

Cardiocor, un corpus para uso de los traductores de la medicina.¹ Cardiocor, a corpus for use by medical translators

Alain Escarrá Jiménez*

Universidad Central “Marta Abreu” de las Villas, Cuba.

Ismarys Díaz Rivas**

Universidad de Ciencias Médicas “Dr. Serafín Ruiz de Zárate Ruiz”, Cuba.

Las nuevas tecnologías, lejos de ser una amenaza para el traductor, deben ser abordadas como un reto. Sólo representan una amenaza para aquel que vive a espaldas de ellas, sin considerar cómo el mundo de la traducción evoluciona. Para el traductor especializado, la elaboración de sus propios corpus lingüísticos representa una ayuda significativa. A menudo se pierde mucho tiempo en la búsqueda y verificación de posibles traducciones de términos, algo que podría evitarse si se hace una cuidadosa compilación de textos con este fin, lo cual garantiza también la fiabilidad de los resultados. Una vez construido el corpus, el traductor debe utilizar una herramienta informática para trabajar con este. En el presente trabajo se presentará la creación de un corpus comparable, sincrónico y especializado sobre aspectos relacionados con la salud y la enfermedad cardiovascular, que viene a suplir un vacío en esta área, y que se realiza siguiendo la metodología para la creación de corpus lingüísticos especializados. CARDIOPCOR (CARDIO que significa corazón y COR de corpus), está dividido en dos subcorpus o categorías que representan dos tipos de textos diferentes. El subcorpus 1 consiste en textos dirigidos al público no especializado y el subcorpus 2 en textos dirigidos a especialistas. La herramienta informática utilizada es el programa *dtSearch 7.60*, que permite la indexación de documentos y la rápida localización de gran cantidad de textos y palabras. Este corpus representa una ayuda significativa para todos aquellos que traducen textos relacionados con estos temas, en cuanto a la búsqueda de elementos lexicales y terminológicos, así como para el análisis de elementos contextuales y estructurales de textos.

Palabras clave: corpus lingüísticos, traducción médica, tecnologías de la información y la comunicación, subcompetencia instrumental.

New technologies, far from being a threat to translators, should be faced as a challenge. They are a threat only for those who live with their backs turned to them without considering how the world of translation evolves. For the experts in translation, creating their own linguistic corpora represents a significant support. A lot of time is often spent in searching and verifying possible translations for terms, something that could be avoided if a thorough text compilation is made with this purpose, and what would ensure reliability of results. Once the corpus is created, the translator has to use a computer tool to work on it. This study presents the generation of a comparable, synchronic, and specialized corpus on aspects related to health and cardiovascular disease which intends to fill a vacuum in this area and which is created following the methodology to develop specialized linguistic corpora. CORDIOPCOR (CARDIO which means heart and COR taken from corpus) is divided into two sub-corpus or categories that represent two different types of text. Sub-corpus 1 consists of texts intended for a non-specialized audience and sub-corpus 2 consists of texts intended for experts. The computer tool used is called *dtSearch 7.60* and it allows indexing documents and fast localization of an important amount of texts and words. This corpus represents a significant support for all those who translate text on these topics in terms of search of lexical and terminological elements as well as the analysis of contextual and structural elements of texts.

Key words: Linguistic Corpora, Medical translation, Information and communications technology, instrumental sub-competence.

¹ Agradecemos sinceramente la colaboración brindada por el MSc. Dr. Alberto Morales Salinas, especialista de I Grado en Cardiología, por los criterios aportados para la realización de este trabajo.

* Correo electrónico: alain@lpn.co.cu

Licenciado en Lengua Inglesa (1997). Máster en Lengua Inglesa para la Comunicación Intercultural (2012). Traductor especializado en medicina. Actualmente labora en la revista *CorSalud*.

** Correo electrónico: idalmis@lpn.co.cu

Doctora en medicina (1986). Especialista de Primer Grado en Medicina General Integral. Profesora instructora.

Introducción

El conocimiento y dominio de las tecnologías de la información y la comunicación (TIC) que puedan aplicarse a la actividad traductora es parte de las competencias que los traductores de hoy deben poseer. Dentro del modelo de la competencia traductora desarrollado por el grupo PACTE (2011), se encuentra la subcompetencia instrumental que es específicamente la relacionada con el uso de la documentación y las TIC aplicadas a la traducción.

Las TIC, según Diéguez y Lazo (2004) se dividen en generales y específicas. Entre las TIC generales se encuentran el uso de internet y su amplia gama de posibilidades, los corpus lingüísticos y los programas generadores de concordancia. Entre las TIC específicas para la traducción están los traductores automáticos y los programas de traducción asistida por computadoras. En el presente trabajo se tratará una de las TIC generales: la creación y utilización de un corpus lingüístico.

¿Qué es un corpus?

Un corpus lingüístico es un conjunto, normalmente muy amplio, de ejemplos reales de uso de una lengua. Estos ejemplos pueden ser textos (típicamente), o muestras orales (normalmente transcritas) (Wikipedia, 2012). Otra definición más explícita es la dada por Sinclair (1996, citado en Gallego Hernández 2010), “una colección de apartes de una lengua que se seleccionan y clasifican de acuerdo a criterios lingüísticos explícitos, con el fin de ser utilizados como una muestra de esta”

Tipos de corpus

Hay varias clasificaciones de corpus pero nosotros adoptamos la de Diéguez y Lazo (2004), quienes expresan que los corpus lingüísticos informatizados, pueden ser de dos tipos. Por una parte, están los *corpus monolingües*, como por ejemplo el CREA (Corpus de Referencia del Español Actual) de la Real Academia Española, el COCA (the Corpus of Contemporary American English), y el BNC (British National Corpus).

Otro tipo de corpus son los bilingües o de traducción, que se dividen en corpus paralelos y corpus comparables.

Corpus paralelos

Un corpus paralelo lo conforman textos originales en una lengua con las traducciones de los mismos hacia una o varias lenguas (Laviosa, 2002, p. 37, citado en Jiménez-Crespo, 2009b, p. 3). Este tipo de corpus también se ha denominado “*corpus de traducción*” (Johansson, 1998) o “*bitexts*” (Bowker, 2002, citado *ibíd.*). El mejor ejemplo de este tipo de corpus son los *Canadian Hansards*, los cuales almacenan las actas del diario de sesiones del Parlamento Federal de Canadá. Los corpus paralelos son de gran utilidad para comprobar cómo se ha traducido previamente un término o unidad fraseológica, una funcionalidad que comparte con las memorias de traducción. A pesar de su amplio uso en la comunidad traductora, de entre las posibles desventajas de este tipo de corpus se encuentra que no presenta una variedad natural de la lengua (McEnery y Xiao, 2007), y que puede representar la idiosincrasia de un sólo traductor (Teubert, 1996, p. 247, citado en Jiménez-Crespo, 2009b, p. 3)

Corpus comparables

Finalmente, los corpus comparables están formados por textos redactados originalmente en dos o más lenguas, los cuales presentan una serie de semejanzas entre sí. Al respecto, Fernández y Rabadán (2002, p.53, cit. en Diéguez y Lazo 2004, p. 57) señalan que este tipo de corpus se podría considerar la “documentación paralela” del traductor. Según Corpas Pastor (2004, p. 137) los corpus comparables son documentos escritos originalmente en la lengua meta (LM), similares al texto de origen (TO) en cuanto a la temática, la función comunicativa, la tipología textual y el nivel de especialización. Si en un principio tales colecciones contenían documentos impresos fundamentalmente, en la actualidad se tiende a manejar documentos electrónicos, generalmente compilados ad hoc a partir de CD-ROM y otros recursos en red. El paso del corpus escrito al corpus electrónico constituye, pues, un estadio más de la evolución de la traducción como actividad profesional, dentro de las profundas transformaciones que esta ha venido experimentando en los últimos años.

Corpus ad hoc

Como expresa Gallego Hernández (2010b, p. 3, trad. a.) la elaboración de corpus ad hoc (también conocidos como corpus virtuales, DIY corpus, etc.) y su explotación como recurso para

la traducción especializada ha sido discutido por varios autores (Aston, 1997; Maia, 1997; Varantola, 1997; Zanettin, 2002; Corpas Pastor, 2001, 2004a, 2004b citado Ibíd.; Sánchez Gijón, 2004). Hablando a grandes rasgos, el diseño y compilación de estos corpus comprende la recopilación en un período de tiempo no muy largo de una serie de textos paralelos, cuya información conceptual, fraseológica y terminológica pueda ayudar al traductor a resolver problemas y dificultades específicas de la traducción.

Esto significa encontrar los textos paralelos, almacenarlos en el disco duro y convertirlos a formatos descifrables para el programa informático específico que extraiga del corpus la información que se necesita para resolver cada problema o dificultad específica.

En términos generales, las fuentes básicas a partir de las cuales los traductores pueden encontrar textos paralelos son CD-ROM, textos escaneados y por supuesto, internet. Según explican estos autores, las búsquedas en internet para elaborar corpus ad hoc se relacionan de manera general a tres tipos de búsquedas presentadas por Austermühl (2001, citado en Gallego Hernández 2010b, p. 4, trad. a.): búsquedas institucionales, búsquedas temáticas y búsquedas por palabras clave. Con respecto a las búsquedas institucionales y temáticas, Corpas Pastor (2001; 2004; 2004b, citado Ibíd.) ha indicado, en el contexto de sus cursos de traducción médica, diferentes tipos de recursos a partir de los cuales ella elabora sus corpus ad hoc: portales y directorios, organizaciones médicas, revistas especializadas, bibliotecas virtuales, bases de datos bibliográficas, artículos escaneados y enciclopedias.

Varios autores proponen programas específicos para bajar los archivos de internet, para su cambio de formato como para su explotación una vez formado el corpus.

Gallego Hernández (2010b, p. 5, trad. a.) continúa explicando que es verdad que las etapas básicas descritas por estos autores incluyen la utilización de diferentes recursos o herramientas para elaborar un corpus aun cuando este software no haya sido desarrollado específicamente para este propósito.

Apoyados en la idea de este autor queremos enfatizar que los traductores no nos tenemos que frustrar si no tenemos los programas propuestos o generalmente utilizados para bajar, convertir o utilizar el corpus. Podemos buscar variantes que sean más sencillas sin comprometer la confiabilidad y utilidad del corpus compilado, que es lo más importante.

Nuestra propuesta para CARDIOCOR. Metodología y Descripción del Corpus

Corpus comparable, sincrónico y especializado, dividido en dos subcorpus o categorías.

Subcorpus 1: Un corpus comparable y especializado con textos sobre la salud y la enfermedad cardiovascular escritos por especialistas dirigido al público no especializado (comunicación semi-especializada). Este a su vez está subdividido en subcorpus 1. Ing y subcorpus 1. Esp. En el primero se encuentran los textos en idioma inglés y en el segundo en idioma español.

Subcorpus 2: Un corpus comparable y especializado con textos sobre la salud y enfermedad cardiovascular escritos por especialistas dirigido a otros especialistas (comunicación especializada). Este a su vez está subdividido en subcorpus 2. Ing y subcorpus 2. Esp. En el primero se encuentran los textos en idioma inglés y en el segundo en idioma español.

Subcorpus 1, dirigido al público no especializado

Está compuesto por 432 textos escritos originalmente en inglés y 432 textos escritos originalmente en español. Los textos tratan sobre la anatomía, la fisiología, la salud y las enfermedades cardiovasculares, así como tratamientos medicamentosos, procedimientos cardiacos, etc.

Criterio de inclusión

1. Textos escritos originalmente en inglés y en español sobre temas afines.
2. Recientes y sincrónicos.
3. Textos en formato electrónico tomados de enciclopedias médicas digitales.

Identificación de los textos en el subcorpus 1

Este subcorpus posee artículos identificados mediante:

1. Un número. Por ejemplo, 1. La numeración es consecutiva en ambos idiomas y va desde el 1 al 432.
2. El idioma. Por ejemplo, 1. Ing. ó 1. Esp.

3. Una clasificación por área de estudio. Estas se dividen en dos categorías:
 - a. Morfo-fisiología (mf.): que incluye anatomía, histología, embriología y fisiología del sistema cardiovascular.
 - b. Enfermedades y tratamientos (et.): es una categoría más amplia e incluye las diferentes enfermedades, su diagnóstico, tratamientos medicamentosos, y procedimientos cardiacos.

4. Una clasificación por palabra clave. Por ejemplo,

296. Ing. mf. atrial fibrillation.

Este texto sería identificado en su totalidad como:

296. Ing. mf. atrial fibrillation.

Su texto comparable en español sería recuperado como:

296. Esp. mf. fibrilación auricular

Subcorpus 2 dirigido a especialistas

Está compuesto por 150 artículos científicos escritos originalmente en inglés y 150 artículos escritos originalmente en español.

Criterio de inclusión

1. Artículos de revistas escritos originalmente en inglés y en español sobre temas afines.
2. Recientes y sincrónicos (2000-2012)
3. Revistas en formato electrónico.
4. Artículos tomados de revistas de alto factor de impacto sobre temas cardiovasculares.
5. Los artículos comprenden diferentes géneros médicos tales como reporte de casos, artículos originales, revisiones bibliográficas, artículos especiales, imágenes en cardiología, editoriales y cartas al editor.

Herramienta informática utilizada para trabajar con el corpus

El programa utilizado es dtSearch 7.60 que permite la búsqueda y localización de Terabytes de textos en un segundo. Esto lo logra realizando una indexación que guarda la localización de cada palabra de sus archivos.

Ejemplos de uso

Supongamos que estamos haciendo la traducción inversa (del español al inglés) de un artículo cuyo tema principal es la fibrilación auricular. Pronto nos damos cuenta que tenemos problemas con la comprensión desde el punto de vista temático o podemos encontrar varios problemas de traducción. Para esto abrimos el programa dtSearch y escribimos esta palabra clave (fibrilación auricular) en el buscador. Instantáneamente obtendremos varios textos donde aparece esta combinación de palabras. Sin embargo debo buscar por el nombre del texto, el artículo principal, que sería por ej. “296. *Esp. mf. fibrilación auricular*”. Una vez localizado en español, entonces procederíamos a buscar su texto paralelo en inglés y para facilitar esta tarea, el texto en español tiene la palabra clave traducida al inglés. Ahora procedemos a realizar una nueva búsqueda, escribimos en el buscador “296. *Ing. mf. atrial fibrillation*” y enseguida tendremos los dos artículos en forma paralela y con ellos podríamos ganar en comprensión del tema a traducir o solucionar los problemas de traducción.

Conclusiones

El presente trabajo pretende confirmar la importancia del conocimiento y uso de las nuevas tecnologías de la información y la comunicación aplicadas a la traducción, específicamente mediante la confección de corpus lingüísticos ad hoc. El corpus propuesto, CARDIOCR, puede ser de ayuda a traductores del área de la medicina y a estudiantes de traducción. El subcorpus 1, con textos destinados al público no especializado puede utilizarse, entre otras cosas, para la solución de problemas lexicales, temáticos y terminológicos. El subcorpus 2, con textos dirigidos a especialistas, puede ser utilizado para el análisis de elementos contextuales (pragmáticos, comunicativos y semióticos) así como de elementos estructurales e intratextuales. En el momento de esta ponencia se ha concluido la primera parte del corpus, es decir, la recopilación de los textos para el subcorpus 1. Una vez terminado el corpus completo podría ponerse a disposición de todos los traductores del sistema de salud cubano.

Recomendaciones

Considerar la creación de corpus lingüísticos ad hoc como una herramienta valiosa para el trabajo del traductor especializado. La inversión de tiempo y esfuerzo en la creación de este tipo de corpus se ve ampliamente compensada por su posterior utilidad y beneficios.

Referencias bibliográficas

- Aston, G., 1997. "Small and large corpora in language learning", in Lewandowska- Tomaszczyk and Melia, pp. 51-62.
- Beeby, A., et. al. 2011 PACTE (2011). "Results of the Validation of the PACTE Translation Competence Model: Translation Project and Dynamic Translation Index", in: O'Brien, Sharon (ed.) IATIS Yearbook 2010, Londres: Continuum.
- Corpas Pastor, G. (2001). "Compilación de un corpus ad hoc para la enseñanza de la traducción inversa especializada", *Trans* 5, pp. 155-184. En línea <
http://www.trans.uma.es/Trans_5/t5_155-184_GCorpas.pdf >
- Corpas Pastor, G. (2004a). *La traducción de textos médicos especializados a través de recursos electrónicos y corpus virtuales*. En L. González y P. Hernández. (eds.). 2004. *Las palabras del traductor. Actas del II Congreso Internacional "El español, lengua de traducción", 20 y 21 de mayo, Toledo 2004*. Bruselas: Comisión Europea/ESLETRA. 137 164.
- Diéguez, M.I y Lazo, R.M. (2004). *Tecnologías de la Información y de la Comunicación (TIC) al Servicio del Traductor Profesional*. Onomázein: Revista de Lingüística, Filología y Traducción, 9 (1), 54-74.
- Fundación Wikipedia. (2012). Corpus lingüístico. Consultado el 23 de julio de 2012.
- Gallego Hernández, D. (2010b). *Acquiring instrumental sub-competence by building do-it-yourself corpora for business translation*. Universidad de Alicante, España. Disponible en:
<http://www.lancs.ac.uk/fass/projects/corpus/UCCTS2010Proceedings/papers/Hernandez.pdf>

- Gallego Hernández, D. (2010c). *COMENEGO (Corpus Multilingüe de Economía y Negocios): corpus estable vs. metodologías ad hoc (web as/for corpus) aplicadas a la práctica de la traducción económica, comercial y financiera*. Universidad de Alicante, España.
Disponible en: http://acorn.aston.ac.uk/RK-publications/COMENEGO_Corpus_Multilingue_060511.pdf
- Jiménez-Crespo, M, A. (2009b). *El uso de corpus textuales en la localización*. Tradumática. Traducció i Tecnologies de la Informació i la Comunicació Número 07. Diciembre.
Disponible en: <http://www.fti.uab.cat/tradumatica/revista>.
- Maia, B., 1997, "Do-it-yourself corpora... with a little bit of help from your friends!", in Lewandowska-Tomaszczyk and Melia, pp. 403-410.
- McEnery, A. y R. Z. Xiao. (2007). "Parallel and comparable corpora: what are they up to?", en *Incorporating Corpora: Translation and the Linguist*. Translating Europe. Clevedon: Multilingual Matters.
- PACTE (2011). "Results of the Validation of the PACTE Translation Competence Model: Translation Project and Dynamic Translation Index", in: O'Brien, Sharon (ed.) *IATIS Yearbook 2010*, Londres: Continuum (aceptada y en prensa).
- Sánchez-Gijón, P. (2004). «La extracción de conocimiento y terminología a partir de corpus ad hoc: el uso de documentos digitales de la web pública», *Linguistica Antverpiensa* 3, p. 179-202.
- Varantola, K., 1997. "Translators, dictionaries and text corpora", available online, <http://www.sslmit.unibo.it/cultpaps/varanto.htm>
- Zanettin, F. (2002). "DIY Corpora: The WWW and the Translator", en *Training the Language Service Provider for the New Millennium*, B. Maia, J. Haller y M. Urlrych, (eds). Porto: Faculdade de Letras, Universidade do Porto. En línea <http://www.dedericozanettin.net/DIYcorpora.html>.