

## El corpus Iberia como recurso para la traducción especializada\*

(The Iberia corpus as a resource for specialized translation)

IGNACIO AHUMADA LARA  
(CSIC, Madrid)

Fecha de recepción: 26 de marzo de 2010

Fecha de aceptación: 17 de abril de 2010

**Resumen:** El corpus Iberia es una herramienta informática creada entre 2008 y 2009 en el seno del Consejo Superior de Investigaciones Científicas para los estudios sobre el español, lengua de la ciencia en la sociedad del conocimiento. Se trata de un corpus textual panhispánico sobre el discurso científico-técnico del español (1985-hoy). La cuidada selección de los textos, tanto por su calidad como por su representatividad, avalan sobradamente las conclusiones a las que puedan llegar en el estudio de estos materiales tanto lingüistas como traductores de lenguas especializadas.

**Palabras clave:** Lingüística Aplicada, Lexicografía, Traducción, Lenguaje científico, Corpus Iberia.

**Abstract:** The Iberia Corpus is a software computing tool first created between 2008 and 2009 within the “Consejo Superior de Investigaciones Científicas” for the use of Spanish Studies, a language of science in the society of knowledge. It is a Panhispanic textual corpus of scientific-technical texts in Spanish (from 1985 hitherto). The carefully chosen selection of texts, under the working criteria of good and representative quality fully validate the conclusions to which both linguists and translators of speciality languages may jump to through the study of the present material.

**Key words:** Applied Linguistic, Lexicography, Translation, Scientific Language, Corpus Iberia

### 0. INTRODUCCIÓN

Entre los principales recursos con los que cuenta el traductor —amén del dominio tanto de la lengua de partida como de la lengua de llegada— el diccionario ocupa, sin la menor duda, el lugar más destacado. He dejado sin especificación el sustantivo *diccionario*, porque veo el fortín del buen traductor pertrechado no solo del imprescindible *diccionario bilingüe*, sino acompañado, además, del *diccionario general* y, al menos, de un *diccionario especializado*, así monolingüe como bilingüe. Cosa bien distinta es que estos diccionarios, sean cuales fueren, respondan a las expectativas del

usuario, esto es, hayan sido concebidos y redactados pensando en los traductores.

Como adenda a recursos tan tradicionales como efectivos, la Red ha propiciado en las últimas décadas la consulta en línea de los más diferentes repertorios lexicográficos y otras herramientas que facilitan las labores de traducción, bien que teniendo en cuenta las preceptivas cautelas ante la información suministrada.

En otro orden de cosas, quiero añadir que veo la Lingüística de corpus como uno de los grandes logros de la Lingüística descriptiva. Antes que una subdisciplina de la Lingüística moderna, nos encontramos ante una metodología regida por unos parámetros capaces de generar datos reales y contextualizados sobre el uso de la lengua. En definitiva, la Lingüística de corpus se ocupa de los principios metodológicos por los que se rige una de las herramientas más productivas que lingüistas y traductores puedan manejar. Su utilidad en estos campos es solo comparable, salvando las distancias, al diccionario con el que unos (lingüistas) y otros (traductores) soñamos. Como es bien sabido, la Lingüística de corpus hunde sus raíces en el primer estructuralismo, pero no afianza su presencia como metodología lingüística hasta que la Informática irrumpe en las diferentes áreas de los estudios humanísticos, esto es, en los años setenta y con la lengua inglesa como eje de sus investigaciones.<sup>1</sup>

Dentro del grupo de investigación *El español, lengua de la ciencia* (ELCI), en el seno del Consejo Superior de Investigaciones Científicas (CSIC), hemos desarrollado una herramienta, el corpus Iberia, que además de proporcionarnos los datos necesarios para nuestros trabajos lexicológicos y lexicográficos, puede llegar a prestar no pocos buenos servicios a los traductores especializados.<sup>2</sup>

## 1. EL CORPUS IBERIA

### 1.1. *Estándares y protocolos*

Es evidente que las diferentes herramientas que se diseñan para los trabajos de descripción lingüística, deben responder esencialmente a los objetivos que persiga el proyecto que los orienta; de otra manera, los resultados podrían ser bien diferentes a los planteados en la hipótesis de trabajo. Pesa, cómo no, en el diseño de estas herramientas un mayor

---

\* Este trabajo resume lo expuesto por su autor en la conferencia plenaria del VII Congreso Internacional de Traducción, Texto e Interferencias, celebrado en la Facultad de Filología-Universidad de Sevilla, los días 13, 14 y 15 de mayo de 2010.

<sup>1</sup> Para una introducción a la Lingüística de corpus puede consultarse con provecho Parodi, 2010.

<sup>2</sup> Para conocer las características técnicas del corpus Iberia véase Ahumada/Porta/Rosal, en prensa.

aprovechamiento de los materiales ofrecidos, no en vano van dirigidos a un grupo en particular, pero también a la comunidad científica en general.

Cuando en ELCI nos propusimos desarrollar un corpus sobre el español como lengua de la ciencia en la sociedad del conocimiento, acudimos, como no podía ser de otra manera, a los trabajos que se habían desarrollado en la Comisión Europea dentro de *Expert Advisory Group on Language Inginnering Standards* (EAGLES). Atendiendo a estas directrices, las recomendaciones para la construcción de corpus podrían resumirse en los siguientes puntos:

EXTENSIÓN.- Debe recoger el mayor número posible de textos.

VARIEDAD.- Debe procurar la incorporación de una elevada diversidad de materiales.

CLASIFICACIÓN.- Debe estar presidido por una clara y nítida clasificación de los mismos.

REPRESENTATIVIDAD.- Debe procurar en todo momento el equilibrio entre sus elementos.

CLARIDAD.- Debe ceñirse a la incorporación de documentos con garantías reconocidas.

Esto nos llevará a la construcción de un corpus fiable y representativo, un corpus que nos permita llegar a conclusiones veraces y con bajos índices de dispersión.

### 1.2. *Definición y objetivos*

El corpus Iberia es una herramienta informática creada entre 2008 y 2009 en el seno del CSIC para los estudios sobre el español como lengua de la ciencia en la sociedad del conocimiento. Se trata de un corpus textual panhispánico sobre el discurso científico-técnico en español (1985-hoy). Si bien los primeros textos tienen su origen en 1985, la mayor profusión de los mismos datan del presente siglo. Los documentos anteriores a 2000 se han incorporado como referencia para los estudios de carácter lexicológico.

Los objetivos que persigue son los siguientes: (a) recopilar el mayor número posible de documentos científicos escritos en español, (b) actualizar permanentemente con la producción científica de última hora, (c) indexar los documentos incorporando la mayor cantidad de información lingüística posible y (d) crear una herramienta útil a través de una interfaz de consulta lo más versátil, potente y sencilla posible.

El corpus Iberia está concebido como un exponente de primera magnitud sobre la situación y uso del español como lengua científico-técnica en su dimensión panhispánica. Durante los dos últimos años el equipo de investigación ha diseñado y alimentado el corpus con documentos científicos cuya propiedad intelectual corresponde al CSIC. Así se han incluido los textos digitalizados de las treinta y tres revistas que edita la institución y los textos digitalizados del repositorio institucional Digital\_CSIC.<sup>3</sup> Se trata, en consecuencia, de una representación del español científico que se produce en España. Dada la concepción panhispánica del proyecto, el grupo de investigación persigue el estudio, digitalización e indexación de las revistas científicas más representativas del español de América, entendiendo este segundo proceso como una segunda fase del proyecto desarrollado en los dos años anteriores para el español de España. A esta justificación, hemos de sumar la representatividad y el equilibrio a que aspira el corpus, no ya en cuanto a representatividad y equilibrio geográfico, sino también representatividad y equilibrio lingüístico y disciplinar. Valga como ejemplo la falta de documentación científica en nuestro corpus en disciplinas y subdisciplinas que no se desarrollan en España como puedan ser el cultivo del café, la fauna autóctona u otras de semejante especificidad.

En la actualidad contiene 6595 documentos y setenta millones de palabras aproximadamente referidos al español científico-técnico del español de España.

### 1.3. Clasificación de los materiales

---

<sup>3</sup> DIGITAL\_CSIC.- En 2006 el CSIC se suma a la Declaración de Berlín, «uno de los hitos fundacionales del movimiento internacional del acceso abierto (*open acces*)». La Declaración de Berlín «animó a los centros europeos de investigación a maximizar la visibilidad e impacto de la ciencia mediante la creación y desarrollo de archivos abiertos institucionales» (Digital\_CSIC, 2010: 1).

Los trabajos de Digital\_CSIC comienzan en 2008. Hasta finales de 2009 se había almacenado una selección de 17 480 documentos (Digital\_CSIC, 2010: 1).

La red de centros de investigación se divide en ocho grandes áreas del saber que comprenden: Humanidades y Ciencias Sociales, Biología y Medicina, Recursos Naturales, Ciencia y Tecnologías Físicas, Ciencia y Tecnología de Materiales, Ciencia y Tecnología de Alimentos, Ciencia y Tecnología Químicas y Ciencias Agrarias

En 2009 los centros e institutos de investigación del CSIC publicaron 11 691 artículo científicos que fueron publicados en las más prestigiosas revistas internacionales, además de 368 libros y 793 tesis doctorales. En el mismo año, los proyectos de investigación de todas las áreas del conocimiento fueron financiados con más de 525M euros (cf. Digital\_CSIC, 2010: 2-3).

Los documentos están sujetos a la siguiente clasificación y subclasificación por áreas y campos. Para ello seguimos la nomenclatura UNESCO, adoptada por la Administración española desde 1983:

#### 1.3.1. Clasificación por áreas

- I Ciencias Exactas y Naturales
- II Ingeniería y Tecnología
- III Ciencias Médicas
- IV Ciencias Agrarias
- V Ciencias Sociales
- VI Humanidades

#### 1.3.2. Clasificación 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

#### 1.4. *Novedad y relevancia*

Algunos grupos de investigación han reunido en los últimos años distintos corpus de léxico científico-técnico, pero a partir de una disciplina concreta (Economía, Medicina, etc.) y referido a un área geográfica concreta (España, México, Chile, etc.), el corpus Iberia, por el contrario, es multidisciplinar (campos y áreas) y panhispánico (el español de España y español de América). El corpus Iberia frente al resto de los corpus existentes, además, de proporcionarnos las características concordancias y su contexto, ofrece los documentos completos para la correcta interpretación. Proporciona, además, la distribución léxica por áreas y campos, datos que nos permiten conocer bien la exclusividad bien la diversidad de un mismo término. Además, de la distribución del término por áreas y campos, nos facilita la vigencia de uso, el nacimiento y en su caso, la muerte de los términos a partir de los índices cronológicos que hemos incorporado. El estudio contrastivo, como es evidente, nos permitirá determinar tanto la unidad como la diversidad del español como lengua de la ciencia.

#### 1.5. *Usuarios*

El corpus Iberia está dirigido fundamentalmente a estudiosos de la Lingüística y a expertos en traducción especializada. Más en concreto, el corpus Iberia va dirigido a los investigadores en disciplinas como la Lexicografía, puesto que se trata de diccionarios en bruto; a los estudiosos de la Lexicología, puesto que recoge el uso, vigencia, nacimiento y mortandad del léxico científico-técnico; de igual modo que a quienes se ocupan del análisis del discurso científico en general.

#### 2. EL CORPUS IBERIA Y LA TRADUCCIÓN ESPECIALIZADA

Al comienzo de mi intervención he hecho referencia a las expectativas de los traductores ante las obras lexicográficas. Pues bien, ocupémonos ahora de la función del diccionario de especialidad y de la utilidad del corpus Iberia como complemento y base de las insuficiencias que presentan estos diccionarios.

Son muy notables las diferencias que distinguen un diccionario monolingüe de otro bilingüe, como aquellas que distinguen un diccionario general de uno especializado. El diccionario general monolingüe, al tratarse del repertorio lexicográfico por excelencia, se somete con una rigidez sin límites a la teoría lingüística y a la metalexicográfica. Son muy pocas las quebras que se le toleran, especialmente en el campo de la definición y en de la normalización de su microestructura. El diccionario bilingüe, sin embargo, debido fundamentalmente a los consabidos problemas de equivalencia y en aras de una mayor precisión significativa, permite al redactor una mayor operatividad a la hora de expresar el contenido. Frente

al diccionario especializado, el diccionario general no puede renunciar, desde mi punto de vista, a la llamada definición lingüística, en tanto que el diccionario de especialidad puede optar por una mayor diversidad definicional: definición científica, terminológica o lingüística.

Las críticas a un diccionario de lengua, escolar o de aprendizaje arrecian cuando pretende cubrir un espectro de usuarios diverso. Los diccionarios de especialidad persiguen no sólo satisfacer las necesidades de los expertos en esa disciplina, sino que parejamente se dirigen a los legos, de aquí el marbete de pedagógicos, y, en última instancia, a los traductores al incorporar un vocabulario bilingüe anexo al cuerpo de la obra. La lexicografía de especialidad, en mi opinión, debe apostar decididamente por lo que llamamos el *diccionario multifuncional*.

Veamos un ejemplo de reciente. La Real Academia de Farmacia, en cuyo haber lexicográfico sólo se contaba un repertorio monolingüe en dos volúmenes publicado en 1865, ha optado directamente por acogerse a la lexicografía bilingüe: *Diccionario terminológico de las ciencias farmacéuticas-A Terminological Dictionary of the Pharmaceutical Sciences. Inglés-Español/Español-Inglés* (2007). El equipo formado por Alfonso Domínguez-Gil, Enrique Alcaraz y Raquel Martínez ha sido el responsable, junto con la sanción de los académicos, del nuevo diccionario. Me interesa destacar, amén del ya mencionado carácter bilingüe, aquella innovación más relevante desde el punto de vista de la técnica lexicográfica. Se trata de los nuevos elementos incorporados a la tradicional microestructura del artículo lexicográfico. Para el admirado y tristemente fallecido profesor Alcaraz, la sola presencia de los equivalentes en un diccionario bilingüe es una de las insuficiencias que presentan los actuales diccionarios, en consecuencia, «es preciso delimitar el significado con una explicación, una frase contextualizadora, palabras relacionadas y sobre todo dónde se ubica la palabra dentro de los 25 campos semánticos que para nosotros constituyen la primera aproximación en el campo de las ciencias de la Farmacia». Así lo resumía en el acto de presentación del diccionario el pasado 10 de mayo de 2007 en la sede de la Real Academia. Esta es la principal innovación de este diccionario, que no dudamos apreciarán cuantos lo consulten, muy en especial, los traductores.

El pasado año, ya sin la dirección de Enrique Alcaraz, pero indudablemente bajo su magisterio, José Mateo Martínez ha publicado el *Diccionario de términos de la banca*. El autor, como no podía ser de otra manera, incorpora la frase contextualizadora de la que habla su maestro.

El corpus Iberia –acabo de indicarlo– muestra las concordancias en su contexto. Nada de particular tiene esta característica, porque la mayoría de los corpus al uso contextualizan el índice de concordancias que muestran en la primera consulta. Al encontrarnos, sin embargo, ante textos

especializados, y concededores de las dificultades que plantea la interpretación tanto de los términos al uso como de los neónimos, el corpus Iberia redirige la consulta del término en su contexto al documento completo, de manera que la interpretación del significado goza de un grado de seguridad mucho más elevado que con la simple contextualización.

He aquí la interfaz de Iberia con las posibles combinaciones de *láser* + *adjetivo*:

**Corpus textual «Iberia»**

**Consulta**

Sencilla | Avanzada

Exacta	láser	Forma	<Categoría morfo-sintáctica>
Exacta	Aquí el segundo...	Forma	Adjetivo
Exacta	Aquí el tercero...	Forma	<Categoría morfo-sintáctica>
Exacta	Aquí el cuarto...	Forma	<Categoría morfo-sintáctica>

Parámetros globales

<Área> | <Campo>

<Resalte> | <Idioma del párrafo> | Intervalo temporal | año | año

**EJECUTAR**

**Resultados**

Concordancias | Colocaciones | Estadísticas

Ancho de la concordancia: 5 | 5 | REFRESCAR

1 a 11 de 55 | anteriores | siguientes

-3▼▲ -2▼▲ -1▼▲ 0▼▲ +1▼▲ +2▼▲ +3▼▲

grabación la tecnología de el **láser azul** y cuya capacidad de almacenamiento comprobando con eficacia mediante microscopía **láser confocal** de barrido ( CSLM ) la microscopía de scáner por **láser confocal** y la espectrometría de ma estudio se trabajó con un **láser continuo** de diodo Rofin-Sinar DL 27 mediante SEM y microscopía óptica **láser confocal**. En el estudio mediante de limpieza con microscopio óptico **láser confocal**, las muestras de los ; y el microscopio óptico **láser confocal** Leica TCS SP2 AOBS con La observación a microscopio óptico **láser confocal** (Figura 10) ,

Figura 10 . Imagen en **láser confocal** mostrando el contraste entre la , tinciones especiales , microscopía **láser confocal**, aplicaciones de biofilms experimentales óptica de fluorescencia y el **láser confocal** ayudan a visualizar y cuantificar

Consulta realizada con éxito. Informe para impresión

Estamos ante la materia prima de la que dispone el lexicógrafo para desarrollar su labor: la papeleta lexicográfica (las unidades objeto de análisis en su contexto).

Y de fundamental importancia para nuestra tarea de interpretación y fijación del significado especializado: (a) los índices de claridad del corpus, esto es, se trata de documentos con garantías reconocidas, (b) la cronología abarca de 1985 al día de hoy, especialmente a partir del año



2000, por lo que facilitara sobremanera la catalogación de neologismo y (c) la clasificación en las distintas áreas y campos para, en función del texto al que nos enfrentamos, establecer correctamente su uso en el contexto.

Además de esto, y dada la diversidad de lenguas que se contempla en la literatura científica que recoge el corpus, las tareas de contrastividad nos aseguran una mayor precisión en el establecimiento del significado. Valga como simple muestra las estadísticas que nos ofrecen las preposiciones: *für*, *pour* y *for*:

PREPOSICIONES	CONCORDANCIAS	DOCUMENTOS	PORCENTAJE DE DOCUMENTOS SOBRE EL TOTAL	DISTRIBUCIÓN POR ÁREAS DE CONOCIMIENTO
<i>FÜR</i>	132	72	1,09%	HUMAN. (29,40%) SOCIALES (10,68%) INGENIER. (9,72%) AGRARIAS (2,38%) EXAC./NAT. (1,31%) VARIA (46,51%)
<i>POUR</i>	482	163	2,47%	HUMAN. (31,60%) AGRARIAS (22,77%) SOCIALES (13,80%) EXAC./NAT. (13,57%) INGENIER. (9,60%) VARIA (8,63%)
<i>FOR</i>	16093	2681	40,65%	AGRARIAS (18,46%) INGENIER. (16,39%) MÉDICAS (16,08%) SOCIALES (15,67%) EXAC./NAT. (12,90%) HUMAN. (7,21%) VARIA (13,29%)

La presencia de la lengua inglesa en el corpus Iberia no deja lugar a dudas sobre las preferencias idiomáticas de nuestros científicos. Y esto

tanto en la lengua en la que divulgan sus investigaciones como en el soporte bibliográfico que avala sus trabajos.

### 3. ¿EL ESPAÑOL, LENGUA DE LA CIENCIA?

#### 3.1. *En los orígenes del español como lengua de la ciencia*

«Assí que non se falla que del rey Tolomeo acá ningún rey nin otro omne tanto fiziesse por ello commo él». Las palabras del infante Don Juan Manuel sobre su tío Alfonso el Sabio, recogidas en el prólogo a su obra *El libro de la caza* (c 1325), no son más que el reconocimiento escrito a la incansable labor del monarca castellano por reunir y divulgar la ciencia de su tiempo. Para mayor gloria nuestra, nos encontramos ante la primera adaptación de una lengua moderna de Europa a la literatura científica. «Et tanto cobdiçió quelos de sus regnos fuesen my sabidores, que fizo tralladar eneste lenguaje de Castiella todas las ciencias, tan bien la teología commo la lógica, e todas las siete artes liberales como toda la arte que dizen mecánica». Como es bien sabido, el traslado se hacía del árabe al primitivo castellano y de este a la lengua de la ciencia por excelencia hasta el siglo XVIII: el latín.

El monopolio comercial de España con el nuevo continente, la llamada carrera de Indias, revoluciona por completo la concepción tradicional del comercio marítimo. En los astilleros españoles se gestan, se desarrollan y se perfeccionan nuevas técnicas de navegación. Una situación de desarrollo tecnológico tan próspera genera una no menos importante literatura científica en lengua española. Los primeros tratados marítimos de Europa se escriben y se difunden en español. Primero, manuscritos; después, impresos. Y junto a ellos, los primeros glosarios de términos marineros: Alonso de Chaves (1538), Juan de Moya (1564), Nuño Verdugo (1582) y Andrés de la Poza (1585), así hasta la primera obra impresa: Diego García de Palacio y su *Instrvcción návthica* (1587). De esta manera arranca nuestra lexicografía de especialidad, generada palmariamente por una investigación de primer orden.

Valga esta breve introducción solo como nota para comparar los orígenes de nuestra literatura científica con el momento actual.

El grupo de investigación en bibliometría del CSIC, dirigido por la profesora Gómez Caridad, ha publicado recientemente un estudio bibliométrico sobre las publicaciones científicas españolas y, en concreto, del CSIC en la *Web of Science* entre 2004 y 2008. De los datos que ofrece este trabajo, seleccionamos aquellos que mejor nos van a permitir comprender la situación actual del español como lengua de la ciencia.

## CORPUS TEXTUAL «IBERIA»

Evolución de la producción científica del CSIC y de España por trienios  
(WoS 1990-2008)

Trienios	N. Documentos		Incrementos	
	CSIC	España	CSIC	España
1990-1992	5.393	38.239		
1991-1993	6.293	43.040	16,69	12,56
1992-1994	7.275	47.520	15,60	10,41
1993-1995	8.139	52.077	11,88	9,59
1994-1996	9.205	57.026	13,10	9,50
1995-1997	10.283	63.093	11,71	10,64
1996-1998	11.327	68.928	10,15	9,25
1997-1999	12.065	74.215	6,52	7,67
1998-2000	13.116	77.750	8,71	4,76
1999-2001	14.196	80.503	8,23	3,54
2000-2002	15.381	84.181	8,35	4,57
2001-2003	16.061	88.995	4,42	5,72
2002-2004	17.144	95.874	6,74	7,73
2003-2005	18.323	103.368	6,88	7,82
2004-2006	19.958	113.450	8,92	9,75
2005-2007	21.769	124.626	9,07	9,85
2006-2008	22.692	130.264	4,24	4,52

Nota: los datos del periodo 1990-1999 fueron obtenidos por consulta directa del WoS

(Gómez y otros, 2010: 23)



Prácticamente se ha cuadruplicado en las dos últimas década nuestra presencia en WoS, prueba palpable de que nuestra producción científica, de forma paulatina, cuenta con una mayor visibilidad a nivel mundial.

## CORPUS TEXTUAL «IBERIA»

**Distribución de la producción científica de España por sectores institucionales (WoS 2004-2008)**

Sectores Institucionales	2004	2005	2006	2007	2008	Total	%
Universidad	20.900	22.504	25.037	27.128	26.014	121.583	60,09
Sector Sanitario	8.936	9.677	10.772	12.067	10.937	52.389	25,89
CSIC	6.204	6.603	7.151	8.015	7.526	35.499	17,55
Entidades sin ánimo de lucro	1.716	2.112	2.646	3.326	3.666	13.466	6,66
Empresas	1.327	1.553	1.744	2.023	1.836	8.483	4,19
Otros OPI	1.023	1.088	1.307	1.827	2.946	8.191	4,05
Administración	932	1.001	1.479	1.775	1.868	7.055	3,49
Organismos Internacionales	195	204	237	271	257	1.164	0,58
Otros	327	236	296	275	226	1.360	0,67
<b>Sumatorio</b>	<b>41.560</b>	<b>44.978</b>	<b>50.669</b>	<b>56.707</b>	<b>55.276</b>	<b>249.190</b>	
<b>Total real</b>	<b>34.601</b>	<b>37.463</b>	<b>41.386</b>	<b>45.777</b>	<b>43.101</b>	<b>202.328</b>	

**Nota:** el sumatorio es mayor que el total real porque existe colaboración entre sectores institucionales

(Gómez y otros, 2010: 24)



El comentario anterior no pierde su validez para esta segunda tabla, si bien ahora nos centramos en exclusiva en el período correspondiente al objeto de estudio: 2004-2008.

Ante cifras tan halagüeñas y como lingüistas y traductores que somos la pregunta inevitable no puede ser otra que interrogar por las lenguas en las que los científicos españoles difunden el resultado de sus investigaciones:

## CORPUS TEXTUAL «IBERIA»

Tabla 7. Distribución de la producción científica del CSIC y España por idioma de publicación (WoS 2004-2008)

Idiomas	CSIC		España	
	Total	%	Total	%
Inglés	34.482	97,14	185.234	91,55
Español	948	2,67	16.362	8,09
Francés	30	0,08	438	0,22
Portugués	8	0,02	107	0,05
Alemán	5	0,01	95	0,05
Italiano	4	0,01	25	0,01
Catalán	16	0,05	23	0,01
Rumano	1	0,00	11	0,01
Chino	2	0,01	7	0,00
Polaco	0	0,00	5	0,00
Ruso	0	0,00	4	0,00
Checo	0	0,00	3	0,00
Croata	0	0,00	3	0,00
Eslovaco	0	0,00	3	0,00
Árabe	2	0,01	2	0,00
Galés	1	0,00	2	0,00
Eslovenio	0	0,00	1	0,00
Holandés	0	0,00	1	0,00
Latín	0	0,00	1	0,00
Serbio	0	0,00	1	0,00
<b>Total real</b>	<b>35.499</b>		<b>202.328</b>	

Nota: no se incluye el sumatorio porque coincide con el total real



(Gómez y otros, 2010: 26)

Al perseguir un mayor impacto, nuestro científicos optan por hacerlo en revistas de gran difusión, revistas cuya lengua de comunicación es el inglés: 185.234 artículos en inglés frente a 16.362 en nuestra propia lengua, es decir, un 91,55% frente a 8,09%.

Sigamos adelante. ¿Cuál es la situación, por el contrario, del inglés en revistas científicas españolas?

CORPUS TEXTUAL «IBERIA»

Tabla 12. Idioma de los documentos del CSIC en revistas españolas (Total de documentos) (WoS 2004-2008)

Idiomas	2004	2005	2006	2007	2008	Total	%
Español	193	171	116	238	156	874	66,72
Inglés	81	61	70	99	90	401	30,61
Catalán	0	0	0	16	0	16	1,22
Francés	2	0	0	6	0	8	0,61
Alemán	1	1	0	2	0	4	0,31
Italiano	0	0	0	3	0	3	0,23
Portugués	0	0	0	1	1	2	0,15
Árabe	2	0	0	0	0	2	0,15
<b>Sumatorio</b>	<b>279</b>	<b>233</b>	<b>186</b>	<b>365</b>	<b>247</b>	<b>1.310</b>	

(Gómez y otros, 2010: 32)



Las propias revistas científicas españolas abren sus páginas a colaboraciones en lengua inglesa en una proporción aproximada del 40% para la lengua inglesa frente a un 60% para la lengua española.

#### 4. A MODO DE CONCLUSIÓN

Las cifras resultan dramáticamente elocuentes y poco satisfactorias para la lengua española. Lo mismo nos da atender a las publicaciones de impacto (*WoS*) que a nuestras propias revistas. La realidad es incontestable: la lengua inglesa es la lengua de preferencia por los científicos españoles. ¿Significa esto que la ciencia española, a muy corto plazo, acabará difundándose en esa lengua, aun dentro de las propias revistas españolas, y no en español? Podríamos pensar que esto solo pudiera afectar a determinadas áreas de conocimiento, como la Biomedicina, la Biología y aquellas en las que nuestra investigación es de

primera líneas, y que siempre el español quedaría reservado para disciplinas con menor impacto internacional. ¿La Historia, por ejemplo, la Lingüística, la Psicología, la Pedagogía...en definitiva, las Ciencias Humanas?

Cuando al comienzo de mi intervención he declarado los principios que nos movían al crear y desarrollar el corpus Iberia, me he reservado a conciencia uno de los más importantes. A estas alturas y con los datos en la mano sé que no se les oculta a ustedes. Y no es otro que este: al ser el castellano del siglo XIII la primera lengua moderna de Europa en recrear la ciencia greco-latina y árabe, así como en el siglo XVI la divulgadora de los primeros tratados de ingeniería náutica, procurar a toda costa que el español no pierda la cuota actual de impacto, sino que la mantenga y acreciente al ritmo que avanza la investigación en España y en las repúblicas hispanoamericanas. De lo contrario, el español quedaría circunscrito a áreas de conocimiento muy concretas. Mucho me temo que quedaría reducido a la literatura vernácula y al uso familiar, esto es, a la lengua de la calle. El panorama es desolador.

¿Si dejamos de producir ciencia en español, no creen que perderíamos una parte esencial de nuestra conciencia como pueblo? Lamentablemente, sí. Aunemos entonces esfuerzos para que estas previsiones no sean una realidad.

##### 5. REFERENCIAS BIBLIOGRÁFICAS

- AHUMADA, Ignacio/PORTA, Jordi/ROSAL, Emilio Del (en prensa): «Design and development of Iberia: a corpus of scientific spanish», en *Corpora. Edinburgh University Press*.
- DIGITAL\_CSIC (2010): *Memoria 2009*, Madrid, Consejo Superior de Investigaciones Científicas.
- EAGLES (1996): *Synopsis and comparison of morphosyntactic phenomena encoded in lexicons and corpora. A common proposal and applications to European languages*, Pisa, ILC-CNR.
- EAGLES (1996a): *Preliminary recommendations on subcategorization*:  
<<http://www.ilc.cnr.it/EAGLES96/synlex/synlex/html>>.
- GÓMEZ, Isabel/BORDONS, María/MORILLO, Fernanda/MORENO, Luz/GONZÁLEZ-ALBO, Borja/CANDELARIO, Antonia/HERRERO, María (2010): «La actividad científica del CSIC a través del *Web of Science*: estudio bibliométrico del período 2004-2008», en Repositorio Digital\_CSIC: <<http://hdl.handle.net/10261/22941>>.
- PABLO NÚÑEZ, Luis (2010): «Fuentes de información y otros recursos documentales para traductores: la labor del CSIC», en Repositorio Digital\_CSIC:

<http://digital.csic.es/handle/10261/24626>.

PARODI, Giovanni (2010): *Lingüística de corpus. De la teoría a la empiria*, Madrid-Frankfurt am Main, Iberoamericana-Vervuert.