

WEBLESP: CORPUS DE COMUNICACIÓN DIGITAL ESPECIALIZADA EN ESPAÑOL. ARQUITECTURA, COMPILACIÓN Y USOS

WEBLESP: CORPUS OF DIGITAL SPECIALISED COMMUNICATION IN SPANISH. DESIGN, COMPILATION AND USE

SARA PICCIONI

Università “G. D’Annunzio” di Chieti-Pescara, Italia
sara.piccioni@unich.it

GIANLUCA PONTRANDOLFO

Università di Trieste, Italia
gpontrandolfo@units.it

RESUMEN

El siguiente artículo presenta el Corpus electrónico WebLesp que reúne textos representativos de la comunicación digital especializada en español en cuatro ámbitos del saber: ciencias ambientales, derecho, economía y medicina. La creación del corpus se enmarca en un proyecto de investigación encaminado a explorar las dinámicas de recontextualización del conocimiento especializado. Su foco radica en la comunicación divulgativa que se realiza en la web para poner conocimientos especializados a disposición de un público lego. Después de ilustrar las premisas teóricas en que se sustenta el diseño del corpus, se describen su arquitectura, los métodos utilizados para su etiquetado y las herramientas que permiten su consulta. En la segunda parte del artículo se propone una serie de ejemplos de uso del corpus, disponible en una plataforma abierta de interrogación que utiliza el programa *Sketch Engine*: utilizando las principales herramientas de análisis proporcionadas por la lingüística de corpus (listados de frecuencia, colocaciones, concordancias, *keywords*), se ilustran las potencialidades del corpus para el estudio de los fenómenos divulgativos.

Palabras clave: WebLesp, lingüística de corpus, divulgación, comunicación especializada.

ABSTRACT

This paper presents WebLesp, an electronic corpus of texts representative of digital specialized communication in Spanish in four fields of knowledge: environmental sciences, law, economics, and medicine. The creation of the corpus is part of a research project aimed

at exploring the dynamics of recontextualization of specialized knowledge, with a special focus on popularised communication carried out on the web in order to communicate specialised topics to non-expert audiences. After illustrating the theoretical foundations guiding corpus design, the paper describes corpus architecture, the methods used for its annotation and the tools that allow its online interrogation. In the second part of the article, we propose a series of examples of corpus use through the open Sketch Engine platform: by using the main corpus analysis tools (wordlists, collocations, concordances, keywords), we illustrate how the corpus can be a valuable tool for the study of popularised discourse.

Keywords: WebLesp, corpus linguistics, popularised discourse, specialised communication.

Recibido: 17/06/2020. *Aceptado:* 07/05/2021.

1. INTRODUCCIÓN

El objetivo del presente artículo es presentar y dar a conocer el corpus WebLesp (Corpus de comunicación digital especializada en español), un corpus electrónico que reúne textos representativos de la comunicación digital especializada en español¹. La creación del corpus se enmarca en el ámbito de un proyecto de investigación (Pontrandolfo y Piccioni, 2018; Pontrandolfo y Piccioni, en prensa) encaminado a explorar las dinámicas de recontextualización del conocimiento especializado en distintos contextos y para distintos propósitos (Linell, 1998; Ciapuscio, 2003; Calsamiglia y López, 2003; Gotti, 2014). En esta perspectiva, el corpus y su diseño permiten describir los mecanismos lingüísticos y discursivos que varían cuando cambia el contexto (participantes, medio, propósitos, etc.) en el que se transmiten los conocimientos especializados. En particular, el foco recae en la comunicación digital, incluyendo en esta etiqueta los géneros discursivos publicados y “consumidos” en la web.

El interés por la divulgación de saberes especializados en la web está justificado por el creciente protagonismo que esta ha ido adquiriendo en nuestras sociedades. Cada día nos vemos convertidos en receptores o “usuarios” de textos que se hacen vehículo de algún tipo de conocimiento especializado en los ámbitos del saber más varios, a través de una constelación de modos y medios de comunicación. La variedad de contextos y de formas en que se realiza la divulgación ha dado pie a una plétora de estudios sobre el fenómeno (Ciapuscio, 2003; Calsamiglia, 2003; Calsamiglia y López, 2003; Calsamiglia y Van Dijk, 2004; Myers, 2003; Myers,

¹ El proyecto “WebLesp - Corpus de comunicación digital especializada en español” ha sido financiado por la Universidad “G. D’Annunzio” de Chieti-Pescara (Italia); nace de la colaboración entre esta universidad y la Universidad de Trieste (Italia), con el soporte técnico e informático de la Universidad de Bolonia (Italia).

2010; Gotti, 2014; Parodi y Ferrari, 2007; Rodríguez Tapia, 2016). Por otra parte, la emergencia Covid-19 en la que nos encontramos sumidos en estos momentos no hace sino confirmar el papel central que la comunicación entre expertos y legos tiene para el bienestar de la sociedad en su conjunto, gracias a su capacidad de guiar las políticas de salud pública e incidir en las conductas individuales.

Coherentemente con los objetivos del proyecto y con sus premisas teóricas y metodológicas, la arquitectura del corpus se guía por una concepción amplia de comunicación especializada, que abarca tanto la comunicación experto-experto como la comunicación entre expertos/divulgadores y el público lego. De esta manera, se pretende poner a disposición de la comunidad investigadora una herramienta que permita explorar el llamado *continuum* divulgativo reflejado en una variedad de géneros, registros y repertorios (Pontrandolfo y Piccioni, en prensa).

WebLesp se inserta en el amplio abanico de corpus lingüísticos de lengua española y responde al creciente interés por aproximaciones basadas en corpus al estudio del español como lengua materna, lengua segunda y lengua extranjera, en particular referencia al ámbito de las lenguas de especialidad². Frente a otros corpus de comunicación especializada en español (el corpus GENTT, García Izquierdo, 2005; Corpus DICIPE, Parodi y Ferrari, 2007; Corpus PUCV-2006 del Español, Parodi et al., 2009; COSPE, Pontrandolfo, 2016; Corpus CT-IULA, Cabré y Bach, 2004), WebLesp se diferencia por: (a) presentar una variedad de géneros capaces de dar cuenta del *continuum divulgativo*; (b) incluir solo géneros publicados en la web. Además, está disponible para consultas en línea a través de la plataforma *Sketch Engine* (Rychlý, 2007)³.

En el resto de este artículo se ilustran las premisas teóricas en que se sustenta el diseño del corpus (sección 2), para luego pasar a describir su arquitectura (sección 2.1), los métodos utilizados para su etiquetado y las herramientas que permiten su consulta (sección 2.2). En la segunda parte del artículo se propone una serie de ejemplos de uso del corpus y de la plataforma de interrogación para el análisis basado en corpus de los fenómenos divulgativos (sección 3) y se concluye el trabajo con algunas reflexiones finales (sección 4).

2. WEBLESP: PREMISAS TEÓRICAS Y METODOLÓGICAS

El diseño del corpus WebLesp responde a una serie de premisas teóricas y metodológicas acerca de la comunicación divulgativa, que nos han permitido identificar

² Entre los corpus de español cabe mencionar los corpus académicos de referencia (CREA, CORDE y CORPES) y el Corpus del Español diseñado por Mark Davies, además de los corpus de interés sociolingüístico PRESEEA (Moreno Fernández, 2006) y ESLORA (Vázquez Rozas et al., 2020).

³ El corpus es accesible en: <https://corpora.unich.it/sito/corpus-weblesp-es.html>

dos principales criterios de recopilación: el género discursivo y la identidad de los destinatarios de los textos, es decir expertos o legos. En esta sección ilustramos cómo las premisas teóricas nos han llevado a la identificación de los criterios de recopilación.

Superando la tradicional concepción de la divulgación como mera reformulación del discurso especializado (Cheveigné, 1997)⁴, el proyecto en el que se enmarca la creación de WebLesp concibe la divulgación como recontextualización de saberes expertos (Ciapuscio, 2003; Calsamiglia y López, 2003). Esto es, se considera que la divulgación se concreta en eventos comunicativos adaptados a las condiciones de adecuación propias del nuevo contexto y a las restricciones del nuevo medio (Gotti, 2014), de tal manera que describir los mecanismos divulgativos implica observar cómo la comunicación queda definida por elementos tales como:

- La identidad de los participantes (expertos, semi-expertos, no expertos, mediadores, etc.).
- Las relaciones que los participantes mantienen entre sí (las dinámicas de poder).
- Los propósitos que cada participante persigue (desde educar(se), informar(se), entretener(se), hasta proyectar y negociar las imágenes de expertos y legos, difundir una específica idea de la ciencia y de su rol en la sociedad, etc.).
- El objeto mismo de la comunicación (ya que los temas tratados en los textos reflejan los propósitos de sus participantes, por ejemplo, enfocando el objeto mismo de un específico hallazgo en textos más especializados y privilegiando las implicaciones que el hallazgo pueda tener para la sociedad, la salud o las conductas de los pacientes, etc. en textos dirigidos a un público lego).
- El medio en que se realiza la comunicación (en medios que permiten distintos niveles de interacción entre los participantes determinando el relativo protagonismo de expertos y no expertos, con las dinámicas de empoderamiento relacionadas).

Como se puede apreciar, los factores que aquí se asocian a la caracterización discursiva de la divulgación son elementos que quedan enmarcados en la noción de *género discursivo*, considerando este como “una clase de eventos comunicativos, cuyos miembros comparten un conjunto de propósitos comunicativos [y cuyos] ejemplares [...] comparten algunos patrones en términos de estructura, estilo, contenido, y audiencia prevista (Swales, 1990). De ahí que se haya establecido como principio organizador primario del corpus WebLesp el criterio genérico, configurando el corpus como una colección balanceada de géneros de la comunicación especializada capaces de reflejar la variación lingüística y textual determina-

⁴ Para una discusión de los límites de la concepción reformulativa de la divulgación, véanse Moirand (2003), Myers (2003), Parodi y Ferrari (2007), Pontrandolfo y Piccioni (en prensa).

da por los elementos contextuales anteriormente mencionados.

El criterio genérico adoptado ha permitido incluir en el corpus tanto géneros exclusivamente dirigidos a un público experto (artículos científicos, tesis doctorales, guías clínicas, etc.), como géneros propios de la comunicación dirigida a un público lego o semi-lego (artículos en prensa generalista o especializada, sitios web de asociaciones o empresas que se dirigen a sus miembros / clientes, etc.). Sin embargo, para una serie de géneros (entre los cuales mencionamos en particular los blogs y los informes) la caracterización de los textos en el *continuum* divulgativo no es posible sin una especificación del destinatario, ya que existen blogs e informes dirigidos a expertos y a legos, cada uno con sus propios rasgos y peculiaridades (Pontrandolfo y Piccioni, 2018)⁵. Para que la selección y clasificación de los textos resultara más sólida, se ha decidido utilizar un segundo criterio de recopilación, distinguiendo entre textos especializados y textos divulgativos a partir de la identidad de los destinatarios de los textos como expertos o legos. Somos conscientes de que la distinción experto-no experto puede resultar simplista a la luz de la profesionalización de las disciplinas y la parcelación de los ámbitos de conocimiento que hacen que sea muy difícil definir qué es un experto (¿experto en qué?) (Myers, 2003). Por otra parte, numerosos estudios han evidenciado que, a menudo, incluso el destinatario lego puede poseer conocimientos especializados, como puede ser el caso de un paciente que –a través de su experiencia de una dada enfermedad– ha adquirido un alto nivel de conocimientos acerca de ella (Willems, 1992; Gülich, 2003). Sin embargo, frente a la necesidad de establecer criterios unívocos para la clasificación de los textos, se ha considerado experto al destinatario que consume los textos en cuanto miembro de una comunidad profesional (el científico que lee un artículo de investigación).

Por último, otra decisión determinada por consideraciones teóricas corresponde a la necesidad de restringir (en la medida de lo posible) la selección de textos a una específica variedad del español, en este caso la del español peninsular. Esto se debe al hecho de que para muchas disciplinas (en particular, el derecho) la variedad lingüística está asociada a una específica cultura y específicas modalidades de intercomunicación establecidas dentro de comunidades discursivas nacionales.

2.1. Arquitectura del corpus

A partir de las premisas y de los criterios expuestos en §2, la arquitectura del corpus se configura según la estructura ilustrada en la Tabla I. Se trata de una es-

⁵ Una cuestión que queda abierta es la clasificación del blog, ya que algunos autores los consideran un género, mientras que para otros se trataría de un medio o canal (Cabrera Méndez, 2012; Cerezo, 2006).

estructura flexible, que no excluye la posibilidad de añadir nuevas secciones y nuevos subcorpus según los criterios adoptados (Torruella y Llisteri, 1999).

Tabla I. Arquitectura del corpus WebLes.

Ámbito de conocimiento	Tema	Subcorpus divulgativo	Subcorpus especializado
Ciencias ambientales	Cambio climático	[Ciencia DIV] - Blogs divulgativos - Informes - Sitios web Dimensión: 859.506 tokens	[Ciencia ESP] - Artículos científicos - Informes - Tesis doctorales Dimensión: 657.752 tokens
Derecho	Divorcio en España	[Derecho DIV] - Blogs divulgativos - Artículos en prensa generalista - Sitios web Dimensión: 759.598 tokens	[Derecho ESP] - Textos normativos (artículos del Código Civil español y leyes) - Textos interpretativos (artículos académicos, tesis doctorales) - Textos aplicativos (sentencias) Dimensión: 761.528 tokens
Economía	Economía española vs. economía UE	[Economía DIV] - Blogs divulgativos - Artículos en prensa generalista Dimensión: 752.781 tokens	[Economía ESP] - Informes - Artículos académicos Dimensión: 852.409 tokens
Medicina	Cáncer de mama	[Medicina DIV] - Artículos en prensa generalista - Artículos en prensa especializada - Sitios web Dimensión: 908.853 tokens	[Medicina ESP] - Resumen - Artículos científicos - Blogs especializados - Guías clínicas - Tesis doctorales Dimensión: 925.460 tokens

Como se puede apreciar, el corpus cubre cuatro macro-ámbitos de conocimiento y, dentro de cada uno de ellos, un tema específico: a saber, ciencias ambientales (con el tema del cambio climático); derecho (divorcio en España); economía (relación entre la economía nacional española y las políticas económicas de la Unión Europea); medicina (cáncer de mama). La identificación de un tema específico nos permite limitar el nivel de variación genérica y terminológica observado en cada macro-ámbito de conocimiento.

Los corpus correspondientes a cada ámbito de conocimiento comprenden un subcorpus divulgativo, con textos dirigidos a un público lego, y un subcorpus especializado, con textos dirigidos a destinatarios expertos que participan en la comunicación en cuanto miembros de una comunidad profesional. Dentro de cada subcorpus se han seleccionado textos pertenecientes a determinados géneros discursivos propios de la comunicación en la web. Así pues, para las ciencias ambientales, se han seleccionado blogs divulgativos, informes y sitios web para el subcorpus divulgativo, mientras que el subcorpus especializado comprende artículos científicos, informes dirigidos a expertos y tesis doctorales. La lista completa de los géneros incluidos en el corpus está en la Tabla I.

Para asegurar el equilibrio y la comparabilidad de las varias secciones del corpus se ha decidido mantener cierta homogeneidad en las dimensiones, aunque admitiendo un mínimo de flexibilidad entre distintos ámbitos del conocimiento. De esta forma, cada sección varía entre los 657.752 tokens (del subcorpus especializado de las ciencias ambientales) y los 925.460 tokens (del subcorpus especializado de la medicina). En su conjunto el corpus consta de 6.477.887 tokens.

2.2. Recopilación, etiquetado y consulta

La recopilación del corpus se ha llevado a cabo a través de una combinación de métodos manuales y semi-automáticos. Estos se han adoptado para aquellos géneros y temas –como por ejemplo los artículos de prensa sobre cáncer de mama– para los cuales el uso de combinaciones de palabras clave adecuadas (cáncer, mama, células, metástasis, tratamiento, etc.) ha permitido descargar automáticamente artículos a partir de los motores de búsqueda de los periódicos seleccionados. La recopilación automática ha sido revisada posteriormente, eliminando manualmente de la lista de los textos descargados aquellos que no respondían a los criterios establecidos. Para otros géneros y otros temas (los géneros divulgativos de derecho y economía) se ha hecho necesaria una selección manual de los textos.

Una vez recopilados los textos, se han sometido a etiquetado morfosintáctico y lematización utilizando el software TreeTagger (Schmid, 1994)⁶. El corpus

⁶ En el momento en que se escribe, el corpus está a punto de migrar a una nueva interfaz de

puede ser consultado en la plataforma *Sketch Engine* (Rychlý, 2007), una versión limitada del programa de concordancias *Sketch Engine* (Kilgarriff y otros, 2014)⁷, cuyo uso y aplicación al estudio basado en corpus de los lenguajes especializados se ilustran a continuación.

3. APLICACIONES DE WEBLESP

Esta sección proporciona una panorámica de las principales tipologías de análisis asistidos por ordenador que se pueden realizar interrogando un corpus (Corpas Pastor, 2008; Cruz Piñol, 2012; Rojo, 2021; Zanettin, 2012) como WebLes. Las técnicas de análisis que se examinan a continuación solo representan ejemplos que muestran cómo se puede acceder a la información recogida en un corpus textual para que el usuario pueda identificar patrones lingüísticos significativos.

3.1. Listas de frecuencia (*Wordlists*) y datos lexicométricos

Conocer cuáles son las palabras más frecuentes de un corpus es una de las primeras tareas que se pueden llevar a cabo en WebLes para obtener datos sobre el contenido general de una colección textual, con la consiguiente aplicación que esto tiene para la localización automática de información (Cruz Piñol, 2012).

La Tabla II muestra la lista de frecuencia del subcorpus científico especializado. La interrogación se realiza fácilmente mediante el botón *Word list*, tras seleccionar el subcorpus (*corpus: Ciencia ESP*) y elegir como atributo “palabras” (*search attribute: word*).

Tabla II. Wordlist de Ciencia ES.

Palabra	Frecuencia	Palabra	Frecuencia	Palabra	Frecuencia
de	39.554	como	2.089	lo	895
la	19.939	más	1.946	efectos	863
y	17.445	o	1.614	e	826
en	14.655	La	1.610	este	801
el	12.058	sobre	1.607	agua	782
los	10.937	El	1.563	mayor	772

Sketch Engine, donde etiquetado morfosintáctico y lematización se realizarán con FreeLing (Padró, 2011).

⁷ <https://www.sketchengine.eu/Sketch-engine/>

Continuación Tabla II.

que	8.419	En	1.532	modelos	756
del	7.707	su	1.516	clima	730
las	7.607	no	1.482	sus	703
a	5.918	et	1.471	Las	701
se	5.654	son	1.329	años	680
para	4.510	entre	1.209	otros	679
con	3.346	Los	1.146	precipitación	669
al	3.302	producción	1.115	ser	631
por	3.280	ha	1.053	desarrollo	625
un	3.232	temperatura	1.033	agricultura	619
cambio	2.799	especies	1.017	sistemas	613
climático	2.696	cambios	970	nivel	612
una	2.659	han	967	variables	610
es	2.552	cultivos	910	puede	597

Si se eliminan las palabras vacías (*stopwords*), o sea, palabras que no tienen valor semántico autónomo como artículos, preposiciones, pronombres, etc., los sustantivos que aparecen en este listado clarifican inmediatamente el tema principal del subcorpus, es decir, el cambio climático (“cambio”, “climático”, “temperatura”, “especies”, “cultivos”, “efectos”, “agua”, “clima”, “precipitación”, “agricultura”, etc.). A partir de estas palabras frecuentes, se puede empezar a estudiar los patrones colocacionales de estos términos y seguir realizando búsquedas adicionales, como se demostrará en las siguientes secciones.

Cabe señalar que (como ya se ha mencionado en §2.2) WebLesp es un corpus anotado morfosintácticamente, lo cual quiere decir que también es posible llevar a cabo búsquedas por categorías gramaticales (*search attribute: tag*). Este tipo de información metalingüística se revela muy interesante a la hora de comparar, por ejemplo, las características de cada subcorpus (DIV vs. ESP). Para hacer esto, se pueden utilizar las palabras clave (*keywords*) como se demuestra en §3.2.

3.2. Palabras clave (*Keywords*)

Una comparación entre listas de frecuencias de distintos corpus puede proporcionar información muy interesante, evidenciando patrones lingüísticos característicos de un determinado corpus (Zanettin, 2012). La composición de WebLesp

permite efectivamente comparar dos tipologías de comunicación, la especializada y la divulgativa, a través de sus subcorpus.

En la Tabla III se comparan las palabras clave del subcorpus científico divulgativo con el subcorpus científico especializado con vistas a identificar cuáles son las palabras distintivas de DIV que se encuentran raramente en ES. La interrogación prevé los siguientes pasos: a) selección de *Word list*; b) elección del corpus: en este caso, Ciencia DIV; c) individuación del atributo: *lemma*⁸; d) selección de Keywords (*output type: Keywords*); e) elección del subcorpus de referencia con el que contrastar las listas de palabras de Ciencia DIV, que en este caso es Ciencia ESP (*Reference subcorpus: Ciencia ESP*).

Tabla III. Keywords de los lemas del subcorpus Ciencia DIV (vs. Ciencia ESP).

lema	<i>Ciencia DIV</i>		<i>Ciencia ESP</i>		Estad.
	Frec	Frec/mill	Frec	Frec/mill	
fabricar	127	147,8	0	0,0	148,8
batería	124	144,3	0	0,0	145,3
bolsa	105	122,2	0	0,0	123,2
verter	243	282,7	1	1,5	112,6
superación	389	452,6	2	3,0	112,3
aglomeración	94	109,4	0	0,0	110,4
invento	84	97,7	0	0,0	98,7
Extremadura	82	95,4	0	0,0	96,4
autonómico	78	90,7	0	0,0	91,7
plástico	432	502,6	3	4,6	90,6
tu	76	88,4	0	0,0	89,4
botella	75	87,3	0	0,0	88,3
euro	172	200,1	1	1,5	79,8
apostar	66	76,8	0	0,0	77,8
ciudadanía	64	74,5	0	0,0	75,5
Valladolid	64	74,5	0	0,0	75,5
contenedor	162	188,5	1	1,5	75,2

⁸ *Sketch Engine* permite realizar la comparación de Keywords también por palabras (*word*) o etiquetas morfosintácticas (*tag*), como se demuestra en este mismo apartado. La ventaja de utilizar lemas para esta búsqueda de estas keywords radica en la posibilidad de reunir las formas flexionadas de una palabra bajo un único lema (forma base), obteniendo así frecuencias y estadística más sólidas y menos fraccionadas.

Continuación Tabla III.

reciclar	256	297,8	2	3,0	74,0
bandera	159	185,0	1	1,5	73,8
imaginar	61	71,0	0	0,0	72,0
reutilizar	60	69,8	0	0,0	70,8
baño	60	69,8	0	0,0	70,8
dispositivo	59	68,6	0	0,0	69,6
incumplir	58	67,5	0	0,0	68,5

Es interesante observar cómo, entre los lemas que destacan por ser frecuentes en Ciencia DIV y (casi) ausentes en Ciencia ESP, se encuentran sustantivos como “batería” (124 vs. 0), “bolsa” (105 vs. 0), “superación” (389 vs. 2), “aglomeración” (94 vs. 0), “plástico” (432 vs. 3), “botella” (75 vs. 0), que remiten a una dimensión práctica del tema del cambio climático en comparación con los términos técnicos y científicos que caracterizan el discurso especializado. La presencia de verbos como “verter” (243 vs. 1), “apostar” (66 vs. 0), “reciclar” (256 vs. 2), “reutilizar” (60 vs. 0) e incluso “imaginar” (61 vs. 0) confirman que en Ciencia DIV el tema del cambio climático se aborda desde una perspectiva material y concreta, indicando las pautas prácticas que ciudadanos e instituciones deberían seguir para limitar su impacto en el medioambiente. En Ciencia ESP, en cambio, el interés de los expertos recae en el cambio climático como proceso científico. El análisis de las concordancias (§ 3.3) de todos estos lemas contribuye a aclarar el contraste entre estas dos dimensiones de la comunicación especializada.

Como ya se ha señalado anteriormente, *Sketch Engine* permite elaborar listas no solo de palabras clave sino también de etiquetas/partes del discurso (*tags*). Si se buscan, por ejemplo, las *keywords* basadas en etiquetas morfosintácticas de Derecho DIV (frente a Derecho ESP) se descubren interesantes elementos que confirman algunos mecanismos divulgativos realizados por los autores de los textos (Tabla IV).

Tabla IV. Keywords de los *tags* del subcorpus Derecho DIV (vs. Derecho ESP).

tag	<i>Derecho DIV</i>		<i>Derecho ESP</i>		Estad.
	Frec	Frec/mill	Frec	Frec/mill	
INT	1.472	1891,6	170	223,2	8,4
SLASH	55	70,7	9	11,8	5,6
PERCT	1.135	1458,5	222	291,5	5,0
VCLIfin	97	124,6	28	36,8	3,3

Continuación Tabla IV.

SYM	617	792,9	221	290,2	2,7
PREP/DEL	39	50,1	14	18,4	2,6
VEfin	1.871	2404,3	868	1139,8	2,1
PPX	4.217	5419,0	2.190	2875,8	1,9
VSfin	10.206	13115,0	5.512	7238,0	1,8
QT	8.278	10637,5	4.673	6136,3	1,7
CCAD	1.767	2270,6	1.012	1328,9	1,7
PPC	3.151	4049,1	1.858	2439,8	1,7
VMinf	257	330,3	160	210,1	1,6
VLinf	18.180	23361,8	11.632	15274,4	1,5
VHfin	5.323	6840,2	3.532	4638,0	1,5
VCLinf	3.797	4879,2	2.612	3429,9	1,4
ADV	22.099	28397,8	15.204	19964,9	1,4
CSUBX	8.813	11324,9	6.202	8144,0	1,4
VEinf	277	356,0	212	278,4	1,3
PPO	7.762	9974,4	5.952	7815,8	1,3
VMfin	6.470	8314,1	4.986	6547,3	1,3
VLfin	45.065	57909,7	34.733	45609,0	1,3
CSUBI	4.929	6333,9	3.953	5190,8	1,2
DM	7.132	9164,8	5.792	7605,7	1,2
VSinf	1.294	1662,8	1.061	1393,2	1,2
FS	26.541	34105,9	21.805	28632,8	1,2
NEG	7.146	9182,8	5.934	7792,1	1,2
PE	449	577,0	374	491,1	1,2
VEadj	37	47,5	31	40,7	1,2
CQUE	22.375	28752,5	19.439	25526,0	1,1
QU	10.820	13904,0	10.044	13189,1	1,1
ART	88.741	114034,5	85.555	112345,0	1,0
SE	10.563	13573,7	10.209	13405,8	1,0

En la primera posición del listado se encuentran los interrogativos; en efecto, las oraciones interrogativas directas e indirectas contribuyen a reducir la dificultad cognitiva que los ciudadanos pueden experimentar a la hora de encontrar información en los textos normativos (“¿qué es la patria potestad?” “¿en qué consiste la declaración de nulidad matrimonial?” “¿qué es el divorcio notarial?”). Las preguntas forman parte de las estrategias para involucrar al lector en los textos divulgativos (Pontrandolfo, 2017) y cumplen una función específica: ayudar al lector a entender los tecnicismos del derecho y a saber qué hacer en caso de divorcio.

Otra categoría gramatical que resalta en DIV vs. ESP son los imperativos con pronombres clíticos (*VCLIfn*): “Cuéntenos su caso y le informaremos de manera personal la manera menos perjudicial para enfrentarse a dicha separación”, “Ponte en contacto con nosotros si quieres resolver cualquier duda”, “Póngase en nuestras manos y olvídese del resto, nosotros nos encargamos”, etc. Se trata de una clara estrategia empleada en la comunicación divulgativa para involucrar al lector: los ejemplos, que se refieren sobre todo a contextos extraídos de blogs, muestran cómo el abogado especialista autor del blog ofrece su ayuda al lector que consulta sus páginas web. Referirse al lector mediante una implicación directa del destinatario en el texto es una de las estrategias de divulgación que se evidencia, a nivel de corpus, mediante el uso de pronombres personales o adjetivos y pronombres posesivos en segunda persona singular o plural o tercera persona singular o plural como forma de cortesía (*tú, tu, tuyo; usted, ustedes, su, suyo; vosotros, vuestro*).

3.3. Concordancias (*Concordances*)

Frente a la descontextualización que suele caracterizar las listas de frecuencias léxicas, las concordancias presentan la palabra objeto de investigación (*keyword*) en su contexto, de modo que proporcionan mucha información sobre el funcionamiento de esa *keyword* en la lengua real y en el nivel textual (Cruz Piñol, 2012). La coaparición de palabras (Corpas Pastor, 2008) es uno de los métodos principales de búsqueda con corpus⁹.

Retomando una de las palabras clave en la Tabla III (“bolsa” en Ciencia DIV) se puede observar el comportamiento colocalacional de dicha palabra en el subcorpus divulgativo, como muestra la Figura 1.

⁹ La interfaz *Sketch Engine* permite elaborar concordancias a partir de palabras, lemas, etiquetas morfosintácticas y secuencias de combinaciones de estos elementos (buscando las combinaciones del lema *cáncer* seguido de adjetivo).

Query **bolsa** 105 (122.16 per million)

[First](#) | [Previous](#) Page of 6 [Next](#) | [Last](#)

#238068	plástico Alternativas para sustituir las bolsas de plástico Por Beatriz González . 27 noviembre
#238079	Beatriz González . 27 noviembre 2018 La bolsa de plástico es uno de los productos más
#238163	entre 150 y 1000 años . De modo que , las bolsas de plástico no solo son las protagonistas
#238196	mejores alternativas para sustituir las bolsas de plástico sigue leyendo el siguiente
#238211	EcologíaVerde . Alternativas para reducir las bolsas plásticas Impacto medioambiental de las
#238217	plásticas Impacto medioambiental de las bolsas de plástico Ya hemos explicado que el impacto
#238229	explicado que el impacto medioambiental de las bolsas de plástico es enorme , no obstante , ¿
#238383	ser las alternativas para sustituir las bolsas de plástico . Bolsas biodegradables Son
#238387	para sustituir las bolsas de plástico . Bolsas biodegradables Son bolsas fabricadas a partir de
#238390	de plástico . Bolsas biodegradables Son bolsas fabricadas a partir de una mezcla de materiales
#238517	alternativa mejor en comparación con las bolsas de plástico , tampoco es una solución perfecta
#238529	es una solución perfecta ni definitiva . Bolsas reutilizables de tela Las bolsas de tela
#238534	definitiva . Bolsas reutilizables de tela Las bolsas de tela están fabricadas generalmente de
#238583	evitar este error , o incluso la duda existen bolsas orgánicas hechas a partir de algodón 100
#238620	muy aconsejable . Además , muchas de estas bolsas se pueden encontrar en tiendas de comercio
#238631	encontrar en tiendas de comercio justo . Las bolsas de tela reutilizables garantizan una mayor
#238670	viendo que tienen muchas más ventajas que las bolsas de plástico , resulta incomprensible como
#238758	sin necesidad de adquirir ningún tipo de bolsa en la tienda . Este tipo de comercios no
#238774	comercios no solo permite reducir el consumo de bolsas de plástico de cara los clientes , sino
#238864	envases . Alternativas para reducir las bolsas plásticas La ciencia también colabora día

[First](#) | [Previous](#) Page of 6 [Next](#) | [Last](#)

Figura 1. Concordancias de “bolsa” en Ciencia DIV.

Las líneas de concordancias confirman la centralidad del papel del destinatario como ciudadano empoderado para actuar en la lucha contra el cambio climático mediante un simple gesto como la sustitución de las bolsas de plástico por otros materiales reutilizables como la tela o el algodón. Los colocados de “bolsas” (sobre todo “de plástico”) son verbos como “sustituir” o expresiones clave como “alternativas para reducir” o “reducir el consumo”. La divulgación se convierte en vehículo de concienciación de la sociedad, capaz de posibilitar cambios de actuaciones individuales y de paradigmas sociales. Esta insistencia en el empoderamiento del ciudadano a través de su concienciación corrobora una importante dimensión de la divulgación como forma de activismo social (Pontrandolfo y Piccioni, en prensa).

Query VS.*, como 58 > Sort Node 58 (63.82 per millón)	
Page 1 of 3	Go Next Last Concordance is sorted. Jump to: e
#138802	cáncer . Yo no quise hacerlo . Contarlo era como decir a todos los enfermos : ' Nos vamos
#163841	, en una ocasión le hizo ver lo duro que era como negociador . " Ajuriaguerra me respondió
#242095	conviviendo sin hacerse daño . Aparentemente . Era como llevar encima una bomba química desactivada
#17618	: Sexo . 50 sombras en el cáncer de mama es como se titulaba esta mesa redonda , con eso
#25150	al tacto " , continúa , " pero me toco y es como si presionara cartón . He padecido una
#42385	preservar sus óvulos antes del tratamiento . Es como un seguro de fertilidad , por si la quimioterapia
#109157	ganar . Encontrar nuevos tratamientos no es como ensamblar coches . En mi etapa en empresas
#109875	Va a ser muy complicado porque el cáncer es como un terrorista dentro del cuerpo . Es difícil
#132406	rodeada de gente que trabaja muchísimo . Es como lo de la siesta ... ¿ Conoces a alguien
#132496	estamos tan enfocados en Cataluña que lo demás es como si no existiera . No sé cómo se va a salir
#172119	tratamiento de quimioterapia a sus espaldas . " Es como tener fuego dentro del cuerpo " , recuerda
#178794	tiene el nombre de ' patient navigator ' . Es como el GPS , para las más tecnológicas , o
#210243	frecuentemente a través de un grupo de WhatsApp . " Es como una red de investigadores privados buscando
#231852	principal opción . " El triple negativo es como un cajón desastre y para poder tratarlo
#305241	con cáncer ? Dice que un pecho sin pezón es como una cara sin ojos ; impersonal , difuso
#305569	aproximada que se tarda en hacer el tatuaje es como el viaje de toda una vida . " Pasa que
#310090	que sería un objetivo móvil , porque así es como se va fabricando la verdad en los artículos
#333836	a formaciones fibroquísticas . El quiste es como una bolsa llena de líquido y la fibrosis
#449919	posibles sublevaciones entre sus filas . Así es como funciona nuestro sistema inmunitario para
#471506	pecho a mí niño . Y ese pecho ya no existe . Es como si te desposeyeran del vínculo más estrecho
Page 1 of 3	Go Next Last Concordance is sorted. Jump to: e

Figura 2. Concordancias del patrón verbo SER + como en Medicina DIV.

Las Figuras 2 y 3 muestran otro ejemplo de la utilidad de las concordancias para el análisis contrastivo de la comunicación especializada en su doble vertiente: el uso en contexto del patrón verbo SER + como en el subcorpus médico divulgativo (58 ocurrencias) y en el subcorpus médico especializado (8 ocurrencias). Para realizar la búsqueda se han empleado las denominadas CQL (*Corpus Query Language*), un lenguaje de búsqueda de corpus basado en algunos códigos especiales de interrogación. La plataforma *Sketch Engine* permite utilizar este lenguaje para buscar patrones léxicos o gramaticales complejos o para usar criterios de búsqueda difícilmente utilizables con la interfaz estándar del programa de concordancia¹⁰. Así pues, buscando el verbo ser ([tag="VS.*"]) seguido por la palabra como ([word="como"]) se pueden obtener concordancias interesantes.

¹⁰ <https://www.sketchengine.eu/documentation/corpus-querying/>

Query VS.º, como 8 > Sort Node 8 (8.37 per million)			
#211083	aunque la forma más frecuente de presentación	es como	carcinoma escamoso3 , la dificultad se
#386733	mastografía , ya que su presentación más frecuente	es como	una tumoración con densidad similar al
#576932	detección precoz del cáncer de próstata [es como	hacer diagnóstico precoz de cáncer de páncreas
#437277	combinada , en intervalos de 14 o 21 días , ya	fuera como	terapéutica complementaria primaria o contra
#136068	supervivencia del paciente oncológico , ya	sea como	consecuencia directa de la infección o
#302572	papel esencial en la actividad maligna , ya	sea como	supresores de tumores o como oncogenes
#455349	QT basada en platino como re-inducción ya	sea como	monodroga o poli quimioterapia . El primero
#310146	una preocupación del gobierno federal a	sido como	proporcionar ayuda a un importante sector

Figura 3. Concordancias del patrón verbo SER + como en Medicina ES.

Se trata de un método eficaz para encontrar metáforas y analogías empleadas con frecuencia en la comunicación divulgativa. Estas estrategias permiten al lector entender más fácilmente los temas a través de una comparación con un área y objetos conocidos por el lector (Lakoff y Johnson, 1980). Muchas de las líneas en la Figura 2 evidencian interesantes símiles y metáforas a las que recurren los divulgadores para explicar el cáncer de mama y su tratamiento (Navarro, 2016; Deignan, Semino y Littlemore, 2013; Williams Camus, 2009). Los símiles pueden servir para describir: el dolor (“me toco y es como si presionara cartón”, “es como tener fuego dentro del cuerpo”), la enfermedad (“era como llevar encima una bomba química desactivada”, “el cáncer es como un terrorista dentro del cuerpo”), el tratamiento (“encontrar nuevos tratamientos no es como ensamblar coches”) o la percepción del nuevo cuerpo (“un pecho sin pezón es como una cara sin ojos”, “es como si te desposeyeran del vínculo más estrecho [...]). Muchas de estas líneas reflejan las mismas palabras de las pacientes que han sufrido esta enfermedad y contribuyen a aclarar las maneras en que se conceptualiza la experiencia de tener cáncer.

No extraña la escasa frecuencia de estos patrones en la comunicación especializada: en Medicina ESP estos patrones no se usan metafóricamente sino literalmente (“la forma más frecuente de presentación es como carcinoma escamoso 3”, “su presentación más frecuente es como una tumoración con densidad similar al tejido mamario”).

Las concordancias pueden ser un método eficaz para encontrar y comparar ejemplos de denominaciones en los dos subcorpus. La denominación es una de las estrategias retóricas para adaptar el conocimiento especializado y consiste en la introducción de nuevos objetos, eventos o términos en correspondencia de expresiones especializadas. Por eso, suele identificarse mediante verbos como *llamar*, *definir*, *denominar*, *nombrar*, *decir*, generalmente empleados en la forma impersonal (“se define”, “se denomina”, “se llama”, etc.) o en participio pasado (“denominado”, “llamado”, “definido”), antes o después del término especializado.

Query **denominar** 188 > Sort **Node** 188 (206.85 per million)

Page of 10 | Concordance is sorted. Jump to: ▾

#11093	cuidar esta parte de tu cuerpo , lo que denomina " los diez superalimentos para tus senos
#16876	paciente no tiene ya cáncer , es lo que se denomina respuesta patológica completa . Y en un
#58354	y que algunos conocen como tatuaje , se denomina en realidad micropigmentación de la areola
#127639	personas . Pero la industria farmacéutica los denomina fármacos diana , porque se dirigen a un
#142713	grandes logros recientes es aplicar lo que se denomina tratamiento neoadyuvante , que consiste
#184009	pacientes en esta situación -que la oposición denomina " el limbo de la lista de espera"- pasara
#191011	, instituciones y población general . Se denomina ejercicio físico oncológico aquel ejercicio
#192599	diagnóstico es menor , favoreciendo lo que se denomina " falsos negativos " , lo que , unido a
#249534	ganglio que sea representativo del resto . Se denomina ganglio centinela a aquel que de manera
#256792	en función de los genes del paciente se denomina farmacogenética o farmacogenómica . Pero
#277610	pacientes con cáncer precoz de mama , que se denomina terapia one step , permite en un solo acto
#304827	extirpado quirúrgicamente el tumor , esto se denomina adyuvancia . Por su parte , la neoadyuvancia
#337642	leche tras el parto , y a las que se les denomina lóbulos y lobulillos . Los lóbulos se encuentran
#383432	, instituciones y población general . Se denomina ejercicio físico oncológico aquel ejercicio
#388801	Javier Márquez . A estos quirófanos los denomina "Márquez" de realidad aumentada " porque
#394805	Cuando se extirpa una parte del tejido , se denomina biopsia por incisión . Ésta suele realizarse
#566697	empleados en este tipo de tratamiento se les denomina fármacos antineoplásicos o quimioterápicos
#578503	del estudio o una sustancia placebo . Se denomina ensayo doble ciego al estudio experimental
#592547	momento puede caer , este sentimiento se le denomina " Síndrome de Damocles " y es el más frecuente
#594355	Factores de riesgo de cáncer de mama Se denomina factor de riesgo a aquellas situaciones

Page of 10 | Concordance is sorted. Jump to: ▾

Figura 4. Concordancias del lema “denominar” en Medicina DIV.

Las Figuras 4 y 5 demuestran el uso diferente que hacen los expertos y los divulgadores de la construcción con el verbo “denominar” en el ámbito médico (DIV: 188 vs. ESP: 50).

Query denominar 50 > Sort Node 50 (52.33 per million)			
Page 1	of 3	Go Next Last	Concordance is sorted. Jump to: d
#252569	citosina y adenina . Cuando la IM es \geq 30 % se	denomina	IM alta y confiere un mejor pronóstico
#280595	diferentes reacciones emocionales , que Torres	denomina	reacciones de duelo.77 Éstas experimentan
#281058	separarlos de los de sus pacientes , algo que	denomina	autoconciencia.79 La transmisión de la
#504321	estos órganos ; sin embargo , a estos se les	denomina	infiltración , y no propiamente enfermedad
#926489	dando un resultado entre 0 a 100 que se	denomina	" recurrence score " (RS) (Figura 17
#282944	técnica frecuente utilizada en TMO es la	denominada	trasplante autólogo de células madre (
#302116	sitios de unión como la región no traducida	denominada	UTR del extremo 3 ' del ARN mensajero 5
#443069	una nueva subespecialidad de la oncología	denominada	Psico-oncología . Este término se impuso
#545151	Sr . Director : La enfermedad de Zuska ,	denominada	así en honor a su descubridor , se considera
#624641	de la revisión de la literatura (sección	denominada	' Principales resulta-dos de la revisión
#624669	pregunta clínica en un resumen (sección	denominada	' resumen de la evidencia ') en el que
#624698	clínicos aleatorizados eva-luados (sección	denominada	' Resumen de la evidencia ') . La calidad
#625191	de la revisión de la literatura (sección	denominada	' Principales resultados de la revisión
#625219	pregunta clínica en un resumen (sección	denominada	' Resumen de la evidencia ') en el que
#625246	clínicos aleatorizados evaluados (sección	denominada	' Resumen de la evidencia ') . 17 1 .
#304414	normales y en lesiones precursoras de PDAC	denominadas	PanIN I y , con el desarrollo de la enfermedad
#331835	conformado por glándulas productoras de leche	denominadas	lóbulos , y por pequeños túbulos que la
#531380	suspensión formando colonias flotantes ,	denominadas	mamosferas , estaban enriquecidas en CMT
#206924	mercado del anticuerpo monoclonal anti-HER-2 ,	denominado	trastuzumab (4) . -El CMI " Basal-like
#225432	permiten destacar a un subgrupo de CMTN ,	denominado	borderline . Habitualmente la mitad de
Page 1	of 3	Go Next Last	Concordance is sorted. Jump to: d

Figura 5. Concordancias del lema “denominar” en Medicina ES.

Si en el subcorpus divulgativo la construcción se usa básicamente para definir un tecnicismo médico (ejemplo, “se denomina ejercicio físico oncológico aquel ejercicio [...]”, “se denomina ganglio centinela a aquel”, “se denomina factor de riesgo a aquellas situaciones”), muchas veces introducida por una explicación (simple) (“[...] que el paciente no tiene ya cáncer, es lo que se denomina respuesta patológica completa”, “cuando se extirpa una parte del tejido, se denomina biopsia por incisión”), en el subcorpus especializado se usa para referirse a contenidos y tecnicismos médicos (“dando un resultado entro 0 a 100 que se denomina ‘recurrence score’ (RS)” “[...] sitios de unión como la región no traducida denominada UTR del extremo 3 del ARN mensajero 5”, etc.) o como mecanismo de cita intratextual (“sección denominada ‘Principales resultados’”, “sección denominada ‘resumen de la evidencia’”, etc.). Además, en Medicina DIV se le está ofreciendo al lector información que enriquece su conocimiento de la materia, incrementando de forma artificial su grado de conocimiento compartido. En Medicina ESP, en cambio, los expertos se sirven de la denominación como recurso argumentativo e interpretativo de sus estudios clínicos.

3.4. Colocaciones (*Collocations*)

Las concordancias pueden servir para mostrar con qué palabras se combina una determinada *keyword* (Cruz Piñol, 2012), permitiendo de esta forma identificar las colocaciones y los denominados *clusters*, o sea, conjuntos de palabras que aparecen secuencialmente en un texto. En particular, las colocaciones proporcionan al usuario el listado completo de los lemas que, estadísticamente, se asocian con frecuencia con el término clave objeto de análisis.

Tabla V. Candidatos a colocación del término “abogado” en Derecho DIV.

	Frecuencia	T-score	MI	logDice
especialista	164	12,750	7,833	11,369
familia	214	14,474	6,564	11,334
procurador	106	10,254	7,943	10,833
divorcio	411	19,574	4,858	10,718
especializado	92	9,544	7,665	10,621
Familia	98	9,810	6,793	10,575
tú	141	11,606	5,469	10,530
con	314	17,008	4,637	10,459
un	653	24,286	4,333	10,378
necesitar	79	8,795	6,569	10,280
abogado	128	10,980	5,085	10,275
Madrid	74	8,511	6,551	10,199
contacto	70	8,302	7,021	10,199
nuestro	100	9,749	5,316	10,179
Bilbao	62	7,836	7,681	10,103
ayudar	64	7,920	6,638	10,036
él le	88	9,131	5,230	10,033
equipo	59	7,625	7,090	9,987
despacho	57	7,487	6,911	9,924
.	726	25,014	3,803	9,907
Valencia	54	7,308	7,514	9,906
en	585	22,366	3,731	9,817
asistencia	50	7,034	7,591	9,808
<unknown>	459	19,773	3,697	9,755
contar	54	7,237	6,044	9,735
acudir	50	6,980	6,272	9,678

El análisis de los candidatos a colocación detectados automáticamente por el programa¹¹ en Derecho DIV y Derecho ESP del lema “abogado” (DIV: 1.672 vs. ESP: 111 ocurrencias) evidencia dos tendencias distintas (Tablas V y VI).

Tabla VI. Candidatos a colocación del término “abogado” en Derecho ES.

	Frecuencia	T-score	MI	logDice
intervención	18	4,236	9,330	10,956
procurador	31	5,555	8,770	10,804
obligatorio	9	2,997	9,859	10,662
preceptivo	7	2,643	9,812	10,398
presencia	7	2,642	9,544	10,316
asistir	6	2,443	8,490	9,765
labor	5	2,231	8,833	9,752
obligatoriedad	3	1,731	10,205	9,487
representar	7	2,632	7,569	9,344
acudir	6	2,437	7,658	9,318
asistencia	4	1,992	8,020	9,232
debido	3	1,728	8,650	9,193
respectivo	4	1,990	7,597	9,028
general	7	2,621	6,716	8,731
expediente	4	1,983	6,838	8,586
incluso	4	1,982	6,806	8,565
familia	6	2,419	6,331	8,380
notario	5	2,203	6,077	8,125
Señor	3	1,705	5,995	7,881
realizar	3	1,703	5,892	7,807
solo	3	1,695	5,543	7,546
nuevo	4	1,953	5,408	7,520
parte	11	3,222	5,134	7,428
demandar	4	1,949	5,296	7,426
proceso	3	1,690	5,378	7,418
necesario	3	1,687	5,277	7,337
estado	3	1,687	5,250	7,316
esposo	3	1,680	5,060	7,161

¹¹ *Sketch Engine* permite detectar los candidatos a colocaciones a partir de la búsqueda de las concordancias del término (en este caso, contaminación) (*Concordance*); tras realizar la búsqueda (*Make concordance*), se selecciona el botón *Collocations* y se elige, como en los casos anteriores, entre *word*, *lemma* o *tag*.

En el subcorpus divulgativo, el término se coloca más frecuentemente con sustantivos como “especialista”, “especializado”, “[de] familia”, “[de] divorcio”, así como con verbos como “necesitar”, “ayudar”, “contar”, “acudir”, mientras que en el subcorpus especializado el término se coloca con sustantivos más neutros como “intervención”, “presencia”, “obligatoriedad”, “asistencia” y verbos como “asistir” y “representar”. Si en el primero el abogado cumple una función importante para el ciudadano, o sea, ayudarlo a entender y actuar según derecho, en el segundo su presencia se describe principalmente desde el punto de vista técnico (los contextos en los que es obligatoria por ley la presencia de un abogado en un procedimiento). En Derecho DIV los divulgadores hacen hincapié en la importancia de contar con un abogado experto y en la autopromoción de sus servicios legales (“Sea cual sea su caso es imprescindible que lo examine un abogado especializado en divorcios, por los siguientes motivos”, “Es importante contar con un abogado especializado en la materia, para que pueda orientarnos de manera adecuada”), mientras que en Derecho ESP el foco recae en el papel jurídico del abogado (“Es llamativo como para el juicio verbal, que en ciertos supuestos no es preceptiva la intervención de abogado y procurador, se exige demanda y para los procesos de separación y divorcio consensual, en que siempre es preceptiva la intervención de abogado y procurador, no habla de demanda sino de solicitud”).

3.5. Asociaciones semánticas

Son muchos los conceptos que se han empleado para definir las distintas tipologías de asociaciones entre palabras o patrones lingüísticos (Zanettin, 2012). Además de las colocaciones, término habitual para referirse a la frecuencia con la que determinadas formas tienden a coaparecer, es posible identificar dos niveles de abstracción que derivan de patrones recurrentes, o sea, la preferencia y la prosodia semántica (Sinclair, 2004). La preferencia semántica hace referencia a la coaparición de palabras con determinadas selecciones semánticas (Corpas Pastor, 2008), mientras que la prosodia semántica se refiere al significado pragmático y funcional, que forma parte inherente de la unidad léxica (Sinclair, 2004).

WebLesp permite identificar numerosos casos de prosodia semántica. Si se compara el uso del término “ambiente” en la sección Ciencia del corpus WebLesp se aprecian matices semánticos interesantes (función *Concordance*); en el subcorpus divulgativo abundan verbos, pronombres o adjetivos positivos como “respetar”, “cuidar”, “nuestro”, etc., mientras que en el subcorpus especializado la actitud del especialista es mucho más neutral (ejemplo, “medio ambiente”, “desarrollo”, “natural”, etc.). También la sección Medicina del corpus WebLesp presenta ejemplos interesantes de asociaciones semánticas: además de casos evidentes como el término “cáncer” que tiende obviamente a colocarse con verbos como

“padecer”, “sufrir”, “diagnosticar”, se puede señalar el caso interesante del término “célula”. Si en Medicina ESP, “células” tiene más colocados neutros, en DIV muchos de los colocados son negativos: verbos como “destruir”, “matar” o “infestar” o adjetivos como “asesina” o “separatista”, ejemplos de las metáforas bélicas usadas para referirse a la enfermedad (Pontrandolfo y Piccioni, en prensa).

4. CONCLUSIONES

El corpus WebLesp ofrece un amplio abanico de posibilidades para explorar la compleja relación entre comunicación especializada y divulgación, tema que justifica su arquitectura global y que resulta muy actual en la sociedad moderna. La plataforma Sketch Engine proporciona una interfaz flexible y funcional que brinda innumerables potencialidades al usuario, que puede interrogar de forma rápida y eficaz el corpus.

Es evidente que en el presente trabajo se han explorado solamente algunas aplicaciones basales del corpus y se han propuesto limitadas lecturas preliminares de los datos empíricos obtenidos; como bien se sabe, interpretar los datos es una tarea del analista, que puede proponer también lecturas de los datos diferentes y complementarias.

Interrogar WebLesp puede revelarse una operación útil no solo para los lingüistas interesados en algunos de los ámbitos de la comunicación especializada seleccionados para el presente proyecto, sino también para redactores técnicos o traductores que necesiten confirmación empírica de sus propuestas escritas. Además, la arquitectura del corpus subraya también la importancia del género discursivo en la comunicación digital, puesto que *Sketch Engine* permite seleccionar y filtrar las búsquedas por géneros (por ejemplo, interrogando solamente, en el subcorpus Ciencia DIV, los blogs).

Por lo que se refiere a futuras trayectorias de investigación, un análisis sistemático de los rasgos típicos de la divulgación podría llevar a la elaboración de una especie de “gramática local” –en la acepción de Gross (1993)– de la divulgación a raíz de los resultados obtenidos por corpus, que permita poner de relieve los mecanismos propios de la recontextualización divulgativa. Esto abre una interesante perspectiva de investigación para el futuro: comprobar si los criterios establecidos en la literatura para definir un sublenguaje (también Pontrandolfo y Goźdz-Roszkowski, 2014) –1) un ámbito de especialidad bien limitado; 2) restricciones léxicas, sintácticas y semánticas; 3) normas gramaticales ‘atípicas’; 4) alta frecuencia de determinadas construcciones; 5) estructuras textuales específicas; 6) usos de determinados símbolos– se cumplen en los subcorpus divulgativos de WebLesp.

Otro reto para el futuro será, sin duda, la ampliación del corpus incluyendo otros ámbitos de especialidad y otros géneros. En este sentido, WebLesp podría

concebirse como un corpus abierto o monitor, o sea, como un conjunto textual que se va renovando cíclicamente mediante la inclusión de nuevos textos (por ejemplo, nuevos géneros del discurso digital 2.0, como foros, tuits, redes sociales, etc.) y exclusión de otros más antiguos, consiguiendo de este modo un corpus vivo y dinámico (Torruella y Llisterri, 1999). Desde esta perspectiva, en WebLesp podrán confluír también textos y géneros discursivos pertenecientes a otras lenguas o variedades lingüísticas muy relevantes en el caso del español, como las de América Latina.

REFERENCIAS

- Cabré, T., y Bach, C. 2004. El corpus tècnic del IULA: corpus textual especializado plurilingüe. *Panacea*, 5(16), 173-176.
- Cabrera Méndez, M. 2012. Blogs. En M. Tascon (coord.), *Escribir en Internet. Guía para los nuevos medios y las redes sociales* (217-227), Fundéu/BBVA, Barcelona, España: Galaxia Gutenberg.
- Calsamiglia, H. 2003. Popularization discourse. *Discourse Studies*, 5(2), 139-146.
- Calsamiglia, H. y López Ferrero, C. 2003. Role and Position of Scientific Voices: Reported Speech in the Media. *Discourse Studies*, 5(2), 147-173.
- Calsamiglia, H., y Van Dijk, T. 2004. Popularization Discourse and Knowledge about the Genome. *Discourse & Society*, 15(4), 369-389.
- Cerezo, J. (coord.) 2006. *La blogosfera hispana: pioneros de la cultura digital*. Madrid, España: Fundación France Telecom España.
- Cheveigné, S. 1997. Introduction. La science dans une société médiatisée. *Hermès*, 21, 15-22.
- Ciapuscio, G. E. 2003. Formulation and Reformulation Procedures in Verbal Interactions between Experts and (Semi-)laypersons. *Discourse Studies*, 5(2), 207-233.
- Corpas Pastor, G. 2008. *Investigar con corpus en traducción: los retos de un nuevo paradigma*. Bern, Suiza: Peter Lang.
- Cruz Piñol, M. 2012. *Lingüística de corpus y enseñanza del español como 2/L*. Madrid, España: ArcoLibros.
- Deignan, A., Semino, E., y Littlemore, J. 2013. *Figurative Language, Genre and Register*. Cambridge, Reino Unido: Cambridge University Press.
- García Izquierdo, I. (ed.) 2005. *El género textual y la traducción. Reflexiones teóricas y aplicaciones pedagógicas*. Bern, Suiza: Peter Lang.
- Gotti, M. 2014. Reformulation and recontextualization in popularization discourse. *Ibérica*, 27, 15-34.
- Gross, M. 1993. Local grammars and their representation by finite automata. En M. Hoey (ed.), *Data, Description, Discourse: Papers on the English Language in*

- Honour of John McH. Sinclair* (26-38). Londres, Reino Unido: Harper Collins.
- Gülich, E. 2003. Conversational Techniques Used in Transferring Knowledge Between Medical Experts and Non-experts. *Discourse Studies*, 5(2), 235-263.
- Lakoff, G., y Johnson, M. 1980. *Metaphors we live by*. Chicago, EEUU: University of Chicago Press.
- Linell, P. 1998. Discourse across boundaries: On recontextualizations and the blending of voices in professional discourse. *Text*, 18, 143-157.
- Moirand, S. 2003. Communicative and Cognitive Dimensions of Discourse on Science in the French Mass Media. *Discourse Studies*, 5(2), 175-206.
- Moreno Fernández, F. 2006. Información básica sobre el *Proyecto para el estudio sociolingüístico del español de España y América* – PRESEEA (1996-2010). *RSEL*, 36, 385-391.
- Myers, G. 2003. Discourse studies of scientific popularization: questioning the boundaries. *Discourse Studies*, 5(2), 265-279.
- Myers, G. 2010. Stance-taking and Public Discussions in Blogs. *Critical Discourse Studies*, 7(4), 263-275.
- Navarro, I. 2016. Metaphorical aspects in cancer discourse. En Ordóñez-López y N. Edo-Marzá (eds.), *Medical Discourse in Professional, Academic and Popular Settings* (125-148). Bristol, Reino Unido: Multilingual Matters.
- Padró, L. 2011. Analizadores multilingües en FreeLing. *Linguamática*, 3(2), 13-20.
- Parodi, G., y Ferrari, S. 2007. El discurso divulgativo de la ciencia y la tecnología en la prensa escrita chilena: Una mirada al Corpus DICIPE – 2004. En G. Parodi (ed.), *Lingüística de Corpus y Discursos Especializados: Puntos de mira* (259-277). Valparaíso, Chile: Ediciones Universitarias de Valparaíso.
- Parodi, G., Ibáñez, R., y Venegas, R. 2009. El Corpus PUCV-2006 del Español: identificación y definición de los géneros discursivos académicos y profesionales. *Estudios Filológicos*, 44, 123-147.
- Pontrandolfo, G. 2016. *Fraseología y lenguaje judicial. Las sentencias penales desde una perspectiva contrastiva*. Roma, Italia: Aracne.
- Pontrandolfo, G. 2017. Géneros divulgativos de la comunicación jurídica: el caso de los blawgs. *Círculo de Lingüística Aplicada a la Comunicación*, 72, 111-200.
- Pontrandolfo, G., y Piccioni, S. 2018. Comunicación simétrica y asimétrica en los blogs de divulgación jurídica: entre modalidad epistémica y modalidad deóntica. *HERMES – Journal of Language and Communication in Business*, 58/2018, 175-191.
- Pontrandolfo, G., y Piccioni, S. En prensa. *Comunicación especializada y divulgación en la red: aproximaciones basadas en corpus*. Londres, Reino Unido: Routledge.

- Pontrandolfo, G., y Goźdź-Roszkowski, S. 2014. Exploring the Local Grammar of Evaluation: The Case of Adjectival Patterns in American and Italian Judicial Discourse. *Research in Language*, 12(1), 71-91.
- Real Academia Española. Banco de datos (CREA) [en línea]. *Corpus de referencia del español actual*. Disponible en <https://www.rae.es/banco-de-datos/crea> [Consulta: 28/5/2021].
- Real Academia Española. Banco de datos (CORPES XXI) [en línea]. *Corpus del Español del Siglo XXI (CORPES)*. Disponible en <http://www.rae.es> [Consulta: 28/5/2021].
- Rodríguez Tapia, S. 2016. El grado de especialización textual: caracterización a partir de la percepción sociocognitiva de la terminología y las relaciones discursivas. Tesis doctoral. Córdoba, España: Universidad de Córdoba.
- Rojo, G. 2021. *Introducción a la lingüística de corpus en español*. Londres, Reino Unido: Routledge.
- Rychlý, P. 2007. Manatee/bonito - a modular corpus manager. *1st Workshop on Recent Advances in Slavonic Natural Language Processing*, 65-70.
- Schmid, H. 1994. Probabilistic Part-of-Speech Tagging Using Decision Trees. *Proceedings of International Conference on New Methods in Language Processing*, Manchester, Reino Unido.
- Sinclair, J. 2004. *Trust the Text: Language, Corpus and Discourse*. Londres, Reino Unido: Routledge.
- Swales, J. 1990. *Genre Analysis: English for Academic and Research Settings*. Cambridge, Reino Unido: Cambridge University Press.
- Torruella, J., y Llisterri, J. 1999. Diseño de corpus textuales y orales. En J. M. Blecua, G. Clavería, C. Sánchez y J. Torruella (eds.), *Filología e informática. Nuevas tecnologías en los estudios lingüísticos* (45-77). Barcelona, España: Editorial Milenio.
- Vázquez Rozas, V., Barcala, F., Domínguez, E., Fernández Sanmartín, A., Rojo, G., y Santalla del Río, M. 2020. Codificación y anotación del habla en un contexto bilingüe: el corpus ESLORA del español del Galicia. *Verba*, 20, 189-224.
- Willems, D. 1992. Susan's Breathlessness. The Construction of Professionals and Laypersons. En J. Lachmund y G. Stollberg (eds.), *The Social Construction of Illness: Illness and Medical Knowledge in Past and Present* (105-114). Stuttgart, Alemania: Steiner.
- Williams Camus, J. 2009. Metaphors of cancer in scientific popularization articles in the British press. *Discourse Studies*, 11(4), 465-495.
- Zanettin, F. 2012. *Translation-Driven Corpora. Corpus Resources in Descriptive and Applied Translation Studies*. Manchester, Reino Unido: St Jerome.