



El papel de las herramientas TAO en la documentación técnica multilingüe

Lidia Cámara, STAR Servicios Lingüísticos

Introducción

Las montañas de información digitalizada ante las que nos vemos confrontados en la actualidad y la necesidad de publicar la información en diferentes idiomas precisa de una reflexión cuyo resultado nos ayude a catalogarla y clasificarla como producto de la comunicación global multilingüe.

Las actividades determinantes de la comunicación global multilingüe más significativas son: la elaboración de documentación técnica multilingüe (DTM), la localización de productos, la realización de campañas corporativas y publicitarias a escala mundial, la creación de diccionarios especializados y la comunicación oral. Estas actividades y las relaciones que establecen entre sí dan cuenta de su complejidad y de la necesidad de sincronizarlas.

En este artículo nos centraremos en describir las diferentes características asociadas a la producción de DTM, que indudablemente comparten rasgos con todas las actividades integradas en el marco de la comunicación global.

El análisis de los problemas y desafíos vinculados a la DTM contribuirá a entender el porqué del desarrollo cada vez más sofisticado de las herramientas de traducción asistida por ordenador (TAO). Asimismo, se pondrá de manifiesto por qué estas herramientas son una solución imprescindible para cumplir con las exigencias impuestas por la producción de DTM.

La filosofía en la que se apoyan las herramientas TAO, recogida en los tres conceptos de **reutilización, integración e intercambio de información**, y que a su vez debería ser aplicada a la producción de DTM, deriva en el planteamiento fundamental de estructurar la información mediante estándares (XML, TMX, MARTIF, OLIF) a fin de lograr un denominador común.

Para finalizar perfilaremos algunas diferencias conceptuales y funcionales de las herramientas TAO con el objetivo de hacernos con unos criterios que nos ayuden a decidir qué modelo se adapta mejor a nuestras necesidades.

Exigencias y desafíos asociados a la producción DTM

"The distinction between what is product, what is marketing, and what is technical material is becoming increasingly blurred"
(Brooks: 1997)

Las necesidades y exigencias del mercado mundial se han visto modificadas en los últimos años en el marco de la Sociedad de la Información, donde se ha hecho absolutamente ineludible la adaptación de la capacidad técnica a las necesidades de la comunicación global, que precisa de la superación del uso del inglés como lingua franca incorporando otras lenguas determinantes en diferentes mercados a fin de desfronterizar la información.

La creciente producción de documentación técnica multilingüe (DTM) no es simplemente fruto de la voluntad particular de instituciones o empresas que desean estar presentes en todos los mercados mediante, por ejemplo, sitios web multilingües. En muchos casos responde a los requerimientos legislativos de comunidades de mercados internacionales.

La Unión Europea, por ejemplo, exige desde hace ya varios años que todos los productos comercializados dentro del continente estén marcados con la etiqueta CE de certificación de calidad que corresponde al cumplimiento de la normativa estipulada para cada producto. Además de los requisitos de calidad impuestos, se obliga a que todos los productos que se introduzcan en un mercado europeo contengan sus especificaciones descriptivas y técnicas en el idioma del país en cuestión.

Este nuevo panorama no sólo afecta a los países de la Unión Europea, sino también a otras comunidades económicas internacionales como los países integrantes del Tratado de Libre Comercio norteamericano (TLC) firmado por Estados Unidos, Canadá y México, o el del MERCOSUR, acordado entre los países del Cono Sur Americano (Argentina, Uruguay, Chile, Paraguay y Brasil), por citar algunos ejemplos.

Esta creciente exigencia de documentación técnica multilingüe requiere de las empresas, organizaciones e instituciones de carácter regional e internacional contar con estrategias y herramientas que ayuden a gestionar volúmenes inmensos de información, que precisan de sistematización para su posterior traducción.

Un análisis previo de los aspectos que confluyen y que afectan a la producción de DTM en la actualidad nos ayudará a situarnos ante la dificultad que ésta representa para poder perfilar más acertadamente una posible estrategia que contribuya a optimizar esa producción.

En primer lugar, la gestión corporativa de las empresas debe entender la generación de DTM como parte del desarrollo de sus productos, debido a las ya mencionadas exigencias legislativas impuestas por algunos mercados. Esto incide en los plazos previstos para el lanzamiento al mercado.

Pero, además de cumplir plazos cada vez más apretados, la gestión documental maneja volúmenes enormes de información muy repetitiva. La producción de especificaciones técnicas minuciosamente descritas cumple la normativa legislativa, además de permitir al usuario explotar todas las posibilidades funcionales de cualquier producto. No se le debe presuponer al usuario ningún conocimiento relacionado con el producto con el que debe familiarizarse. Así la información se construye siempre desde algo básico, encadenándose con el tiempo instrucciones archirrepetidas en la documentación.

Para poder manejar estos grandes volúmenes de información en un corto plazo de tiempo se requiere de una estructuración de los recursos lingüísticos, especialmente terminológicos, a fin de favorecer la coherencia del contenido para su posterior traducción.

Otro de los desafíos a los que se enfrenta la producción de DTM está relacionado con los numerosos formatos y plataformas de almacenamiento incompatibles entre sí en los que se halla almacenada digitalmente la información que debe ser tratada. La cadena de producción puede verse alterada ante formatos de documentos desconocidos por el usuario que debe realizar la traducción o simplemente porque éste carezca de los programas específicos que permiten editarla.

La convergencia de diferentes tipos de documentación y la participación conjunta de sus canales de distribución es otro de los retos planteados en el marco de la producción de DTM. En las empresas de estructura tradicional se observa que cada departamento, por ejemplo los de mercadotecnia, ventas y servicio postventa, tiene su propio tipo de documentación, su propio formato de almacenamiento y diferentes canales de distribución de la información para comunicarse con el cliente. El reaprovechamiento de la información se ve dificultado por la falta de comunicación documental entre los diferentes departamentos, que provoca problemas de armonización estilística y terminológica a veces difíciles de resolver.

Cada vez es mayor la necesidad de publicar la documentación técnica en diferentes medios: en papel, en la Red, en ayudas en línea, en CD-ROM o en microfilmes cuyo valor informativo se intersecta, lo que hace difícil su reaprovechamiento debido no sólo a los diferentes tipos de almacenamiento usados, sino también a la diferencia que existe a la hora de estructurar la información, por ejemplo en un sistema de ayuda o en una página web.

En el caso concreto de productos informáticos uno de los retos concretos a los que se enfrenta la producción de DTM es desarrollarlos de forma que las especificaciones técnicas puedan traducirse a otras lenguas para mercados diferentes sin hacer ningún cambio que afecte a la ingeniería del producto, o reduciéndolos al mínimo. Esta tarea, conocida como **internacionalización**, es una de las condiciones para la optimización de la producción de DTM de un producto, y simplifica la adaptación y traducción de los productos a un mercado específico respetando las convenciones y aspectos culturales del último, proceso conocido como **localización**.

La inmersión de aplicaciones informáticas en multitud de otros desarrollos técnicos obliga a extrapolar estas consideraciones referentes a la internacionalización de esos productos. Pensemos, por ejemplo, en los visores interactivos incorporados en los móviles de última generación o los manuales de reparación en formato CD-ROM que publican la mayoría de las multinacionales relacionadas, por ejemplo, con el mundo del automóvil.

El escenario descrito muestra a grandes rasgos la complejidad de los factores que actualmente intervienen en la producción de DTM. Esta situación ha generado la necesidad de contar con la producción simultánea de documentación y de herramientas para el tratamiento multilingüe de la documentación en todas sus facetas, desde la generación del documento hasta su manipulación multilingüe y su publicación, como se expondrá en el siguiente apartado.

Condiciones para el desarrollo de la ingeniería lingüística: las herramientas TAO

Dentro de las herramientas integradas en la ingeniería lingüística, son las herramientas TAO las que más han proliferado en los últimos 15 años, ya que son decisivas en la producción de DTM. En este sentido, no obviemos la importancia de los sistemas de tratamiento de textos que, sin lugar a dudas, fueron las primeras herramientas que se beneficiaron también de las aportaciones relacionadas con las disciplinas vinculadas a la IL (introducción de correctores ortográficos y de estilo, incorporación de módulos de sinónimos y antónimos, etc.). Sin embargo, consideramos estos sistemas como una de las aplicaciones asociadas a la Tecnología de la Información que quedan fuera del alcance del presente trabajo.

Estas herramientas se encuadran dentro de la disciplina ya conocida como ingeniería lingüística (IL, también denominada informática aplicada a la lingüística y tecnología del lenguaje), que, a su vez, tiene carácter multidisciplinar. La IL se aprovecha del conocimiento desarrollado en el marco informático del procesamiento del lenguaje natural (PLN) y del marco lingüístico nutrido por las disciplinas de la traducción, de la terminología y de la lingüística computacional, tanto en sus vertientes teóricas como aplicadas.

El auge de la ingeniería lingüística y de las aplicaciones tecnológicas asociadas, especialmente las relacionadas con la traducción asistida por ordenador (TAO) y la gestión de recursos lingüísticos multilingües, especialmente terminológicos, no es producto de la casualidad, sino el resultado de la previsión de compañías proveedoras de servicios que supieron ver el ejercicio de la traducción y de la ingeniería lingüística como un negocio muy próspero, siempre y cuando se optimizara el flujo de trabajo con herramientas como las que nos ocupan. Estas empresas reciben denominación de industrias de la lengua y sus proyectos de desarrollo reciben subvenciones institucionales de hasta un 50%. Las ayudas de investigación que absorben las industrias de la lengua son el reflejo de la necesidad también institucional de poner en marcha estrategias que resuelvan los problemas de tiempo y coste relacionados con la producción de DTM.

Ventajas de las herramientas TAO

Las herramientas TAO ofrecen una optimización de la productividad en el proceso de traducción gracias a la automatización de los procesos repetitivos, lo que permite reducir considerablemente la velocidad del flujo de trabajo. Así mismo incrementan la productividad gracias al reciclaje (reutilización) de la información ya digitalizada obtenida mediante traducciones previas y por el reaprovechamiento de la estructura y el formato de los documentos originales, generados automáticamente en las versiones traducidas.

Dependiendo del tipo de sistema TAO utilizado, puede optimizarse igualmente la calidad del producto final. Estos sistemas pueden integrar módulos y funciones cuya aplicación afecta directamente a la calidad de trabajo gracias a la mejora de la homogeneidad del estilo y la terminología en grandes volúmenes de documentos. En el apartado Comparación conceptual y funcional de las herramientas TAO se explica ampliamente la relevancia de la integración modular en un entorno al servicio de las necesidades de la traducción.

Uno de los problemas ya mencionados al describir las características actuales de la producción de DTM está relacionado con la diversidad de los formatos utilizados para su elaboración. Las herramientas TAO solucionan este problema mediante la incorporación de filtros de conversión a un entorno siempre familiar para el usuario de estos sistemas.

Los sistemas TAO facilitan el trabajo intelectual y mecánico relacionado con la traducción de una manera tan eficiente que sobrepasa las capacidades de los humanos. Esto, no obstante, no va en detrimento de la calidad final de la producción, de la que por ahora carecen las herramientas de traducción automática, ya que se reaprovecha la generada por un profesional humano.

Más allá del uso de estas herramientas en la producción de DTM, cada vez se está prestando más importancia a la gestión de memorias de traducción, debido a sus otras posibles aplicaciones. Por ejemplo, se están llevando a cabo proyectos para la realización de diccionarios bilingües basados en corpus paralelos que no son otra cosa que memorias de traducción, así como los últimos esfuerzos para la integración de estos datos en los sistemas de traducción automática con futuro, basados en corpus paralelos generados con memorias de traducción. Ya se han desarrollado interfaces que integran los sistemas TAO con los de TA. Por ejemplo, la empresa Logos de traducción automática y STAR han creado una interfaz de estas características con ayuda del sistema Transit, donde se constata que la independencia de plataformas y el uso consecuente de estándares contribuye a la mejor integración de las memorias de traducción con los sistemas TA y viceversa.

Intercambio de memorias de traducción y necesidad de estándares

Cuando describíamos las características asociadas a la DTM apuntábamos la necesidad de la multidireccionalidad de los flujos de información entre los diferentes departamentos de una empresa u organización, es decir, entre los diversos tipos de documentación y sus canales de distribución. También anotábamos que la DTM hoy en día se publica en diferentes medios: en papel, en la Red, en ayudas en línea, en CD-ROM y en microfilmes.

Es evidente que se está requiriendo una estrategia de **reutilización, integración e intercambio de información**.

Cualquier error o retraso a la hora de acceder a los datos en una corporación cuyo capital se basa en la disponibilidad de su conocimiento y en la eficiencia en la de gestión para su recuperación, como por ejemplo en una agencia de traducción, puede ser devastador. Por tanto, invertir en tecnología que garantice el reaprovechamiento de los datos ya procesados, en este caso de las memorias de traducción y de las bases de datos terminológicas con capacidad de ser intercambiados independientemente de la plataforma, resulta rentable.

Estos requerimientos se han visto satisfechos en los últimos años gracias a los inmensos desarrollos de las tecnologías de la información (TI), con su logro más espectacular: Internet. La transferencia de información cuenta por tanto con un medio revolucionario que ha supuesto un hito en la comunicación global por ser el vehículo accesible verdaderamente orientado hacia el intercambio comercial, cultural y científico a escala mundial. Para ello, se ha contado con un formato de intercambio de información HTML que no es otra cosa que una Definición de Tipo de Documento (DTD) desarrollada a partir del metalenguaje SGML.

El concepto de DTD tiene una marcación descriptiva específica de los rasgos que caracterizan a un determinado grupo de interés; es decir, que se utiliza para representar de forma intuitiva las etiquetas disponibles para el tratamiento de la información con una demarcación temática y funcional determinada.

El metalenguaje de marcas SGML es una de las respuestas a las necesidades creadas por la Sociedad de la Información. Se trata de una gramática universal para organizar la estructura general de un documento. Este metalenguaje internacional estándar para lenguajes etiquetados se incorpora como norma ISO en 1986 con el fin de administrar información públicamente sin restringirse a un modelo o fabricante determinado, ya que todos entienden esta codificación, que se reutiliza y permite optimizar el proceso de digitalización de los recursos lingüísticos.

Gracias a esta solución de lenguaje de etiquetas comunes, también se hace posible el intercambio de información de intranets y extranets facilitado por la sincronización de contenidos.

Las estrategias técnicas empleadas en la Red están en continuo crecimiento, en función de las aplicaciones cada vez más integradas que los usuarios van encontrando en este medio. Debido a esta creciente necesidad, HTML se está quedando corto para representar, por ejemplo, la información con archivos multimedia anidados. Por eso, en 1996 un grupo de trabajo bajo los auspicios del consorcio del WWW desarrolló XML, una forma simplificada de SGML, es decir, una metalengua con posibilidades abiertas de crear DTD con fines específicos. El formato XML no sólo está pensado para abarcar las necesidades de transferencia de información en la Red, sino en todos los entornos.

Así, por ejemplo, la información codificada en XML puede publicarse en cualquier formato sin pérdida de datos y sin necesidad de complejas conversiones, además de ser almacenada para su posterior manipulación.

El empleo de los formatos estándar de intercambio es la estrategia de trabajo de la Industria de la Lengua por la que se está apostando con más rigor en los últimos tres años, gracias a los resultados que ofrece en la reutilización, integración e intercambio de información.

En 1998, con el apoyo del consorcio LISA (Localisation Industry Standards Association, es decir, la Asociación para la estandarización de la industria de la localización), donde participan los fabricantes y usuarios más importantes de herramientas TAO y otros relacionados con la ingeniería lingüística, nació la especificación del formato Translation Memory eXchange (TMX). Este formato estándar de intercambio de memorias de traducción es un DTD escrito utilizando XML. Hoy en día los productos más importantes de este sector dan soporte a este estándar, aunque no todos al mismo nivel.

La gestión terminológica afecta a todo el proceso de redacción y traducción. Conscientes de ello, los miembros del consorcio LISA también trabajan en el establecimiento de un formato de intercambio de bases de datos terminológicas, denominado TBX. Este DTD, también escrito en XML, reúne el DTD estándar MARTIF (ISO 12200) para el intercambio de bases de datos terminológicas orientadas a humanos, en el que se inscribe el formato OLIF, nacido del proyecto Otelo con el fin de convertirse en estándar para el intercambio de bases de datos de lexicones orientados a la traducción automática.

La aplicación consecuente de estándares en el desarrollo de las herramientas TAO es y debe ser uno de los retos actuales no sólo de todas las empresas desarrolladoras de estos sistemas, sino también de todas las empresas que generan documentación o la traducen y adaptan a otros idiomas en beneficio de la comunicación global.

Mediante un estándar común que garantice un reaprovechamiento de todas las memorias de traducción con independencia de la herramienta utilizada para su generación y gestión, se advierte un incremento del valor capital de la información debido a su intercambiabilidad y su simplificación de manipulación, al tratarse de ficheros con el mismo tipo de formato.

El proceso total de esta producción de documentación técnica, en el que participan traductores a diferentes lenguas, correctores, técnicos especialistas de la materia, editores y diseñadores gráficos, se beneficia de estas especificaciones, consiguiendo mayor agilidad en el flujo de trabajo y haciendo más transparente e inequívoco el intercambio de información.

Mediante el uso de estándares de intercambio, en el futuro no sólo el cliente tendrá siempre mayor libertad para decidir quién gestiona, por ejemplo, las actualizaciones de sus productos, sino que las empresas de traducción o los traductores autónomos no estarán obligados a adquirir todos los sistemas de traducción asistida del mercado, sino cualquiera que dé soporte por lo menos a TMX y MARTIF.

En resumen, las empresas deben considerar su arsenal de información como un capital inherente a la propia empresa, con una capacidad de rendimiento creciente en función de lo intercambiables que sean sus recursos lingüísticos con otras plataformas, así como de lo reutilizables que sean para su publicación en diferentes soportes.

Comparación conceptual y funcional de las herramientas TAO

Aunque en principio se pudiera creer que todas las herramientas de traducción asistida son iguales porque se venden bajo la misma rúbrica, y aunque los sistemas TAO han ido paulatinamente introduciendo desarrollos cada vez más similares, es importante tener en cuenta las clases de herramientas existentes en el mercado a fin de evaluar con criterio las ventajas e inconvenientes de cada tipo.

- **Integración del editor en el sistema TAO**

El **editor de traducción** puede ser **independiente** de cualquier programa de procesamiento de textos e **integrado** en el sistema TAO, como es el caso de la mayoría de los sistemas de traducción asistida. O puede estar integrado en un procesador de textos estándar, como en el programa TAO de la empresa Trados.

- **Integración del programa de gestión terminológica en el sistema TAO**

El gestor terminológico puede hallarse totalmente integrado en el editor de traducción, como en el programa TermStar integrado en Transit. Este sistema está constantemente buscando términos de forma sincronizada con la unidad de traducción en la que se encuentra el traductor. Las búsquedas pueden ser de equivalencias para términos en la lengua de destino o multilingües. Así, por ejemplo, si el traductor está traduciendo un texto del inglés al catalán y no cuenta con un término en la lengua de destino, pero sí con el equivalente en español o en francés, puede derivar por asociación idiomática el correspondiente en catalán. El traductor no sólo puede incorporar al editor de traducción los términos encontrados en los diccionarios especializados tan sólo pulsando una tecla, sino que además puede introducir nueva terminología a golpe de ratón.

- **Motor usado por la memoria de traducción**

Con respecto al tipo de **motor usado por la memoria de traducción** existen en el mercado tres tipos de gestores de memorias de traducción: El **estático** que sólo identifica las equivalencias coincidentes totales, es decir de un 100%. El **interactivo** que identifica equivalencias parciales con coincidencia ajustable en porcentajes y la **combinación** de estático e interactivo. Es obvio que el modelo combinado es el que ofrece mejores resultados.

- **BD en comparación con el método abierto XML**

Otro de los aspectos que más puede afectar a la hora de tomar la decisión de adquirir una herramienta TAO es saber qué forma de almacenamiento utilizan las memorias de traducción, siempre y cuando se sea consciente de las ventajas y desventajas de los dos actualmente usados: el basado en base de datos, utilizado por la mayoría de los sistemas de traducción asistida, y el basado en archivos XML, usado por Transit, que además no utiliza un formato interno propio para sus archivos, sino el estándar XML con Unicode para el formato de texto.

El estándar Unicode de codificación de 16 bits es la solución para que las máquinas de todo el mundo se entiendan. Es independiente de la plataforma, del programa y del idioma usados. Los fabricantes más importantes de hardware y software (Apple, HP, IBM, JustSystem, Microsoft, Oracle, SAP y Sun) ya han adoptado este sistema de codificación, que se ha convertido, además, en el requisito para los estándares modernos tales como XML o Java.

La forma de almacenamiento utilizada puede afectar al flujo de trabajo. La posibilidad de que las bases de datos se vean alteradas, con el consiguiente riesgo de pérdida de datos, es frecuente, mientras que los archivos XML no corren ese peligro.

El modelo XML ofrece datos independientes de cualquier plataforma, lo que hace viable cualquier manipulación de estos datos para otros fines. Este lenguaje permite, a escala metasemántica, incluir datos lógicos de clasificación de los conocimientos en el flujo de datos, por ejemplo, para utilizar las memorias de traducción codificadas en XML integrándolas en otros sistemas de objetivos lingüísticos diferentes al ámbito de la traducción. Además de ser el lenguaje del futuro inmediato en Internet, como hemos explicado antes. Por tanto ofrece mayor libertad para anidar en archivos de texto paquetes de datos como pistas de audio o de vídeo, datos gráficos, etc. o vincularlos para facilitar los efectos multimedia.

La administración y gestión de las bases de datos requiere un perfil especial, ya que, por una parte, la estructuración de las bases de datos puede ser bastante compleja para un usuario, y por la otra se complica la actualización en caso de querer descartar las malas traducciones de las memorias de traducción almacenadas en bases de datos. Otra de las desventajas del modelo de bases de datos está relacionada con la imposibilidad de acceder al contexto original de las memorias de traducción almacenadas en las mismas, mientras que el modelo XML permite el acceso directo a las memorias de traducción contextualizadas.

La velocidad del flujo de trabajo es notablemente mayor con el modelo XML que con el modelo de bases de datos. Al tratarse de archivos abiertos independientes de la plataforma, pueden integrarse en otras de las tareas que repercuten directamente en todo el proceso de trabajo, como la administración y gestión.

A modo de conclusión

El análisis sistemático del entorno específico de la DTM ayuda a establecer el marco en el que se encuadra la actividad y constata la relación cada vez más estrecha con las otras actividades vinculadas a la comunicación multilingüe, así como la necesidad de aplicar la idea de **reutilización, integración e intercambio de información** en todos los ámbitos. Para poderlo llevar a cabo es necesario estructurar la información con formalismos estándar de gran flexibilidad por su independencia de plataforma y fabricante y de gran difusión. XML es la solución para homogeneizar todas las estructuras de información. TMX y MARTIF son los DTD de XML específicos que describen las

estructuras que participan en el proceso de la DTM.

Las industrias de la lengua tienden a incorporar los estándares para estructurar la información en las herramientas TAO. Estos sistemas son una de las soluciones más eficaces que dan respuesta a las necesidades actuales de DTM. La eficacia de su producción ha llevado a considerar la aplicación de las memorias de traducción como estratégica en procesos de TA o en la creación de diccionarios bilingües basados en corpus paralelos.

El grado de integración de todos los procesos que interfieren en la producción de DTM y, por tanto, de posible automatización se convertirá en uno de los retos más ambiciosos a la hora de evaluar el sistema TAO en el entorno de la producción profesional.