

Clara Inés López-Rodríguez & Miriam Buendía-Castro

En busca de corpus *online* a la carta en el aula de traducción científica y técnica

Corpora Online à la carte in a Scientific and Technical Translation Course – Abstract

Corpora are rich information sources that can provide the translator with both linguistic and conceptual knowledge that is not found in dictionaries. The question that arises within this context is whether the web can be considered as a corpus. In this paper, we compare and evaluate the two main corpus approaches to the web in the context of a scientific and technical translation course at university level: (i) Web for Corpus (WfC), in which the web is used as a source of texts in digital format for the subsequent implementation of an offline corpus; and (ii) Web as Corpus (WaC), which uses the web directly as a corpus. To that end, we asked two groups of students (control and experimental) in the Translation and Interpreting Degree Program at the University of Granada to carry out a technical translation assignment. Our objective was to find out whether the Web as Corpus approach was able to compensate for the lack of subject field knowledge of the experimental group.

1 Introducción

Internet ha traído consigo una nueva forma de organizar y obtener la información.¹ Dada la gran cantidad de información que ofrece, constituye “a fabulous linguists’ playground” (Kilgarriff/Grefenstette 2003: 333). La cantidad de datos que circulan en Internet en un día cualquiera es mucho mayor que toda la información disponible en el siglo XIX (Austermühl 2001: 7). El inglés continúa monopolizando la red al representar el 45% del número total de páginas web. Otras lenguas europeas con un porcentaje destacado de páginas web son el alemán (5,90%), el francés (4,41%), el español (3,80%), el italiano (2,66%) y el portugués (1,39%), según la Unión Latina (2007). Sin embargo, el incremento del número de usuarios de Internet que hablan otras lenguas está motivando un cambio en la situación (Fletcher en prensa; *Internet World Stats* 2009).

¹ Este trabajo se ha elaborado dentro del marco del Proyecto I+D ECOSISTEMA: Espacio Único de Sistemas de Información Ontológica y Tesauro sobre el Medio Ambiente (FFI2008-06080-C03-01/FILO), financiado por el Ministerio de Ciencia e Innovación de España. Este artículo es la versión ampliada en lengua española de un trabajo presentado en el Congreso Using Corpora in Contrastive and Translation Studies (UCCTS), celebrado en Edge Hill University (Liverpool, Reino Unido) en julio de 2010.

Internet ha hecho posible que se recopilen corpus de millones de palabras en poco tiempo. Hace algunas décadas, compilar un corpus constituía una labor tediosa. Los lingüistas se pasaban horas y horas en las bibliotecas intentando buscar la información que necesitaban. Hoy día, muchos estudiantes y profesionales de la Traducción siguen creyendo que compilar un corpus requiere mucho tiempo a corto plazo. No obstante, con el advenimiento de Internet, cientos de textos pueden compilarse en unos minutos, lo que era impensable hace unos años.

La Traducción no ha permanecido ajena a este desarrollo, de ahí que la metodología utilizada en las fases de documentación y vaciado terminológico ha evolucionado sustancialmente. Si en el pasado los traductores se apoyaban casi exclusivamente en recursos lexicográficos y terminográficos —como los diccionarios generales y especializados— para desempeñar su trabajo, ahora los corpus, cuando están correctamente diseñados, proporcionan al traductor información lingüística y conceptual que no siempre es posible encontrar en los diccionarios. De hecho, los corpus especializados resultan de gran utilidad en la traducción puesto que proporcionan terminología especializada y actualizada, a la vez que aportan información sobre el uso combinatorio de los términos y su frecuencia.

Aunque la Lingüística de corpus se remonta a la década de 1970, su aplicación a la enseñanza de la traducción ha sido posterior. Desde 1997, *Corpus Use and Learning to Translate (CULT)* ha sido un área de investigación muy fructífera (Beeby/Rodríguez/Sánchez 2009) que defiende que tanto los corpus como Internet constituyen herramientas esenciales para los traductores, junto con recursos lexicográficos y terminográficos, memorias de traducción, etc. Sin embargo, la utilización de corpus para la Traducción va más allá de compilar un corpus de grandes dimensiones:

The difficulty of using corpora is that they rarely provide immediate answers to a translator's problems. Unlike translation memory or machine translation systems, they do not instantly present a preferred candidate for the user to accept, modify or reject. Corpus data has to be interpreted and evaluated comparatively to reach conclusions, and this requires not only technical skill [...] but above all critical thought. (Aston 2009)

En cualquier caso, en el contexto de los estudios de Traducción, son muchos los investigadores que han puesto de manifiesto las ventajas pedagógicas de usar, en el aprendizaje de lenguas extranjeras o en la traducción, un *do-it-yourself-corpus* (DIY corpus), es decir, una colección de documentos provenientes de Internet que se compilan *ad hoc* como respuesta a un encargo específico de traducción (Zanettin 2002: 242). Este tipo de corpus que se compila para un encargo específico de Traducción también ha recibido el nombre de *ad hoc, virtual corpus* (Ahmad et al. 1994) o *disposable corpus* (Varantola 2003), puesto que los textos se compilan para satisfacer necesidades transitorias, y no para formar parte de un corpus permanente.

La pregunta que cabe hacerse en este contexto es si la Red podría considerarse como un corpus propiamente dicho. Basándonos en la distribución hecha por De Schryver (2002), nos unimos a la idea de que existen dos enfoques para usar la Red en la investigación en Lingüística de corpus, y opinamos que ambos son aplicables al ámbito de la Traducción:

- la Web para recopilar un corpus (*Web for Corpus, WfC*), en la que la Red se utiliza como fuente de textos en formato electrónico para la posterior compilación de los corpus *off-line*.
- la Web como corpus (*Web as Corpus, WaC*), que utiliza la Red directamente como si fuera un corpus propiamente dicho (Kilgariff/Grefenstette 2003; Fletcher 2004, 2007; Baroni/Bernardini 2006).

En el presente estudio establecemos una comparación y una evaluación de estos dos enfoques en el contexto de un curso universitario de traducción científico-técnica. Pedimos a dos grupos de estudiantes de la Licenciatura en Traducción e Interpretación de la Universidad de Granada que llevaran a cabo un encargo de traducción científica hacia el español. A un grupo se le pidió que lo hiciera siguiendo una metodología WaC (grupo experimental) y al otro, usando métodos convencionales de la WfC (grupo de control).

2 La Red desde la perspectiva de lingüistas y traductores que buscan corpus

2.1 La Web para recopilar un corpus (WfC)

Tal y como hemos mencionado anteriormente, la Web para recopilar un corpus (*Web for Corpus, WfC*) es la técnica que ha venido usándose tradicionalmente para compilar textos en formato digital para la posterior implementación de los corpus *off-line*. A este respecto, en el campo de la Traducción, la noción de *DIY (do-it-yourself) corpus* (Zanettin 2002: 242) se ha utilizado para describir la colección de documentos provenientes de Internet compilados *ad hoc* como respuesta a un encargo específico de traducción.

Autores como Sinclair (2005) se manifiestan claramente a favor de este enfoque tradicional. No obstante, a pesar de reconocer la gran utilidad que supone Internet para cualquier lingüista, destaca que la Red no es realmente un corpus porque no se ha diseñado siguiendo criterios lingüísticos.

El enfoque de la WfC conlleva navegar por Internet de forma manual en búsqueda de información de calidad. El usuario puede empezar, bien introduciendo una serie de palabras clave en un motor de búsqueda, o introduciendo directamente una URL en el navegador, y a partir de ahí, visitará distintos sitios web con el objetivo de descargar los textos de su interés y poder procesarlos mediante un programa de análisis léxico como WordSmith Tools (Scott 1999):

El hecho de que, hoy por hoy, Internet sea la principal fuente de textos para la compilación de corpus, significa que el concepto de calidad está directamente relacionado con la calidad de los sitios web. Por este motivo, la selección de los textos resulta crucial para el desarrollo de un corpus representativo que será una fuente de datos fiables. Tal y como señala Austermühl:

Finding data on the World Wide Web is no problem at all. But finding reliable information is rather a difficult task. And finding the information you really need can be very time-consuming and often frustrating. (Austermühl 2001: 52)

Varantola incluso considera que conocer los criterios para compilar corpus *ad hoc* forma parte de la competencia traductora:

I would even go a step further and claim that the knowledge of how to compile and use corpora is an essential part of modern translational competence and should therefore be dealt in the training of prospective professional translators. (Varantola 2003: 56)

Se ha escrito mucho acerca de los parámetros que determinan la calidad de los recursos digitales. Buendía y Ureña (2009) hicieron una revisión de estos parámetros y establecieron un protocolo compuesto de tres parámetros: (a) *autoridad*, que evalúa principalmente el prestigio y experiencia de los autores; (b) *contenido*, que incluye la cobertura informativa, precisión, objetividad, actualización y audiencia; y (c) *diseño*, que abarca los subparámetros ayudas a la navegación, accesibilidad, y presentación y gestión de la información.

2.2 La Web como Corpus (WaC)

A día de hoy, no existe un consenso con respecto al significado de la expresión *Web as Corpus (WaC)*. Siguiendo la clasificación propuesta por Bernardini, Baroni y Evert (2006), consideramos tres maneras de acercarse a la Red como corpus desde una perspectiva lingüística: (a) la Web como sustituta del corpus (*Web as Corpus Surrogate*); (b) la Web como mega-corpus o mini Web; y (c) la Web como un *supermercado* de corpus (*Web as Corpus Supermarket*).

2.2.1 La Web como sustituta del corpus

Este primer enfoque considera la Web en sí misma como un gran corpus y se implementa con sistemas informáticos que ofrecen normalmente una interfaz de búsqueda —en la que se introducen los términos a buscar—, para posteriormente devolver los resultados a modo de concordancias, como si estuviésemos analizando un corpus descargado en nuestro ordenador mediante un programa de análisis léxico, pero con la diferencia de que el corpus está en línea.

Este tipo de sistemas se diferencia de buscadores convencionales como Google o Altavista, en que pre-procesan las preguntas antes de enviarlas a los buscadores, post-procesan los resultados y los ofrecen enfocados a facilitar los estudios de tipo lingüístico. Algunos de los sistemas más conocidos son WebCorp² (Kehoe/Renouf 2002), KWicFinder³ (Fletcher 2001), Linguist's Search Engine (Elkiss/Resnik 2004), WebCorp⁴ (Fletcher 2007) o Corpeus⁵ (Leturia et al. 2007).

No obstante, estos sistemas que pre-procesan y post-procesan los datos también presentan ciertas limitaciones, que coinciden en gran parte con las limitaciones de los motores de búsqueda convencionales.

En primer lugar la cantidad de cibertexto que buscan está limitado por restricciones de tiempo, por lo que la recuperación (*recall*) puede ser pobre. Estos

² <http://www.webcorp.org.uk/>

³ <http://www.kwicfinder.com/KWicFinder.html>

⁴ <http://webascorpus.org/searchwac.html>

⁵ <http://www.corpeus.org>

sistemas de pre- y post-procesamiento ofrecerán, en general, menos resultados que los buscadores porque tienen que filtrar aquellos que no se ajustan a la ecuación de búsqueda del usuario.

Una segunda limitación la constituye el hecho de que la proporción de textos web potencialmente relevantes esté limitada por los criterios de búsqueda de los motores de búsqueda. WebCorp o cualquier otra herramienta similar no tienen ningún control sobre el *ranking* de Google. Al realizar una búsqueda, lo ideal sería que se nos ofreciera una muestra al azar de páginas fiables que incluyeran los términos de búsqueda; sin embargo, el motor de búsqueda devolverá una lista de páginas en función de criterios como la popularidad o la proximidad geográfica, lo que carece de interés lingüístico. De esta forma, incluso si introducimos la misma ecuación de búsqueda en el mismo buscador, los resultados ofrecidos serán diferentes dependiendo, por ejemplo, de si lanzamos la búsqueda desde el Reino Unido o desde Estados Unidos (Hundt/Nesselhauf/Biever 2008: 2-3). Por otra parte, la popularidad de un término en Internet no coincide con la frecuencia de uso lingüístico. Y así, Fletcher señala que los resultados ofrecidos por los motores de búsqueda no se corresponden con la frecuencia de esas expresiones lingüísticas en corpus. En sus palabras: "most widespread does not necessarily mean 'preferred' in linguistic terms" (Fletcher en prensa).

En tercer lugar, los buscadores son frágiles por su propia naturaleza. La información se actualiza con tanta rapidez, que los resultados de las búsquedas nunca son replicables. Fletcher (2007: 37) habla de la volatilidad de la Red y afirma que los resultados que nos ofrece un buscador pueden variar, no sólo si se realiza la búsqueda en lugares o periodos de tiempo diferentes, sino incluso en la misma sesión de usuario. Ntoulas et al. (2004) llevaron a cabo una investigación en torno a la dinamicidad y volatilidad de la Web a través del estudio de 154 páginas web, de cuyo análisis se desprende que las nuevas páginas web aparecen en una proporción de 8% a la semana. No obstante, *nuevo* no significa *adicional* o *novedoso*; según el estudio, el número total de páginas y el tamaño de éstas permanece relativamente constante, ya que las páginas *viejas* desaparecen y el contenido nuevo, en realidad, sólo representó en torno al 5%.

Dadas estas limitaciones, se hace necesario desarrollar un buscador específico para lingüistas que no sea tan dependiente de buscadores convencionales (Lüdeling/Evert/Baroni 2007), que es lo que se pretende con el enfoque que siguen aquellos que conciben la Red como un mega corpus.

2.2.2 La red como mega-corpus o mini-Web

Esta interpretación de la expresión "la Web como corpus" se refiere al intento de crear un nuevo objeto, una especie de "mini-web" o "mega-corpus" adaptado a la investigación lingüística. Este nuevo objeto tendría tanto características derivadas de la Web, como propias de los corpus. Así pues, al igual que la Web, sería muy grande y estaría relativamente actualizada; contendría material textual derivado de la Red; y proporcionaría una interfaz basada en la Web que permitiera acceder a los datos de forma fácil y rápida. Asimismo, al igual que un corpus, este mega-corpus o mini-Web

estaría anotado (anotación morfosintáctica, de lemas, sintáctica y semántica), sería relativamente estable y permitiría búsquedas sofisticadas.

Se trata de un proyecto ambicioso y de un gran valor, ya que daría respuesta a la necesidad cada vez mayor por parte de la comunidad de lingüistas computacionales, profesores de lengua y traducción, y resto de profesionales del lenguaje, de recursos que combinen la fiabilidad y flexibilidad de los corpus y sus herramientas de análisis, con el tamaño, variedad y rapidez de la Web. Este nuevo *motor de búsqueda para lingüistas* (Volk 2002; Kilgariff 2003; Fletcher 2004) podría beneficiar tanto a aquellos que quisieran investigar aspectos de la lengua a través de la Red, como a los interesados en investigar aspectos de la Red a través de la lengua (Bernardini/Baroni/Evert 2006: 14).

Las posibilidades de que los motores de búsqueda convencionales se interesen por este proyecto son escasas, pero no se trata de una utopía. Prueba de ello es que varios grupos de investigación se encuentran ya trabajando en la implementación de estos sistemas, entre los que se destaca el proyecto WaCky (*Web as Corpus kool ynitivative*)⁶ de la Universidad de Bolonia-Forlì.

El principal objetivo de WaCky es compilar corpus de grandes dimensiones (aproximadamente dos mil millones de palabras) a partir de Internet, y ofrecer herramientas para procesar y explotar dichos corpus. A día de hoy, ya existen varios corpus WaCky disponibles: *deWaC* (para el alemán), *itWaC* (para el italiano), *ukWaC* (para la lengua inglesa), *frWaC* (para el francés) y actualmente trabajan en el corpus en español. Nuestros alumnos han hecho uso de estos corpus, que se encuentran integrados en Sketch Engine, tal y como se verá en adelante.

2.2.3 La Web como supermercado de corpus

Por último, la Red puede percibirse como si fuera un *supermercado* en el que se pueden adquirir diferentes corpus. Los que siguen este enfoque se dirigen a la red para buscar textos mediante un motor de búsqueda, gracias a las facilidades que éstos ofrecen, seleccionan los textos y los descargan creando un corpus en línea. Este enfoque es el que siguen especialmente los traductores, y tiene mucho en común con la metodología tradicional de compilación de corpus (*Web for Corpus*), con la diferencia de que en la Web como supermercado de corpus, las fases iniciales del proceso se llevan a cabo de forma semi-automática y el usuario tiene control sobre los textos que le interesan.

Hay herramientas útiles para la traducción que permiten compilar automáticamente corpus de la Red de una forma rápida. Para nuestros objetivos de investigación, las herramientas BootCat (Baroni/Bernardini 2004) se configuran como unas de las más relevantes. Las herramientas BootCat (acrónimo para *bootstrapping corpora and terms from the web*) nacieron con el propósito de ofrecer una respuesta rápida a los traductores que necesitaban encontrar la terminología propia de un dominio de especialidad. Esta información no suele aparecer en los diccionarios generales y, si aparece en diccionarios especializados, éstos son difíciles de localizar, mantienen

⁶ <http://wacky.sslmit.unibo.it>

precios elevados y, cuando salen al mercado, ya suelen estar *desfasados* (Baroni et al. 2006) por el tiempo que necesitan para producirse y publicarse, y por lo rápido que evoluciona el lenguaje especializado.

Hay un servicio web de las herramientas BootCat, que se denomina WebBootCat (Baroni et al. 2006), al que se accede a través de la herramienta de análisis de corpus SketchEngine⁷ (cf. Kilgarriff et al. 2004). En el presente estudio, nuestros estudiantes hicieron uso de Sketch Engine como herramienta de apoyo a la traducción.

Sketch Engine (también conocido como *SkE* o *Word Sketch Engine*) es un sistema de análisis de corpus que incorpora *word sketches*,⁸ relaciones gramaticales y un tesoro distribucional. Tal y como puede verse en la Fig. 1, una cuenta de Sketch Engine ofrece al usuario:

- Corpus ya compilados (de entre 60 millones y 2.000 millones de palabras) especialmente para el inglés, francés, alemán, japonés, ruso, italiano y español, aunque también para otras lenguas como el árabe, chino, neerlandés, croata, griego, hebreo, hindi, persa, polaco, portugués, rumano, serbio, esloveno, sueco y vietnamita.
- Acceso a WebBootCat, que permite compilar corpus de miles de palabras en unos minutos a través de los "seed terms" ('términos para sembrar') que el usuario introduce en la ecuación de búsqueda. Asimismo, ofrece la posibilidad de descargar el corpus a nuestro ordenador, añadir nuevos documentos a nuestro corpus desde la Web o desde nuestro disco duro, extraer las palabras clave de un dominio, ver los textos tanto en formato plano como en formato vertical (es decir, anotado morfológicamente y por lemas) o abrir el corpus a través del programa de análisis léxico propio de Sketch Engine para trabajar con el corpus desde el punto de vista lingüístico – generar concordancias, listas de frecuencias, colocaciones o word sketches (Figura 2) –.
- Un compilador de corpus que permite subir e instalar nuestro propio corpus desde nuestro disco duro y trabajar con él desde un punto de vista lingüístico tal y como se ha descrito más arriba.

⁷ <http://www.sketchengine.co.uk/>. Ahora SketchEngine se va a integrar en un nuevo sistema llamado new Corpus Architect corpus management system.

⁸ Las *word sketches* son el resumen automático de una página basado en el corpus del comportamiento gramatical y colocacional de una palabra (Kilgarriff et al. 2004).

The screenshot shows the Sketch Engine interface. On the left is a navigation menu with options like 'Create corpus', 'WebBootCaT', and 'Support'. The main area is titled 'Corpora' and contains a table of available corpora. Below this is a section for 'My corpora' which is currently empty.

Corpus name	Language	Size
Internet-ZH	Chinese, Simplified	277,931,664
British National Corpus	English	112,181,850
ukWaC v1.0 old	English	1,526,599,198
French web corpus	French	126,850,281
deWaC	German	1,627,169,557
JpWaC	Japanese	409,384,405
Russian web corpus	Russian	187,965,822
Spanish web corpus	Spanish	116,900,060

user: Miriam Buendía
 free: 1,000,000 tokens
 days left: 16

Corpora
[Create corpus](#)
[WebBootCaT](#)
 Configuration templates
 Sketch grammars
 User groups
 Settings
 Log out

Support
 Help
 Report a bug

My corpora

Corpus ID	Corpus name	Language	Size
no corpora			

[Show 33 more corpora](#)
[Create corpus](#) / [WebBootCaT](#)

Figura 1: Interfaz de Sketch Engine con sus principales aplicaciones

gripe Spanish web corpus freq = 952

object of 77 1.0	subject of 105 3.1	modifies 276 1.6	v o 29 0.9
contraer <u>7</u> 22.78	progresar <u>3</u> 14.89	virus <u>59</u> 44.71	neumococo <u>2</u> 19.22
temer <u>4</u> 16.13	ser <u>26</u> 12.89	pandemia <u>21</u> 42.97	neumonía <u>3</u> 19.13
tener <u>20</u> 13.83	poder <u>12</u> 12.39	epidemia <u>23</u> 35.11	tosferina <u>2</u> 19.06
coger <u>3</u> 11.07	llegar <u>6</u> 11.77	brote <u>8</u> 21.34	virus <u>2</u> 10.2
provocar <u>3</u> 9.24	actualizar <u>2</u> 9.17	caso <u>28</u> 19.92	
combatir <u>2</u> 9.08	haber <u>8</u> 8.72	síntoma <u>11</u> 19.18	
llamar <u>4</u> 8.91	matar <u>2</u> 8.15	causante <u>4</u> 15.6	
prevenir <u>2</u> 8.83	estar <u>6</u> 7.8	vacuna <u>6</u> 15.31	
padecer <u>2</u> 8.61	representar <u>2</u> 5.82	incidencia <u>6</u> 15.28	
transmitir <u>2</u> 8.32	era <u>2</u> 4.61	cepa <u>4</u> 14.9	
parecer <u>3</u> 7.91	tener <u>6</u> 4.59	pandemias <u>2</u> 13.91	
denominar <u>2</u> 7.63	deber <u>2</u> 3.71	variante <u>4</u> 12.94	
venir <u>2</u> 6.4	hacer <u>2</u> 1.82	expansión <u>5</u> 12.68	
evitar <u>2</u> 6.34	n modifier 138 0.8	llegada <u>4</u> 11.2	
ir <u>2</u> 4.82	aviar <u>41</u> 58.71	furia <u>2</u> 9.88	
	aviaria <u>18</u> 50.43	ataque <u>4</u> 9.42	
		estallido <u>2</u> 9.37	

Figura 2: Word sketch correspondiente al lexema gripe en el corpus español del proyecto WaCky

Castagnoli (2006) muestra las ventajas y limitaciones de la utilización de BootCat en un curso de Terminología y Lengua para fines específicos (LSP). De su estudio se desprende que la utilidad de los corpus compilados automáticamente depende principalmente de la familiaridad del usuario con el dominio de especialidad en cuestión, y de su habilidad para evaluar los textos/términos encontrados (Castagnoli 2006: 171).

3 La WfC y la WaC en una clase de Traducción científica y técnica

Muchos estudios han aplicado la lingüística de corpus a la didáctica de la Traducción (Bowker 1998, 2000; López 2002; Zanettin 1998; Zanettin/Bernardini/Stewart 2003), puesto que la utilización de corpus ayuda a los estudiantes a encontrar las palabras apropiadas para un contexto y un tipo de texto determinado y, por ende, incrementa su autonomía en el proceso de aprendizaje (López/Tercedor 2008). Sin embargo, sería interesante estudiar cuál de los dos enfoques principales descritos anteriormente facilita más el aprendizaje en el contexto de clases de traducción reales.

Para ello, llevamos a cabo el siguiente estudio en el marco de una clase de Traducción científica y técnica. Nuestro cometido era comprobar si el uso de SketchEngine (enfoque de la Web como corpus, WaC) ofrecía más ventajas a los estudiantes de traducción que el enfoque tradicional de utilizar la web de forma manual para la compilación de textos *offline* (WfC). Queríamos comprobar si el enfoque de considerar a la Red como corpus podía compensar la falta de conocimiento especializado de un grupo de estudiantes sobre el tema de un encargo de traducción (la gripe A), para que la calidad de sus traducciones fuera más o menos similar a la de otro grupo de estudiantes, que ya habían traducido algunos textos sobre esa temática.

3.1 Hipótesis

De todos es sabido que cuando se traduce un texto, el conocimiento previo en la materia, así como la experiencia de haber traducido textos sobre el mismo tema, facilitan el proceso de traducción. Asimismo, traducir en un entorno nuevo con herramientas informáticas con las que no estamos familiarizados puede empeorar el resultado de la traducción. Con lo cual, si estas premisas son ciertas, se podría formular la siguiente hipótesis nula:

Un grupo de estudiantes que traduce un texto sobre la gripe A con la ayuda de Google, después de haber traducido ya dos textos de la misma materia (Grupo de control), realizará mejor el encargo de traducción en condiciones de examen, que un grupo de estudiantes sin acceso a Google y que no haya traducido previamente ningún texto acerca de la gripe A (Grupo experimental).

No obstante, si proporcionamos al grupo experimental una herramienta para compensar su falta de experiencia previa en la traducción de textos de la gripe A y la limitación de no tener acceso a Google, y con esta herramienta, la calidad de sus traducciones es similar a la del grupo de control, podremos concluir que este recurso

resulta de gran eficacia para los traductores. Esta presuposición nos lleva a formular una hipótesis alternativa:

El enfoque de la Web como Corpus a través del uso de Sketch Engine puede compensar la falta de conocimiento de la materia en cuestión y de su terminología, por lo que resulta de gran utilidad para los traductores.

3.2 Diseño del estudio

Pedimos a dos grupos de alumnos de tercer curso de la Licenciatura en Traducción e Interpretación de la Universidad de Granada que tradujeran, en condiciones de examen, un fragmento de un artículo de investigación titulado "Insights from Investigating the Interaction of Oseltamivir (Tamiflu) with Neuraminidase of the 2009 H1N1 Swine Flu Virus" (Wang et al. 2009). En dos horas tenían que traducir 350 palabras, permitiéndoseles el acceso a Internet y a materiales de referencia como diccionarios electrónicos y de papel, y a los corpus que previamente habían compilado para el encargo. No obstante, establecimos algunas diferencias entre los dos grupos, de forma que uno fuera el grupo experimental y el otro, el grupo de control:

- Los estudiantes del grupo experimental nunca habían traducido textos acerca de la gripe A, en contraposición con los estudiantes del grupo de control, que ya habían traducido dos textos acerca de la misma temática.
- A los estudiantes del grupo experimental se les pidió que tradujeran sin Google, es decir, no podían utilizar Google ni para comprobar la terminología ni para documentarse.
- Los estudiantes del grupo experimental recibieron dos sesiones para aprender a manejar Sketch Engine, y compilaron un corpus acerca de la gripe A usando Sketch Engine una semana antes del examen.

Finalmente se les repartió un cuestionario tanto al grupo experimental como al de control para recabar información acerca de su formación previa, sus hábitos en la traducción y sus impresiones.

3.2.1 El cuestionario

Preparamos un cuestionario con tres secciones bien diferenciadas y un total de 34 preguntas. El grupo experimental tenía que dar respuesta a las tres secciones, mientras que el de control únicamente debía responder a las dos primeras. El cuestionario completo se incluye en el apéndice.

La primera sección proporciona detalles acerca de la formación y características de los sujetos con el objetivo de saber si estos datos podían haber sesgado los resultados del examen de traducción: edad, lengua materna y lenguas extranjeras, nivel de inglés, rapidez con el teclado y estudios universitarios previos (es lógico asumir que un licenciado en Biología con un buen dominio de la lengua inglesa traducirá mejor el texto).

La segunda sección incluye 9 preguntas acerca del conocimiento previo de la temática abordada en el encargo de traducción (la gripe A), y acerca de sus destrezas y hábitos de documentación. Les preguntamos si solían compilar corpus para sus

encargos de traducción y, en caso afirmativo, qué fases solían seguir en el proceso y si analizaban el corpus de forma manual o con programas de análisis léxico. Asimismo, incluimos dos preguntas para saber qué diccionarios habían empleado en el encargo de traducción.

Por último, en la tercera sección recogimos la opinión de los estudiantes del grupo experimental acerca de la utilidad de Sketch Engine, tanto para este encargo concreto, como para tareas de documentación y traducción en general. Además, les pedimos que destacaran las principales ventajas e inconvenientes de esta herramienta.

3.2.2 La muestra: edad, lenguas, destrezas de documentación y conocimiento previo de la materia

La muestra estaba compuesta por dos grupos de 12 estudiantes cada uno de tercer curso de la Licenciatura en Traducción e Interpretación de la Universidad de Granada que cursaban la asignatura de Traducción científica y técnica de inglés a español, siendo el inglés su primera lengua extranjera. Su edad se situaba entre los 20 y 38 años; la media fue 27 años y la moda, 21.

En una escala de 1 a 5, su nivel de inglés se situaba en 4 (Avanzado/C1/CAE). Todos los estudiantes, excepto uno, tenían el español como lengua materna y, además del inglés, todos poseían conocimientos de una segunda lengua extranjera (francés, alemán, chino, griego, árabe y ruso), un factor que facilita la interpretación de textos.

En concordancia con el hecho de que los sujetos cursaban una licenciatura en Traducción, el 75% decía estar acostumbrado a compilar textos en formato electrónico para las fases posteriores de documentación y extracción terminológica. También tenían experiencia en las búsquedas por Internet y en la utilización de recursos lexicográficos digitales, como los sitios web que proporcionan acceso a diccionarios, foros y redes sociales. De hecho, la mayoría consultó material lexicográfico en línea y diccionarios electrónicos instalados en sus ordenadores, en vez de diccionarios en papel.⁹ Para este encargo de traducción concreto, los recursos más utilizados por los estudiantes para documentarse quedan recogidos en la Tabla 1:

⁹ Esta tendencia y las nuevas reacciones de los estudiantes hacia los diccionarios de papel tradicionales se ha descrito en Tercedor, López y Faber (en prensa).

Diccionarios en línea:
- WordReference.com: http://www.wordreference.com/es/
- Diccionario Reverso en línea: http://dictionary.reverso.net/
- Free Dictionary.com: http://www.thefreedictionary.com/
- Diccionario de inglés Oxford en línea: suscripción de la Universidad de Granada
- Merriam Webster en línea: http://www.merriam-webster.com/
- Diccionario de la Real Academia Española: http://www.rae.es
Diccionarios electrónicos instalados en ordenador:
- Diccionario Oxford bilingüe inglés-español, español-inglés
- Diccionario Collins bilingüe inglés-español, español-inglés
Bases de datos bibliográficas:
- Medline
Diccionarios en papel:
- Diccionario crítico de dudas inglés-español de medicina de Fernando Navarro.

Tabla 1: Recursos de documentación más utilizados por los estudiantes

3.2.3 Utilización de Sketch Engine

El grupo experimental asistió a dos sesiones (de 2 y 1 hora, respectivamente) destinadas a explicar las diferencias entre los dos enfoques (WfC y WaC) y a iniciarlos en el uso de Sketch Engine, el programa de análisis descrito en el apartado 2.2.3.

En la primera sesión, los estudiantes se registraron en Sketch Engine, consultaron algunos corpus WaCky y empezaron a compilar un corpus de su propia elección para familiarizarse con la interfaz. A continuación se estableció un debate en el que los estudiantes nos hicieron partícipes de sus dudas y pudimos darles algunas recomendaciones para mejorar su rendimiento con Sketch Engine siguiendo a Castagnoli (2006).

En la segunda sesión, aclaramos dudas adicionales acerca del uso de Sketch Engine y pedimos a los estudiantes que compilaran dos corpus sobre la gripe A, uno en inglés y otro en español. Les comentamos que iban a utilizar dichos corpus para traducir un texto acerca de la gripe A en condiciones de examen. Si bien no tendrían acceso al texto hasta el día del examen, les proporcionamos algunas palabras clave que podían utilizar como "semillas" (*seeds*) iniciales para compilar el corpus en inglés: *flu, influenza A, influenza virus, H1N1, drugs, oseltamivir*. Les pedimos que anotaran el número de palabras de sus corpus, y las *semillas* que cada uno proponía para el corpus en español. Asimismo, les recomendamos que leyeran textos para familiarizarse con el tema objeto de traducción.

3.2.4 Evaluación de la calidad de las traducciones

Para evaluar los 24 exámenes seguimos el baremo de corrección de Robinson (1998) y Robinson, López y Tercedor (2006). Dicho baremo representa un enfoque holístico en la evaluación de traducciones y permite identificar las principales áreas en las que ocurren los errores de traducción, vinculándolas a las dos fases principales del proceso de traducción: la descodificación del texto origen (errores de significado) y la codificación del texto meta. Estas áreas son: 1) contenido/sentido, 2) registro,

vocabulario, terminología, 3) adecuación a la lengua y texto meta y 4) expresión escrita.

Sin embargo, al corregir las traducciones no aplicamos todo el baremo, ya que asumimos que Sketch Engine podría mejorar principalmente la calidad de las traducciones de los estudiantes en lo que se refiere a: (i) contenido; y (ii) elección del vocabulario o terminología pertinente (áreas 1 y 2). Es decir, aplicamos sólo las columnas del baremo relacionadas con estas dos facetas, a cada una de las cuales se le otorgó una puntuación de 0 a 5 (véase Tabla 2).

	A. Contenido	B. Registro, vocabulario, terminología
0	El texto no cumple el nivel mínimo aceptable	El texto no cumple el nivel mínimo aceptable
1	Comprensión insuficiente. Existen errores de contenido de importancia. Se omite una cantidad importante del contenido.	Registro inadecuado e incoherente. Vocabulario limitado y errores básicos. Conocimiento limitado de la terminología adecuada para el texto.
2	Conocimiento parcial del tema. Errores de contenido de menor importancia. Algunas omisiones del contenido del texto original.	En alguna ocasión, el registro es inadecuado e incoherente. Faltas ocasionales de vocabulario básico. Conocimiento de la terminología, aunque se producen algunos errores.
3	Dominio suficiente del tema. Omisiones de poca importancia de contenido del texto original menos relevante. Se producen fenómenos de sobretraducción o infratraducción que distorsionan el contenido del TO o causan ambigüedad.	Registro adecuado y coherente en general. Vocabulario eficaz, aunque se produzcan algunas faltas. Terminología adecuada, aunque en alguna ocasión podría haber realizado una mejor elección.
4	Buen dominio del tema. La sobretraducción o infratraducción no distorsionan el contenido del TO ni causan ambigüedad.	Registro adecuado y coherente. Vocabulario eficaz, aunque en alguna ocasión podría haber realizado una mejor elección. Terminología adecuada a pesar de algunos errores.
5	Dominio completo del tema. La transferencia del contenido y los matices de significado del texto origen se han realizado con una precisión absoluta.	Registro adecuado, eficaz y coherente a lo largo del texto. Elección minuciosa y eficaz del vocabulario. Terminología adecuada y completamente precisa.

Tabla 2: Criterios para evaluar la calidad de la traducción atendiendo al contenido y registro/terminología (adaptado del baremo de corrección de Robinson/López/Tercedor 2006)

Entre los errores que afectan al contenido se incluyeron no sólo cambios en el significado (falsos sentidos), sino también ejemplos en los que no existe cohesión o en los que los datos del texto origen se han cambiado. En cuanto a la evaluación del vocabulario y terminología del texto, tuvimos en cuenta la riqueza léxica y el grado de adecuación entre el léxico empleado y el registro del texto.

Finalmente, puntuamos cada traducción de 0 a 10, sumando las notas en la columna A (contenido) con las de la columna B (registro, vocabulario, terminología).

4 Resultados

Después de corregir los exámenes y analizar los cuestionarios, llegamos a unos resultados que permiten rechazar la hipótesis nula y aceptar la hipótesis alternativa.

A pesar de que el grupo de control ya había traducido previamente dos textos sobre la gripe A y utilizó Google, contraposición con el grupo experimental, sorprendentemente, la calidad de sus traducciones no difirió mucho (7,5 frente a 7,4). En términos generales, las puntuaciones correspondientes al contenido y léxico fueron muy similares:

	CONTENIDO (notas medias 0-5)	VOCABULARIO/ TERMINOLOGÍA (notas medias 0-5)	PUNTUACIÓN GLOBAL
Grupo control	3,7	3,8	7,5
Grupo experimental	4,0	3,4	7,4

Tabla 3: Nota media de los dos grupos en lo que se refiere a contenido y vocabulario/terminología

No obstante, al llevar a cabo un análisis más detallado, llegamos a dos resultados inesperados. En primer lugar, independientemente del enfoque que habían adoptado, los estudiantes que utilizaron diccionarios en papel obtuvieron, en general, mejores resultados que aquellos que simplemente se sirvieron de material electrónico. Para corroborar este hecho, ofrecemos a modo de ejemplo la entrada del concepto *neuraminidase* (neuraminidasa) en el *Diccionario crítico de dudas inglés-español de medicina* de Fernando Navarro (Tabla 4):

<i>NEURAMINIDASE</i>
Esta enzima es muy conocida por ser uno de los dos <i>antígenos de superficie</i> de los virus de la gripe – <i>hemaglutinina y neuraminidasa</i> –, que permiten clasificarlos en H1N1, H3N2, etc. (...) esta hidrolasa que <i>escinde</i> los enlaces glucosídicos entre un <i>residuo de ácido siálico</i> y uno de hexosa o hexosamina (...).

Tabla 4: Definición de *neuraminidase* en el *Diccionario crítico de dudas inglés-español de medicina* (énfasis añadido)

Esta breve descripción de un diccionario especializado en papel nos proporciona terminología específica que aparece en el texto origen como *antígeno* (antigen), *hemaglutinina* (hemagglutinin), *neuraminidase* (neuraminidase), *escindir* (cleave), o *residuo de ácido siálico* (sialic acid moieties), así como pistas para usar determinadas colocaciones. Por ejemplo, el diccionario indica que se dice *antígenos de superficie* y no *antígenos de la superficie* o cualquier otra construcción.

En segundo lugar, los estudiantes encontraron más problemas en la traducción de palabras de la lengua general que se usan en textos especializados como *target* o *insights*, que en la traducción de términos internacionales especializados como *neuraminidase*, *oseltamivir*, *zanamivir*, *Tamiflu*, *NA*, *HA*, *hemagglutinin*, o *virions*. En cualquier caso, el uso de Sketch Engine y de buenos diccionarios especializados como el que acabamos de mencionar, resultaron útiles a la hora de captar el significado del texto, con lo que se redujo el número de errores que afectan al sentido. Esta es la razón por la que el grupo experimental obtuvo ligeramente mejores resultados en lo que se refiere al contenido que el grupo de control.

Del análisis de las respuestas contenidas en la sección 3 del cuestionario, que versaban sobre el uso de Sketch Engine para traducir el texto, obtuvimos interesantes reflexiones para el uso lingüístico y traductológico de Sketch Engine. Las resumimos a continuación.

En primer lugar, nos llamó la atención que el 52% de los estudiantes no use normalmente diccionarios en papel. El resto se decanta por el uso de diccionarios en línea y electrónicos, y las búsquedas con Google, aunque el 50% declaró que suele dispersarse en sus búsquedas con Google, con lo que pierde bastante tiempo.

Por otra parte, el 81% de los estudiantes señaló que usaría de nuevo Sketch Engine en condiciones de examen y que, ahora que saben de la existencia de esta herramienta, no desearían tener que seguir traduciendo con métodos tradicionales. Todos coincidieron, no obstante, en que la mejor manera de sacarle mayor partido a Sketch Engine era utilizarlo en combinación con Google.

El 50% declaró que Sketch Engine resultaba de más utilidad para el inglés que para el español. Por un lado, porque el sistema contenía más corpus en inglés y con un mayor número de palabras, y por otro lado, porque al utilizar WebBootCat, el número de textos que Sketch Engine recuperaba era mucho mayor en inglés que en español, dado que la lengua inglesa está mucho más representada en la Red.

En lo que se refiere al tiempo empleado en llevar a cabo la traducción, el 33% concluyó que tardó menos, mientras que el resto señaló que tardó lo mismo que siguiendo métodos convencionales; ninguno señaló que hubiese tardado más.

Entre las principales ventajas de Sketch Engine, el 100% de los estudiantes destacó que al utilizar Sketch Engine obtenían textos más fiables que los ofrecidos por Google. También señaló su rapidez para compilar los corpus y las facilidades que ofrece para analizar los corpus, especialmente a través de concordancias. El 43% destacó asimismo la utilidad de Sketch Engine para documentarse y para adquirir conocimiento experto en la materia.

En cuanto a los inconvenientes, los estudiantes destacaron el hecho de que la licencia gratuita caduca en 30 días; su novedad, que hizo que no estuvieran totalmente familiarizados con todas las aplicaciones que ofrece y, por tanto, no utilizaron. Tampoco les gustó que no se puedan comparar equivalentes¹⁰ de traducción; y que Sketch Engine, a veces, no ofrezca todos los resultados requeridos. A menudo, elimina textos que podrían haber resultado de utilidad; Google, en general, ofrece más resultados.

En definitiva, los resultados muestran que Sketch Engine puede compensar la falta de conocimiento sobre la materia objeto de la traducción y resulta de gran utilidad para los traductores. De todas formas, con el experimento se ha constatado que los métodos sometidos a evaluación (WfC y WaC) deben ser complementarios y que los estudiantes deberán decidir en función de sus propias necesidades, es decir, en función del encargo de traducción, la novedad de la traducción, la direccionalidad, la especificidad de la traducción, el tiempo disponible y el nivel de análisis requerido. En cualquier caso, este estudio pone de manifiesto que el uso de corpus debe complementarse con el uso de fuentes lexicográficas de calidad.

5 Conclusiones

Tal y como hemos mencionado en este artículo, la documentación y la extracción terminológica constituyen unas de las fases más destacadas del proceso de traducción, fases en las que los corpus representan un papel fundamental. El hecho de que Internet sea hoy en día la principal fuente para la compilación de corpus, plantea la cuestión de si la web podría considerarse en sí misma como un auténtico corpus. En este estudio hemos pretendido demostrar cuál de los dos enfoques propuestos por De Schryver (2002), la web para compilar corpus (WfC) o la web como corpus (WaC), es más apropiado en la clase de Traducción.

Hemos comparado y evaluado estos dos enfoques en dos grupos de alumnos de la asignatura obligatoria de Traducción científica y técnica de la Licenciatura en Traducción e Interpretación de la Universidad de Granada, que realizaron una traducción especializada sobre la gripe A. El grupo experimental utilizó la web como corpus mediante la aplicación Sketch Engine, y el que hizo de grupo de control compiló un corpus de manera convencional.

Tras evaluar las traducciones aplicamos el baremo de corrección de Robinson (Robinson, López y Tercedor 2006), observamos que, a pesar del hecho de que uno de los grupos ya había traducido textos sobre la gripe A, en contraposición con el otro grupo, la calidad de sus traducciones no difería mucho. Así pues, la presente investigación corrobora que el enfoque de utilizar la web como corpus, mediante el uso de Sketch Engine, puede compensar la falta de conocimiento sobre la materia objeto de traducción y ayudar en la labor terminológica y es, por tanto, una herramienta de gran utilidad para los traductores y para cualquier lingüista interesado en el uso *in vivo*

¹⁰ En Google, por ejemplo, se puede buscar un término en español en un sitio web inglés añadiendo, por ejemplo, el dominio uk en la búsqueda, de forma que es posible comparar equivalentes.

del lenguaje en textos auténticos. Los resultados de esta investigación también ponen de manifiesto que, independientemente del enfoque adoptado, los estudiantes que utilizaron diccionarios en papel obtuvieron, en general, mejores resultados que aquellos que únicamente se sirvieron de recursos electrónicos.

Del análisis del cuestionario se desprende que la mayoría de los estudiantes resultaron gratamente sorprendidos con las aplicaciones y beneficios que Sketch Engine ofrecía a los traductores, ya que les permitió tardar menos tiempo en llevar a cabo el encargo de traducción, y les ayudó a la hora de documentarse y adquirir conocimiento experto. Todos coincidieron en la fiabilidad de los textos proporcionados por Sketch Engine y en que la mejor manera de aprovecharlo al máximo era utilizarlo en combinación con Google.

Este estudio puede servir de base a futuras investigaciones para averiguar las ventajas y los inconvenientes de ambos enfoques y sus aplicaciones didácticas. Y así, podría ser interesante que, al comienzo de las asignaturas de traducción, los estudiantes aprendieran a trabajar con ambas metodologías de forma complementaria, que es lo más provechoso para su aprendizaje, conforme se desprende de este estudio. Una vez familiarizados con ambas, se les repartiría un cuestionario para evaluar las ventajas e inconvenientes de cada enfoque. Asimismo, se les pediría que realizaran, en condiciones de examen, una traducción hacia el inglés (puesto que Sketch Engine proporciona más recursos en lengua inglesa) sobre un tema con el que estuvieran familiarizados. Al finalizar la traducción, los estudiantes reflejarían en un cuestionario para qué habían utilizado cada uno de los enfoques, sus ventajas e inconvenientes. También se les pediría que elaboraran dos tablas. En la primera de ellas, los estudiantes tendrían que incluir las referencias bibliográficas de los textos recopilados con cada metodología, de cara a investigar qué enfoque aporta textos de mayor fiabilidad.¹¹ En una segunda tabla, propondrían su traducción para unidades terminológicas especialmente complejas, indicando con qué metodología habrían llegado a dicha traducción.

Finalmente, este estudio abre el debate acerca de la necesidad de actualizar los programas de la asignatura de Terminología en los nuevos grados de Traducción e Interpretación, incorporando el aprendizaje de estas nuevas herramientas que son, sin lugar a dudas, de extrema utilidad para traductores y terminólogos.

Bibliografía

Ahmad, Khurshid; Paul Holmes-Higgin, Sibte Raza Abidi (1994): "A Description of Texts in a Corpus: 'Virtual' and 'Real' Corpora." Willy Martin, Willem Meijs, Margriet Moerland, Elsemiek ten Pas, Piet van Sterkenburg, Piek Vossen (eds): *Proceedings of EURALEX*. Amsterdam: Vrije Universiteit, 390-402

¹¹ Para evaluar la fiabilidad de los textos empleados para la documentación, puede ser interesante el cuestionario de Barker (2006).

- Aston, Guy (2009): "Foreword." Allison Beeby, Patricia Rodríguez Inés, Pilar Sánchez Gijón (eds): *Corpus Use and Translating: Corpus Use for Learning to Translate and Learning Corpus Use to Translate*. (Benjamins Translation Library 82.) Amsterdam/Philadelphia: Benjamins, ix-x
- Austermühl, Frank (2001): *Electronic Tools for Translators*. Manchester: St. Jerome
- Barker, Joe (2006): "Evaluating Web Pages: Techniques to Apply & Questions to Ask." Berkeley: University of California Library – <http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/Evaluate.html> (08/06/2011)
- Baroni, Marco; Silvia Bernardini (eds) (2006): *Wacky! Working Papers on the Web as Corpus*. Bologna: GEDIT
- Baroni, Marco; Adam Kilgarriff, Jan Pomikálek, Pavel Rychlý (2006): "WebBootCat: Instant Domain-Specific Corpora to Support Human Translators." André Lynam, Lena Korsnes (eds): *Proceedings of EAMT 2006 – 11th Annual Conference of the European Association for Machine Translation*. Oslo, 247-252
- Baroni, Marco; Silvia Bernardini (2004): "BootCaT: Bootstrapping Corpora and Terms from the Web." Maria Teresia Lino, Maria Francisca Xavier, Fátima Ferreira, Rute Costa, Raquel Silva (eds): *Proceedings of LREC*. Lisbon, 1313-1316 – http://sslmit.unibo.it/~baroni/publications/lrec2004/bootcat_lrec_2004.pdf (30/06/2010)
- Beeby, Allison; Patricia Rodríguez-Inés, Pilar Sánchez-Gijón (eds) (2009): *Corpus Use and Translating: Corpus Use for Learning to Translate and Learning Corpus Use to Translate*. (Benjamins Translation Library 82.) Amsterdam/Philadelphia: Benjamins
- Bernardini, Silvia; Marco Baroni, Stefan Evert (2006): "A WaCky Introduction." Marco Baroni, Silvia Bernardini (eds): *WaCky! Working Papers on the Web as Corpus*. Bologna: GEDIT, 1-32
- Bowker, Lynne (1998): "Using Specialized Monolingual Native-Language Corpora as a Translation Resource: A Pilot Study." *Meta* 43 [4]: 631-651
- Bowker, Lynne (2000): "Towards a Methodology for Exploiting Specialized Target Language Corpora as Translation Resources." *International Journal of Corpus Linguistics* 5 [1]: 17-52

trans-kom

ISSN 1867-4844

trans-kom ist eine wissenschaftliche Zeitschrift für Translation und Fachkommunikation.

trans-kom veröffentlicht Forschungsergebnisse und wissenschaftliche Diskussionsbeiträge zu Themen des Übersetzens und Dolmetschens, der Fachkommunikation, der Technikkommunikation, der Fachsprachen, der Terminologie und verwandter Gebiete.

Beiträge können in deutscher, englischer, französischer oder spanischer Sprache eingereicht werden. Sie müssen nach den Publikationsrichtlinien der Zeitschrift gestaltet sein. Diese Richtlinien können von der **trans-kom**-Website heruntergeladen werden. Alle Beiträge werden vor der Veröffentlichung anonym begutachtet.

trans-kom wird ausschließlich im Internet publiziert: <http://www.trans-kom.eu>

Redaktion

Leona Van Vaerenbergh
Artesis Hogeschool Antwerpen
Vertalers en Tolken
Schilderstraat 41
B-2000 Antwerpen
Belgien
leona.vanvaerenbergh@scarlet.be

Klaus Schubert
Universität Hildesheim
Institut für Übersetzungswissenschaft
und Fachkommunikation
Marienburger Platz 22
D-31141 Hildesheim
Deutschland
klaus.schubert@uni-hildesheim.de

- Buendía Castro, Miriam; José Manuel Ureña Gómez-Moreno (2009): "Parameters of Evaluation for Corpus Design." *International Journal of Translation* 21: 73-88
- Castagnoli, Sara (2006): "Using the Web as a Source of LSP Corpora in the Terminology Classroom." Marco Baroni, Silvia Bernardini (eds): *Wacky! Working Papers on the Web as Corpus*. Bologna: GEDIT, 159-172
- De Schryver, Gilles-Maurice (2002): "Web for/as Corpus: A Perspective for the African Languages." *Nordic Journal of African Studies* 11 [2]: 266-282 – <http://tshwanedje.com/publications/webtocorpus.pdf> (30/06/2010)
- Elkiss, Aaron; Philip Resnik (2004): "The Linguist's Search Engine User's Guide." – <http://lse.umiacs.umd.edu/lseuser.pdf> (30/06/2010)
- Fletcher, William H. (2001): "Concordancing the Web with KWicFinder." Paper Presented at the *3rd North American Symposium on Corpus Linguistics and Language Teaching* (Boston, 23-25 March 2001) – <http://kwicfinder.com/FletcherCLLT2001.pdf> (30/06/2010)
- Fletcher, William H. (2004): "Facilitating the Compilation and Dissemination of Ad-hoc Web Corpora." Guy Aston, Silvia Bernardini, Dominic Stewart (eds): *Corpora and Language Learners*. Amsterdam: Benjamins, 275-302
- Fletcher, William H. (2007): "Concordancing the Web: Promise and Problems, Tools and Techniques." Marianne Hundt, Nadja Nesselhauf, Carolin Biewer (eds): *Corpus Linguistics and the Web*. Amsterdam: Rodopi, 25-45
- Fletcher, William H. (en prensa): "Corpus Analysis of the World Wide Web." Carol A. Chapelle (ed.) *Encyclopedia of Applied Linguistics*. London: Wiley-Blackwell – <http://webascorpus.org/Corpus Analysis of the World Wide Web.pdf> (30/06/2010)
- Hundt, Marianne; Nadja Nesselhauf, Carolin Biewer (2007): "Corpus Linguistics and the Web." Marianne Hundt, Nadja Nesselhauf, Carolin Biewer (eds): *Corpus Linguistics and the Web*. Amsterdam: Rodopi, 1-5
- Internet World Stats (2009): "Internet World Users by Language. Top ten Languages." – <http://www.internetworldstats.com/stats7.htm> (30/06/2010)
- Kehoe, Andrew; Antoinette Renouf (2002): "WebCorp: Applying the Web to Linguistics and Linguistics to the Web." David Lassner, David De Roure, Arun Iyengar (eds): *Proceedings of the WWW 2002 Conference*. New York: ACM – <http://www2002.org/CDROM/poster/67> (25/06/2010)
- Kilgarriff, Adam (2003): "Linguistic Search Engine." Kiril Simov, Petya Osenova (eds): *Proceedings of the Workshop on Shallow Processing of Large Corpora*. Lancaster, 53-58
- Kilgarriff, Adam; Gregory Grefenstette (2003): "Introduction to the Special Issue on the Web as Corpus." *Computational Linguistics* 29 [3]: 333-347
- Kilgarriff, Adam; Pavel Rychly, Pavel Smrz, David Tugwell (2004): "The Sketch Engine." Geoffrey Williams, Sandra Vessier: *Proceedings Euralex*. Lorient, 105-116
- Leturia, Igor; Antton Gurrutxaga, Iñaki Alegria, Aitzol Ezeiza (2007): "CorpEus, a 'Web as Corpus' Tool Designed for the Agglutinative Nature of Basque." Cédric Fairon, Adam Kilgarriff, Hubert Naets, Gilles-Maurice De Schryver (eds): *Building and Exploring Web Corpora*. Louvain-la-Neuve: Cahiers du Central, 69-82
- López Rodríguez, Clara Inés (2002): "Training Translators to Learn from News Report Corpora: the Case of Anglo-American Cultural References." Belinda Maia, Johann Haller, Margherita Ulrych (eds): *Training the Language Services Provider for the New Millenium*. Porto: Faculdade de Letras Universidade do Porto, 213-222
- López Rodríguez, Clara Inés; Maribel Tercedor Sánchez (2008): "Corpora and Students' Autonomy in Scientific and Technical Translation Training." *Journal of Specialised Translation* 9: 2-19 – http://www.jostrans.org/issue09/art_lopez_tercedor.php (30/06/2010)

- Lüdeling, Anke; Stefan Evert, Marco Baroni (2007): "Using Web Data for Linguistic Purposes." Marianne Hundt, Nadja Nesselhauf, Carolin Biewer (eds): *Corpus Linguistics and the Web*. Amsterdam/New York: Rodopi, 7-24
- Ntoulas, Alexandros; Junghoo Cho, Christopher Olston (2004): "What's New on the Web? The Evolution of the Web from a Search Engine Perspective." Stuart I. Feldman, Mike Uretsky, Marc Najork, Craig E. Wills (eds): *Proceedings of the 13th International Conference on World Wide Web*. New York: ACM Press, 1-12
- Robinson, Bryan; Clara Inés López-Rodríguez, Maribel Tercedor-Sánchez (2006): "Self-Assessment in Translator Training." *Perspectives: Studies in Translatology* 14 [2]: 115-138
- Robinson, Bryan (1998): "Traducción Transparente: Métodos Cuantitativos y Cualitativos en la Evaluación de la Traducción." *Revista de Enseñanza Universitaria*, Número extraordinario: 577-589
- Scott, Mike (1999): *Wordsmith Tools*. Version 3. Oxford: Oxford University Press – <http://www.lexically.net/wordsmith> (08/06/2011)
- Sinclair, John (2005): "Corpus and Text-basic Principles." Martin Wynne (ed.): *Developing Linguistic Corpora: A Guide to Good Practice*. Oxford: Oxford Books, 1-16
- Tercedor-Sánchez, Maribel; Clara Inés López-Rodríguez, Pamela Faber (en prensa): "Working with Words: Research Approaches to Translation-oriented Lexicographic Practice." *TTR: traduction, terminologie, rédaction* 23
- Unión Latina (2007): "Lenguas y culturas en la Red. Estudio 2007." – http://dtiil.unilat.org/LI/2007/es/resultados_es.htm (30/06/2010)
- Varantola, Krista (2003): "Translators and Disposable Corpora." Federico Zanettin, Silvia Bernardini, Dominic Stewart (eds): *Corpora in Translator Education*. Manchester: St. Jerome, 55-70
- Volk, Martin (2002): "Using the Web as Corpus for Linguistic Research." Renate Pajusalu; Tiit Hennoste (eds): *Tähendusepüüdja. Hatcher of the Meaning. A Festschrift for Professor Haldur Õim*. Tartu: University of Tartu, 1-10
- Wang, Shu-Qing; Qi-Shi Dub, Ri-Bo Huang, Da-Wei Zhang, Kuo-Chen Chou (2009): "Insights from Investigating the Interaction of Oseltamivir (Tamiflu) with Neuraminidase of the 2009 H1N1 Swine Flu Virus." *Biochemical and Biophysical Research Communications* 386: 432-436 – <http://download.thelancet.com/flatcontentassets/H1N1-flu/virology/virology-36.pdf> (08/06/2011)
- Zanettin, Federico; Silvia Bernardini, Dominic Stewart (eds) (2003): *Corpora in Translator Education*. Manchester-Northampton: St Jerome
- Zanettin, Federico (2002): "DIY Corpora: The WWW and the Translator." Belinda Maia, Jonathan Haller, Margherita Urlrych (eds): *Training the Language Services Provider for the New Millennium*. Porto: Faculdade de Letras, Universidade do Porto, 239-248
- Zanettin, Federico (1998): "Bilingual Comparable Corpora and the Training of Translators." *Meta* 43 [4]: 616-630

Apéndice

A. INFORMACIÓN SOBRE EL ESTUDIANTE

1. NOMBRE	
2. Edad	
3. Lengua materna	4. Primera lengua extranjera
5. Segunda lengua extranjera	
6. Tercera lengua extranjera	
7. Mi nivel de inglés: <input type="checkbox"/> Level 2 / Low intermediate level of English / B1 / PET <input type="checkbox"/> Level 3 / High intermediate level of English / B2 / FCE <input type="checkbox"/> Level 4 / Advanced level of English / C1 / CAE <input type="checkbox"/> Level 5 / Proficient in English / C2 / CPE	
8. Velocidad/precisión con el teclado: <input type="checkbox"/> Muy buena <input type="checkbox"/> Normal <input type="checkbox"/> Regular <input type="checkbox"/> Mala	
9. Accedí a 2º ciclo de traducción desde otros estudios:	
10. Estudios universitarios previos:	

B. CONOCIMIENTO PREVIO SOBRE EL TEMA Y DOCUMENTACIÓN

11. ¿He leído textos paralelos sobre la gripe A antes de realizar la traducción?
12. ¿Había traducido algún texto sobre gripe A en otra asignatura?
13. ¿He usado Google para comprobar la terminología o documentarme a la hora de hacer esta traducción?
14. ¿Suelo recopilar un corpus en soporte electrónico para traducir textos?
15. ¿Qué herramientas utilizo para recopilarlo y qué pasos sigo?
16. ¿Suelo analizar dicho corpus con un programa de análisis léxico (que genera líneas de concordancia, listas de frecuencia, etc.) para documentarme y extraer información sobre terminología o fraseología?
17. ¿Analizo dicho corpus <u>de forma manual</u> para documentarme sobre el tema o utilizo <u>software</u> de análisis léxico o de corpus?
18. Diccionarios electrónicos consultados
19. Diccionarios en papel consultados

C. USO DE SKETCH ENGINE PARA TRADUCIR EL TEXTO PROPORCIONADO

20. Corpus en inglés: tiempo en recopilarlo y nº de palabras del corpus
21. Corpus en español: tiempo en recopilarlo y nº de palabras del corpus
22. <i>Seeds</i> de mi corpus en español
23. ¿He recopilado un corpus adicional en español a raíz del texto original? Tiempo en recopilarlo y nº de palabras del corpus <i>Seeds</i>
24. ¿Utilizaría de nuevo Sketch Engine en condiciones de examen?
25. ¿En condiciones de examen prefiero seguir haciendo las traducciones como antes?
26. ¿Si continuara mi licencia, seguiría utilizando Sketch Engine para mis encargos de traducción?
27. ¿Prefiero utilizar Sketch Engine en combinación con Google?
28. ¿Suelo dispersarme mucho en las búsquedas con Google?
29. Independientemente del tiempo empleado en recopilar el corpus ¿he tardado menos en traducir este encargo?
30. ¿Sketch Engine es más útil para encargos hacia una lengua extranjera?
31. ¿Sketch Engine es útil independientemente de la direccionalidad de la traducción?
32. ¿Antes del examen he practicado algunas aplicaciones avanzadas de Sketch Engine?
33. ¿Qué aplicaciones de Sketch Engine he empleado? <input type="checkbox"/> Concordance <input type="checkbox"/> Word List <input type="checkbox"/> Word Sketch <input type="checkbox"/> Thesaurus <input type="checkbox"/> Sketch-Diff
34. ¿En qué sentido(s) me ha resultado útil Sketch Engine? <input type="checkbox"/> Para localizar textos más fiables y relevantes que los que proporciona Google <input type="checkbox"/> Para documentarme sobre el tema y adquirir conocimiento experto <input type="checkbox"/> Para decidir sobre qué término es más apropiado <input type="checkbox"/> Para comprobar patrones de complementación <input type="checkbox"/> Otro (especifique)
35. Ventajas de Sketch Engine
36. Inconvenientes de Sketch Engine
37. Otras observaciones que quiero hacer constar

Autoras

Clara Inés López Rodríguez es Profesora Titular de Universidad en Traducción científica y técnica inglés-español en el Departamento de Traducción e Interpretación de la Universidad de Granada, donde obtuvo su doctorado con una tesis sobre la relación entre cohesión léxica, tipología textual y traducción biomédica. Su investigación actual versa sobre terminología y la aplicación de la lingüística de corpus y las nuevas tecnologías a la traducción científico-técnica, aunque también tiene trabajos sobre accesibilidad web y estudios culturales. Ha participado, entre otros, en los proyectos "Oncoterm: Sistema bilingüe de información y recursos oncológicos", "Marcos de conocimiento multilingües en la gestión integrada de zonas costeras" y "Ecosistema: Espacio único de sistemas de información ontológica y tesauro sobre el medio ambiente" (<http://ecolexicon.ugr.es/visual/index.html>).

Correo electrónico: clarailr@ugr.es

Web: [http://www.ugr.es/~dpto ti/](http://www.ugr.es/~dpto_ti/); <http://lexicon.ugr.es/lopezrodriguez>

Miriam Buendía Castro es Licenciada en Traducción e Interpretación por la Universidad de Granada y obtuvo el título de Máster en investigación en Traducción e Interpretación por dicha universidad. Está realizando su tesis doctoral en el grupo de investigación LexiCon de la Universidad de Granada, cuya investigadora responsable es la Dra. Pamela Faber. Sus líneas de investigación giran en torno a la Lingüística de corpus y la anotación semántica. Actualmente cuenta con una beca de formación de profesorado universitario (FPU), concedida por el Ministerio de Educación español, y se encuentra realizando una estancia de investigación en la Erasmushogeschool Brussel (Bruselas) bajo la dirección de la Dra. Rita Temmerman.

Correo electrónico: mbuendia@ugr.es

Web: <http://lexicon.ugr.es/buendia>

Neuerscheinungen bei Frank & Timme

FFF – Forum für Fachsprachen-Forschung

Herausgegeben von
Prof. Dr. Dr. h. c. Hartwig Kalverkämper

Julia Neu: **Mündliche Fachtexte
der französischen Rechtsprache.**
ISBN 978-3-86596-351-2

Sylvia Reinart: **Kulturspezifik in der
Fachübersetzung.** Die Bedeutung der
Kulturkompetenz bei der Translation
fachsprachlicher und fachbezogener Texte.
ISBN 978-3-86596-235-5

TRANSÜD. Arbeiten zur Theorie und Praxis des Übersetzens und Dolmetschens

Herausgegeben von
Prof. Dr. Dr. h. c. Hartwig Kalverkämper
und Prof. Dr. Larisa Schippel

Radegundis Stolze: **The Translator's
Approach.** An Introduction to Translational
Hermeneutics with Examples from
Practice. ISBN 978-3-86596-373-4

Pekka Kujamäki/Leena Kolehmainen/
Esa Penttilä/Hannu Kemppanen (eds.):
**Beyond Borders – Translations Moving
Languages, Literatures and Cultures.**
ISBN 978-3-86596-356-7

Agnès Welu: **Neuübersetzungen ins
Französische – eine kulturhistorische
Übersetzungskritik.** Eichendorffs
„Aus dem Leben eines Taugenichts“.
ISBN 978-3-86596-193-8

Miriam Paola Leibbrand: **Grundlagen
einer hermeneutischen Dolmetsch-
forschung.** ISBN 978-3-86596-343-7

Gisela Thome: **Übersetzen als inter-
linguales und interkulturelles Sprach-
handeln.** Theorien – Methodologie –
Ausbildung. ISBN 978-3-86596-352-9



F Frank & Timme

Verlag für wissenschaftliche Literatur

Wittelsbacherstraße 27a, D-10707 Berlin
Telefon (0 30) 88 66 79 11, Fax (0 30) 86 39 87 31
info@frank-timme.de, www.frank-timme.de

F Frank & Timme
Verlag für wissenschaftliche Literatur

Frank & Timme GmbH

Wittelsbacherstraße 27a, 10707 Berlin

Telefon: (0 30) 88 66 79 11

Fax: (0 30) 86 39 87 31

info@frank-timme.de

www.frank-timme.de