Traducción automática y memoria de traducción

MARJORIE LEÓN*

Servicio de Traducciones
Organización Panamericana de la Salud
Washington DC, EE.UU.
Tel. 202 974 3590 Fax: 202 974 3620 leonmarj@paho.org

Resumen

Los intentos de proponer nuevas herramientas informáticas a los traductores han encontrado bastante resistencia en la mayoría de las organizaciones internacionales. En cambio, la Organización Panamericana de la Salud utiliza un sistema propio de traducción automática desde hace más de veinte años. Gracias al esfuerzo de sus traductores, lingüistas y programadores, el software y sus diccionarios extensos se han convertido en importantes bienes de capital. Se describe el método de trabajo del Servicio de Traducciones y una iniciativa reciente que promete agilizar aun más la labor del traductor.

Palabras clave: traducción automática, memoria de traducción, diccionarios electrónicos, terminología, revisión de la traducción.

1 Introducción

Los traductores con aptitudes para *teclear* tienen muchas herramientas a su disposición. Entre ellas figuran procesadores de textos, bases terminológicas, memorias de traducción, traducción automática, bases de datos con textos completos y motores de búsqueda de Internet. Las funciones más básicas de «cortar y pegar» y «buscar y reemplazar» son probablemente las que más tiempo hacen ganar.

Las bases terminológicas y las memorias de traducción son sistemas de almacenamiento con diferentes mecanismos para agregar y recuperar contenido. Dado que están vacías cuando se adquieren, no son de utilidad inmediata, sino que hay que añadir el esfuerzo sustancial de introducir

datos al desembolso inicial. En cambio, cuando se compra un sistema de traducción automática con buenos diccionarios, es como comprar un piso amueblado o un frigorífico bien repleto. Se puede sacarle provecho de inmediato.

Un sistema de traducción automática puede ser un asistente valioso, que puede funcionar sin exigir la presencia en tiempo real del traductor. El programa no se cansa, ni se queja, ni exige que se le paguen horas extraordinarias. Recuerda las instrucciones, no le ofende la crítica y aprende lo que se le enseña.

2 Antecedentes

La Organización Panamericana de la Salud (OPS) desarrolló su sistema de traducción automática para satisfacer sus necesidades internas de traducción. SPANAM® (de español a inglés) está en uso desde 1980 y ENGSPAN® (de inglés a español) desde 1985. En sus comienzos era una aplicación instalada en un ordenador central. En 1992 salió una versión para PC, y la versión para Windows vio la luz en 2000. El software funciona con Windows 95/98/NT y con Windows 2000/XP. Puede trabajar con documentos preparados en Word y WordPerfect (archivado en RTF), ficheros de HTML, SGML, XML y texto. La instalación independiente total requiere 100 MB de almacenamiento en disco duro; la instalación de LAN requiere 140 MB. Se recomienda un PC con un procesador, al menos, Pentium II y un mínimo de 64 MB de RAM. El software está protegido con un dispositivo de seguridad.

En la sede de la OPS, el software está instalado en la LAN y en la Intranet. También está instalado en la Organización Mundial de la Salud (OMS) en Ginebra, en treinta oficinas de campo y centros regionales de la OPS, y en ocho instituciones nacionales de salud del continente americano. Cuarenta y dos clientes externos han adquirido la licencia de utilización.

Los sistemas de traducción automática de la OPS son completamente automáticos y disponen de diccionarios plenamente adaptables a las necesidades del cliente.

3 Cómo se usa la traducción automática en la OPS

Nuestra unidad de traducción tiene una plantilla de dos traductores para el español y una traductora para el inglés. Como tenemos tan poco personal interno, damos al exterior una gran cantidad de trabajo de posedición a traductores independientes. En 2001, la unidad tradujo un total de 1 470 000 palabras del inglés al español, el 93% de ellas partiendo de la traducción automática. Los traductores externos hicieron un 61% de los trabajos y un 85% del número total de palabras.

La OPS usa la traducción automática para traducir muy diversos tipos de texto. Si bien es verdad que nuestros diccionarios contienen muchos términos médicos, el sistema funciona bien con textos de muchos otros ámbitos. Traducimos habitualmente textos que tratan muchos aspectos de la salud pública, incluidos temas políticos, económicos y sociales. Los textos son redactados para públicos variados: nuestros Cuerpos Directivos, otros foros internacionales, organismos financieros, instancias decisorias nacionales, científicos, profesionales de la salud, estudiantes, pacientes y medios de comunicación. Es posible definir microglosarios para manejar términos pertenecientes a diferentes materias o registros.

Los diccionarios de traducción automática de la OPS son el resultado de más de 25 años de trabajo de sus traductores y lingüistas. El léxico se ha ido recogiendo entre los más de 50 millones de palabras que se han traducido con este software y que han poseditado nuestros traductores, de plantilla y contratados. Cada conjunto de diccionarios contiene aproximadamente 100 000 entradas, expresiones y reglas contextuales. Los diccionarios se actualizan diariamente, partiendo de la información comunicada por los traductores.

El sistema espera encontrar una sintaxis ordinaria. No se requiere revisión previa del texto original. Antes de lanzar la traducción automática, nuestro personal de secretaría comprueba el fichero de entrada por si hay problemas de formateo, comprueba que la copia electrónica coincida con la copia impresa, procede a la corrección ortográfica y bloquea secciones del texto que no deban traducirse. En su caso, el operador puede especificar un tipo de texto para decirle al programa que va a encontrarse con un tipo especial de sintaxis.

Nuestros traductores realizan en pantalla la posedición de la traducción automática bruta. Producen traducciones de la mayor calidad que el tiempo permita, teniendo en cuenta el uso que se dará al texto.

También los traductores de plantilla revisan en pantalla el trabajo de los externos. El recurso a la traducción automática reduce el tiempo requerido para revisar el trabajo de un traductor externo, porque la traducción bruta ya contiene la mayor parte de la terminología correcta. Los nombres de entidades, iniciativas, enfermedades y medicamentos aparecen con marcas de fiabilidad. Además, el sistema de traducción automática no se salta involuntariamente porciones del texto, transpone números ni se equivoca con los nombres propios. En algunos casos, el traductor de plantilla da indicaciones previas en la versión bilingüe, en papel, que acompaña al fichero electrónico, o puede haber hecho algunos cambios clave antes de enviar el texto al exterior.

4 Uso del sistema de traducción automática de la OPS con memoria de traducción

Hasta hace poco, la unidad de traducción de la OPS no había probado un sistema de memoria de traducción. Nuestra renuencia se basaba en el carácter no repetitivo de los documentos que solemos traducir y en el elevado costo de la tecnología, tanto en cuanto a la adquisición de la licencia como al esfuerzo inicial. Pensábamos que nuestros traductores, tanto de plantilla como externos, tal vez no quisieran familiarizarse con una interfaz nueva, sino seguir trabajando en MS Word. Reinaba también cierto escepticismo en cuanto al hipotético ahorro que supondría, pues seguiría habiendo bastantes segmentos no traducidos, y sería necesario controlar las traducciones aproximadas y aun las traducciones exactas para verificar que se estuvieran utilizando en el contexto adecuado.

No obstante, nos dimos cuenta de que una memoria de traducción podría ayudarnos con algunos de nuestros documentos de trabajo que van siendo objeto de modificaciones conforme se presentan a tres distintos Cuerpos Directivos. Buscamos un paquete de bajo costo que pudiera vincularse a nuestro sistema de traducción automática, y dimos con Wordfast, una herramienta de memoria de traducción que puede descargarse gratuitamente de www.champollion.net. El productor de Wordfast, Yves Champollion, que ha sido traductor, ha ideado una interfaz de manejo muy sencillo. Fue fácil conectarle el componente de traducción automática de la OPS para que tradujera automáticamente todos los segmentos que no encontrara en memoria. Acabamos de empezar nuestro experimento, de manera que no puedo informar sobre los beneficios

obtenidos a partir de las memorias de traducción específicas de los documentos elaborados por nuestros traductores. Sí puedo decir, no obstante, que a nuestros traductores de plantilla les gusta la revisión posterior en el entorno Wordfast.

Con Wordfast también se puede generar un archivo de salida que contiene la frase original seguida bien por la traducción que se encontraba en memoria, o bien generada por traducción automática. Cada segmento del idioma de llegada está claramente marcado para saber cómo se originó. Un traductor externo puede trabajar con este archivo de salida en el entorno Wordfast sin necesidad de disponer del sistema de traducción automática ni de la memoria de traducción.

También hicimos una prueba con un texto «pretraducido» con Translator's Workbench de Trados. En el archivo pretraducido, Trados usa diferentes colores y códigos de idioma para los segmentos recuperados de la memoria y los que no tienen traducción. Estos códigos hacen que nuestro sistema de traducción automática sólo modifique los segmentos no traducidos que llevan el código correspondiente al idioma original. El resultado es un archivo que puede mandarse al exterior para ser revisado en Translator's Workbench o en entorno Wordfast.

Nos fue grato constatar que se puede combinar la traducción automática y la memoria de traducción sin necesidad de exportar los segmentos no traducidos, procesarlos con traducción automática e importarlos a la memoria, como sugiere Trados. Se eliminan los pasos de exportar e importar, y los segmentos traducidos automáticamente no se añaden a la memoria hasta después de haber sido revisados.

Pensamos que el sistema de traducción automática de la OPS ayuda a superar al menos dos de las deficiencias de los sistemas existentes de memoria de traducción. En vez de buscar oraciones completas repetidas, la traducción automática coteja los segmentos a nivel de la palabra y de la frase. Y la traducción automática ya tiene incorporada información sintáctica, semántica y contextual que sólo ahora están empezando a adoptar algunos sistemas de memoria de traducción.

3 Conclusión

Al combinar la traducción automática y la memoria de traducción, esperamos añadir aún más velocidad y exactitud a nuestro proceso de traducción. La demanda de traducción va en aumento, pero los presupuestos institucionales disminuyen. En la OPS, buscamos un efecto

multiplicador del talento y las aptitudes de nuestros traductores, poniendo a su disposición herramientas de trabajo cada vez más poderosas.

Agradecimientos

Como jefa del Servicio de Traducciones de la OPS, he tenido el privilegio de trabajar con un grupo de traductores competentes, entusiastas y entregados a su labor. Quiero expresar el reconocimiento a la valiosa contribución de las siguientes personas, actuales o antiguos miembros de la unidad: Gustavo Silva, Eleana Gómez , Miguel Turrión, Maxine Siri y Laura Vlasman. Son también imprescindibles para nuestro funcionamiento los muchos traductores independientes que, trabajando bajo contrato, proceden a la posedición de nuestras traducciones automáticas. La responsabilidad del desarrollo del programa ha recaído en Julia Aymerich y en mí misma.

* Doctora en Lingüística computacional; responsable de los servicios de traducción de la Organización Panamericana de la Salud y de los programas de traducción automática SPANAM® and ENGSPAN® de esta organización.