

Herramientas de traducción automática... mito y realidad

Cristina Márquez Arroyo

Herramientas de traducción automática... mito y realidad

Durante la última década se ha producido un fenómeno en el campo de la traducción que ha despertado y despierta aún los fervores más decididos y las más enconadas críticas. Se trata de los programas informáticos relacionados con el proceso de traducción. ¿Quién de nosotros no ha sentido preocupación ante la amenaza de ser reemplazado por un CD ROM de fácil instalación en una computadora personal? ¿Cuántas veces hemos pensado que el próximo sistema operativo anunciado por Bill Gates incluirá una versión completa de algún programa mágico que le brinde a cualquier usuario, y lo que es peor, a nuestros clientes, la posibilidad de prescindir de nuestros servicios?

A decir verdad, hace ya más de 20 años que empecé a escuchar rumores de programas con inteligencia artificial que permitirían reemplazar al traductor en forma eficaz y rentable. Y no puedo negar que en su momento, fueron rumores alarmantes. Aunque mirando hacia atrás, me doy cuenta de que en parte eran rumores impulsados por los ingenieros de sistema a quienes obligábamos a revisar las traducciones producidas por el Centro Latinoamericano de Traducciones de IBM, en cuya creación y organización tuve el honor de formar parte, bajo la guía del ingeniero Luis Di Masi, un verdadero visionario totalmente desconocido en el mundo de la traducción, que sin embargo, impulsó en Argentina y Latinoamérica la tecnología y el uso de memorias y bases de datos para agilizar el proceso de traducción.

Aunque ningún colega duda de que nuestro trabajo será realizado cada vez más frecuentemente utilizando memorias de traducción, aún hay muchos profesionales que se niegan a aceptar la existencia de estas herramientas, por otro tan importantes para facilitar su tarea. En un excelente artículo publicado en el boletín del ITI (*Institute of Translation and Interpreting*), el traductor británico Michael Benis destaca que, según la encuesta de tarifas y honorarios realizada por dicha institución en 1998 (*1998 Rates and Salaries Survey*), sólo el 15% de los encuestados manifestaron utilizar memorias de traducción, pero esta cifra aumentó al 40% en el grupo que tenía ingresos de 72000 a 80000 dólares anuales, y al 50% en el grupo cuyos ingresos superaban los 108.000 dólares anuales.¹

La expresión «herramientas de traducción asistida por computadora» (CAT Tools) describe en forma amplia distintos tipos de aplicaciones y tecnologías informáticas que automatizan o ayudan a realizar el proceso de traducción de un idioma a otro. En general, se pueden agrupar en tres categorías de aplicaciones:

- Traducción automática (MT)
- Gestión de terminología (TM)
- Traducción humana asistida por computadora (CAHT)

1- Michael Benis, *Forget Me Not*, artículo publicado en el boletín de ITI (www.ITI.org.uk)

En esta ponencia me voy a referir a las dos últimas categorías por incluir los productos que nos permiten mejorar la calidad y productividad de nuestro trabajo. No obstante, en forma breve me gustaría mencionar que los programas de traducción automática (MT) pueden ser aceptables bajo ciertas circunstancias, y en ningún caso ponen en peligro nuestras fuentes de trabajo y de ingresos. Sus beneficios se limitan a usuarios que por un motivo u otro sólo quieren tener una idea general del texto en el idioma extranjero, o necesitan identificar textos dentro de grandes cantidades de documentación a fin de determinar cuáles de ellos requieren traducción humana. Es decir, traducciones de consumo interno.

Herramientas basadas en Memorias de traducción o Gestión de memorias

Las herramientas que utilizan memorias de traducción funcionan sobre una idea muy simple: en vez de tener que volver constantemente al glosario, a las notas de la traducción o al diccionario (que además sistemáticamente hemos cerrado), para recordar cómo tradujimos el término o la frase que se nos escapa de la mente, guardamos la información en un espacio de la computadora (a veces en el disco duro y otras en memoria), y cuando aparece nuevamente, una instrucción de la aplicación que estemos utilizando lo trae a la pantalla, nos lo muestra y hasta nos lo incluye automáticamente en la traducción.

Uno de los problemas con todos estos productos es que son caros. Algunos más que otros, pero en general, todos representan una inversión considerable. Por eso es importante que nos formulemos algunas preguntas básicas antes de desembolsar el dinero.

En primer lugar, este tipo de programas sólo trabaja con documentos en formato electrónico. Ya sea en editores de texto comunes (como *Word* o *WordPerfect*), o en programas de diseño de textos o maquetación (como *FrameMaker*, *QuarkXpress* o *PageMaker*). Aunque la exploración digital de documentos ha avanzado enormemente en los últimos tiempos, aún requiere una gran cantidad de correcciones manuales, las cuales reducen considerablemente el beneficio de una mayor productividad.

La segunda condición para que un programa de gestión de memoria sea eficaz es que el texto a traducir tenga cierto nivel de repetición. Por este motivo, de nada nos servirán si nuestra especialización es la traducción literaria. Sin embargo, si tenemos tres o cuatro clientes para quienes traducimos manuales de instrucciones, documentación de *software*, formularios repetitivos o de seguros, evaluaciones médicas, etc., es muy probable que recuperemos la inversión a corto plazo.

Otro punto importante a considerar cuando tomamos la decisión de usar programas de gestión de memoria es que nos dediquemos a proyectos de localización

de software, aunque no realicemos la porción técnica del trabajo sino únicamente las tareas de traducción. En estos proyectos, la coherencia terminológica entre los distintos componentes de la aplicación (interfaz del usuario, sistema de ayuda, mensajes y documentación) es absolutamente crucial. Y la ayuda de la memoria automática es una ventaja invaluable. Además, el tiempo transcurrido entre el anuncio del producto original en inglés y su lanzamiento en el mercado internacional es clave para el éxito del producto. Los programas de gestión de memoria permiten comenzar a trabajar cuando el producto está aún en las últimas etapas de su diseño ya que los pequeños cambios que a esa altura se producen pueden incorporarse automáticamente en la documentación, disminuyendo así el tiempo de localización y aumentando la productividad. Además, los productos de software sufren periódicas actualizaciones, que se pueden incorporar rápidamente al producto ya traducido. Por último, es probable que más de una vez debamos traducir productos que son actualizaciones de otros cuya traducción ya ha sido realizada. En estos casos, es sumamente útil el uso de una memoria ya existente o su creación mediante la alineación de los textos traducidos para el programa original. Las ventajas incluyen no sólo la traducción automática o la pretraducción de los textos repetitivos, sino también la verificación de la terminología a utilizar sea exactamente la misma de los productos similares o de la misma empresa. En el mercado actual, hay varios productos de este tipo. Los más conocidos son *Translation Manager* de IBM, *DejáVú* de Atril, *Star* de Transit, *SDLX* de SDL y *Translator's Workbench* de Trados.

Todos ellos ofrecen las mismas funciones básicas, aunque utilizan metodologías diferentes:

- Gestión de memoria
- Alineación
- Modificación de la base de datos
- Gestión de terminología
- Editor de texto
- Filtros

La gestión de memoria es la función más importante de estos programas, ya que es la que analiza el texto fuente, lo compara con la base de datos de referencia o la memoria, detecta las posibles traducciones existentes y trae a la pantalla las coincidencias encontradas, ya sea parciales o totales.

La alineación es una función que permite utilizar las traducciones ya existentes de un tema para crear una o más memorias de traducción. Compara dos textos –fuente y meta– y crea las equivalencias entre las unidades de traducción.

La modificación de la memoria, mal llamada «mantenimiento», permite acceder a los archivos que forman la memoria y modificar manualmente los términos ya incluidos.

La gestión de terminología incluye la creación de glosarios y en algunos casos el uso de diccionarios incorporados en la base de datos del programa (*Translation*

Manager de IBM). Esta función permite importar glosarios creados previamente y también visualizar el término que necesitamos con sólo hacer clic sobre él.

El **editor de texto** es el programa que nos permite introducir la traducción. La mayoría de los programas CAT utilizan una tabla con una columna que muestra el texto fuente y otra para introducir el texto meta, y paneles adicionales que permiten observar la unidad de traducción, el glosario y otros componentes con que se está trabajando en ese momento. La excepción es Trados, que permite trabajar directamente con *Microsoft Word* mediante plantillas y macros que se añaden al programa.

Los **filtros** son pequeñas aplicaciones que permiten importar los archivos con cualquier atributo y extraer el texto para su traducción. En la mayoría de los casos, estos filtros bloquean los códigos internos contenidos en el texto creando «fichas» o tags protegidas que el traductor no puede modificar accidentalmente.

Los programas CAT utilizan una jerga particular, en la que se pueden destacar los siguientes términos:

Unidades de traducción:	Fragmentos del texto que se visualizarán en la pantalla para su traducción.
Match	Coincidencia entre el texto fuente y el texto meta.
Full o 100% Match	Coincidencia total entre el texto fuente y el texto meta.
Fuzzy Match	Coincidencia parcial entre el texto fuente y el texto meta.
Assembled Match	Coincidencia prearmada mediante la función de gestión de memoria.
Multiple Exact Matches	Coincidencias entre un texto fuente y varios textos meta, según el contexto de la traducción.
% of Fuzziness	Porcentaje de coincidencia deseado para aceptarla en forma parcial para su confirmación o corrección.

La siguiente es una breve reseña de las funciones y ventajas de los tres programas más conocidos y utilizados por la comunidad profesional: *Translator's Workbench*, *DejáVù* y *Transit*.

Translator's Workbench

El programa preferido de los traductores, ofrece la ventaja exclusiva de permitir trabajar con el procesador de texto *Microsoft Word* al que estamos acostumbrados. Es decir, funcionalidad WYSIWYG (*What You See Is What You Get*). Además, permite comenzar a trabajar prácticamente de inmediato, con sólo conocer las funciones de los botones que lo ponen en marcha, que se añaden mediante una plantilla a la barra de herramientas del editor de texto.

Claro que el uso de las demás funciones requiere de capacitación más profunda para lograr el mayor aumento de productividad posible, pero familiarizarse con el programa lleva realmente poco tiempo. Es también la herramienta preferida de las empresas de traducción (no así de las de localización), por lo cual se ha convertido en el programa líder de la industria. Puestos a decidir acerca de un solo producto para comprar, el traductor puede apostar por *Workbench* con la seguridad de que va a recuperar su inversión rápidamente.

Trados (como todo el mundo lo conoce) viene en distintas versiones que cubren desde las sofisticadas necesidades de las empresas de traducción hasta las más sencillas de los traductores independientes. Estos pueden obtener desde hace dos años una versión llamada *Freelance* que incluye los componentes más importantes a un precio reducido. Los componentes a que me refiero son *MultiTerm* (la aplicación de gestión de terminología), *Translator's Workbench* (el programa de gestión de memoria), *WinAlign* (la herramienta para alineación de textos), *TagEditor* (un editor para archivos en formato HTML, XML o SGML), *S-Tagger* (para archivos *FrameMaker* e *Interleaf*) y *T-Windows* (para archivos *PowerPoint*).

Translator's Workbench funciona con una base de datos en segundo plano en la que compila la traducción, a medida que se la introduce. Es decir, guarda la unidad de traducción en el idioma fuente con sus respectivas traducciones. A medida que se avanza en la traducción y van apareciendo nuevas unidades, el programa las compara con las unidades ya traducidas y cuando identifica un texto idéntico o similar, lo «trae» a la pantalla para que el traductor lo acepte, lo complete o lo corrija. El programa puede configurarse para que acepte automáticamente las unidades idénticas y pase a la siguiente oración.

Además, incluye funciones prácticas como la conversión automática de fechas, números y medidas, la concordancia, el manejo de notas de pie de página y entradas de índice. Vale la pena destacar el análisis de archivos para determinar la cantidad exacta de texto repetitivo y texto nuevo, y por lo tanto el tiempo y el costo involucrado. *Workbench* es la única herramienta que permite realizar este análisis antes de empezar el proyecto.

La función de concordancia permite encontrar todas las unidades que incluyen una palabra o frase determinada, con sólo un clic del ratón. El resultado se muestra en una ventana adicional que muestra todas las traducciones existentes hasta ese momento, con la indicación de la fecha en que fueron realizadas y el nombre del usuario que las introdujo. *Workbench* utiliza una ventana con dos paneles que muestran la traducción a realizar en la parte superior, y la traducción existente en la parte inferior. Si se utiliza *MultiTerm*, se lo puede acomodar en un tercer panel. Debido a que no utiliza un editor propio y la traducción se realiza directamente en *Word* es necesario acomodar los dos programas (*Workbench* y *Word*) para que puedan visualizarse simultáneamente, a fin de lograr una mayor eficacia y productividad.

WinAlign es el producto que realiza la función de alineación. Si bien lo hace en forma automática, requiere que el traductor se deslice a través de los textos alineados para «retocar» la alineación manualmente. La diferencia de estructura gramatical del español y el inglés obliga frecuentemente a realinear las unidades de traducción, pero para hacerlo sólo hay que mover los vectores que enlazan unas con otras. Es decir, se trata de una auténtica herramienta visual de alineación que puede utilizarse en forma sencilla y rápida, y que permite capitalizar todas las traducciones existentes guardadas en formato electrónico. La única desventaja de esta aplicación es que sólo funciona con archivos en formato RTF (*Rich Text Format*) o HTML (*Hypertext Markup Language*).

Por último, y aunque no sea una aplicación demasiado sólida desde el punto de vista técnico, vale la pena mencionar *TagEditor*, un nuevo editor diseñado por *Trados* para la traducción de archivos HTML. Una de sus ventajas es que permite ver la traducción tal como se verá en la web, pero el producto todavía no es demasiado estable. Con las debidas precauciones de guardar el archivo a menudo, esta es una herramienta ideal para los traductores que recién se inician en la traducción de textos que contienen código HTML.

DejáVù

Éste es el producto promocionado como la herramienta para traductores diseñada por traductores. En primer lugar, cuenta con el respaldo de un grupo de expertos que «pone la cara» y está siempre a disposición de los usuarios para resolver sus dudas, contestar sus preguntas y escuchar sus sugerencias y necesidades, las cuales en muchos casos se incorporan a las actualizaciones del producto. Emilio Benito, el creador de este producto, se dedica permanentemente a actualizar y mejorar el programa, a fin de que los traductores dispongan permanentemente de una herramienta que incluye la más amplia variedad de funciones, filtros y características del mercado. El producto utiliza una memoria residente en una base de datos que incluye las referencias y traducciones existentes sobre el proyecto en que se está trabajando. Por el momento sólo permite utilizar una base de datos, pero el equipo de apoyo técnico está trabajando para que pueda utilizar múltiples bases de datos, incluyendo *Oracle* y *SQL*.²

El programa incluye un tutorial que permite comenzar a utilizarlo en unas pocas horas, y además, la empresa "Language Partners International" ofrece cursos completos de dos días de duración para la capacitación de los usuarios en las funciones avanzadas del producto. Inicialmente diseñado como una herramienta para traductores independientes, *DejáVù* ha ido incorporando funciones que lo convierten en una herramienta ideal para empresas de traducción y de localización que deben manejar grandes proyectos en múltiples idiomas. Es importante destacar que la asistencia técnica es gratuita, como también las actualizaciones de una misma versión.

2- *Apuntes*, publicación de SpanSIG, Volumen 8 N.º 4, Otoño del 2000.

DejáVú incluye funciones de alineación, de gestión de terminología, correctores ortográficos y filtros para los distintos programas de procesamiento de texto y maquetación (*FrameMaker*, *QuarkXpress*, *PageMaker*, etc.). Una de sus características más atractivas es la facilidad con que maneja los archivos de recursos, inevitables componentes de los proyectos de localización de SW. Con sólo pulsar un par de teclas, estos complejos archivos con una enorme cantidad de código interno, se convierten en tablas simples que además pueden exportarse en formato texto en caso de que el traductor trabaje con colaboradores que no tienen el producto. La fragmentación es sumamente flexible, y aunque reduce la fluidez de la lectura como todos los productos de este tipo, permite separar o unir los segmentos manualmente y sobre la marcha.

La porción de gestión de memoria, llamada *DVI (DejáVú Interactive)* es el entorno de traducción en sí mismo. Su aspecto es similar al del resto de los productos (a excepción de *Workbench*). Es decir, un panel dividido en dos columnas con el texto fuente y el texto meta, y dos paneles ubicados en la parte inferior de la pantalla, el primero con el segmento a traducir y el segundo con el espacio para introducir la traducción. La función *Pretranslate* permite importar todas las coincidencias completas y las coincidencias parciales, además de las coincidencias parciales prearmadas y las coincidencias exactas múltiples. Esta función es una de las ventajas más notables de *DejáVú*. No obstante, el programa ofrece también una amplia gama de opciones para el manejo de códigos incorporados, uso del cursor para cortar y pegar, uso del diccionario y de los glosarios disponibles, diseminación de texto repetitivo, propagación de la traducción sobre la marcha (manual o automática), creación y actualización de la base de datos terminológica, búsqueda de términos exactos o por aproximación (*Learn/Scan*), etc.

Las funciones de control de calidad también son excelentes, ya que no sólo permiten verificar la corrección ortográfica sino también la terminología utilizada, comparándola con la existente en la base de datos.

Si bien incluye funciones avanzadas que pueden prolongar la curva de aprendizaje, el traductor puede usar el producto de manera sencilla, como una herramienta para aumentar la productividad en trabajos repetitivos o relacionados entre sí. Y al mismo tiempo, el localizador de software encuentra todas las funciones sofisticadas que le permiten realizar la compleja tarea técnica de preparación de los archivos que forman parte de los proyectos de localización.

DejáVú es sin duda uno de los mejores programas CAT disponibles en el mercado. Y a un precio sumamente razonable.

Transit

Este programa, diseñado por Star, se distingue del resto porque no trabaja con una memoria de traducción ni con una base de datos de referencias. En cambio,

utiliza directamente los archivos que contienen traducciones de temas relacionados realizadas anteriormente con la aplicación. Estos archivos se pueden guardar en distintas carpetas de referencia, y hasta en una estructura que permita consultarlos según el nivel de semejanza con el texto a traducir. Es decir, *Transit* permite seleccionar el material que se desea utilizar como referencia, y luego crea un enlace para extraer los textos a utilizar, llamado *Associative Network*. La mayor ventaja de este sistema es que no debemos lidiar con el mantenimiento y los problemas inherentes a la base de datos. Y además, que podemos modificar las referencias fuente tantas veces como sea necesario, sobre la marcha.

Transit presenta una pantalla donde se pueden incluir hasta cinco paneles para el texto fuente, el texto meta, las referencias, el diccionario y las notas. Este último panel es el equivalente del bloc anotador que normalmente tenemos sobre el escritorio para garabatear todo lo que seguramente nos olvidaremos en unos pocos minutos. Claro que en este caso, los garabatos pueden encontrarse mucho más fácilmente que en el anotador y no se pueden perder. El formato de los paneles es flexible, y cada traductor lo adapta a sus preferencias personales, incluida la posibilidad de tener el diccionario a la vista o no.

La función de gestión de terminología tiene la peculiaridad de que presenta el formato de un auténtico diccionario de tapa dura. En él se pueden cargar múltiples diccionarios e incorporar los términos en la traducción con un sólo clic del ratón. *Transit* permite también incorporar automáticamente las coincidencias encontradas en los archivos de referencia para armar las traducciones, a fin de proporcionar una coincidencia total. Esto puede ser útil o no, y depende del nivel de precisión establecido para las coincidencias parciales en las opciones de configuración. En caso de que el resultado obtenido no sea satisfactorio o requiera demasiado tiempo para su corrección, se puede sencillamente borrar y sobrescribir.

El programa ofrece mayor flexibilidad en cuanto al diseño de la interfaz del usuario, y permite simular un entorno de *Word* o *WordPerfect*. Dispone también de protección para las fichas (o tags) con códigos del texto, y también permite dividir o unir los segmentos que componen las unidades de traducción.

Uno de los problemas de *Transit* es la falta de documentación que permita instalar el programa y comenzar a usarlo en un tiempo más o menos razonable. Para contrarrestar, existe un foro de discusión en el que se pueden plantear todo tipo de consultas con respecto al uso del producto. En mi experiencia, un par de horas bastan para obtener respuestas al problema más complicado.

Entre sus ventajas vale la pena destacar los filtros, que funcionan en forma transparente, sin necesidad de importar o exportar archivos a ningún otro formato, e incluyen la posibilidad de utilizar memorias de traducción creadas con *Workbench*. Funciona con la mayoría de los programas de maquetación disponibles (*FrameMaker*,

PageMaker, *Ventura*) y dispone de una función de control de calidad que incluye el corrector ortográfico, la terminología y la codificación del texto.

Transit es especialmente útil para proyectos de localización de software, traducción de páginas web, y todo otro tipo de texto con codificación interna.

Ésta ha sido una reseña muy breve de los tres programas más utilizados en este momento. Para obtener más información sobre ellos, se recomienda visitar los respectivos sitios web, en las siguientes direcciones:

<i>Atril (DéjàVù)</i>	http://www.atril.com
<i>Star (Transit)</i>	http://www.star-ag.ch
<i>Trados (Translator's Workbench)</i>	http://www.trados.com

Una última recomendación: antes de comprar cualquiera de estos productos, es conveniente transferir una demo de cada uno de ellos, disponibles en los sitios web respectivos, ya que ésta es la única forma de poder probar el producto y evaluar su conveniencia y facilidad de uso a nivel personal.