



La traducción del software libre

MEMORIAS COMPARTIDAS

Hablábamos en la segunda entrega del momento fácil de la enumeración destructiva de problemas. Es la hora de enfrentarnos a los que tenían que ver con la necesidad de compartir esfuerzos, herramientas y resultados.

Vamos a examinar las respuestas. **POR JUAN RAFAEL FERNÁNDEZ GARCÍA**

Hay en todo proyecto de traducción libre una tensión entre la necesidad de facilitar la colaboración de personas sin experiencia que están aprendiendo a

traducir y a usar las herramientas de traducción, y la necesidad de que las traducciones se revisen, sean correctas, coherentes en su terminología y su estilo, y de calidad. Este problema

de flujo de trabajo, que se da en cualquier grupo de traducción, se agudiza en los proyectos basados total o parcialmente en el trabajo desinteresado, en los que no se puede seleccionar

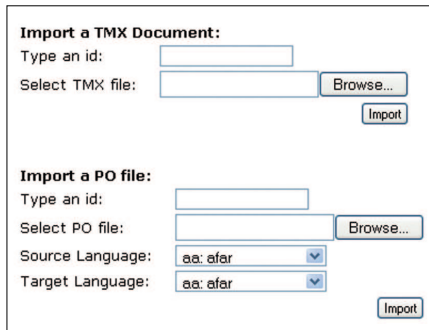


Figura 1: Subiendo memorias a un servidor tumatxa.

previamente a los participantes ni se puede exigir más allá del compromiso ético. Y hoy día, cuando el software libre se ha profesionalizado, las circunstancias son aún más apremiantes: hay que cumplir plazos, alcanzar estándares de calidad, y quizás trabajar codo con codo con traductores que participan porque es su trabajo.

¿Cómo compartir memorias?

Llevamos dos secciones refiriéndonos a las memorias de traducción sin haber realizado una presentación sistemática. Quizás sea el momento. Una memoria de traducción es un tipo especial de corpus multilingüe. Un corpus segmentado, alineado y anotado constituye una memoria de traducción.

Las memorias de traducción tienen su origen no en las investigaciones del campo de la traducción ayudada por ordenador, sino en el de la traducción automática. El origen está en la llamada traducción «por ejemplos»: parte de la idea estadística de que en un corpus suficientemente grande y representativo de un lenguaje especializado se dará un

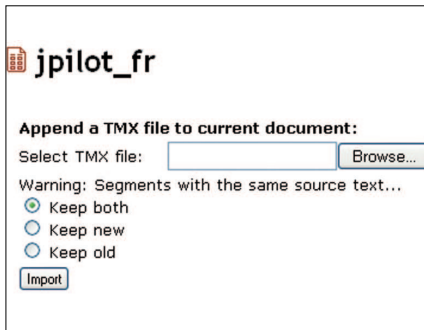


Figura 2: Fusión de memorias de traducción con tumatxa.

número significativo de reapariciones de fórmulas y expresiones. Como escribe Joseba Abaitua en «Memorias de traducción en TMX compartidas por Internet. Traducción automática en el umbral del siglo XXI: sistemas y herramientas»[1] «Son muy adecuados para textos que contengan un alto porcentaje de expresiones formulaicas y giros idiomáticos, como es el caso de los textos de especialidad. No sirven para textos creativos o expresivos, para los que de todas formas tampoco dan buenos resultados los métodos basados en reglas y requieren traducción humana».

Claro, en el mundo del software es evidente que la interfaz de usuario encaja como candidata perfecta (¿Desea usted guardar el archivo? es una expresión que se repite aplicación por aplicación); ¿qué ocurre con la documentación? Que para llegar a coincidencias significativas es imprescindible un proceso previo de segmentación y alineación adecuados.

¿Desde cuándo existen memorias de traducción en el mundo del software libre? Memorias personales, desde el principio: cada actualización



Figura 3: Edición web de segmentos en tumatxa.

de un fichero PO reutiliza la traducción anterior. Y es posible crear un compendium que reúna los ficheros PO de un usuario. Y ya hemos examinado herramientas (klabel, OmegaT) que nos permiten sacar provecho de memorias PO y TMX. Este aspecto está resuelto.

La pregunta es... ¿desde cuándo se comparten las memorias? Cuando uno se inició en el campo de la traducción con gettext, allá por el lejano siglo pasado, y a la vieja usanza, fue amablemente remitido a la lectura de la documentación, creyó ingenuamente en el mito del gran compendio de las traducciones libres. Ese compendio, pese a intentos más ingeniosos aún de crearlo, ni existe ni es posible. ¿En un solo fichero? ¿descargable y mantenido por la internet de los modems por puerto serie? ¿actualizado? ¿correcto? ¿con calidad revisada? ¿en esta vida?

La solución está en internet.

Tumatxa era demasiado... ¿avanzado?

Tumatxa ZTMX es un producto libre desarrollado por Roberto Quero (copyright de 2003; licencia GNU GPL), de la empresa vasca CodeSyntax, con la ayuda de instituciones como el Departamento de eusquera de la Diputación Foral de Guipúzcoa y la asistencia del

Cuadro 1: Porqué no usar rosetta

Porque Rosetta (<https://launchpad.net/rosetta>) no es libre. Leemos en <https://help.launchpad.net/RosettaFAQ>:

"¿Es Rosetta software abierto o libre?"

No, Rosetta no es software de fuente abierta o libre en estos momentos. Rosetta será de fuente abierta en algún momento del futuro pero no hemos puesto fecha, aunque algunas partes de la Launchpad ("pista de lanzamiento", la estructura base sobre la que se construye Ubuntu) ya han sido liberadas bajo la licencia GPL por Canonical Ltd.". ¡Qué extraña respuesta! ¿Dónde está el porqué?

Porque no es la única utilidad de traducción en línea. Quizás incluso podría ser, por ahora, la mejor, o la más utilizada (también podría ser la que tiene mejor marketing). Pero ¿no sabemos ya que es el carácter libre del software el que permite que entre todos lo mejoremos? ¿Cómo admitir que la piedra angular de un sistema libre se fundamente sobre software cerrado?

Por cierto, estoy dispuesto a reescribir todo este artículo cuando (si) rosetta se libere como software libre. Si ese día sigue siendo un proyecto vivo.

grupo DELi, de la Universidad de Deusto[2].

El origen del nombre está, como es habitual en el software libre, en una broma: según su web es la pronunciación *euskerizada* de «too much», jugando con las siglas TMX. Sería algo así como el *pasota* «demasiado» que se oía en tono laudatorio hará veinte años (no sé si este párrafo se comprenderá en latinoamérica).

Pero, ¿qué es *tumatxa*? Un gestor y un repositorio de memorias de traducción, escrito en python, y que se monta sobre un sitio Zope. Permite trabajar con TMX y con ficheros PO. Divide a los usuarios en tres categorías: visitantes anónimos, traductores (con permisos de escritura) y los llamados *managers* (se nota que es Deusto; serían los administradores del almacén de memorias).

Vamos a ser subjetivos y a hablar de impresiones personales. No hemos tenido ocasión de confirmarlo (la hipótesis es fácilmente falsable), pero siempre que repaso las ideas que sustentan la construcción de *tumatxa* no puedo dejar de hacer una asociación con el artículo de Abaitua, que debió de escribirse a finales del año 2000 y comienzos del 2001. La idea (al parecer es de Minako O'Hagan) es «convertir Internet en un inmenso depósito abierto de traducciones». Los dos avances que lo iban a permitir eran el formato TMX (que independizaba la alimentación de las memorias de la aplicación con la que se realizara) y la libre disponibilidad de los textos que sólo permiten las licencias libres.

Con *tumatxa* las memorias pueden compartirse en un entorno web, bien sea una Intranet o servidor local, o a través de la WWW. Como este humilde divulgador no administra sitio Zope ninguno, ha descargado y examinado la fuente, pero no ha instalado *tumatxa* y se limitará a glosar la documentación presente en la web y en las fuentes. En primer lugar, la imagen 1 nos muestra la interfaz para subir nuestras memorias al servidor *tumatxa* (directamente si son TMX, señalando

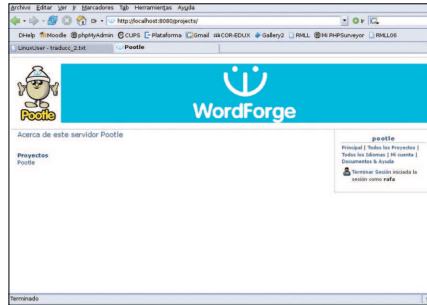


Figura 4: Instalación local de *pootle*.

las lenguas origen y destino si son PO); a partir de este momento las memorias son indexadas, y se convierten en buscables y editables. Una memoria dada puede ampliarse con nuevos segmentos (figura 2). El usuario identificado puede editar directamente los segmentos vía web (figura 3), no tiene más que pulsar sobre cualquiera de ellos. Finalmente, el sistema permite buscar y seleccionar memorias diversas, para empaquetarlas luego y descargárselas como un fichero único.

La versión descargable es la 0.1.2, de agosto de 2004. Salvo la utilización por parte del gobierno vasco, no conocemos ejemplos de uso en el mundo del software libre ni desarrollos posteriores. Un servidor estuvo (¿está?) suscrito a la lista de distribución de Tumatxa; sin embargo, el último correo recibido es también de agosto de 2004. Nos atreveríamos a concluir que la idea de *tumatxa* cuajó en tanto que idea (gestión web de repositorios compartidos), pero no su implementación.

¿Y la terminología?

Hemos hablado en la introducción de coherencia en la terminología. No es únicamente una cuestión de estilo. El problema es más grave: la terminología como disciplina no trata de describir usos, sino de determinar el uso correcto[3].

Aunque este capítulo aparezca después de analizar las memorias de traducción, el del trabajo terminológico es un proceso conceptualmente anterior (como debería ser anterior estudiar las herramientas de ayuda al trabajo del traductor en tanto que escritor: revisores orto-tipográficos,

gramaticales, de estilo...). Un equipo de traducción debe (debería) consensuar sus decisiones terminológicas en glosarios antes de proceder a traducir, y desde luego antes de consolidar las decisiones en memorias de traducción.

El estándar **TBX de LISA/OSCAR** (TermBase eXchange)[4] es el estándar abierto basado en XML (¡otra vez!) creado por la industria para el intercambio y almacenamiento de datos terminológicos. La gran ventaja que proporciona es que facilita la importación y exportación de las bases de datos terminológicas creadas con cualquier programa (libre o privativo) y, al ser abierto, garantiza la permanencia del acceso a estas bases de datos.

¿Cómo lograr que este enfoque normativo sea seguido por todos los participantes en la traducción? No ha habido hasta muy recientemente una utilización sistemática de los estándares relativos a la terminología ni un trabajo especializado en este campo dentro del mundo del software libre. Pero la consciencia del problema está ahí. ¿Quieren ejemplos de avances incipientes? La necesidad de que los programas libres, como OmegaT, puedan utilizar bases de datos terminológicas TBX se manifiesta continuamente en las distintas listas. Los glosarios de OmegaT, rudimentarios, en algún momento serán glosarios TBX. Francisco Javier Fernández Serrador, coordinador del equipo de traducción al castellano de Gnome ha redactado un documento[5] de reflexión y consenso, con contenidos normativos como los del siguiente ejemplo

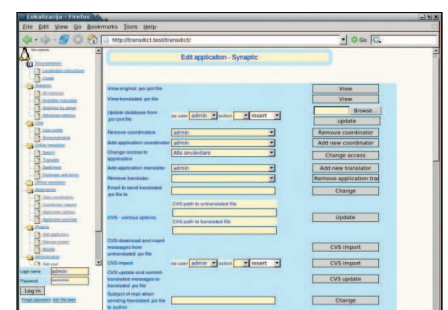


Figura 5: Transdict.

10.4. Prestablecido/**Z**
 predeterminado/por defecto/**Z**
 por omisión
 Se traduce de la forma
 siguiente:
 Inglés Español
 default predeterminado
 by default por omisión

Nunca usar «por defecto»
 (aunque es correcto, pero hace
 que las memorias de traducción
 sean poco eficientes).

Y, *last but not least*, *pootle* puede utilizar glosarios TBX (¡y los crea internamente!).

Pootle

¿Qué es pootle?[6] Pootle es (figura 4) «un portal web de ayuda para hacer más fácil la traducción». Permite la traducción en línea, sin que el traductor tenga que descargar ningún fichero (aunque es posible hacerlo), y sin que tenga ninguna necesidad de aprender a utilizar herramientas especializadas (si no quiere).

Pootle no es el único representante de esta nueva categoría de software (<http://translate.sourceforge.net/wiki/guide/tools/online> lista varias). Es una herramienta comparable a **rosetta**. El porqué no documentamos aquí *rosetta* lo explicamos en el cuadro 1 (si el agudo lector no lo adivina). Del mismo género es también interesante **transdict**[7], de Eddy Petrisor, Denis Lackovic y Robert Sedak. Al parecer es un proyecto muy vivo, aunque aún no ha alcanzado la fase estable; en la práctica no ha sido utilizado más que por el equipo de traducción croata (pronto podremos verlo alojado en Extremadura, ver la nota [11]). Hay capturas preparadas localmente por Daniel Nylander (figura 5).

Entrans [8], de Khader Abbebe y R. Vigneswaran, es otra herramienta web de traducción colaborativa de ficheros PO. Fue desarrollada específicamente para la comunidad de la India (espero traducir correctamente el *Indic* de la documentación en inglés), y tiene características, como el editor de

entrada integrado, que puede hacer su trabajo más fácil a los traductores de las lenguas de la India. La versión más reciente es la 0.3.1, de abril de 2006. En la lista de correo leo que también está siendo utilizada por equipos de Georgia.

La última versión de **kartouche**[9], de Kevin Donnelly, sería la 0.2, de 28 de octubre de 2003 (aunque es posible descargar la 0.2.3 desde Berlios). Leo en la documentación que se ha utilizado en el mundo real para la traducción de los programas de KDE al galés.

Un nuevo proyecto es **phptranslator**, de Fernando Monllor, *wunslov*[10], del Grupo de Usuarios de Linux de Alicante, que ha presentado una versión 1.0.0 en agosto de 2006.

Volvamos a *pootle*. Pongámonos el gorro de pitoniso: *pootle* es la aplicación con más futuro de su categoría. Frente a aplicaciones congeladas y a aplicaciones que llegan tarde, y en contraste con aplicaciones de oscuro desarrollo (perdón, quería decir cerrado). Señales: sigue los estándares de la industria (XLIFF, TMX, TBX); la colaboración del proyecto Debian con la fundación WordForge decidida en la Debconf6 de mayo de 2006, continuada con el proyecto *Google Summer of Code 2006* llevado a cabo por el lituano Gintautas Miliauskas «mejoras en la arquitectura del servidor Pootle: separación de la aplicación de fondo (backend, se trata del almacenamiento de las traducciones) de la interfaz (frontend)» y confirmada en la reunión de Extremadura de septiembre de 2006[11]; la existencia desde el mes de julio de 2006 de paquetes *pootle* y *translate-toolkit* en los repositorios de Debian Testing (tras su paso obligado por Inestable). En el momento de escribir estas líneas la versión disponible es la 0.10 (liberada el 30 de agosto de 2006).

Pongamos las cosas en su sitio. *Pootle* es una de las aplicaciones producidas por el Proyecto WordForge («forja de palabras»), a su vez una de las tareas que se ha impuesto la Fundación WordForge para el logro de su misión: «la

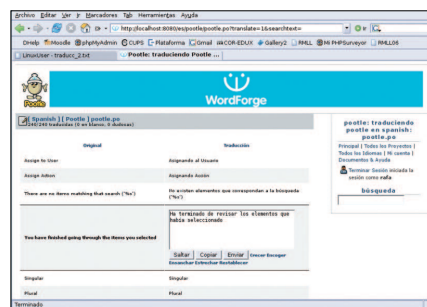


Figura 6: Interfaz de traducción de pootle.

defensa de las lenguas minoritarias que -a causa del progreso tecnológico- están siendo sustituidas poco a poco por idiomas mayoritarios en las escuelas y puestos de trabajo de todo el mundo»[12]: las lenguas minoritarias están en *peligro digital*. El Proyecto WordForge se esfuerza por simplificar el proceso de localización del software, por medio de la creación de una serie de aplicaciones: el servidor *pootle* (el nombre deriva de *PO-based Online Translation / Localization Engine*), el conjunto de filtros y conversores *translate toolkit*, etc. No obstante, y esto puede llevar a confusión, el origen de *pootle* y del *translation toolkit* está en el Proyecto Translate, utilizado en Sudáfrica por varios equipos de traducción; sólo después se generalizó el proyecto hasta incluirlo en WordForge.

Gracias al paquete *pootle* hoy es fácil instalar un servidor de traducción, basta con dedicarle un poco de tiempo a la configuración y tras un `sudo /usr/sbin/PootleServer` obtendremos la figura 6. ¡Nuestro servidor de traducciones! ¿Qué nos proporciona *pootle*?

- Un servidor de ficheros traducidos o traducibles en formato PO o XLIFF
- Un conversor a y desde formatos PO y XLIFF
- Un editor en línea de traducciones
- Una aplicación que permite compartir memorias de traducción XML (¡al fin!) y proporciona sugerencias a los traductores
- Una aplicación que permite compartir glosarios TBX y ayuda al rigor terminológico
- Una aplicación de control del proceso de traducción (cuándo,

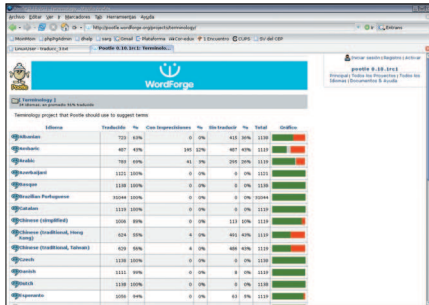


Figura 7: Estadísticas en pootle.

quién, fecha límite, errores detectados...)

- Una aplicación que proporciona estadísticas generales y detalladas del estado de las traducciones (vean en la figura 7 el estado de las traducciones de un servidor *pootle*).

En resumen, *pootle* es un sistema distribuido de gestión de traducciones. Vamos a dejar para la siguiente

entrega los distintos papeles (traductor, revisor, responsable de equipo) posibles en el sistema y cómo se enfrenta al problema del control de calidad del proceso.

Y en el próximo número...

Hay muchos detalles del campo de las herramientas de la traducción que por falta de espacio no tienen cabida aquí y que nos vamos a limitar a enumerar: los diccionarios, las herramientas de corrección ortotipográfica, gramatical y estilística, las colocaciones (cuando sea posible, es una promesa, volveremos sobre el tema)... pero hay dos cuestiones sin las cuales esta serie quedaría definitivamente coja: el formato XLIFF y un examen de la dinámica de los principales grupos de traducción al castellano pensando sobre todo en

cómo responden a las exigencias de calidad en su producción. No olvidemos para terminar que empezamos planteándonos acercar dos mundos distanciados, el de los profesores que saben idiomas y el de los traductores de software libre. Vamos a intentar establecer puentes.

EL AUTOR

Juan Rafael Fernández García fue uno de los coordinadores de Lucas/TLDP-ES, es socio de OFSET, profesor de educación secundaria y tiene una larga experiencia en la traducción y documentación del software libre. Ha sido coordinador de uno de los Centros que participan en la experiencia andaluza de integrar las TIC en la educación y actualmente trabaja como asesor de formación del profesorado.

NOTAS

- [1] Sobre Abaitua, <http://es.wikipedia.org/wiki/Usuario:JosebaAbaitua>. El artículo es fácilmente localizable en la red, por ejemplo en http://www.serv-inf.deusto.es/ABAITUA/konzeptu/ta/ehu_uda01.htm. Recordemos de la segunda entrega que ya vimos que TMX es el estándar industrial para el formato de intercambio de memorias de traducción: todos los programas *serios* de traducción deben de ser capaces de importar y exportar TMX.
- [2] La página web de Tumatxa es <http://www.tumatxa.com/>; la lista de créditos está en <http://www.tumatxa.com/tmx/credits>. Las capturas que ilustran la aplicación proceden de la web y pueden tener derechos reservados.
- [3] El Informe Final del proyecto POINTER (<http://www.computing.surrey.ac.uk/ai/pointer/report/section1.html#2>) nos aclara el concepto de terminología: «Frente a la lexicología, estudio de los términos en general, la terminología es el estudio de las palabras o términos de los lenguajes especiales relacionados con áreas particulares de conocimiento especializado (...) en las terminologías la prescripción (también llamada normalización o estandarización) desempeña un papel esencial». En conclusión, citando a Martínez de Sousa, sub voce terminología, en el Diccionario de lexicografía práctica, ed. Biblograf 1995: «Hoy la terminología es una ciencia bien estructurada que se ocupa en crear los catálogos léxicos propios de las ciencias, las técnicas, los oficios, etc., partiendo de sistemas coherentes establecidos por organismos nacionales e internacionales».
- [4] Ver <http://www.lisa.org/standards/tbx/>. Sobre los estándares terminológicos, el mensaje de Kara Warburton al foro de discusión sobre «localización» y terminología de LISA <http://www.lisa.org/signs/phpBB/viewtopic.php?topic=69&forum=1&1>.
- [5] El documento de Fernández Serrador está publicado en http://www.openshine.com/Members/serrador/gnome_110n_es.pdf/download y se comunicó a la lista de localización de Gnome <traductores@es.gnome.org> el 5 de julio de 2006.
- [6] Pootle y el Translate Toolkit (translate.sourceforge.net) son parte del proyecto WordForge. Existe también un wiki: <http://translate.sourceforge.net/wiki/> y unas especificaciones: http://translate.sourceforge.net/wiki/wordforge/functional_specificaions.
- [7] El proyecto se aloja en <http://sourceforge.net/projects/transdict> y no tiene una página web internacionalizada. Es posible descargarlo por *subversion*. Capturas en http://sourceforge.net/project/screenshots.php?group_id=147229.
- [8] <http://sourceforge.net/projects/entrans/>.
- [9] Al parecer el proyecto ha pasado de <http://110n.kde.org/tools/kartouche/> a alojarse en <https://developer.berlios.de/projects/kartouche>. En el momento de escribir este artículo la aplicación está en <http://www.dotmon.com/kartouche/kartouche-v0.2.3.tar.gz>.
- [10] <http://sourceforge.net/projects/phptranslator>. Sitio de wunslov: <http://www.wunslov.com/>. Hay también hilo en Barrapunto, de julio de 2006 (<http://barrapunto.com/article.pl?sid=06/07/03/169217>).
- [11] Como se ha informado en <http://lists.debian.org/debian-i18n/2006/09/msg00114.html> la llamada *Debian i18n task force*, el núcleo del grupo de internacionalización de Debian, y la Junta de Extremadura (representada por César Gómez Martín) acordaron dedicar un servidor a las actividades de internacionalización de Debian. La máquina aloja un servidor Pootle y entornos *chrooted* para la instalación de software alternativo (como el mencionado servidor *transdict*).
- [12] Información obtenida en <http://www.wordforge.org/drupal/>. Proyecto Translate: <http://translate.sourceforge.net/>.