

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 IULA¹ o los incluidos en IntelliText de la Universidad de Leeds²— o permiten buscar directamente en las páginas web a través de los metadatos —como la versión 2 del corpus de Davies³—.

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 for a RAR file. On the left, there is a file icon labeled 'RAR'. Below it, the document type is listed as 'Otros' and the date as '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 citation:

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 otro⁴, 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 Ahumada⁵ 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 bibliométricos⁶: 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)⁸. El observatorio de neonomía o del neologismo científico es otro de sus componentes⁹. 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 hispano¹⁰.

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)¹¹, en 2015 alcanzaba ya los 280 millones (66 584 documentos)¹² y en 2018 alcanzó los 318 millones (317 931 294 formas en 77 951 documentos con límite cronológico del 2016)¹³. Esto lo hacía similar en extensión al CORPES de la RAE¹⁴, 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

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).

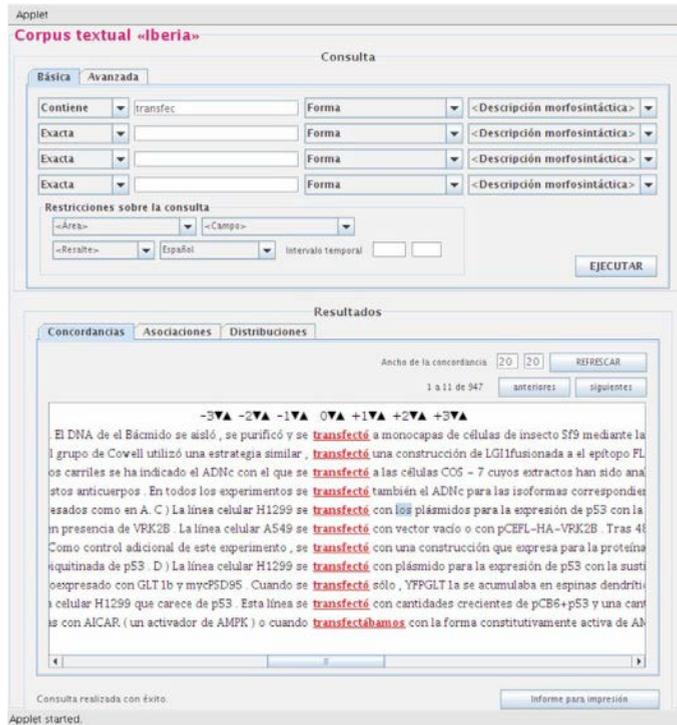


Figura 3. Corpus IBERIA, versión interna, año 2010, ejemplo de búsqueda.
Fuente: Porta *et al.* (2011: 154); *cf.* asimismo Pablo (2010: 62-65).



Figura 4. Corpus IBERIA, versión actual, interfaz en línea.

2.2. Documentos y representatividad

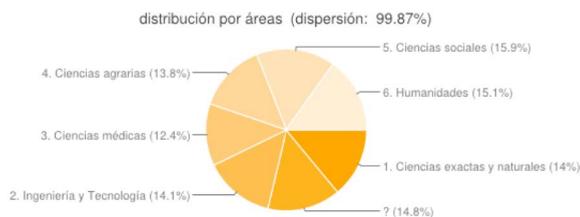
IBERIA recopila los textos aparecidos principalmente en revistas académicas de distintos ámbitos científicos. Así, se nutre fundamentalmente de las revistas y documentos del

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 recientemente¹⁵). 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 Científico¹⁶, más otras revistas médicas del grupo Elsevier¹⁷, 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.

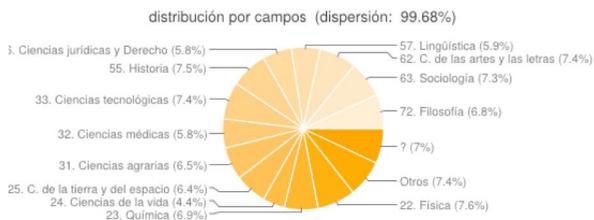
No obstante, en 2013 se respondía a la cuestión: la representatividad que aportaba entonces cada área científica al corpus no era superior a un 15 % del total:



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 textuales¹⁸).

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, CORPES¹⁹), el corpus del español especializado del IULA²⁰, 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 recursos²¹. 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 γ* : “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 μm) 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”²².

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 the first item being 'rayo gamma'. The 'Restricciones sobre los textos' section is empty. Below the search criteria 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 interacciona con el cristal centelleador. Dicho es...', 'de interacción del rayo gamma a lo largo del mismo. El desconoci...', '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 search results are sorted by 'área' in ascending order.

Figura 8. Corpus IBERIA, búsqueda de rayo gamma.

La búsqueda de rayos γ 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 γ) (“Telefonía móvil: ¿representa algún riesgo para la salud?”, *Revista Medicina clinica*, 132/14)

The screenshot shows the 'Corpus IBERIA' search interface. The search criteria are set to 'coincide con' for six items, with the first item being 'rayo gamma'. The 'Restricciones sobre los textos' section is empty. Below the search criteria 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 search results are sorted by 'palabra por la izquierda' in ascending order.

Figura 9. Corpus IBERIA, búsqueda de rayo γ .

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):

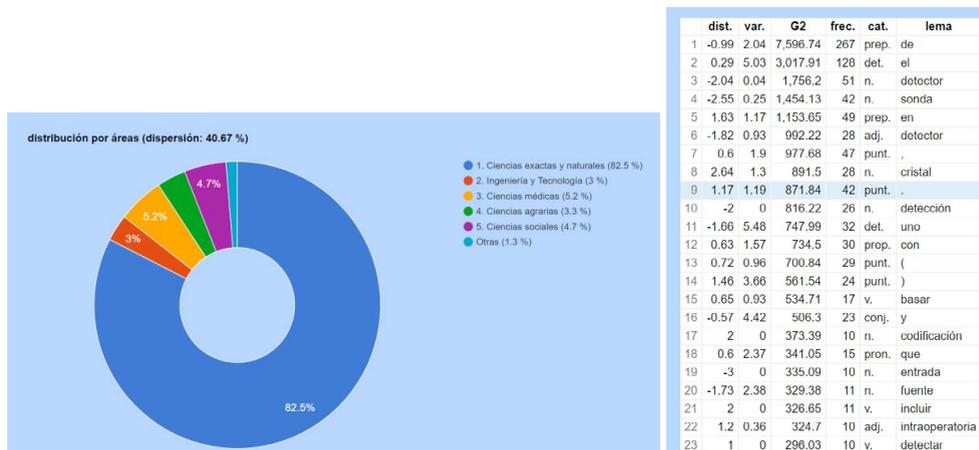


Figura 10. Corpus IBERIA, distribución por áreas y asociaciones de *rayos gamma*.

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*).

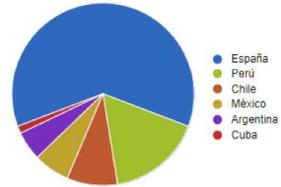
La documentación en CREA para “rayos gamma” es mayor: 88 veces en 55 documentos. Aquí sí es patente el uso técnico, por la preponderancia de los textos científicos (Ciencia y tecnología: 72,72 %); como “rayo gamma” solo hay cuatro concordancias, una de 1980, de la obra *Einstein o la armonía del mundo*, 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

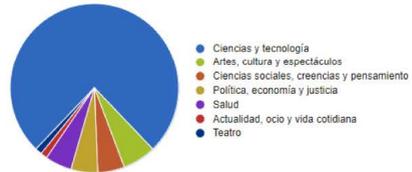


Distribución Tema

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.

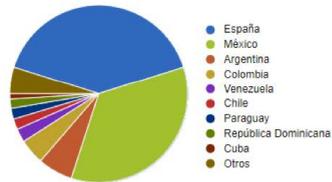
En cuanto a CORPES, si la versión beta 0.91 señalaba 356 casos en 132 documentos, la versión más reciente (0.94, julio 2021) recoge 410 en 156 documentos, lo que indica no solo un enriquecimiento del corpus, sino también un uso mucho más extendido de este término, con un predominio apabullante del ámbito 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

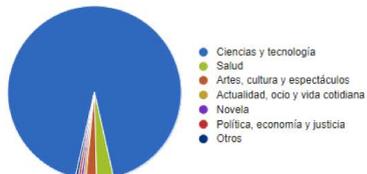


Distribución Tema

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



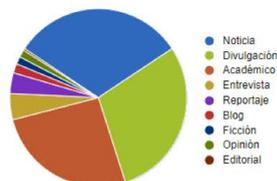
Figuras 13 y 14. Distribución de *rayos gamma* en CORPES.

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

CSIC

Ítem o secuencia de ítems: Restricciones sobre los textos:

1. forma | coincide con | cromosfera | cualquier análisis | área: |
 2. forma | coincide con | | cualquier análisis | campo: |
 3. forma | coincide con | | cualquier análisis | año inicial: | año final: |
 4. forma | coincide con | | cualquier análisis |
 5. forma | coincide con | | cualquier análisis |
 6. forma | coincide con | | cualquier análisis |

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 USMR ha estado t...
 2007 días, de unos 4.000°C, que vemos oscuros 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; estas zonas juegan un important	** 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	cromosfera; 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 cromosfera 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 cromosfera , 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 cromosfera , 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 cromosfera , 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 cromosfera 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 (cromosfera y corona). El plasma solar es un gas muy caliente que está parcialmente ionizado
7	2002 Ur.	La cromosfera 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 cromosfera 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 cromosfera 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 cromosfera 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 cromosfera solar revelan la existencia de enigmáticos picos de polarización en la señal de
12	2003 Esp.	aficionados pidiendo disculpas porque no pudo obtener unos anisados espectros de la cromosfera 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 cromosfera , en la que se proyectan gases a muy altas temperaturas y de la que sobresalen las
14	2004 Esp.	Tanto la cromosfera como las protuberancias pueden verse directamente en los momentos de la totalidad
15	2004 Esp.	Por encima de la cromosfera está la corona (figura 7) (Capas del Sol. Graiño 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 cromosfera 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) Cromosfera 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 (cromosfera , corona) tienen temperaturas mucho más altas. Muchos fenómenos muy interesantes
19	2009 Méx.	Cromosfera
20	2009 Méx.	La atmósfera solar está formada por dos capas llamadas cromosfera y corona que se encuentran encima de la fotosfera. En la cromosfera 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

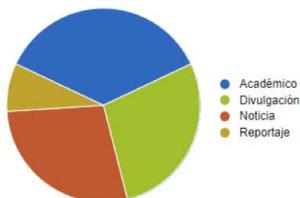


Figura 20. CORPES, voz *cromosfera*, distribución por tipología textual.

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):

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.

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: < >

consultar limpiar

concordancias asociaciones distribuciones ayuda

(concordancias: 36 de 317,931,294 / documento: 26 de 77,951)

crisis	2%	dirección	coincide con	ordenar
9111994	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...	
9111994	941% empujante. Los 4 meses siguientes sin patológicos. Se ver...		mastodinia o palpación de nudo mamario 92,33% en patológicas (4 tumores mamario...	
9111998	de consulta más frecuente fue la revisión de rutina (44,8...		mastodinia (55,2%), siendo mucho menos frecuentes otros motivos: US, 72,3% de las mug...	
9111999	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...	
9111999	manejo >valorada por el informe manográfico. Resultados: De las...		mastodinia en el seguimiento, ni se detectaron modificaciones en la inspección ni en la p...	
9111999	diminución de la ansiedad y uno (16,67%) persistió. De las 37 pac...		mastodinia en controles sucesivos, y en ninguna se detectó modificaciones en la inspecc...	
9111999	Se cree que el paso de progesterona hacia el torax circulato...		mastodinia, náuseas, edemas, etc. (3). El objetivo de este trabajo, es revisar las seque...	
9111999	hacer 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...	
9112000	atracción del pezco, úlcera o enfermedad de Paget, en 1/8 casos (...)		mastodinia en descrita por 118 pacientes (9,51 por 100) y la presencia de secreción a trav...	
9112004	de un complejo sintomático abigarrado con un gran número de sint...		mastodinia, etc.). Entorno, en realidad, era un verdadero síndrome causado por la hiper...	
9112007	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...	
9112008	actual inicia los preparados del fide de agnóstico para el tratamie...		mastodinia. Todas estas indicaciones están reconocidas por el Comité de El Ministerio...	
9112011	en caso de vómitos profundos, puede administrarse la dosis por vía...		mastodinia (hasta el 47%), que suele durar pocos días. Otros efectos secundarios, más raro...	
9112011	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...	
9112012	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...	
9112015	antecedentes de tabaquismo, que ingresó en el hospital por deterio...		mastodinia. En la exploración destacaba una ginecomastia bilateral dolorosa. En la radiog...	
9112015	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...	
9112016	no se produjeron variaciones del peso corporal en las mujeres del A...		mastodinia y alteraciones musculares (dismetrias, amiotrofias, etc.) en las mujeres que c...	
9112016	determina la evolución con la combinación de sistemas ("ambo") de ...)		mastodinia, spotting) y de la temperatura basal ("termica"). Así se consigue una mayor e...	
9112016	mensual. Solo aparecer amenorrea en el 6,20% de los casos. Los a...		mastodinia, náuseas, insomnio, ganancia de peso) aparecen muy eventualmente.	
9112016	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	
9112016	de acción antagónica, por lo que se eliminan los molestos efec...		mastodinia, disminución de la libido, disfunción erectil) que el Altablex® ocasiona en l...	
9112016	alleviación el dolor, hay que hablar de mastalgia como la expresión...		mastodinia se utiliza para referirse al dolor de origen mamario funcional. Lo que si se ut...	
9112016	hegosa por vía parenteral. Coincidiendo con la detección del VHC...		mastodinia bilateral sin galactomas, difusión erectil o disminución de la libido. Se desca...	
9112016	plazo, lo que obligó a su suspensión (como efecto adverso...		mastodinia. 10 sujetos presentaron disfunción erectil, 2 sujetos presentaron enanismo y ot...	
9112016	este efecto cosmopolítico. En estos pacientes se ha descrito dolor...		mastodinia, demencia) y molestias en los pliegues submamarios, además de retención...	
9112016	manejada por el Servicio de Traumatología o por fisioterapia...		mastodinia otros relacionados con la hiperfrotia de las mamas, desalojo de evolución. En e...	
9112016	en 74 libros normal (7,8-19 mg/dl). La paciente refiere un episodio...		mastodinia el año anterior. Fue estudiada en su país, pero no disponía de informes medicos...	
9112016	esta terapia ha sido tan efectiva como terapia y danzari en ensay...		mastodinia y depresión. Se ha descrito en su uso prolongado periodos de más osea, que se...	
9112016	antecedentes. Síntesis de evidencia >No encontramos con 3 posib...		mastodinia en pacientes que realizan tratamiento de depresión hormonal para el cáncer...	
9112016	amoxicilina 20 mg/d que ocasiona una disminución importante de l...		mastodinia, dada su alta incidencia, hacen que la exploración física sea un arma funcio...	
9112016	20 mg/d constituye la mejor opción para el tratamiento y la preven...		mastodinia, mientras que en el caso de la ginecomastia establecida de larga evolución la c...	
9112016	antecedentes. Síntesis de evidencia >No encontramos con 3 posib...		mastodinia en pacientes que realizan tratamiento de depresión hormonal para el cáncer...	
9112016	amoxicilina 20 mg/d que ocasiona una disminución importante de l...		mastodinia, dada su alta incidencia, hacen que la exploración física sea un arma funcio...	
9112016	20 mg/d constituye la mejor opción para el tratamiento y la preven...		mastodinia, mientras que en el caso de la ginecomastia establecida de larga evolución la c...	
9112016	apropiación de hematomas, seromas, la secreción de los teg-los, la e...		mastodinia, la dificultad diagnóstica en mamas densas y muchas veces, solació de ver de l...	

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 that is a navigation menu with options like 'Concordancias', 'Coapariciones', 'Configuración', 'Ayuda', 'Estadística', 'Modo de cita', 'Sugerencias', and 'Preguntas frecuentes'. The main search area shows the word 'mastodinia' entered in the 'Forma' field. Below the search results, there are two entries for 'mastodinia' with their respective RRF (Recuperación de Frecuencia) values and classification details.

RRF (Clasificación, país)	CONCORDANCIA	Ordenar por	Año ascendente	sin criterio
1 2014 Esp.	Cambios en la sed y el apetito, deseos intensos de comer ciertos alimentos.			
2 2014 Chile	síntomas como irritabilidad, edema (hinchazón provocada por la retención de líquido);			

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
- 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.^a 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.^a 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)