

# Análisis cuantitativo de erratas del *Diccionario Terminológico de las Ciencias Farmacéuticas Inglés-Español/Spanish-English* (Ariel, 2007)

Santiago Rodríguez-Rubio Mediavilla\*

**Resumen:** En los textos del ámbito biosanitario, la corrección del lenguaje es fundamental. Por otro lado, un diccionario debe mostrar un nivel de corrección muy elevado al tratarse de una obra de referencia. Un análisis manual pormenorizado del *Diccionario Terminológico de las Ciencias Farmacéuticas, Inglés-Español/Spanish-English* (Ariel, 2007) reveló una incidencia muy significativa de erratas. El presente artículo pretende demostrar la necesidad de corregir este diccionario en papel, así como la conveniencia de adaptarlo al entorno digital.

**Palabras clave:** ciencias farmacéuticas, Damerau, diccionario médico, discordancia de género, discordancia de número, error tipográfico, lenguaje médico, procesamiento del lenguaje natural (PLN).

## **A quantitative analysis of typographical errors in *A Terminological Dictionary of the Pharmaceutical Sciences, English-Spanish/Español-Inglés* (Ariel, 2007)**

**Abstract:** Language accuracy is essential in medical texts. Furthermore, a dictionary, as a reference work, must feature a very high degree of correctness. A thorough manual analysis of *A Terminological Dictionary of the Pharmaceutical Sciences, English-Spanish/Español-Inglés* (Ariel, 2007) revealed a large number of typographical errors. This article aims to demonstrate that this paper dictionary must be corrected, and that it would be advisable to adapt it to the digital environment.

**Keywords:** pharmaceutical sciences, Damerau, medical dictionary, gender disagreement, number disagreement, typographical error, medical language, natural language processing (NLP).

Panace@ 2018; XIX (47): 76-88

Recibido: 27.IV.2018. Aceptado: 28.V.2018

## 1. Introducción

Un análisis manual pormenorizado del *Diccionario Terminológico de las Ciencias Farmacéuticas Inglés-Español/Spanish-English* (Ariel, 2007), en adelante «DTCF», reveló una incidencia muy significativa de dos tipos de erratas: por un lado, erratas independientes del contexto o *non-word errors* por sustracción, adición, sustitución, transposición o repetición de letras (a partir de Damerau, 1964); por otro lado, erratas dependientes del contexto (o *real-word errors*), incluyéndose en esta última categoría sustituciones de palabras que en algunos casos implican discordancias de género o número.

Es habitual que los profesionales hispanohablantes del campo del procesamiento del lenguaje natural (PLN) empleen los términos ingleses *non-word error* y *real-word error*, incluso cuando escriben en español. No obstante, en el referido campo, las expresiones españolas «independiente del contexto» (o «de palabra no válida») y «dependiente del contexto» (o «de palabra válida») están sancionadas por el uso para referirse a *non-word* y *real-word*, respectivamente. En este artículo emplearemos las denominaciones «errata independiente del contexto» y «errata dependiente del contexto», por ser las más sintéticas.

El análisis del DTCF forma parte de un trabajo de revisión mucho más amplio que abarca un corpus de veintidós diccionarios especializados bilingües en papel de las editoriales Ariel y Gestión 2000, ambos sellos pertenecientes al Grupo Planeta (véase relación de diccionarios analizados en el Anexo). El corpus, de más de 15 000 páginas de extensión, fue elegido según criterios de relevancia y representatividad. De ello da buena muestra el hecho de que trece de los dieciocho diccionarios de Ariel analizados (incluido el DTCF) son objeto de un capítulo<sup>1</sup> en *The Routledge Handbook of Lexicography*, importante obra de referencia en el campo de la lexicografía a nivel mundial.

Por otro lado, el trabajo global de revisión realizado es un estudio relevante, novedoso y exhaustivo en el campo de la corrección formal de diccionarios especializados bilingües españoles en papel, fundamentalmente en la combinación de lenguas inglés-español. Los hallazgos demuestran la existencia de un claro paradigma de erratas en el corpus de referencia. Este paradigma podría constituir una sólida base para ciertas aplicaciones, no solo en el campo de la lexicografía, sino también en determinadas parcelas de la lingüística computacional, como puede ser el aprendizaje automático (*machine learning*) de correctores.

\* Filólogo y traductor independiente (España). Dirección para correspondencia: [santirrm@hotmail.com](mailto:santirrm@hotmail.com).

### 1.1. Importancia de la corrección en el lenguaje médico

La corrección del lenguaje es fundamental en los textos del ámbito biosanitario. Aleixandre *et al.* (2015: 396) sostienen:

El lenguaje médico debe caracterizarse por su precisión, neutralidad emocional y estabilidad. La comunicación efectiva de los resultados de los trabajos científicos depende del cumplimiento de las normas vigentes de redacción y estilo, de manera que en un texto con defectos formales puede quedar mermado el interés de los hallazgos.

Los autores abundan en la cuestión: «La comunicación científica debe ser correcta en el fondo y en la forma...». En relación con las cuestiones formales, indican la necesidad de evitar los errores ortográficos, entre otros aspectos (2015: 397).

Aleixandre e Iscla sostienen que los defectos formales no suponen la invalidación total de un texto científico, «pero lo desvirtúan, quedando en entredicho el contenido entero». A continuación, se refieren a un intangible fundamental: «El autor que descuida las formas inspira justa desconfianza» (1997: 5).

En el prefacio de la séptima edición del diccionario *Lexicon Medicum; or Medical Dictionary* (Robert Hooper) el editor señalaba que en su época no se solía tomar nota de los errores tipográficos, lo cual conllevaba que dichos errores se transmitieran de una edición a otra, hasta el punto de afectar al propio significado. El editor indicaba a continuación: «In this edition of Hooper's Dictionary, such errors, of whatever kind, as are likely to mislead, are noticed in a table of errata» (1839: vii).

En el campo del PLN, se desarrollan técnicas y herramientas para detectar erratas y errores automáticamente. Como veremos en este artículo, esas herramientas siguen teniendo a día de hoy limitaciones, de modo que el factor humano sigue siendo indispensable. Más allá de que los avances en el referido campo puedan ir desarrollando herramientas que corrijan erratas de difícil detección en los textos médicos de referencia, la corrección del lenguaje empleado en estos últimos es fundamental para que la información pueda ser recuperada y transmitida de manera eficaz. Los mecanismos de indexación y de recuperación son clave en la gestión de la información en la era digital.

### 1.2. Diccionarios como obras de referencia

Un diccionario es una obra de referencia, una de cuyas funciones principales es la de satisfacer las necesidades de información de sus usuarios. En ese sentido, el nivel de corrección formal de un diccionario debe ser muy elevado. La presencia de erratas en lemas del DTCF (unida a la incidencia significativa de erratas detectada en muchas otras ubicaciones) contravendría un consenso tácito en virtud del cual la consulta de un diccionario debe garantizar que el usuario accederá a información correcta, no solo desde el punto de vista del contenido, sino también desde el de la forma.

En el número 17 de los Cuadernos de la Fundación Dr. Antonio Esteve, San Andrés menciona entre otras fuentes de

información el uso de diccionarios para redactar textos médicos (2009: 67). En la misma línea, Aleixandre e Iscla se refieren a la conveniencia de usar diccionarios y otras obras de referencia, dado que pueden «ayudar a una mayor corrección en el discurso y en la escritura de textos médicos» (1997: 5).

En el campo de la farmacología, más concretamente del «Sistema Español de Farmacovigilancia de Medicamentos de Uso Humano», encontramos la base de datos FEDRA, cuya versión más moderna usa obras de referencia, como pueda ser el diccionario médico electrónico MedDRA<sup>2</sup> (Madurga *et al.*, 2010: 134).

### 1.3. Reto de adaptación de la lexicografía especializada bilingüe española a la era digital

Más allá de que se deba corregir el DTCF dada la alta incidencia de erratas, sería conveniente adaptar el diccionario a un contexto en el que la preeminencia de lo digital es, y seguirá siendo, cada vez mayor.

Las restricciones espacio-temporales no afectan de la misma forma a los diccionarios electrónicos y a los diccionarios en papel. Las obras de referencia electrónicas pueden ser actualizadas y corregidas de manera más o menos continua, mientras que, en el caso de un diccionario en papel, se ha de esperar a la siguiente reedición para implementar dicha actualización o corrección. Cuestiones distintas serían que esa revisión continua a la que se puede (y debe) someter a los diccionarios electrónicos se lleve o no a la práctica, y con qué periodicidad o grado de exhaustividad se haga en cada caso.

En relación con los diccionarios españoles bilingües del campo de la economía, Mateo (2014: 44) indica que, aunque en España los diccionarios en papel siguen siendo las herramientas de consulta predominantes entre los traductores profesionales españoles, esta tendencia podría cambiar, dadas las ventajas que ofrecen los diccionarios digitales en términos de inmediatez y dada su creciente fiabilidad. A continuación, el autor se refiere al punto de inflexión en el que se encuentra la lexicografía bilingüe especializada en España, en el campo de la economía:

The current financial situation of many publishing houses in Spain may also have a decisive role in this shift of tendency. Most of them are struggling to survive not only the present economic crisis but also the increasing presence of the eBook and other electronic reading devices. This condition is reflected, in my experience, in the publishers [sic] conservative and reluctant attitude to publish and update specialized bilingual dictionaries on paper often enough to meet current needs and new challenges.

Mateo (2018: 434) también explica la reticencia de las editoriales a la hora de dedicar recursos a la lexicografía electrónica especializada:

Logically, printed lexicographical works should be abandoned and replaced by more accessible electronic versions. However, in our opinion, online specialised

bilingual dictionaries today are not in all cases as good and stable as they should be. Publishing houses are still reluctant to allocate the necessary resources in electronic specialised lexicography. They require a constant update which demands full- or part-time staff and, more importantly, a large client base willing to pay for this service.

Leroyer (2013: 129-138) ilustra la transformación de los diccionarios tradicionales en herramientas lexicográficas de información (*lexicographic information tools*) mediante cuatro herramientas de distintas ramas del conocimiento. La primera de esas herramientas pertenece al ámbito biosanitario y se denomina *Lexonco (lexique oncologique)*. Se trata de un recurso terminológico concebido para solucionar problemas de información que puedan surgirles a pacientes de cáncer y a sus familiares (2013: 130).

Conforme avance el siglo XXI, irá aumentando el número de usuarios que demandarán recursos lexicográficos adaptados a su tiempo. En ese sentido, el trabajo de revisión de veintidós diccionarios que se ha realizado ilustra el reto de adaptación a la era digital de los diccionarios especializados bilingües españoles en papel.

En el ámbito biosanitario existen a día de hoy importantes obras de referencia electrónicas (monolingües, bilingües o multilingües), ya sean obras digitales de nueva planta o versiones digitales de anteriores obras en papel. Entre esas obras se encuentran el *Diccionario de términos médicos* (Real Academia Nacional de Medicina, RANM), el *Diccionario crítico de dudas inglés-español de medicina* o «Libro rojo» (Navarro), el recurso multilingüe arriba mencionado MedDRA y el *Stedman's Medical Dictionary*, por citar algunos ejemplos. Actualmente se encuentra en desarrollo un proyecto para crear el *Diccionario panhispánico de términos médicos* (RANM y Asociación Latinoamericana de Academias Nacionales de Medicina, España y Portugal), que pretende convertirse en obra de referencia de la lexicografía médica en español.

## 2. Antecedentes

En el pasado, se han realizado estudios de defectos formales en textos del ámbito biosanitario. Sin embargo, no existen investigaciones que se hayan centrado de manera exhaustiva en la detección de erratas en diccionarios médicos españoles.

Ferner y Aronson han realizado un estudio cuantitativo de variaciones de nombres de fármacos en diferentes bases de datos del ámbito biosanitario. Los autores afirman que los errores ortográficos no son inusuales en bases de datos como PubMed y Medline (2016: 1). En el estudio se establecieron tres categorías de nombres: *standard names* (nombres normalizados), *variant names* (nombres cuya ortografía difería de la de nombres normalizados) y *hidden reference variants* (nombres cuya ortografía difería de la de nombres normalizados y que remitían a publicaciones en bases de datos como PubMed, al ser introducidos como términos de búsqueda). Los nombres de la segunda y tercera categoría se generaban, según los autores, por sustitución, omisión, adición, transposición, dupli-

cación [ $l > ll$ ], eliminación de alguna letra duplicada [ $ll > l$ ], o bien por alguna combinación de las anteriores (2016: 1-2). Ferner y Aronson señalan que los autores y editores deberían prestar más atención al hecho de escribir los nombres de los fármacos de manera correcta, así como que los indexadores de bases de datos como PubMed deberían vincular los nombres ortográficamente incorrectos a sus correspondientes nombres normalizados. Los autores concluyen que, cuando los investigadores buscan nombres de fármacos en bases de datos como las referidas anteriormente, deberían incluir nombres escritos de manera incorrecta, puesto que, caso de introducir solo nombres normalizados de fármacos, los resultados de las búsquedas podrían no mostrar las publicaciones indexadas mediante nombres ortográficamente incorrectos (2016: 1).

## 3. Materiales/Métodos

### 3.1. Materiales

El *Diccionario Terminológico de las Ciencias Farmacéuticas, Inglés-Español/Spanish-English* (Ariel, 2007) es un diccionario especializado bilingüe bidireccional en papel (sin versión electrónica), cuyo cuerpo tiene una extensión de 1092 páginas. Los autores del diccionario son Alfonso Domínguez-Gil Hurlé, Enrique Alcaraz Varó y Raquel Martínez Motos. La obra fue editada en 2007 y reimpresa en 2011, siendo esta última impresión la que se ha analizado en la revisión objeto del presente artículo. La obra cuenta con el respaldo de la Real Academia Nacional de Farmacia (RANF).

Se trata, pues, de un referente en el campo de los diccionarios españoles del ámbito biosanitario. En el número 19 de *Panace@* (marzo de 2005), dos de los autores presentaron en primicia una breve descripción y una pequeña muestra del contenido del diccionario que vería la luz dos años después (<<https://goo.gl/CqL8Ng>>).

En cuanto al sello editorial, cabe señalar que, según el sistema orientativo SPI<sup>3</sup>, realizado por el grupo del CSIC «Investigación sobre el Libro Académico (ILIA)», la editorial Ariel figura en los últimos datos publicados (2014) en el segundo puesto del ranquin general de la categoría «Prestigio editorial», en el ámbito de las editoriales académicas españolas.

### 3.2. Métodos

El DTCF se revisó en toda su extensión y de manera exhaustiva (página a página). No solo se analizó el cuerpo (partes «Inglés-Español» y «Spanish-English»), sino también el resto de apartados de la obra (las portadas interiores, el prólogo y la introducción). En los resultados de este artículo únicamente figuran las erratas del cuerpo, por ser esta la parte medular del diccionario.

Por ser especialmente relevantes en términos tanto cuantitativos como cualitativos, en este artículo nos referiremos a dos de las categorías de erratas o correcciones descritas en el trabajo de análisis del DTCF (que incluye, entre otros aspectos, correcciones en el uso de signos de puntuación, de referencias internas, de la negrita, de la cursiva, de la mayúscula y de los espacios). Las dos categorías elegidas son las erratas independientes del contexto o *non-word errors*

(sustracción, adición, sustitución, transposición o repetición de letras) y las erratas dependientes del contexto o *real-word errors* (sustitución, repetición u omisión de palabras).

En el trabajo de análisis del DTCF se aplicó un enfoque cuantitativo porque se consideró que era lo más adecuado para analizar de manera exhaustiva los patrones de erratas presentes en la obra. Se prestó especial atención a los patrones de repetición de erratas, teniendo en cuenta tanto las que se repetían en el marco de una sola de las partes del cuerpo del diccionario («Inglés-Español» o «Spanish-English») como aquellas que lo hacían de una parte a otra. La repetición de una misma errata se indicó mediante el símbolo «=>», mientras que la existencia de una errata similar se indicó mediante el símbolo «~» (véase la tabla 1). Ampliando el foco, en algunos casos se detectaron erratas iguales o similares en otros de los diccionarios del corpus, de manera intertextual.

Las repeticiones de erratas se incluyeron en el cómputo, es decir, cada repetición se contó a su vez como errata. Decidimos no descontar las repeticiones porque el número de erratas repetidas representa un porcentaje bajo (en torno a un 9 %) sobre el total de casos. Así pues, en el DTCF no existe un esquema de pocas erratas que se repiten muchas veces (en cuyo caso sí habría sido conveniente descontar las repeticiones), sino uno de muchas erratas diferentes, algunas de las cuales se repiten relativamente pocas veces.

Para reflejar los patrones de erratas del DTCF, se decidió incluir una serie de fórmulas que expresan relaciones lógicas. Entre esas fórmulas que proponemos están las de analogía (v. gr. repetición de una misma errata o existencia de una errata similar) o las de contraposición. La consecuencia práctica de este método es que, a diferencia de una fe de erratas<sup>4</sup>, en el análisis realizado las erratas similares o contrapuestas aparecen agrupadas mediante uno o varios niveles de indexación. Los niveles de indexación y las expresiones de relaciones lógicas (v. gr. «a contrario sensu», «errata repetida en», etc.) que aparecen en las tablas del presente artículo son elementos que no forman parte del contenido del diccionario, sino que han sido introducidos por el autor. Véase la tabla 1.

Tabla 1. Niveles de indexación y relaciones lógicas (repetición de erratas, erratas similares, etc.)		
Pág. del DTCF	Representación del contenido de la entrada del DTCF <sup>5</sup>	Comentarios
166	<b>code<sup>1</sup> &gt; code<sup>2</sup></b> (V. <i>nacional drug code</i> )	Errata dependiente del contexto (interlingüística). Debería indicarse en inglés
220	~ <b>diet &gt; dietary</b> (V. <i>nutricional</i> )	
434	= <b>nutritional imbalance</b> (V. <i>nutricional balance</i> )	

Tabla 1 (cont.). Niveles de indexación y relaciones lógicas (repetición de erratas, erratas similares, etc.)		
Pág. del DTCF	Representación del contenido de la entrada del DTCF <sup>5</sup>	Comentarios
181	~ a contrario sensu <b>confection<sup>1</sup></b> > <b>confectionery<sup>1</sup></b> ( <i>NUTRITION</i> )	Debería indicarse en español
435	<b>object<sup>1</sup> &gt; objective<sup>2</sup></b>	Errata independiente del contexto (sustracción de una letra) presente de manera intertextual en otros dos de los diccionarios del corpus
435	= errata repetida en <b>object<sup>1</sup></b> (V. <i>objective<sup>2</sup></i> )	Por otro lado, sobra espacio entre término subyacente y superíndice 2
573	~ errata similar por antonimia en <b>subject<sup>1</sup> &gt; subjective</b> ( <i> Symptoms are subjective...</i> )	

### 3.2.1. Consideraciones en torno a erratas independientes del contexto (*non-word errors*)

La expresión *non-word errors* se refiere a erratas (o a errores, en ciertos supuestos<sup>6</sup>) que implican la creación de un término que no figura en ningún diccionario, por ser incorrecto desde el punto de vista idiomático (Ahmed *et al.*, 2009: 39). Estas erratas se originan típicamente por un fallo humano.

De las dos categorías de erratas objeto del presente artículo, la de las erratas independientes del contexto es la más fácil de detectar tanto por medios manuales como automáticos.

Para clasificar los *non-word errors* tomamos como referencia el estudio seminal de Damerau (1964), que establecía cuatro categorías de «errores ortográficos» (*spelling errors*):

1. Error por sustitución de una única letra.
2. Error por sustracción de una única letra.
3. Error por adición de una única letra.
4. Error por transposición de dos letras adyacentes.

Damerau señaló que más del 80 % de los «errores» detectados en el marco de su investigación entraron en alguna de las cuatro categorías arriba indicadas (1964: 171).

En el trabajo de análisis objeto del presente artículo, se indicaron no solo las erratas por sustracción de una única letra, sino también aquellas por sustracción de más de una letra. Lo mismo cabe decir en relación con las erratas por adición. La transposición puede ser de dos o más letras no necesariamente adyacentes. Por otro lado, se incluyó una categoría

que no figuraba en el artículo de Damerau, a saber, la de la repetición de sílabas o grupos de letras. Finalmente, se describieron erratas que combinan varias de las categorías anteriores, pero dichas erratas no figuran en los resultados de este artículo por tener una relevancia cuantitativamente menor. En síntesis, proponemos las siguientes cinco categorías de *non-word errors*:

1. Errata por sustitución de una única letra.
2. Errata por sustracción de una o más letras.
3. Errata por adición de una o más letras.
4. Errata por transposición de dos o más letras no necesariamente adyacentes.
5. Errata por repetición de sílaba o grupo de letras.

### 3.2.2. Consideraciones en torno a erratas dependientes del contexto (*real-word errors*)

La expresión *real-word errors* se refiere a erratas (o a errores, en ciertos supuestos) que implican la creación de una palabra correcta desde el punto de vista idiomático, pero incorrecta desde el punto de vista contextual. En otros términos, la palabra que constituye la errata es una palabra válida del idioma en cuestión, pero no es la palabra correcta teniendo en cuenta el contexto en el que aparece. En algunos casos, la errata también puede implicar la omisión de una palabra idiomática y contextualmente correcta, por ejemplo la omisión de una preposición. Los *real-word errors* pueden tener su origen en un fallo de la máquina o en uno humano.

Kukich (1992: 412) incluye entre los *real-word errors* aquellos que se crean mediante lo que denomina «simples erratas» (v. gr. *from* > *form*), aquellos que se forman mediante faltas gramaticales o sintácticas, incluyendo formas flexivas incorrectas (v. gr. *arrives* > *arrive*) y aquellos que se generan mediante la inserción o la supresión de palabras enteras.

Más específicamente, en el trabajo de análisis del DTCF se describieron dos subcategorías de erratas dependientes del contexto:

1. Erratas que implican la creación de una palabra, de manera interlingüística [v. gr. *nutritional* (ING) > *nutricional* (ESP)].
2. Erratas que implican la creación o la omisión de una palabra, de manera intralingüística [v. gr. *forearm* (ING) > *foreman* (ING)].

La extensa casuística de los *real-word errors* intralingüísticos detectados en el DTCF incluye repeticiones de palabras y omisiones de palabras, en la línea de lo que trasladaba Kukich. A los efectos de este artículo, los *real-word errors* intralingüísticos se dividen en:

1. Erratas que implican discordancia de género.
2. Erratas que implican discordancia de número.
3. Otras erratas intralingüísticas.

Los *real-word errors* constituyen un tema de estudio relevante en el campo del procesamiento del lenguaje natural

(PLN). A pesar de los avances en el campo de la lingüística computacional, la detección automática de erratas dependientes del contexto sigue siendo en buena medida un campo por explorar. Por ejemplo, un determinado *real-word error* será más o menos fácil de detectar automáticamente en función de si lleva aparejado un cambio gramatical o un cambio semántico. En palabras de Samanta y Chaudhuri (2013: 211): «Word error is a major hindrance to the real world applications of Natural Language Processing». Concluyen los autores que usualmente este tipo de erratas (o errores) solo pueden detectarse de manera manual.

Los correctores automáticos suelen detectar determinadas discordancias de género y número, si bien hay casos en los que las discrepancias escapan a su radar. Por ejemplo, una discordancia que implique que un determinante no concuerde con un sustantivo que le sigue inmediatamente será más fácil de detectar automáticamente que una discrepancia que implique una cadena en la que se intercalan elementos correctos y elementos incorrectos. Ng *et al.* (2014: 12) sostienen que, en el terreno de la corrección automática de discordancias gramaticales, existe un amplio margen de mejora, ya que los sistemas de corrección existentes ofrecen unas prestaciones limitadas en cuanto a su precisión.

## 4. Resultados

En la tabla 2 se muestran los resultados del trabajo de análisis del DTCF en lo que se refiere a erratas independientes del contexto (*non-word errors*) y a erratas dependientes del contexto (*real-word errors*), repeticiones de erratas incluidas. Indicamos la incidencia de cada categoría y subcategoría de errata. En la última columna de la tabla, figuran como referencia los datos de frecuencia del corpus de veintidós diccionarios. En el caso del DTCF, se muestran dos datos de frecuencia de erratas:

1. La frecuencia aproximada en función de las entradas, resultado del cociente entre el número total aproximado de entradas del diccionario (~ 28 000) y el número total de erratas. El número total aproximado de entradas viene indicado en la introducción del propio diccionario (p. XX).
2. La frecuencia en función de las páginas, resultado del cociente entre el número total de páginas del cuerpo del diccionario (= 1092) y el número de erratas.

El resultado global del DTCF revela una frecuencia de un caso de alguna de las dos categorías generales de erratas cada 1,58 páginas o cada aproximadamente 40,46 entradas. De manera desglosada, en el diccionario se ha descrito una frecuencia de una errata independiente del contexto cada 3,12 páginas (cada 80 entradas aproximadamente) y de una errata dependiente del contexto cada 3,19 páginas (cada 81,87 entradas aproximadamente).

La frecuencia global del DTCF de una errata cada 1,58 páginas es aproximadamente el doble que la del corpus de veintidós diccionarios (cada 3,63 páginas). La frecuencia del

Tabla 2. Frecuencia de erratas en el DTCF en relación con corpus general

CATEGORÍA ERRATA	N.º CASOS	FRECUENCIA	FRECUENCIA	FRECUENCIA
	DTCF	(X ENTRADAS)	(X PÁGS.)	CORPUS (X PÁGS.)
<b>Independiente del contexto (<i>Non-word error</i>)</b>	<b>350</b>	<b>80,00</b>	<b>3,12</b>	<b>6,34</b>
Sustracción	179	156,42	6,10	12,96
Adición	75	373,33	14,56	29,36
Sustitución una letra	56	500,00	19,50	35,30
Transposición	20	1400,00	54,60	70,76
Repetición grupos	20	1400,00	54,60	253,35
<b>Dependiente del contexto (<i>Real-word error</i>)</b>	<b>342</b>	<b>81,87</b>	<b>3,19</b>	<b>8,49</b>
Intralingüística	210	133,33	5,20	11,64
Disc. número	68	411,76	16,06	53,61
Disc. género	32	875,00	34,13	118,11
Otros	110	254,55	9,93	17,00
Interlingüística	132	212,12	8,27	31,35
<b>Todas las categorías</b>	<b>692</b>	<b>40,46</b>	<b>1,58</b>	<b>3,63</b>

corpus es el resultado del cociente entre el número total de páginas de los cuerpos de todos los diccionarios (= 15 708) y el número total de erratas del corpus (= 4327), repeticiones de erratas incluidas.

Las dos subcategorías de erratas más frecuentes en el DTCF son, por orden descendente, la de las erratas intralingüísticas y la de las erratas por sustracción. Los datos del trabajo global de revisión de veintidós diccionarios confirman esta pauta.

Como se ha mencionado antes, en los resultados expuestos solo figuran las erratas presentes en el cuerpo del diccionario (partes «Inglés-Español» y «Spanish-English»). Así pues, no se han incluido las erratas presentes en otras partes de la hiperestructura de la obra, como las portadas interiores, el prólogo o la introducción. En el grupo de casos que no figuran en los resultados, cabe destacar por su visibilidad y repercusión la errata por adición presente en la parte en inglés del título de ambas portadas interiores: *A TERMINOLOGICAL DICTIONARY OF THE PHARMACEUTICAL SCIENCES* (páginas III y V de la obra). Al formar parte del título de la obra, esta errata trasciende los límites del propio diccionario; aparece por ejemplo en la versión electrónica de la primera edición de *The Routledge Handbook of Lexicography*, más concretamente en el apartado «Referencias» del capítulo «The Alicante Dictionaries» (Mateo). A fecha de presentación del presente artículo, la errata aparece igualmente en la página web de la Biblioteca Virtual de la Real Academia Nacional de Farmacia (RANF), aunque en este caso se indica mediante la expresión «[sic]» (<<https://goo.gl/WRGskt>>).

#### 4.1. Incidencia de erratas independientes del contexto (*non-word errors*)

En la figura 1 se observa que las erratas por sustracción representan aproximadamente la mitad del total de erratas independientes del contexto. La suma de las erratas por sustracción y las erratas por adición representa aproximadamente un 70 % del total. Los datos del trabajo global de revisión de veintidós diccionarios confirman estas pautas.

A continuación, indicamos el porcentaje de casos de cada subcategoría de *non-word error*:

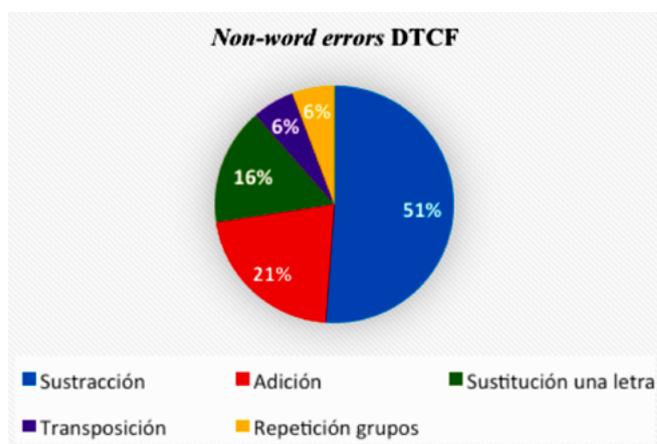


Figura 1. Incidencia de *non-word errors* en el DTCF

En el DTCF, se han descrito cinco erratas independientes del contexto en términos de las cabeceras de página, concre-

tamente en las subcategorías «Adición de una o más letras» y «Sustitución de una única letra» (tablas 4 y 5, respectivamente).

A continuación se indican algunos ejemplos de *non-word errors* del DTCF. Señalamos en algunos casos el patrón de repetición de erratas.

#### 4.1.1. Sustracción de una o más letras

Tabla 3. Ejemplos de erratas por sustracción en el DTCF		
Pág. del DTCF	Representación del contenido de la entrada del DTCF	Comentarios
55 292	<b>anastrozole</b> (V. <i>aromatase</i> <i>ihibitor</i> ) = <b>formestane</b> (V. <i>aromatase</i> <i>ihibitor</i> )	Debería indicarse « <i>inhibitor</i> » Véanse en la tabla 4 ejemplos de erratas por adición con el mismo término subyacente (« <i>inhibitor</i> »)
164	<b>clotrimazole</b> (◇ <i>Clotimazole is an antigungal</i> [sic]...; V. <i>anfungal</i> )	Debería indicarse « <i>Clotrimazole</i> » y « <i>antifungal</i> », respectivamente. Por otro lado, errata por sustitución de una única letra en « <i>antigungal</i> »
357 357 358	<b>JCAH</b> (V. <i>Joint Commision on Accreditation of Hospitals</i> ) = <b>JCAHO</b> (V. <i>Joint Commision on Accreditation of Healthcare Organizations</i> ) = <b>joint<sup>1</sup></b> > <b>Joint Commision on Accreditation of Healthcare Organizations, JCAHO</b>	Debería indicarse « <i>Commission</i> » Variación de la errata en <b>SPANISH-ENGLISH 703: azociclotina</b> (European <i>Comission</i> ) N. B. En otros 9 de los diccionarios analizados, aparece esta errata (o alguna variación de ella)
336 444	<b>immune</b> > <b>immunity</b> ( <i>-foreign substances-</i> ) ~ <b>organo-</b> ( <i>-organic life or substances-</i> )	Debería indicarse « <i>substances</i> » Variación de la errata en <b>SPANISH-ENGLISH 904: insensibilización</b> (... to a drug or other <i>subtances</i> )
762	<b>competividad feroz</b>	Debería indicarse « <b>competitividad feroz</b> ». Errata repetida de manera intertextual en otro de los diccionarios del corpus

#### 4.1.2. Adición de una o más letras

Tabla 4. Ejemplos de erratas por adición en el DTCF		
Pág. del DTCF	Representación del contenido de la entrada del DTCF	Comentarios
220	<b>DICYCLOMINE HYDROCHLORIDE</b>	Debería indicarse « <b>HYDROCHLORIDE</b> ». Errata en término de cabecera de página
265	<b>euro-</b> > <b>European Pharmacopeia</b> (◇ ... <i>and turbidimetric assays are contained in ther European Pharmacopeia</i> )	Adición por partida doble. Debería indicarse « <b>assays</b> » y « <b>the</b> », respectivamente
289 331	<b>flupentixol</b> (... las que cursan con <b>alucionaciones</b> ...) = <b>hypnotic</b> > <b>hypnosis</b> (... es una [sic] estado similar al sueño... inducida [sic]... en el que pueden surgir... <b>alucionaciones</b> ... ◇ ... <i>on a full range of options from aspirin hypnosys</i> [sic])	Debería indicarse «alucinaciones» Además, discordancia de género en «una estado... inducida» y errata por sustitución de una única letra en « <i>hypnosys</i> »
453 595	<b>pad<sup>1</sup></b> > <b>padding<sup>1</sup></b> (◇ ... <i>is necesssary</i> ) = <b>tissue repair</b> (◇ <i>Valine is necesssary</i> ...)	Debería indicarse « <b>necessary</b> ». Errata similar en <b>SPANISH-ENGLISH 775: crédito<sup>1</sup></b> (◇ <i>Los créditos blandos son necessarios</i> ...)
588	<b>tenofovir disoproxil</b> ( <i>-nucleoside reverse transcriptase inhibitors-</i> ... V. <i>nucleoside reverse transcriptase inhibitors</i> )	Debería indicarse « <i>inhibitors</i> ». Por partida doble

4.1.3. Sustitución de una única letra

Tabla 5. Ejemplos de erratas por sustitución de una única letra en el DTCF		
Pág. del DTCF	Representación del contenido de la entrada del DTCF	Comentarios
254	<b>ENALAPRIL HYDROCR</b> LORIDE	Debería indicarse « <b>HYDROCH</b> LORIDE». Errata en término de cabecera de página. En la errata por adición en término de cabecera de página indicada arriba, se añade la «R», mientras que en este caso la «R» sustituye a la letra correcta «H»
316	<b>half &gt; halve</b> (◊ ... <i>risk of developing Alzheimer</i> )	Debería indicarse « <b>Alzheimer</b> ». Implica, a su vez, transposición de letras N. B. En inglés, se suele emplear genitivo sajón: « <i>Alzheimer's</i> » o « <i>Alzheimer's disease</i> »
519	<b>repair</b> <sup>1</sup> (◊ <i>Proteins are necesssry...</i> ; V. <i>retoration</i> [sic])	Debería indicarse « <b>necessary</b> ». Además, errata por sustracción en « <i>retoration</i> »
835	<b>esteroide</b> <sup>1</sup> > <b>esteroide</b> <sup>2</sup> (... for their cortocoid activity... and their use as hormonal subtitutes [sic])	Debería indicarse « <b>cortico</b> id». Además, errata por sustracción en « <i>subtitutes</i> »
902	<b>inhibitorio</b> (PHARMACOLOGY/BIOCHENISTRY)	Debería indicarse « <b>BIOCHEM</b> ISTRY» Errata en denominación de campo semántico. Véase en <b>5. Discusión</b> comentario sobre erratas en campos semánticos
946	<b>mio-</b> > <b>miorrelajente</b>	Debería indicarse « <b>miorrelajante</b> »

4.1.4. Transposición de dos o más letras no necesariamente adyacentes

Tabla 6. Ejemplos de erratas por transposición en el DTCF		
Pág. del DTCF	Representación del contenido de la entrada del DTCF	Comentarios
130 296	<b>candidiasis</b> (◊ <i>Candiadisis can be...</i> )  = <b>funga</b> l > <b>fungu</b> s (V. <i>candiadisis...</i> )	Debería indicarse « <b>Candidi</b> asis». Errata repetida en <b>SPANISH-ENGLISH 726: candida</b> [sic] > <b>candidiasis, candidosis</b> (◊ <i>La candiadisis...</i> ) + 888: <b>hongo</b> (S. <i>candiadisis...</i> )
506	<b>rachis</b> > <b>rachitis</b> <sup>1</sup> , <b>ratic</b> hism, rhachitis	Debería indicarse « <b>rachit</b> ism». Errata repetida en <b>SPANISH-ENGLISH 1009: raquitico</b> > <b>raquitismo</b> ( <i>ratic</i> hism)
607	<b>turbid</b> > <b>turbidimetric assays</b> ( <i>-spectrotophometer-</i> )	Debería indicarse « <b>spectro</b> photometer»
803	<b>disfagia</b> (PHATOLOGY)	Debería indicarse « <b>PATHO</b> LOGY» Errata en denominación de campo semántico. Véase en <b>5. Discusión</b> comentario sobre erratas en campos semánticos

## 4.1.5. Repetición de sílabas o grupos de letras

Tabla 7. Ejemplos de erratas por repetición de sílabas o grupos de letras en el DTCF

Pág. del DTCF	Representación del contenido de la entrada del DTCF	Comentarios
160	<b>clenbuterol</b> ( <i>–obstructive pulmonary diseases–</i> )	Debería indicarse « <i>obstructive</i> »
292	= <b>formoterol</b> ( <i>–obstructive pulmonary diseases–</i> )	Errata repetida en <b>SPANISH-ENGLISH</b> 860: <b>formoterol</b> (and other <b>obstructive</b> pulmonary diseases) + 1055: <b>teofilina</b> (and other chronic <b>obstructive</b> pulmonary diseases)
590	= <b>theophylline</b> ( <i>–chronic obstructive pulmonary diseases–</i> )	
514	<b>red</b> <sup>1</sup> > <b>red heat</b> (... de un [sic] <b>sustancia</b> ...)	Debería indicarse «sustancia». Además, discordancia de género
942	<b>metilina</b> ( <i>–penicilina semisintética–</i> )	Debería indicarse « <i>penicilina</i> »

4.2. Incidencia de erratas dependientes del contexto (*real-word errors*)

En la tabla 2 se observa cómo el número de discordancias de número es sensiblemente superior al de discordancias de género, siendo las erratas intralingüísticas superiores a las

interlingüísticas. Los datos del trabajo global de revisión de veintidós diccionarios confirman estas pautas.

A continuación se indican algunos ejemplos de *real-word errors* del DTCF. Señalamos en algunos casos el patrón de repetición de erratas.

## 4.2.1. Erratas intralingüísticas

## 4.2.1.1. DISCORDANCIAS DE NÚMERO

Tabla 8. Ejemplos de discordancias de número en el DTCF

Pág. del DTCF	Representación del contenido de la entrada del DTCF	Comentarios
215	<b>desiccant</b> (son <b>sustancia</b> muy higroscópicas...)	Debería indicarse «son <u>sustancias</u> ...»
325	<b>hirudin</b> (... glándulas salivarias de <b>la</b> sanguijuelas <i>–salivary glands of leeches–</i> )	Debería indicarse «de <u>las</u> sanguijuelas», por coherencia de número con la correspondiente expresión en inglés, si bien «de <u>la</u> sanguijuela» también sería correcto en español. Errata repetida en <b>SPANISH-ENGLISH</b> 887: <b>hirudina</b> (◇ ... <i>glándulas salivarias de la sanguijuelas</i> )
644	<b>absoluto</b> (◇ <i>El reposo absoluto es <b>imprescindibles</b>...</i> )	Debería indicarse « <i>El reposo absoluto es <u>imprescindible</u>...</i> »
661	<b>adelanto</b> <sup>1</sup> (◇ ... <i>un gran adelanto para <b>las</b> humanidad</i> )	Debería indicarse «... para <u>la</u> humanidad»
996	<b>propiedad</b> <sup>1</sup> > <b>propiedades terapéuticas</b> (◇ <i>Las hojas del laurel <b>tiene</b> importantes...</i> )	Debería indicarse « <i>Las hojas... <u>tienen</u> importantes...</i> »

4.2.1.2. DISCORDANCIAS DE GÉNERO

Tabla 9. Ejemplos de discordancias de género en el DTCF		
Pág. del DTCF	Representación del contenido de la entrada del DTCF	Comentarios
131	<b>capital, cap</b> (–artilugios que ahorran <b>un</b> gran inversión de capital–... –industrias que requieren <b>un</b> gran inversión de capital–)	Debería indicarse «... <u>una</u> gran inversión de capital». Por partida doble Erratas repetidas de manera intertextual en otro de los diccionarios analizados, en la entrada <b>capital, cap</b> de dicho diccionario (exactamente en las mismas frases que figuran en el DTCF)
208	<b>deferoxamine mesilate</b> (... se usa en el tratamiento del envenenamiento... <b>causada</b> por <b>un</b> ingesta accidental...)	Debería indicarse «... <u>causado</u> por <u>una</u> ingesta accidental...»
634	<b>xylitol</b> (... y también en la industria <b>farmacéutico</b> )	Debería indicarse «... en la industria <u>farmacéutica</u> »
800	<b>dieta equilibrada</b> (◊ <i>Un dieta equilibrada es esencial...</i> )	Debería indicarse « <u>Una</u> dieta equilibrada...»
827	= <b>equilibrado</b> (◊ <i>Un dieta equilibrada es esencial...</i> )	

4.2.1.3. OTRAS ERRATAS INTRALINGÜÍSTICAS

Tabla 10. Ejemplos de otras erratas intralingüísticas en el DTCF		
Pág. del DTCF	Representación del contenido de la entrada del DTCF	Comentarios
190	<b>correlate &gt; correlation</b> (... alude a una correspondencia cuantitativa – <i>qualitative correspondence</i> – entre dos entidades comparables – <i>between to comparable entities</i> –)	Debería indicarse «... – <i>between two comparable</i> ...–». Además, « <u>cuantitativa</u> » no concuerda con <i>qualitative</i> (uno de los dos términos es erróneo)
283	<b>filth &gt; filth disease</b> (es una enfermedad <u>causada</u> la contaminación del suelo...)	Omisión de preposición. Debería indicarse «... <u>causada por</u> la contaminación...»
300	<b>gastrostomy tube</b> (◊ ... <i>from a leaking <b>gastronomy</b> tube</i> )	Debería indicarse « <u>gastrostomy</u> »
614	<b>U.S. &gt; U.S. Pharmacopoeia <b>Convection</b>, USP</b> (Farmacopea <b>Farmacopea</b> de los Estados Unidos)	Debería indicarse « <b>Convention</b> ». Por otro lado, repetición del término «Farmacopea»
631	<b>wrist</b> (– <i>foreman and the hands</i> –)	Debería indicarse « <u>forearm</u> »
681	<b>ampolla</b> <sup>1</sup> (... is hermetically sealed by <b>meeting</b> glass)	Debería indicarse « <u>melting</u> »
690	<b>antipsicótico</b> (... as tranquilizers and <b>and</b> sedatives...)	Repetición de palabra

## 4.2.2. Erratas interlingüísticas

Tabla 11. Ejemplos de erratas interlingüísticas en el DTCF

Pág. del DTCF	Representación del contenido de la entrada del DTCF	Comentarios
186	<b>continuous<sup>1</sup> &gt; continuous subcutaneous insulin infusión, CSII</b>	Debería indicarse « <b>infusion</b> ». Errata repetida en <b>SPANISH-ENGLISH 921: lecitina</b> (... by intravenous <b>infusión</b> )
344	= <b>infuse<sup>1</sup> &gt; infusion<sup>2</sup></b> (V. <i>continuous subcutaneous insulin infusión</i> )	Por otro lado, sobra espacio entre término subyacente y superíndice 1
204	<b>DAW &gt; DAW code</b> (V. <i>dispense as <b>griten</b></i> )	Debería indicarse « <b>written</b> »
341	<b>indigestible &gt; indigestion</b> (◇ ... <i>in the treatment of <b>indigestión</b>...</i> )	Debería indicarse « <b>indigestion</b> ». Errata repetida en <b>SPANISH-ENGLISH 804: dispepsia</b> ( <b>dispepsia</b> , gastric <b>indigestión</b> )
345	<b>inicial therapy</b>	Debería indicarse « <b>initial</b> »
353	<b>irritable &gt; irritable bowel <b>síndrome</b></b>	Debería indicarse « <b>syndrome</b> »
958	<b>nitr-, nitro- &gt; nitrato</b> (... <b>sometimos</b> used as <b>anantihypertensive</b> [sic]...)	Debería indicarse « <b>sometimes</b> » Además, problema de repetición de letras o de ausencia de espacio en « <b>anantihypertensive</b> »

## 5. Discusión

Ningún trabajo de revisión manual o automática puede detectar el cien por cien de las erratas o problemas formales de un determinado texto de cierta extensión y complejidad. En este sentido, el nivel «cero-erratas» es una entelequia, pero no se debe renunciar a él como objetivo. Esta lógica se ha de aplicar también a la propia revisión que se ha realizado del *Diccionario Terminológico de las Ciencias Farmacéuticas*. En el caso de las erratas más fáciles de detectar manualmente (v. gr. erratas independientes del contexto) se ha conseguido localizar y registrar una gran parte de ellas. Ahora bien, en un futuro se precisaría, por ejemplo, una revisión específica de todas y cada una de las denominaciones de campos semánticos de las entradas del diccionario, ya que:

1. Tanto en el trabajo de revisión global como en el de revisión del DTCF se han descrito ejemplos de erratas en denominaciones de campos semánticos (véanse tablas 5 y 6 del presente artículo).
2. Ni en el trabajo de revisión global ni en el del DTCF se han podido revisar manualmente todas y cada una de esas denominaciones, por ser una tarea ímproba que requeriría más bien de una revisión automatizada (en

la que se podrían realizar por ejemplo «búsquedas negativas» mediante términos erróneos).

Los resultados del análisis del *Diccionario Terminológico de las Ciencias Farmacéuticas, Inglés-Español/Spanish-English* (Ariel, 2007) no pueden contrastarse con datos previos, ya que no existen estudios que hayan analizado de manera exhaustiva erratas en diccionarios médicos españoles. Pero ello no merma la validez de los resultados aquí expuestos. A falta de estándar, estos resultados pueden servir como referente, dado que el trabajo global de revisión de veintidós diccionarios especializados bilingües españoles en papel que se ha realizado es un estudio relevante, novedoso y exhaustivo en el campo de la corrección formal de los diccionarios especializados bilingües españoles en papel, fundamentalmente en la combinación de lenguas inglés-español.

Cabe reiterar aquí que la frecuencia de erratas del DTCF es claramente superior (aproximadamente el doble) que la frecuencia global del corpus de veintidós diccionarios, siendo la frecuencia global del corpus significativa de por sí.

## 6. Conclusiones

Los textos del ámbito biosanitario deben estar presididos por la corrección formal, ya que los defectos formales pueden

provocar un menoscabo en el plano del contenido. Ello cobra especial relevancia en obras de referencia de primer nivel, como el diccionario analizado.

Los resultados arriba expuestos muestran, de manera cuantitativa, la presencia de un claro paradigma de erratas en el *Diccionario Terminológico de las Ciencias Farmacéuticas, Inglés-Español/Spanish-English* (Ariel, 2007). El trabajo de análisis del diccionario demuestra que es necesario corregir la obra.

Este trabajo puede ser una fuente valiosa de información en el terreno de la corrección de erratas, así como en parcelas de la lingüística computacional, como puede ser el aprendizaje automático (*machine learning*) de correctores.

Los resultados aquí expuestos demuestran la conveniencia de modernizar el diccionario en cuestión, e ilustran la problemática del reto de adaptación a la era digital de la lexicografía especializada bilingüe española en papel. Este reto se ha de afrontar si se quiere estar en condiciones de ofrecer recursos lexicográficos que respondan a las necesidades del usuario del siglo XXI.

### Agradecimientos

El autor desea agradecer a la Dr.<sup>a</sup> del Dpto. de Filología y Traducción de la Universidad Pablo de Olavide (Sevilla) Nuria Fernández Quesada sus aportaciones, que han servido de orientación en la realización de este artículo.

### Notas

1. «The *Alicante Dictionaries*» (Mateo): <<https://goo.gl/ZGm6ry>> [consulta: 14.VI.2018].
2. El recurso MedDRA es propiedad de *International Federation of Pharmaceutical Manufacturers and Associations (IFPMA)*: <<https://goo.gl/pJ4qkY>> [consulta: 14.VI.2018].
3. Scholarly Publishers Indicators In Humanities and Social Sciences: <<https://goo.gl/pPksCj>> [consulta: 14.VI.2018].
4. Véase ejemplo de «Fe de erratas» convencional, en concreto la de la segunda edición (2005) del *Diccionario crítico de dudas inglés-español de medicina* («Libro rojo»), en la página web de TREMÉDICA («Asociación Internacional de Traductores y Redactores de Medicina y Ciencias Afines»): <<https://goo.gl/bKyUAE>> [consulta: 14.VI.2018].
5. En los casos en que la errata se encuentra en una subentrada del diccionario, se indica primero la entrada principal, seguida de la subentrada que contiene la errata, para facilitar su localización.
6. *Non-word errors* y *real-word errors* son expresiones genéricas que pueden referirse tanto a erratas (o equivocaciones materiales) como a errores genuinos (por desconocimiento de las reglas ortográficas o gramaticales de una lengua). Por ejemplo, la sustracción de una letra en «commision» se ha considerado una errata cuando ha aparecido en el DTCF (porque se presupone que los autores del diccionario conocen la grafía correcta), mientras que en un contexto de aprendizaje de lenguas (v. gr. *TEFL*) esa misma forma errónea sería un error genuino si el que incurriera en él fuera un alumno de lengua extranjera que desconociera la grafía correcta.

### Referencias bibliográficas

Ahmed, Farag; Ernesto William De Luca y Andreas Nürnberger (2009): «Revised N-Gram based Automatic Spelling Correction Tool to

Improve Retrieval Effectiveness», *Polibits*, 40 (6): 39-48. <<https://goo.gl/7qHgp8>> [consulta: 14.VI.2018].

Aleixandre, Rafael y Amador Iscla (1997): «Vicios del lenguaje y defectos del estilo científico en las comunicaciones del *IV Congreso Nacional de Documentación Médica*», *Papeles Médicos*, 6 (3): 5-13. <<https://goo.gl/wA8ETa>> [consulta: 14.VI.2018].

Aleixandre, Rafael; J. C. Valderrama y F. J. Bueno-Cañigral (2015): «Utilización adecuada del lenguaje médico: principales problemas y soluciones», *Revista Clínica Española*, 215 (7): 396-400.

Damerau, F. J. (1964): «A Technique for Computer Detection and Correction of Spelling Errors», *Communications of the ACM*, 7 (3): 171-176.

Ferner, Robin E. y Jeffrey K. Aronson (2016): «Nominal ISOMERS (Incorrect Spellings Of Medicines Eluding Researchers)—variants in the spellings of drug names in PubMed: a database review», *BMJ*, 355:i4854. <<https://goo.gl/z9wTqg>> [consulta: 14.VI.2018].

Hooper, Robert (1839): *Lexicon Medicum; or Medical Dictionary*. London: Longman, p. vii. <<https://goo.gl/4kdrMB>> [consulta: 14.VI.2018].

Kukich, Karen (1992): «Techniques for Automatically Correcting Words in Text», *ACM Computing Surveys*, 24 (4): 377-439. <<https://goo.gl/GhMuy5>> [consulta: 14.VI.2018].

Leroyer, Patrick (2013): «Change of Paradigm: From Linguistics to Information Science and from Dictionaries to Lexicographic Information Tools», en Pedro A. Fuertes-Olivera y Henning Bergenholtz: *e-Lexicography*. London/New York: Bloomsbury, pp. 121-140.

Madurga, M.; E. Lázaro, R. Granados y G. Martín-Serrano (2010): «FEDRA. La base de datos del Sistema Español de Farmacovigilancia de Medicamentos de Uso Humano», en Gloria Manso, Agustín Hidalgo, Alfonso Carvajal y Francisco J. de Abajo: *Los primeros 25 años del Sistema Español de Farmacovigilancia de Medicamentos de Uso Humano*. Oviedo: Universidad de Oviedo, pp. 131-144. <<https://goo.gl/qY2jss>> [consulta: 14.VI.2018].

Mateo, José (2014): «Lexicographical and Translation Issues in the Inclusion of English Financial Neonyms in Spanish Bilingual Dictionaries of Economics on Paper», *Hermes – Journal of Language and Communication in Business*, 52: 41-58. <<https://goo.gl/628MU6>> [consulta: 14.VI.2018].

Mateo, José (2018): «The *Alicante Dictionaries*», en Pedro A. Fuertes-Olivera (ed.): *The Routledge Handbook of Lexicography*. Abingdon/New York: Routledge, pp. 421-437.

Ng, HweeTou; Siew Mei Wu, Ted Briscoe, Christian Hadiwinoto, Raymond Hendy Susanto y Christopher Bryant (2014): «The CoNLL-2014 Shared Task on Grammatical Error Correction», *Proceedings of the Eighteenth Conference on Computational Natural Language Learning: Shared Task*: 1-14. <<https://goo.gl/yWgjNX>> [consulta: 14.VI.2018].

Samanta, Pratip y Bidyut B. Chaudhuri (2013): «A simple real-word error detection and correction using local word bigram and trigram», *Proceedings of the Twenty-Fifth Conference on Computational Linguistics and Speech Processing (ROCLING 2013)*: 211-220. <<https://goo.gl/CDHg3n>> [consulta: 14.VI.2018].

San Andrés, Isabel (2009): «Fuentes de información para el redactor de textos médicos», en Fernando Rico-Villademoros y Vicente Alfaro: «La redacción médica como profesión», *Cuadernos de la Fundación Dr. Antonio Esteve*, 17: 53-75.

Anexo. Relación de veintidós diccionarios del corpus		
Título (Fecha edición/reedición)	Editorial	Combinación lenguas
1. <b>Diccionario de Términos de la Banca</b> , 2009	Ariel	ING-ESP-ING
2. <b>Diccionario de Términos de la Bolsa</b> , 2003	Ariel	ING-ESP-ING
3. <b>Diccionario de Términos del Calzado e Industrias Afines</b> , 2006	Ariel	ING-ESP-ING
4. <b>Diccionario Terminológico de las Ciencias Farmacéuticas</b> , 2007 (impresión de 2011)	Ariel	ING-ESP-ING
5. <b>Diccionario de comercio internacional</b> , 2007	Ariel	ING-ESP-ING
6. <b>Diccionario de Términos de (los)* Derechos Humanos</b> , 2008 * Discrepancia entre el título indicado en la Introducción de la obra, en la portada interior de la obra y en el lomo encuadernado de la obra (donde no se indica «los»), y en la portada de la sobrecubierta de la obra y en el lomo de la sobrecubierta (donde sí se indica «los»).	Ariel	ING-ESP-ING
7. <b>Diccionario de términos económicos, financieros y comerciales</b> , 2012 (6.ª ed., impresión de 2014)	Ariel	ING-ESP-ING
8. <b>Diccionario de Fiscalidad Internacional y Aduanas</b> , 2009	Ariel	ING-ESP-ING
9. <b>Diccionario de Informática y Telecomunicaciones</b> , 2001	Ariel	ING-ESP
10. <b>Vocabulario de Informática y Telecomunicaciones</b> , 2002	Ariel	ING-ESP
11. <b>Diccionario de términos jurídicos</b> , 2012 (11.ª ed., impresión de 2014)	Ariel	ING-ESP-ING
12. <b>Diccionario de términos de marketing, publicidad y medios de comunicación</b> , 2005 (2.ª ed.)	Ariel	ING-ESP-ING
13. <b>Diccionario de Términos de la Piedra Natural e Industrias Afines</b> , 2005	Ariel	ING-ESP-ING
14. <b>Diccionario de Términos de la Propiedad Inmobiliaria</b> , 2003	Ariel	ING-ESP-ING
15. <b>Diccionario de Términos de Seguros</b> , 2003	Ariel	ING-ESP-ING
16. <b>Diccionario de términos de turismo y de ocio</b> , 2006 (2.ª ed.)	Ariel	ING-ESP-ING
17. <b>Diccionario de términos jurídicos</b> , 2006 (presentación de 2013) N. B. De este diccionario solo se ha analizado el texto en español, por desconocer el autor de este artículo la lengua árabe.	Ariel	ARAB- ESP
18. <b>Diccionario de términos del turismo</b> , 2009	Ariel	FRA-ESP-FRA
19. <b>Diccionario económico, contable, comercial y financiero*</b> , 2002 * Discrepancia entre el título indicado en la portada interior de la obra (donde se indican las letras iniciales en minúscula), y en la portada exterior de la obra y en el lomo de la obra (donde se indican las letras iniciales en mayúscula).	Gestión 2000	ING-ESP
20. <b>Diccionario de Economía y Empresa</b> , 2002	Gestión 2000	ESP-ING-ESP
21. <b>Diccionario Jurídico</b> , 2003	Gestión 2000	ESP-ING-ESP
22. <b>Diccionario Inglés de Publicidad y Marketing</b> , 2000	Gestión 2000	ING-ESP