alcalá, 175-194.
- Ruiz Fajardo, G. (ed.) (2012). *Methodological Developments in Teaching Spanish as Second and Foreign Language*. Cambridge Scholars.
- Sánchez Lobato, J. y Santos Gargallo, I. (eds.) (2004). *Vademécum para la formación de profesores de español como segunda lengua (L2)/ lengua extranjera (LE)*. SGEL.

**Galindo Merino, M. M. y M. C. Méndez (eds.) (2022).**  
***La lingüística del amor: de la pasión a la palabra.***  
**Pie de página editorial, 322 páginas**

ANASTASIA KHAYLINA  
Universidad de Alicante. Grupo GRIALE, España  
[anastasia.khaylina@ua.es](mailto:anastasia.khaylina@ua.es)  
<https://orcid.org/0000-0002-5699-0096>



El lenguaje es el motor que rige, en cierto modo, nuestro mundo. Las palabras son capaces de alterar realidades, modificarlas e incluso transmitirlos. Así se concibe este libro, como una relación entre la realidad, el amor, y su manifestación, las palabras. Sin embargo, en vez de quedarse en la superficie del concepto *amor*, este libro consi-gue indagar en profundidad en las diversas relaciones sexoafectivas, a través de análisis multimodales, fonéticos, psicológicos, entre otros, que muestran la manera tan peculiar y variada que tenemos como sociedad de expresar estas relaciones a través del lenguaje.

El libro se estructura alrededor de dieciséis capítulos sin enumerar, elaborados por veinte especialistas, con un prólogo y un epílogo de la mano de las dos editoras. Cada capítulo está introducido por un currículum breve de la persona autora y se estructura en varios apartados, las conclusiones y las correspondientes referencias o bibliografía relevante. Esta disposición permite una lectura dinámica y ligera, además de la posibilidad de acudir a las fuentes originales sin tener que adelantarse al final del libro. Si algo cabe remarcar del estilo en el que está redactado el libro en su totalidad es la cohesión, tanto visual como narrativa, donde prepondera la cercanía hacia el/la lector/a, sin emplear un registro demasiado culto, pero manteniendo la distinción científica.

El prólogo lo presenta una de las editoras del libro, doctora en Lingüística aplicada y profesora de Lingüística en la Universidad de Alicante, Mar Galindo. Ella realiza un adelanto breve de las temáticas que se tratarán a lo largo de los capítulos y presenta a los autores que han aportado sus perspectivas como especialistas, invitando a seguir leyendo.

---

**Para citar esta reseña:** Khaylina, A. (2022). Galindo Merino, M. M. y M. C. Méndez (eds.) (2022). *La lingüística del amor: de la pasión a la palabra*. Pie de página editorial, 322 páginas. *ELUA*, (38), 353-358. <https://doi.org/10.14198/ELUA.22806>

Recibido: 25/05/2022 Aceptado: 26/05/2022

© 2022 Anastasia Khaylina



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

Debido a que no podemos hablar de un orden numérico de capítulos, se mencionará el título y el nombre de los autores en cada uno de ellos. El primer capítulo lo presenta la profesora y escritora Tamara Tenenbaum, titulado *De Romeo y Julieta a HBO: El amor y la conversación*. Aquí, la autora nos ofrece un recorrido a lo largo de varias series, musicales y novelas, como *El violinista sobre el tejado* (1971), *Casablanca* (1942) o las obras de Jane Austen, *Orgullo y prejuicio* y *Sensatez y sentimientos*. Estas obras cinematográficas y literarias le permiten hablar sobre la relevancia de la conversación como pilar fundamental de las relaciones románticas, cómo estas influyen en la construcción de estas relaciones y cómo afecta la ausencia de la conversación al desarrollo óptimo de la pareja romántica. Asimismo, se hace énfasis sobre el cambio de paradigma a la hora de considerar la dinámica y los roles dentro de la pareja, el giro hacia una búsqueda de la construcción consciente y deconstruida que observamos, cada vez más, en la sociedad actual.

El siguiente capítulo lo presenta la consultora especialista en habilidades de comunicación personal, Teresa Baró, bajo el título *La conducta no verbal en los rituales de seducción. Cuerpo, escenario y ritmo*. Tras la breve introducción se habla sobre los rituales de seducción, las diferencias de género en el acto del cortejo, ademanes y actitudes para el cortejo y las señales de falta de interés. Durante todo el capítulo, la autora enfatiza la importancia de la comunicación no verbal como parte fundamental en el contacto inicial entre dos personas, realizando paralelamente una comparativa con el ritual del cortejo en el reino animal para señalar precisamente ese predominio del instinto y el impulso que nos guía a la hora de conocer a alguien que nos interesa, o no. Sin embargo, aunque este análisis se realiza desde un punto de vista binario de los roles de género, sin profundizar en otras dinámicas, la propia autora menciona que “el reto que tenemos todos en esta primera mitad del siglo XXI es movernos en una época de cambio, entender la diversidad de opciones y de actitudes, adaptar nuestra propia comunicación a nuevos patrones de relación” (p. 46).

En las siguientes páginas encontramos *Las distintas etapas del amor a través de la voz*, escrito por la profesora titular del departamento de Lengua Española y Lingüística General de la UNED, Nuria Polo. En él, la voz y el habla ocupan el lugar central del análisis. La autora nos guía a lo largo de un análisis, basado en diversos estudios científicos (Klatt y Klatt, 1990; Dabbs y Mallinger, 1999; Giles, 2008; Fitch y Reby, 2011; Geneid *et al.*, 2015; Kirgezen *et al.*, 2017; Lã y Polo, 2020), acerca de cuánta información pueden llegar a transmitir nuestras voces y de la atracción que puede existir hacia la voz de una persona. Para finalizar, la autora proporciona una relación de códigos QR que permiten consultar rápidamente algunas de las referencias relevantes mencionadas a lo largo del capítulo.

También encontramos códigos QR en el capítulo de la catedrática de Lingüística General, Iraide Ibarretxe-Antuñano *La pasión vasca... y otras historias de amor <<incardinado>>*. No obstante, la autora ha añadido también una variedad de imágenes, emoticonos, capturas de videoclips, fotos y tablas que acompañan sus explicaciones sobre las diversas metáforas conceptuales relacionadas con el amor y sus expresiones metafóricas, en diferentes lenguas, tanto orales como de signos, y culturas. El hecho de pensar automáticamente en el corazón como el centro de las emociones y concretamente del amor, se lo debemos, en gran parte, a Descartes y su modelo dualista que separa la cabeza del corazón, y, por consiguiente, el intelecto del corazón. Sin embargo, esto no sucede del mismo modo en todas las culturas del mundo debido a que las expresiones metafóricas son diferentes. Por ejemplo, la autora señala cómo el modelo abdominocéntrico, común en el sudeste asiático, define el

hígado como el centro de las emociones, el pensamiento y el deseo. Mientras que el modelo cardiocentrista de las culturas asiáticas relacionan el corazón con el amor y el intelecto. El euskera, por otro lado, recupera la dualidad del intelecto/corazón, pero los ubica a ambos en un concepto incorpóreo: *gogo*.

El quinto capítulo lo presentan Moisés Fernández Cano, doctorando y miembro del *LGTBIQ+ and Feminist Working Group* en la European University Institute, y Aarón Pérez Bernabeu, doctorando en la Universidad de Alicante y coordinador de formación en línea de la editorial Edinumen, sobre *El lenguaje del deseo queer: códigos de supervivencia tras las bamblinas de un espacio silenciado*. En este capítulo se explora precisamente el funcionamiento de la comunicación sexo afectiva del deseo *queer*: El estudio comienza en la Lingüística Queer, o *Lavender Linguistics*, que debe su denominación al ‘Terror lila’ (extrapolación del término ‘Temor rojo’) o *Lavender Scare*. A través de este marco teórico los autores nos muestran los códigos particulares que se han creado dentro de estos espacios no ‘cisheternormativos’. Estos códigos semióticos, o “universo simbólico queer” (p. 103), incluyen diversos términos sin marcas de género que funcionan en la conversación como marca de alerta para los demás participantes, además de las maneras en las que es posible comunicar a través del silencio o el uso de colores/emoticonos para expresar las preferencias sexuales.

Prosiguiendo con las particularidades del lenguaje, la directora editorial Mar Abad realiza un recorrido a través del lenguaje que se forja entre las parejas en su capítulo titulado *El lenguaje de las parejas*. Como hablantes tendemos a la búsqueda y creación de nuevas palabras cuando sentimos la necesidad de expresar nuevas realidades, tal y como sucede con el lenguaje en las redes. Lo mismo ocurre, consecuentemente, en las relaciones de pareja, donde la intimidad y cercanía potencian la creación de un código común y exclusivo para denominar realidades compartidas, donde las anécdotas conllevan a la preferencia de uso de una palabra frente a otra: ‘cocreatas’ frente a croquetas o ‘tasis’ en vez de ‘taxi’ (p. 117).

Los autores del siguiente capítulo son Clara Planchuelo, investigadora predoctoral y miembro del CINC, y Jon Andoni Duñabeitia, catedrático en la Facultad de Lenguas y Educación de la Universidad Antonio de Nebrija. Este capítulo, *El sexo léxico*, se desarrolla partiendo del planteamiento inicial acerca de cómo los distintos sexos emplean el léxico y cómo los significados sexuales se recogen en el nivel léxico de una lengua. Además, se investiga sobre las valencias y el *arousal* de las palabras relacionadas con el sexo, es decir, su efecto emocional sobre el hablante, a través del *Self-Assessment Manikin* (Bradley y Lang 1994). Finalmente, se plantea la cuestión en torno al uso de las palabras de contenido sexo afectivo en lenguas extranjeras y qué efectos tendría en comparación con su uso en la lengua nativa correspondiente, ¿conllevaría esto a un distanciamiento emocional mayor con respecto a la pareja y su percepción de la relación?

En las siguientes páginas encontramos el capítulo sobre *El lenguaje del sexo en la traducción de novelas eróticas* de la traductora literaria Scheherezade Surià. Aquí la autora hace un recorrido a través del vocabulario implicado en las obras eróticas en inglés y en español agrupado en diversas temáticas: mobiliario, ropa interior, juguetes y accesorios. Las tablas que ha incluido la autora ayudan a una visión clara e ilustrativa, y permiten entrever la profunda investigación lingüística que deben realizar los traductores debido a la diversidad de conceptos existentes. Igualmente destaca temas problemáticos como la censura, el contexto o la moralidad del traductor, ya que los términos sexuales suelen conllevar sus connotaciones que cabe tener en cuenta para lograr el propósito intencional inicial del escritor original.

Tras pasar el ecuador del libro nos encontramos con el capítulo de Ártemis López, traductora e intérprete sanitaria certificada de español-inglés en ambas direcciones. En *Cómo decirle a una persona trans que la quieres sin meter (mucho) la pata* relata la incomodidad del contexto en el que se encuentra repetidamente el colectivo trans durante ciertas conversaciones debido a los temas que se implican. Menciona algunas de las microagresiones (no tan micro) que se sufren a diario, como las preguntas invasivas, la cuestión del nombre y género gramaticales o las conversaciones en torno a los cuerpos trans. Incluso cuando las intenciones parecen ser ‘buenas’ nos vemos abocados al cuestionamiento de la identidad e integridad de la persona con la que tratamos. La manera de evitarlo sería “formándote cada día, intentando evitar las microagresiones (conscientes e inconscientes) y aprendiendo a gestionar tus errores” (p. 179).

En la línea del estudio de las metáforas conceptuales encontramos también a la autora del siguiente capítulo, Reyes Llopis García, que nos introduce en *Las metáforas enamoradas y los abrazos metonímicos* con banda sonora incluida a través del QR que se puede encontrar en la introducción. ¿Qué hay más genuino en el amor que su capacidad para unir las mentes? Así es como se conciben las metáforas conceptuales. Nos permiten expresar experiencias concretas a partir de un dominio conceptual común; el afecto se expresa a través de la temperatura (p. 187), la vida y el amor son expresados como un viaje (p. 190), mientras que el corazón funciona, una vez más, como metonimia a diversidad de expresiones amorosas (pp. 194-196). Tan fuertes son estas metáforas conceptuales que, como nos recuerda la autora, nos han unido en la época de la pandemia, donde los aplausos nos permitieron manifestar el amor o la gratitud que sentíamos como sociedad frente a los/as trabajadores/as sanitarios/as.

El siguiente capítulo lo presentan Carla Botella Tejera, doctora en Traducción e Interpretación por la Universidad de Alicante, y María del Mar Ogea Pozo, doctora en Traducción e Interpretación por la Universidad de Córdoba, sobre *¿Por qué lo llaman amor cuando quieren decir sexo? La traducción de las relaciones amorosas en las series de adolescentes de hoy*. Las autoras se centran en mostrarnos la evolución que ha experimentado el léxico empleado por los jóvenes de las series (ahora principalmente de las plataformas digitales) a la hora de expresarse acerca del sexo, las relaciones o las drogas. Debido al cambio, todavía gradual, en el panorama sociocultural surge la necesidad en los hablantes de crear o acuñar terminología nueva, ya que de la necesidad de cambio nace la innovación. Para mostrar este cambio proponen una propuesta de categorías, que va desde el léxico sobre la amistad hasta las prácticas sexuales (p. 213). Además, realizan una comparativa cuantitativa entre los términos existentes en el idioma inglés y en el español, y que conformarían precisamente ese discurso jergal de los jóvenes.

La siguiente aportación al libro proviene de la lingüista forense, fundadora y directora del Laboratorio SQ-Lingüistas Forenses, Sheila Queralt. No todo lo que reluce es oro, y no todo el amor es bonito. Así nos lo muestra la autora en *Cuando el amor es una estafa* donde habla sobre cómo funcionan y se construyen las estafas amorosas, motivadas por un interés puramente económico bajo pretexto romántico. La autora ofrece una breve definición de la Lingüística forense y una clasificación de las tres áreas principales que la conforman. Para cometer el delito de manera satisfactoria se establece un vínculo rápidamente con la persona estafada mediante la falsa ilusión de confianza o preguntas personales e íntimas. Sin embargo, aunque el lenguaje pueda servir de herramienta para construir este engaño, también les permite a los expertos lingüistas elaborar un perfil lingüístico y recaudar pruebas lingüísticas para construir el caso contra el cibercriminal.

En las siguientes páginas, Pilar Martínez, psicóloga general sanitaria, presenta *Señales lingüísticas de una relación tóxica*. Mediante una exposición de ejemplos de comportamientos tóxicos y preconcepciones equivocadas, la autora explica la complejidad de las relaciones tóxicas y el trabajo individual que uno/a debe realizar a la hora de construir una relación con otra persona para no caer en ello. Remarca términos como la “responsabilidad afectiva” (p. 252), el “sometimiento” (p. 251) que predomina en las relaciones tóxicas o la “profecía autocumplida” (p. 254) que sucede cuando la pareja finalmente se cansa del abuso constante.

El capítulo siguiente, *Claves lingüísticas para ligar en los tiempos del Tinder*, lo han escrito M.<sup>a</sup> Carmen Méndez, una de las editoras del libro y doctora en Lengua española por la Universidade de Vigo, junto con la ayudante doctora en la Universitat de València, Esther Linares Bernabéu. Como objetivo de esta investigación se han propuesto estudiar el lenguaje que se emplea en las descripciones de perfil de los usuarios/as de Tinder, concretamente de orientación heterosexual. A través de un cuestionario online han estudiado qué prefieren las mujeres y los hombres a la hora de leer una descripción de perfil y qué les empuja a crear un *match*. Entre las temáticas más destacables han encontrado la mención a los gustos, el sentido del humor, el factor de la originalidad y la ortografía. Igualmente han observado qué adjetivos son más ‘atractivos’, cómo se debería comenzar correctamente una conversación y la fluidez que debe haber en ella.

Centrándose también en el plano virtual encontramos *Desvinculación a través del ghosting: fallos de comunicación en la era virtual*, escrito por Lucía Álvarez López, psicóloga general sanitaria y sexóloga. La autora señala que el *ghosting* supone un problema para las relaciones interpersonales en la era digital, ya que ocasionan, en la mayoría de los casos, efectos negativos sobre los hablantes que lo sufren. Refleja, además, una falta de empatía y responsabilidad afectiva (consultar capítulo de Pilar Martínez). Por lo tanto, debido a que la comunicación online cada vez cobra más peso en nuestras vidas y en la manera en la que nos relacionamos, es necesario cuidar las formas “para comunicarnos de un modo más ético con las otras personas” (p. 293).

En el último capítulo, *21 palabras para conjurar el amor*, los autores son Susana Rodríguez Barcia, profesora en el departamento de Lengua Española de la Universidade de Vigo, y Bosco Gil de Gárate Hernández, músico e investigador independiente. Curiosamente, al principio del capítulo han incluido un código QR con una lista de Spotify para acompañar la lectura. En este estudio han seleccionado 441 canciones de amor y sexo en inglés de diferentes géneros musicales, como el blues, soul, rock y pop de los años 1950-2020, para extraer las palabras más frecuentes. El resultado es el total de 21 sustantivos, entre los que encontramos *Love, Baby, Time, Heart* y *Eyes* como los cinco más frecuentes del corpus.

Para cerrar, presenta el epílogo la doctora María del Carmen Méndez Santos, bajo el título *El cigarrillo de después*, donde realiza un breve recorrido de clausura a través de todos los capítulos, mencionando las ideas centrales de cada uno y aportando un breve comentario personal a cada uno, para concluir con un agradecimiento al/a la lector/a con las palabras “ahora tocaría lo que todes ya sabemos: un cigarrillo. O si lo queremos enfocar desde un punto de vista más actual del siglo XXI y saludable, un vasito de *kombucha* o una partida al Candy crush” (p. 330).

En conclusión, en esta obra escrita por veintidós especialistas y editada por M.<sup>a</sup> Mar Galindo y M.<sup>a</sup> Carmen Méndez ofrece una visión extensa sobre la conexión de la lingüística con las relaciones sexo afectivas, y muestra la variedad de espacios y contextos en las que

estas interactúan. Intervienen además ramas como la neurolingüística, la sociolingüística, la pragmática, entre otras, para hablar sobre temas como la influencia de las voces a la hora del cortejo, el funcionamiento del lenguaje no verbal al conocerse, las dinámicas *queer*, las metáforas y metonimias del amor, la traducción del lenguaje sexual, la jerga juvenil, la conversación amorosa, el ligue digital, la manipulación... Todo ello gracias al uso del lenguaje y las diferentes estrategias lingüísticas, que tienen el amor por común denominador. En resumen, podemos admitir que nos encontramos frente a un libro con diversidad temática, escrito en un estilo claro y conciso, con un uso acertado del plano visual y con una estructuración fácil de seguir. Esta obra es recomendable para todo/a aquel/aquella que esté interesado/a en la lingüística, pero no sepa por dónde empezar, o para quien desee expandir sus horizontes dentro de este campo tan heterogéneo.







## Artículos

- Sara Pistola e Iria da Cunha, 17  
Antonio-José Silvestre-López, 47  
Marta Sánchez-Saus Laserna y Vanesa Álvarez Torres, 73  
María Ángeles Orts Llopis, 99  
Miguel A. Candel-Mora, 117  
Sofía Zakhir Puig, 137  
Émeline Pierre, 165  
Ana García Toro, 187  
Luis Pablo Núñez, 217  
Eva Lucía Jiménez-Navarro, 241  
Macarena Palma Gutiérrez, 263  
Mario Crespo Miguel, 281  
Eva María Llorens Simón, 301

## Reseñas

- Raquel Rodríguez Fernández, 327  
Neus Gimeno Gimeno, 333  
María Eugenia Conde Noguerol, 339  
Irene Arance Rusillo, 343  
Miguel Ángel Mora Sánchez, 347  
Anastasia Khaylina, 353

Dpto. Filología Española,  
Lingüística General  
y Teoría de la Literatura

2022



Universitat d'Alacant  
Universidad de Alicante

