

# La traducción asistida por ordenador

## Índice

1. Introducción.....	1
1.2. Componentes básicos de un sistema de traducción asistida.....	4
1.2.a Entorno de trabajo.....	4
1.2.b. Consulta a memorias de traducción.....	6
1.2.c. Consulta a glosarios y diccionarios.....	7
1.2.d. Combinación de traducción asistida y traducción automática.....	8
1.2.e. Tratamiento de formatos.....	8
1.3. El proceso de traducción con un sistema de traducción asistida.....	10
1.3.a. Tratamiento del formato.....	10
1.3.b. Segmentación: el formato SRX.....	10
1.3.c. Formatos de proyectos de traducción: el formato XLIFF.....	14
1.3.d. Asignación de recursos a un proyecto de traducción.....	15
1.3.e. Traducción.....	15
1.3.f. Revisión dentro de la herramienta de traducción asistida.....	16
1.3.g. Creación de los documentos traducidos.....	16
1.3.h. Revisión en formato final.....	16
1.3.i. Gestión de los recursos generados.....	17
1.4. Los sistemas de traducción asistida.....	18
1.5. Conclusiones.....	18
1.6. Para ampliar conocimientos.....	19
1.6.1. Herramientas de traducción asistida on-line.....	19
1.6.2. SRX y las expresiones regulares.....	20
1.6.7. El corrector gramatical LanguageTool.....	22
Bibliografía.....	24
Licencia de este documento.....	24

## 1. Introducción

La informática se ha introducido plenamente en todos los sectores productivos y actividades profesionales. La traducción no es una excepción y desde hace ya años el proceso de traducción se lleva a cabo casi siempre utilizando un ordenador. Muy a menudo el uso del ordenador se limita a aplicaciones ofimáticas estándar, especialmente procesadores de textos. Las consultas a fuentes de información habituales en la traducción (diccionarios generales y terminológicos, enciclopedias, etc) se hace ya también de forma generalizada a través de un ordenador, ya sea mediante la consulta de recursos instalados en la máquina o mediante la consulta a sitios webs específicos. El ordenador se ha convertido también el medio principal de comunicación entre el traductor y sus clientes, especialmente mediante el correo electrónico. Pero relevar al ordenador a ser un simple sustituto de una máquina de escribir, un diccionario y un servicio de correos eficiente es menospreciar sus enormes posibilidades.

Los *programas de traducción asistida por ordenador* engloban una serie de aplicaciones informáticas especialmente diseñadas para asistir de manera eficiente al traductor en su tarea.

En un sentido amplio, los sistemas de traducción asistida engloban todas las aplicaciones informáticas diseñadas para tareas específicas del proceso de traducción.

En un sentido más específico, generalmente se habla de sistemas de traducción asistida cuando nos referimos a los programas que ayudan a traducir a partir de consultas a una o varias memorias de traducción y de manera opcional a uno o más glosarios terminológicos. Tanta importancia tiene el concepto de memoria de traducción que a menudo se ha denominado a los sistemas de traducción como memorias de traducción o bien sistemas de gestión de memorias de traducción.

Lippmann (1971) hacía una descripción pionera del concepto de traducción asistida por ordenador:

*Computer-aided translation (CAT) is a storage and retrieval operation carried out on line with a computer during the time in which a translation is produced. A system of dictionary access and updating routines, text-processing facilities, and on-line utilities is designed to telescope the delay between the initiation of a translation and its finished print out. The system does not attempt to simulate the human translator by producing an autonomous translation via programmed algorithms; rather, it serves as an extension of the capabilities of the user, who is able to call on the resources of the computer as needed in the process of translation and get an immediate response. Under the system described, users communicate over ordinary telephone lines with the computer by means of remote terminals. In employing the system, the user can switch back and forth as many times as required among human translation, direct dictionary look up, editing, printing, and system interrogation, and thereby ACHIEVE rapid interaction toward the desired goal, a finished translation.*

Así, uno de los objetivos principales de un sistema de traducción asistida es poner al alcance del traductor de manera automática y rápida todos los recursos que le puedan resultar de utilidad. Para hacer su tarea un traductor habitualmente consulta diccionarios generales, terminológicos y enciclopédicos. Estos diccionarios pueden estar en papel y pueden constituir varios volúmenes. Esto hace que la consulta manual de estos recursos pueda resultar muy costosa en términos de tiempo. Un recurso muy interesante, pero a la vez difícil de gestionar de manera manual, son los ejemplos de traducciones anteriores, ya sean realizadas por el mismo traductor como por otro profesional. A menudo, cuando se traduce una determinada oración se tiene la sensación de ya haberla traducido anteriormente. El hecho de disponer de un registro de traducciones accesible de manera fácil y rápida puede suponer un ahorro de tiempo importante. Los sistemas de traducción asistida, como veremos más adelante, nos dan un acceso automático e inmediato a todos estos recursos.

El proceso de traducción con un sistema de traducción asistida se divide de manera genérica en los siguientes pasos:

- **Tratamiento de formato:** el sistema de traducción asistida nos permitirá crear proyectos de traducción a partir de archivos en diferentes formatos. De este modo, con una única herramienta podemos traducir archivos en diferentes formatos sin la necesidad de disponer un gran conjunto de herramientas específicas.
- **Segmentación:** la traducción del texto original se lleva a cabo en pequeñas unidades. Estas unidades se denominan *segmentos* y en general se puede asimilar un segmento a una única

oración (como veremos más adelante no siempre coincide un segmento con una oración). La finalidad de esta división en segmentos o *segmentación* es básicamente la consulta a memorias de traducción. La probabilidad de encontrar una unidad similar a la unidad que estamos traduciendo es inversamente proporcional al tamaño de esta unidad. Por lo tanto, es más probable encontrar una oración similar en la memoria que un párrafo entero.

- **Asignación de recursos:** una vez creado el proyecto de traducción, y con el fin de sacar provecho de todas las posibilidades de la herramienta de traducción asistida, es imprescindible asignarle una serie de recursos, principalmente glosarios terminológicos y memorias de traducción.
- **Traducción:** este paso se realiza también segmento a segmento y la herramienta muestra la información relevante (coincidencias exactas y parciales en las memorias de traducción, entradas de los glosarios presentes en el segmento a traducir, etc.) de una manera clara.
- **Creación de recursos a medida que se va traduciendo:** durante la traducción de un proyecto es un buen momento para crear nuevos recursos. Las memorias de traducción se crean de manera totalmente automática durante la traducción. Mientras vamos traduciendo podemos incluir nuevas entradas en los glosarios terminológicos. De esta manera nos ahorraremos nuevas consultas en el futuro.
- **Revisión dentro de la herramienta de traducción asistida:** Estas herramientas disponen de diferentes funciones para asegurar la calidad de la traducción. Desde correctores ortográficos y gramaticales hasta la comprobación de etiquetas y marcas de formato. Es importante hacer una primera revisión de la traducción dentro de la herramienta de traducción asistida, antes de la creación de los documentos traducidos.
- **Creación de los documentos traducidos:** De igual manera que la herramienta es capaz de tratar varios formatos para introducirlos en un proyecto de traducción, también es capaz de generar los documentos traducidos en el mismo formato original, y manteniendo en un buen grado la maquetación del archivo.
- **Revisión de los documentos traducidos en su formato original:** Es muy importante hacer una última revisión de la traducción una vez exportada a su formato original.

## 1.2. Componentes básicos de un sistema de traducción asistida

### 1.2.a Entorno de trabajo

Entendemos por entorno de trabajo de un sistema de traducción asistida su interfaz gráfica. Uno de los aspectos más importantes del entorno de trabajo es la disposición de la información en la pantalla. La mayoría de aspectos de esta disposición son personalizables, pero se pueden distinguir algunos grupos generales. La primera distinción se puede hacer entre:

- Programas que disponen de una interfaz de trabajo propia. En este grupo podemos clasificar un gran número de herramientas de traducción asistida: OmegaT, Virtaal, SDL, Déja Vu, WordFast Profesional, etc. En la figura 1 podemos observar la interfaz gráfica de Virtaal.

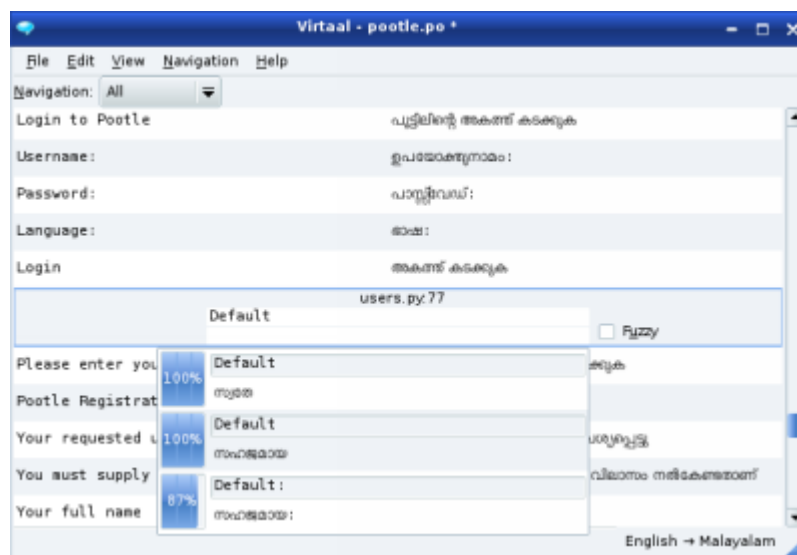


Figura 1: Virtaal

- Programas que se integran en otro programa, generalmente un procesador de textos: Anaphraseus (que se integra dentro de OpenOffice), WordFast Classic y Trados WorkBench (que se integran en Microsoft Word). En la figura 2 podemos ver Anaphraseus integrado en OpenOffice.

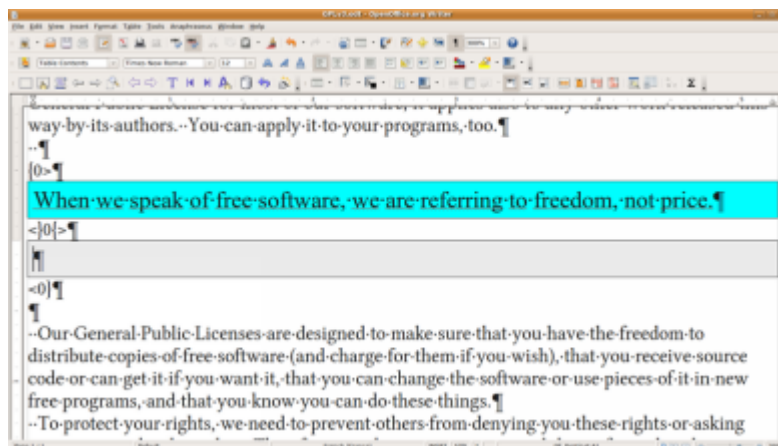


Figura 2: Anaphraseus

Dentro del primer grupo, es decir, de los programas que disponen de una interfaz propia, podemos distinguir diferentes subgrupos atendiendo a la disposición de los segmentos originales y traducidos. Así, podemos encontrar los siguientes subgrupos:

- Segmento traducido bajo el original: como por ejemplo, en OmegaT. Muy a menudo, en esta disposición, los segmentos no activos sólo se muestran en un idioma (generalmente el original si el segmento no está traducido y la traducción en caso de haberse traducido). El segmento que estamos traduciendo se muestra en las dos versiones. En la figura 3 podemos ver la interfaz de OmegaT.

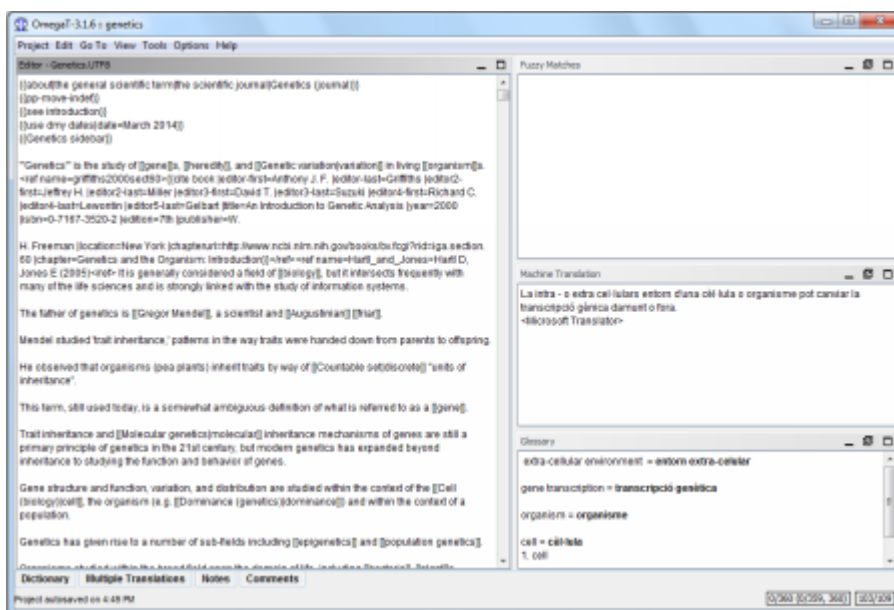


Figura 3: OmegaT

- Disposición en dos columnas: donde generalmente la izquierda corresponde al original y la derecha a la traducción. En la figura 4 podemos ver la interfaz de Déjà Vu X2.

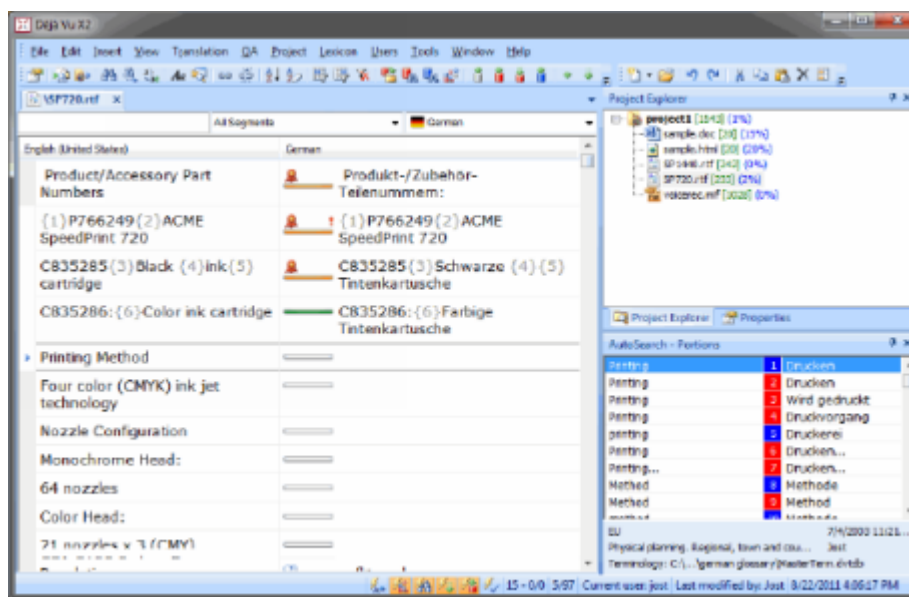


Figura 4: Déjà Vu X2

Además de la disposición del segmento original y traducido también es muy importante la disposición de las ventanas de consulta a memorias de traducción y bases de datos terminológicas. En general, la posición y tamaño de estas pantallas es totalmente configurable. El usuario puede modificar el tamaño y la posición de estas ventanas para ajustarlas a sus necesidades y preferencias.

### 1.2.b. Consulta a memorias de traducción

El sistema de traducción asistida debe ser capaz de hacer consultas a una o más memorias de traducción. El objetivo de la consulta es encontrar segmentos de la memoria que sean iguales o similares al segmento que estamos traduciendo. Si encuentra un segmento igual estamos hablando de una *coincidencia exacta* (*exact match*) y si el segmento es similar hablamos de *coincidencias parciales* (*fuzzy match*).

Esta búsqueda de segmentos similares se hace tanto con los segmentos almacenados en la memoria de traducción como en los segmentos traducidos anteriormente dentro del mismo proyecto. En este caso hablamos de *repeticiones internas*.

El usuario puede ajustar la similitud mínima que debe tener un segmento para que aparezca como coincidencia parcial. El sistema mostrará todas las coincidencias que tengan una similitud igual o superior a la mínima establecida por el usuario, y en orden descendente de similitud, es decir, empezando por la más parecida. Los sistemas de traducción asistida, además, mostrarán en colores las diferencias entre el segmento que estamos traduciendo y lo que aparece en la memoria (Figura 5). Algunos sistemas, además, intentarán combinar fragmentos más pequeños presentes en la memoria para hacer una propuesta de traducción a partir de las traducciones de estos segmentos.

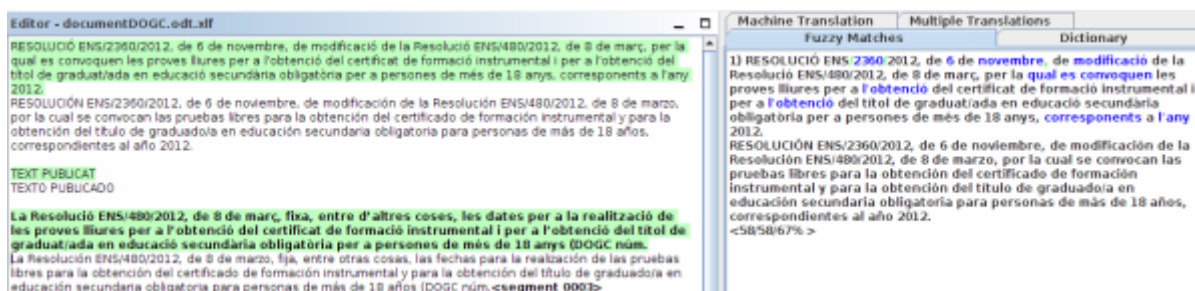


Figura 5. Coincidencias parciales en OmegaT

En el capítulo siguiente, dedicado íntegramente a las memorias de traducción veremos a fondo todos estos aspectos y muchos otros, como el concepto de indexación de las memorias para una búsqueda rápida.

### 1.2.c. Consulta a glosarios y diccionarios

El sistema de traducción asistida también debe ser capaz de hacer consultas a glosarios y diccionarios terminológicos. En este caso, cada segmento que estamos traduciendo puede tener más de una entrada del glosario y se mostrarán todos los resultados. La principal dificultad para la búsqueda a glosarios es hacer que sea capaz de encontrar las entradas de manera independiente a la forma morfológica en la que se encuentre en el segmento original. Es decir, debe ser capaz de encontrar la entrada *allergic reaction* tanto si el texto se encuentra en singular como en plural (*allergic reactions*). Para lenguas morfológicamente más ricas, como por ejemplo el castellano, la tarea se complica, ya que a veces el plural implicar pluralizar todos los elementos (*reacción alérgica - reacciones alérgicas*) y en otros casos sólo uno de ellos (letra de cambio - letras de cambio). En lenguas con una morfología mucho más rica todavía el término puede aparecer en el segmento en muchísimas formas (механическое напряжение, механического напряжения, механические напряжения, механических напряжений ...) y todas deberían mostrar la entrada correspondiente del glosario. Las diferentes herramientas de traducción asistida pueden tener más o menos habilidad en tratar las variantes morfológicas de los términos. En el capítulo 3, dedicado a las bases de datos terminológicas profundizaremos en este tema. En la Figura 6 podemos ver la pantalla de terminología de OmegaT. Si nos fijamos en los resultados, a pesar de disponer del término catalán *prova lliure* no se ha detectado automáticamente ya que en el original aparece en plural (*proves lliures*).

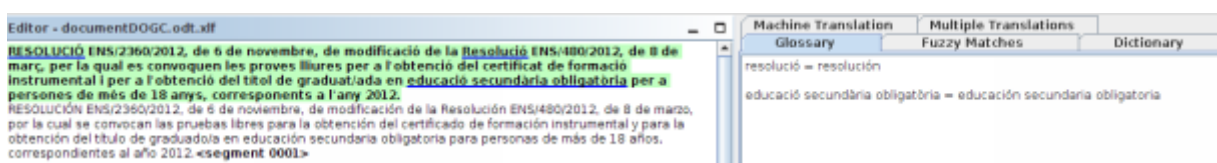


Figura 6. Pantalla de terminología de la herramienta OmegaT

Algunas herramientas de traducción asistida permiten también la consulta automática a diccionarios generales. Esta funcionalidad no es tan interesante para un traductor profesional pero en ciertos casos puede resultar de utilidad. Como la consulta es totalmente automática, si se dispone de esta funcionalidad y de un diccionario adecuado puede ser buena idea activar esta opción en nuestros proyectos. En la figura 7 podemos observar esta funcionalidad en OmegaT. Dado que la información que aparece es muy extensa, si hacemos doble clic en una determinada palabra, OmegaT muestra directamente la información asociada a esa palabra.

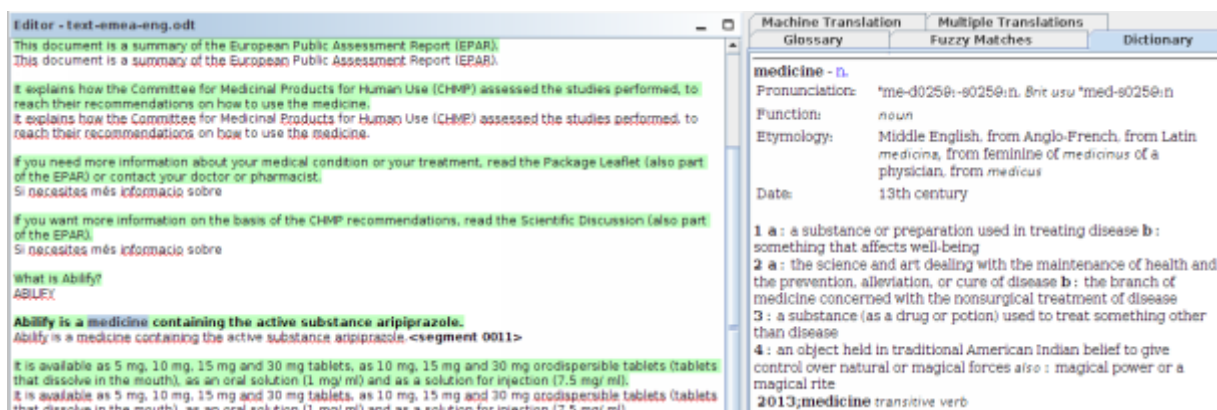


Figura 7. Consulta automática a un diccionario en OmegaT

### 1.2.d. Combinación de traducción asistida y traducción automática

Muchos sistemas de traducción asistida permiten hacer consultas a sistemas de traducción automática, por lo que además de presentar los resultados provenientes de una memoria de traducción presentan también el resultado de traducir el segmento con un sistema de traducción automática. Esta consulta a sistemas de traducción automática puede llegar a ser muy productiva para algunos pares de lenguas, ya que buena parte de las propuestas se pueden aprovechar haciendo algunos cambios mínimos. Esto suele suceder para pares de lenguas suficientemente cercanas, como por ejemplo catalán-castellano, castellano-francés, etc. Si se utiliza esta opción hay que tener mucho cuidado, ya que por las prisas tendremos tendencia a aceptar como totalmente buenas algunas propuestas de traducción que no son del todo correctas.

El capítulo 4 de este libro lo dedicaremos en su totalidad a la traducción automática y volveremos a hablar con más detalle de esta combinación. En la figura 8 podemos observar la pantalla de Traducción automática de OmegaT en funcionamiento.

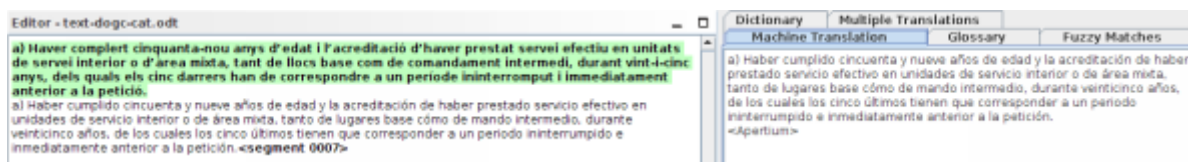


Figura 8. Traducción automática en OmegaT

### 1.2.e. Tratamiento de formatos

Los sistemas de traducción asistida deben ser capaces de tratar un conjunto variado de formatos de archivo. Así, deben ser capaces de importar archivos correspondientes a procesadores de texto (Word, OpenOffice-LibreOffice), html, archivos propios de lenguajes de programación, etc. El proceso de importación consistirá en seleccionar y mostrar sólo el texto que hay que traducir y ocultar el correspondiente a marcas de formato. Una vez traducido y revisado el proyecto, el programa debe ser capaz de exportar la traducción, es decir, crear los archivos traducidos que deben estar en el mismo formato que el original.

Así pues, la herramienta de traducción asistida nos permitirá trabajar en un mismo entorno con diversos formatos que requieren habitualmente programas de edición específicos. En la tabla 1 ofrecemos algunos ejemplos de herramientas y formatos que pueden tratar<sup>1</sup>.

<sup>1</sup> La información se ha extraído de la Wikipedia [[http://en.wikipedia.org/wiki/Computer-assisted\\_translation](http://en.wikipedia.org/wiki/Computer-assisted_translation)] y puede no estar del todo actualizada.



Herramienta	Formatos
OmegaT	Plain text, HTML, XHTML, StarOffice, OpenOffice.org, OpenDocument (ODF), MS Office Open XML, Help & Manual, HTML Help Compiler (HCC), LaTeX, DokuWiki, QuarkXPress CopyFlow Gold, DocBook, Android Resource, Java Properties, Typo3 LocManager, Mozilla DTD, Windows RC, WiX, ResX, INI files, XLIFF , PO , SubRip Subtitles, SVG Images
Open Language Tools	XLIFF , HTML/XHTML, XML, DocBook SGML, ASCII, StarOffice/OpenOffice/ODF, PO , .properties, .java (ResourceBundle), .msg/.tmsg (catgets)
SDL Trados	Features four translation environments: dedicated TagEditor, MSWord Interface, SDLX, the integrated interface SDL Trados Studio 2011. Additional filters for translating with TagEditor available: Word, Excel, PowerPoint, OpenOffice , InDesign, QuarkXPress, PageMaker, Interleaf, Framemaker, HTML, SGML, XML, SVG, .... Includes SDL MultiTerm for terminology management and Project Management Dashboard for automating tasks and tracking.
Virtaal	XLIFF, PO and MO, TMX, TBX, Wordfast TM, Qt ts Many others via converters in the Translate Toolkit
WordFast Classic	MS Word, Excel, PowerPoint (for Windows and Mac); tagged documents

Tabla 1. Formatos tratados por diversos sistemas de traducción asistida

El capítulo 5 de este libro lo dedicamos íntegramente a aspectos técnicos relacionados con el tratamiento de formatos.

## 1.3. El proceso de traducción con un sistema de traducción asistida

En este capítulo explicaremos el proceso genérico de traducción de un documento o proyecto de traducción con un sistema de traducción asistida. Aprovecharemos en este capítulo para hablar en profundidad de algunos aspectos que no quedan recogidos en otros capítulos: la segmentación y el formato SRX, el formato XLIFF para el intercambio de proyectos de traducción y localización y algunos aspectos relacionados con los correctores ortográficos y gramaticales. Aunque estos aspectos no se tratan en otros capítulos del libro, se puede encontrar información adicional en los diferentes enunciados de las prácticas asociadas a este libro.

### 1.3.a. Tratamiento del formato

En nuestro proyecto de traducción deberemos tratar uno o más ficheros que estarán en uno o más formatos informáticos. En la mayoría de los casos, la herramienta de traducción asistida será capaz de tratar los formatos requeridos sin problemas, así que este paso suele ser totalmente transparente para el traductor y lo lleva a cabo sin más preocupaciones.

En algunas ocasiones, nuestra herramienta de traducción asistida no será compatible con el formato del archivo que tenemos que traducir. Un caso habitual es el tratamiento del formato doc de Microsoft Word. No todas las herramientas de traducción asistida pueden importarlo. Por ejemplo, OmegaT no es compatible con este formato. Hay que recordar que es un formato propietario y que además las herramientas que sí son compatibles con doc suelen requerir que tengas el Microsoft Word instalado en el sistema, condición que no se da siempre. En casos como estos, la solución es sencilla y funciona en la mayoría de ocasiones: transformar el archivo doc en un archivo compatible (como el formato ODF empleado para LibreOffice y OpenOffice y considerado como estándar por la ISO, o bien, el docx de Microsoft). La conversión suele funcionar sin problemas. Una vez finalizado el proceso de traducción se podrá hacer la conversión inversa sin mayores inconvenientes.

En otros casos más complejos, es posible que no exista un formato intermedio disponible para nuestra herramienta de traducción asistida. En muchos de estos casos la solución pasará por la utilización del formato XLIFF que se explica en este mismo capítulo en la sección 1.3.c. Muy probablemente existirá algún programa que pueda transformar el formato en cuestión en este formato estándar, el XLIFF, que es compatible con la gran mayoría de herramientas de traducción asistida.

### 1.3.b. Segmentación: el formato SRX

El proceso de segmentación consiste en dividir el texto de entrada en unidades de un tamaño adecuado para poder presentarlas una tras otra al traductor. Las consultas a los diferentes recursos, como por ejemplo las memorias de traducción se harán en estas unidades de texto. No conviene que estas unidades sean demasiado largas (por ejemplo un párrafo entero) porque la probabilidad de encontrar fragmentos iguales o similares a la memoria de traducción es menor si el fragmento es largo. Tampoco conviene que las unidades sean demasiado cortas, por ejemplo, una o dos palabras, ya que la unidad y coherencia del texto se rompería y haría imposible su traducción. Así, la longitud ideal suele ser algo parecido a una frase u oración. En cada una de estas unidades se les denomina *segmento* y al proceso de crearlos

*segmentación*. Ahora bien, las herramientas de traducción asistida no suelen incorporar demasiado conocimiento lingüístico y por este motivo la segmentación se lleva a cabo a partir de elementos textuales, como signos de puntuación, presencia de caracteres en mayúsculas, etc. Lo primero que se nos ocurre es segmentar por puntos ".", Pero esto no siempre funciona. Veamos el siguiente ejemplo:

*El Sr. Martínez vendrá con el A. V. E. de las 15.30 h. Después se reunirá con el Dr. Pérez en nuestro despacho de la av. Diagonal.*

Como podemos observar, una segmentación basada únicamente en puntos "." Produciría una gran cantidad de segmentos.

*El Sr.*

*Martínez vendrá con la A.*

*V.*

*E.*

*de las 15.*

*30 h.*

*Después se reunirá con el Dr.*

*Pérez en nuestro despacho de la av.*

*Diagonal.*

Las reglas de segmentación se suelen definir mediante expresiones regulares que definen puntos de posibles cortes de segmentos y especifican si se debe producir el corte o no.

Una expresión regular (o coloquialmente llamadas *regexp*, acrónimo del inglés *regular expression*) es una representación, según unas reglas sintácticas de un lenguaje formal, de una porción de texto genérico a buscar dentro de otro texto, como por ejemplo unos caracteres, palabras o patrones de texto concretos. [Fuente Wikipedia]

Existe un lenguaje XML estándar para la definición de reglas de segmentación: el formato SRX (*Segmentation Rule eXchange*). A continuación mostramos un SRX muy simple, que sólo define un par de reglas.

```
<?xml version="1.0" encoding="UTF-8"?>
<srx xmlns="http://www.lisa.org/srx20" xmlns:okpsrx="http://okapi.sf.net/srx-extensions" version="2.0">
<body>
<languagerules>
<languagerule languagerulename="default">
<rule break="no">
<beforebreak>([A-Z]\.){2,}</beforebreak>
<afterbreak>\s</afterbreak>
</rule>
<rule break="yes">
<beforebreak>\.</beforebreak>
<afterbreak>\s</afterbreak>
</rule>
</languagerule>
</languagerules>
</body>
```

La primera de las reglas:

```
<rule break="no">
<beforebreak>([A-Z]\.){2,}</beforebreak>
<afterbreak>\s</afterbreak>
</rule>
```

especifica que un conjunto de dos o más letras mayúsculas seguidas de un espacio forman un punto donde no se producirá un corte de segmentos. En cambio, la segunda regla:

```
<rule break="yes">
<beforebreak>\.</beforebreak>
<afterbreak>\s</afterbreak>
</rule>
```

especifica que un punto seguido de un espacio sí forman un punto donde se ha de producir un corte de segmento.

Con nuestro conjunto de dos reglas ahora la segmentación de nuestra frase se haría de la siguiente manera:

*Sr.*

*Martínez vendrá con la A. V. E. de las 15.30 h.*

*Después se reunirá con el Dr.*

*Pérez en nuestro despacho de la av.*

*Diagonal.*

Para poder segmentar correctamente nuestro texto, deberíamos incorporar algunas reglas. Si añadimos como primeras reglas:

```
<rule break="no">
<beforebreak>\b(Sr|Dr|Av)\.</beforebreak>
<afterbreak>\s</afterbreak>
</rule>
<rule break="yes">
<beforebreak>\bh\.</beforebreak>
<afterbreak>[A-Z]+</afterbreak>
</rule>
```

Ahora sí que conseguiremos un texto segmentado correctamente. Hay que tener en cuenta que el orden de las reglas es significativo, ya que si una regla ha producido un corte de segmento, una regla posterior no lo deshará. *Ratel*, una de las herramientas que se distribuye con el paquete *Okapi* (<http://okapi.opentag.com>), permite editar de manera fácil reglas de segmentación y probarlas sobre fragmentos de texto. En la figura 9 podemos ver una captura de pantalla de esta herramienta.

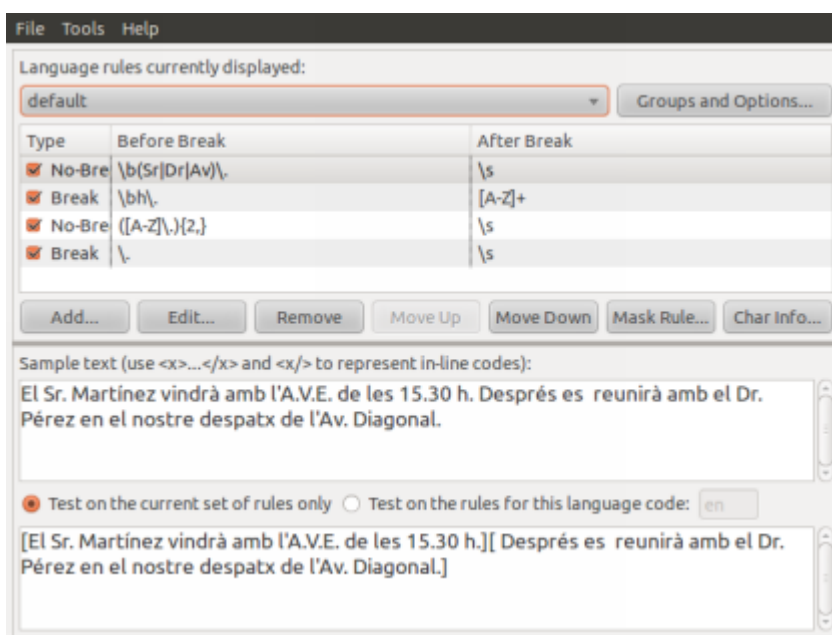


Figura 9. Ratel de Okapi: herramienta para la edición de reglas SRX

Afortunadamente, ya existen conjuntos de reglas para diferentes lenguas y el traductor deberá intervenir muy poco sobre éstas. Sólo en casos especiales deberá añadir o eliminar alguna regla.

La importancia del formato SRX es que permite un intercambio rápido de las reglas. Para un mayor aprovechamiento de las memorias de traducción, conviene que las reglas de segmentación que utilizamos para crear un proyecto sean las mismas que las reglas que se utilizaron para crear el proyecto que ha generado la memoria de traducción que queremos utilizar. De este modo, la probabilidad de encontrar segmentos coincidentes aumenta significativamente.

### 1.3.c. Formatos de proyectos de traducción: el formato XLIFF

Cada herramienta de traducción asistida tiene un formato propio para almacenar los proyectos de traducción. Algunas herramientas almacenan los proyectos como bases de datos, donde en una determinada tabla guardan los segmentos a traducir y los segmentos traducidos, en otra tabla la memoria de traducción ya indexada, etc. Otros formatos incluyen también estructura de carpetas y subcarpetas, donde en cada carpeta se almacena cierta información: en una los documentos traducidos, en otra las memorias de traducción, en otra las bases de datos terminológicas, etc

Esta multiplicidad de formatos dificulta que un traductor con una determinada herramienta pueda traducir proyectos de traducción creados con otra herramienta. Existe un formato estándar para el intercambio de proyectos de traducción o localización: el XLIFF (*XML Localisation Interchange File Format*). Este formato también está basado en XML. A continuación podemos ver un fragmento de un proyecto de traducción en formato XLIFF:

```
<?xml version="1.0" encoding="UTF-8"?>
<xliff version="1.2" xmlns="urn:oasis:names:tc:xliff:document:1.2" xmlns:okp="ok
api-framework:xliff-extensions">
<body>
<trans-unit id="1" restype="x-text:p">
<source xml:lang="en">Only free resources available online have been used.</source>
<target xml:lang="es">Únicamente se han utilizado recursos libres disponibles en
Internet.</target>
</trans-unit>
</body>
</xliff>
```

Como podemos observar, se trata de un proyecto de traducción que sólo tiene un segmento en inglés que ya está traducido al castellano. Si el proyecto estuviera sin traducir el segmento correspondiente al target estaría vacío, o bien tendría copiado el segmento original.

La mayoría de herramientas de traducción asistida pueden traducir documentos XLIFF, aunque no todas las herramientas son capaces de crear estos archivos. Se puede crear fácilmente proyectos de traducción en formato XLIFF con la herramienta *Rainbow* de *Okapi Tools* (<http://okapi.opentag.com>). El *Translate*

*Toolkit* (<http://translate.sourceforge.net>) también proporciona herramientas para crear proyectos de traducción en este formato.

#### 1.3.d. Asignación de recursos a un proyecto de traducción

Una vez creado un proyecto hay que dotarlo de los recursos de consulta disponibles, principalmente memorias de traducción y bases de datos terminológicas. Traducir con un sistema de traducción asistida sin recursos de poco sirve, aunque no es del todo inútil tampoco. Normalmente, cuando empezamos a utilizar sistemas de traducción asistida no disponemos de recursos de consulta. En este caso el sistema sólo será capaz de ayudarnos si dentro del proyecto hay alguna repetición, lo que se conoce como *repeticiones internas*. Además, una vez terminado el proyecto ya habremos generado nuestra primera memoria de traducción. Si además aprovechamos para ir recopilando la terminología a medida que la vamos encontrando en el texto, también iremos confeccionando nuestra primera base de datos terminológica que podremos utilizar en proyectos futuros. En la figura 10 podemos ver la pantalla que permite añadir nueva terminología a medida que traducimos a la herramienta OmegaT.

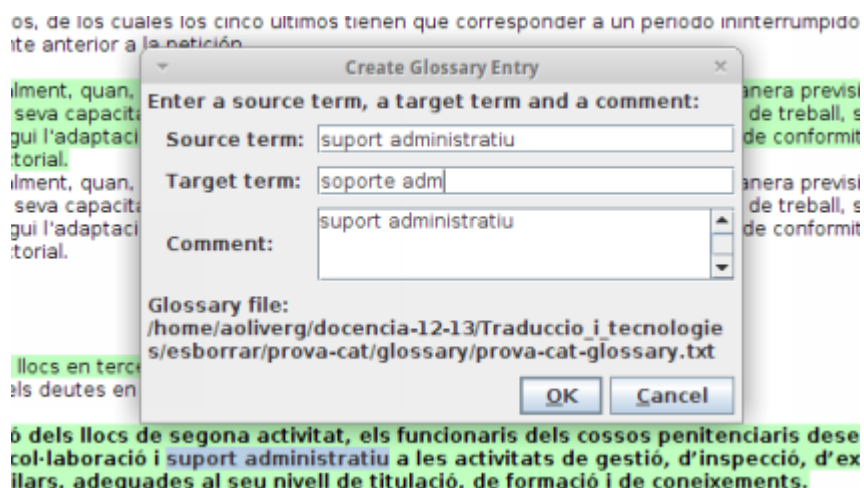


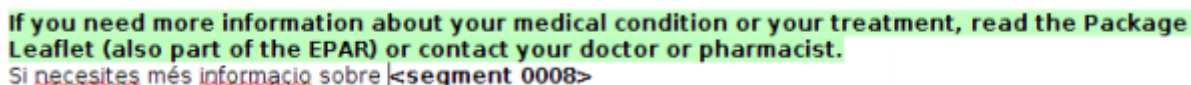
Figura 10. Pantalla de entrada de terminología en OmegaT

#### 1.3.e. Traducción

Una vez creado el proyecto y asignados los diferentes recursos ya podemos empezar a traducir. Como hemos comentado anteriormente, es un buen momento para ampliar nuestras bases de datos terminológicas. Las traducciones que vamos haciendo se almacenarán automáticamente en una memoria de traducción.

### **1.3.f. Revisión dentro de la herramienta de traducción asistida**

Mientras vamos traduciendo dispondremos de varias ayudas adicionales, como correctores ortográficos y gramaticales. En la figura 11 podemos ver el corrector ortográfico de OmegaT en acción:



If you need more information about your medical condition or your treatment, read the Package Leaflet (also part of the EPAR) or contact your doctor or pharmacist.  
Si necesitas més informació sobre <segment 0008>

Figura 11. Corrección ortográfica en OmegaT

Una vez finalizado el proyecto hemos podemos llevar a cabo diversas revisiones:

- Una nueva corrección ortográfica y gramatical automática
- Una releída a fondo de todo el proyecto
- Verificaciones automáticas de consistencia terminológica. Algunas herramientas permiten verificar si la terminología empleada en el proyecto coincide con la de las bases de datos terminológicas asignadas
- Verificación automática de consistencia de etiquetas, para asegurar que los documentos finales aparecerán correctamente en su formato original.

La ventaja de hacer la revisión dentro del sistema de traducción asistida es que todos los cambios que hagamos se verán reflejados en las memorias de traducción generadas en el proyecto. En cambio, es posible pasar por alto algún aspecto relacionado con el formato final de los documentos.

### **1.3.g. Creación de los documentos traducidos**

Una vez hechas las revisiones dentro de la herramienta de traducción asistida ya podremos generar los documentos finales. Estos programas son capaces de generar documentos traducidos que mantienen el formato de los documentos originales.

Un caso especial son los proyectos que hemos creado en formato XLIFF con alguna herramienta específica y los hemos traducido con alguna herramienta de traducción asistida. Por ejemplo, creamos un proyecto en formato XLIFF con Rainbow de Okapi Tools y lo traducimos con OmegaT. En este caso, cuando finalizamos el proyecto con OmegaT y creamos el fichero traducido obtendremos un XLIFF traducido que se tendrá que convertir en el formato correspondiente al original utilizando de nuevo Rainbow.

### **1.3.h. Revisión en formato final**

Es importante hacer una segunda revisión con los documentos traducidos en el formato final. Además de detectar algún error no detectado anteriormente, podremos ver si hay algún problema con el formato de los documentos. El inconveniente es que los cambios que hagamos no se verán reflejados en las memorias de traducción generadas en el proyecto. Si el error es severo, puede ser una buena idea entrar



de nuevo en el proyecto y hacer también la corrección, para asegurarnos de que los cambios aparezcan también en las memorias de traducción.

### ***1.3.i. Gestión de los recursos generados***

Durante la traducción se habrán generado dos recursos importantes: una memoria de traducción y una base de datos terminológica. Una vez finalizado el proyecto es importante gestionar correctamente estos recursos. La idea básica es conservarlos en un lugar accesible y saber en todo momento qué recursos podemos utilizar para cada nuevo proyecto. Dependiendo de la herramienta empleada esta gestión de los recursos se puede hacer de una manera diferenciada. Una de las prácticas asociadas a este libro trata en profundidad este tema.

## 1.4. Los sistemas de traducción asistida

Hay una gran cantidad de herramientas de traducción asistida en el mercado. Muchas de ellas, como por ejemplo OmegaT, son herramientas de software libre y se pueden utilizar de manera libre y gratuita. Muchas otras, en cambio, son herramientas propietarias que requieren adquirir algún tipo de licencia. Para obtener información actualizada sobre las herramientas de traducción asistida existentes recomiendo consultar dos enlaces:

La página de Wikibooks dedicada a CAT-Tools: <http://en.wikibooks.org/wiki/CAT-Tools>

La página de la Wikipedia inglesa dedicada a la traducción asistida:  
[http://en.wikipedia.org/wiki/Computer-assisted\\_translation](http://en.wikipedia.org/wiki/Computer-assisted_translation)

## 1.5. Conclusiones

En este capítulo hemos presentado las principales funcionalidades de los sistemas de traducción asistida por ordenador y hemos detallado las fases de los trabajos de traducción cuando se utilizan este tipo de herramientas. En los próximos capítulos profundizaremos en los principales recursos asociados a las herramientas de traducción asistida: las memorias de traducción y las bases de datos terminológicas.

En la sección *Para ampliar conocimientos* de este mismo capítulo encontraremos información sobre herramientas de traducción asistida on-line y comentarios sobre las ventajas e inconvenientes del uso de este tipo de herramientas. También hay una ampliación del tema de las expresiones regulares, con el objetivo de dominar más el uso del formato SRX para la creación de reglas de segmentación. Por último presentaremos un corrector gramatical libre, el LanguageTool, que se puede integrar en algunas herramientas de traducción asistida.

## 1.6. Para ampliar conocimientos

### 1.6.1. Herramientas de traducción asistida on-line

Tradicionalmente las herramientas de traducción asistida por ordenador han sido unas aplicaciones informáticas que se instalaban en el ordenador del traductor y se ejecutaban desde este mismo ordenador. En los últimos años han aparecido algunas herramientas de traducción asistida que funcionan remotamente en un servidor a las que accedemos mediante un navegador web.

Estas herramientas ofrecen la ventaja de que no requieren ningún tipo de instalación y que tanto la herramienta como los archivos de trabajo están disponibles on-line desde cualquier ordenador. Estas herramientas generalmente permiten un trabajo colaborativo de una manera muy sencilla, simplemente compartiendo el proyecto de traducción entre varias personas. De esta manera todos los traductores pueden trabajar sobre el mismo proyecto, añadir comentarios, hacer revisiones de partes realizadas por otro, etc. El único inconveniente destacable es que para usarlas es imprescindible disponer de una conexión a Internet y si por algún motivo la conexión falla no podemos continuar trabajando en el proyecto.

Entre estas herramientas podemos destacar dos, que son de uso gratuito:

- Google Translator Toolkit (<http://translate.google.com/toolkit/>)
- WordFast Anywhere (<http://www.wordfast.net/?whichpage=anywhere>)

El uso de estas herramientas es bastante sencillo y no requieren de ninguna instalación por lo que son una buena solución en aquellos casos que no disponemos de una herramienta instalada o en los casos que tengamos que colaborar con personas que no disponen de ninguna herramienta.

Hay otras herramientas pensadas para agencias de traducción que permiten hacer una gestión integral del proyecto vía web: desde la creación del proyecto, la asignación a los traductores así como la traducción y revisión de los archivos por parte de los colaboradores. Entre estas herramientas podemos destacar dos:

- GlobalSight (<http://www.globalsight.com/>): se trata de una herramienta de código abierto.
- Memsource (<http://www.memsource.com>): es una herramienta propietaria de pago que tiene una versión limitada de uso gratuito para traductores.

## 1.6.2.SRX y las expresiones regulares

Las reglas de segmentación en formato SRX se expresan mediante expresiones regulares, lo que permite una gran flexibilidad en la definición de las reglas. A continuación podemos observar un resumen de la sintaxis de las expresiones regulares utilizadas en las reglas SRX (resumen de la tabla que se ofrece en <http://www.gala-global.org/oscarStandards/srx/srx10.html><sup>2</sup>:

Carácter	Descripción
\A	Coincide con el principio de la entrada. Difiere de ^ en el hecho que \A no coincide después de una nueva línea dentro de la entrada.
\b, outside of a [Set] \b, fora d'un [Conjunt] \b, fuera de un [Conjunto]	Coincide si la posición actual es el límite de una palabra. Los límites tienen lugar en las transiciones entre caracteres palabra (\w) y no palabra (\W), ignorando las marcas de combinación.
\b, within a [Set] \b, dins d'un [Conjunt] \b, dentro de un [Conjunto]	Coincide con un RETROCESO \u0008.
\B	Coincide si la posición actual no es un límite de una palabra.
\d	Coincide con cualquier carácter de la Categoría Nd (Número, Dígito Decimal) de Unicode.
\D	Coincide con cualquier carácter que no sea un dígito decimal.
\e	Coincide con un ESCAPE, \u001B.
\f	Coincide con un FORM FEED, \u000C.
\G	Coincide si la posición actual es el final de la coincidencia anterior.
\n	Coincide con un LINE FEED, \u000A.
\r	Coincide con un CARRIAGE RETURN, \u000D.
\s	Coincide con un carácter de espacio en blanco. El espacio en blanco se define como [\t\n\f\r\p{Z}].
\S	Coincide con cualquier carácter que no sea el espacio en blanco.
\t	Coincide con una TABULACIÓN HORIZONTAL, \u0009.
\uhhhh	Coincide con el carácter con el valor hexadecimal hhhh
\Uhhhhhhh	Coincide con el carácter con el valor hexadecimal hhhhhhhh. Se tienen que dar exactamente 8 dígitos hexadecimales, aunque el punto de código unicode más largo sea U0010ffff.
\w	Coincide con un carácter palabra. Los caracteres palabra son: [\p{Ll}\p{Lu}\p{Lt}\p{Lo}\p{Nd}].
\W	Coincide con cualquier carácter no palabra.
\x{hhhh}	Coincide con el carácter con valor hexadecimal hhhh
\xhh	Coincide con el carácter con valor de dos dígitos hexadecimales hh
\Z	Coincide si la posición actual está al final de la entrada, pero antes del último terminador de línea, si es que hay alguno.
\z	Coincide si la posición actual está al final de la entrada.
\0nnn	Coincide con el carácter con valor octal nnn

<sup>2</sup> SRX 1.0 Specification. Copyright © The Localisation Industry Standards Association [LISA] 2004. All Rights Reserved.

<code>\n</code>	Referencia. Coincide con el n-ésimo grupo coincidente, n tiene que ser >1 i < que total de grupos de la expresión
<code>[pattern]</code>	Coincide con cualquiera de los caracteres del conjunto.
<code>.</code>	Coincide con cualquier carácter.
<code>^</code>	Coincide con el principio de una línea.
<code>\$</code>	Coincide con el final de una línea.
<code>\</code>	Se tiene que poner una barra invertida delante de ciertos caracteres para referirse a ellos, ya que tienen un significado especial dentro de las expresiones regulares. Estos caracteres son: * ? + [ ( ) { } ^ \$   \ . /

También se pueden utilizar los siguiente operadores (és un resumen de <http://www.gala-global.org/oscarStandards/srx/srx10.html><sup>3</sup>):

Operador	Descripción
	Alternancia. A B coincide con A o con B.
*	Coincide 0 o más veces. Coincide tantas veces como sea posible.
+	Coincide 1 o más veces. Coincide tantas veces como sea posible.
?	Coincide ninguna o una vez. Prefiere una vez.
{n}	Coincide exactamente n veces
{n, }	Coincide como mínimo n veces. Coincide tantas veces como sea posible.
{n,m}	Coincide entre n y m veces. Coincide tantas veces como sea posible, pero no más de m veces.
*?	Coincide 0 o más veces. Coincide las mínimas veces posibles.
+?	Coincide 1 o más veces. Coincide las mínimas veces posibles.
??	Coincide cero o una vez. Prefiere cero.
{n}?	Coincide exactamente n veces.
{n, }?	Coincide como mínimo n veces, pero no más veces de las necesarias para coincidir con toda la expresión.

3 SRX 1.0 Specification. Copyright © The Localisation Industry Standards Association [LISA] 2004. All Rights Reserved.

### 1.6.7. El corrector gramatical LanguageTool

Un corrector ortográfico es capaz de detectar las palabras mal escritas en una lengua comparándolas con un diccionario de palabras existentes en la lengua (con todas sus formas flexionadas).

Por ejemplo: cualquier corrector ortográfico del castellano es capaz de encontrar los errores de la siguiente oración:

Cualquier corrector ortografico del castellano es capaz de encontrar los errores de esta oracion.

En cambio, un corrector simplemente ortográfico no encontrará el siguiente error:

Se a hecho un gran esfuerzo en el diseño de este corrector.

Ya que “a” debería escribirse con “h” ya que es del ver "haber". Como que "a" es una palabra correcta en la lengua (pero no en este contexto) el corrector ortográfico no lo marca como error.

Los correctores gramaticales van más allá y permiten encontrar algunos de estos error además de construcciones gramaticales incorrectas. Por ejemplo, vemos la actuación del corrector LanguageTool en el ejemplo anterior:

Se **a** hecho un gran esfuerzo en el diseño de este corrector.

<b>El verbo 'haber' lleva hache: 'ha'</b>
ha
Ignore this error
Rule implementation...

Vemos que este corrector ha podido detectar que en este caso “a” se debería escribir con “h” ya que corresponde a una forma del verbo haber.

Vemos a continuación la regla de LanguageTool que permite hacer esta detección:

```
<rule id="A_PARTICIPIO" name="a + participio">
  <pattern>
    <token><exception inflected="yes" regexp="yes">ol[oe]r|sab[e]r|gust(o|ar)|aroma|pest.
+</exception></token>
    <marker>
    <token>a</token>
    </marker>
    <token postag="V[ASM]P00SM" postag_regexp="yes"/>
  </pattern>
  <message>El verbo 'haber' lleva hache: <suggestion>ha</suggestion></message>
  <short>El verbo «haber» lleva hache.</short>
  <example correction="ha" type="incorrect">El atleta <marker>a</marker> corrido de forma
espectacular.</example>
  <example type="correct">El atleta <marker>ha</marker> corrido de forma espectacular.</example>
  <example type="correct">Este bacalao huele <marker>a</marker> pescado fresco.</example>
</rule>
```

No entraremos en detalle de implementación pero nos fijaremos sólo en el hecho de que la regla define un patrón (*pattern*) que en este caso tiene tres palabras (*tokens*). La primera define que si se cumple el patrón será una excepción a la regla (y define palabras como oler, saber, gustar, aroma, peste, etc.). La segunda es la propia “a” y la tercera es un verbo en participio (que se define mediante las etiquetas morfosintácticas). Si nos fijamos en los ejemplos que se dan al final de la regla, podremos entender aún mejor su estructura.

A continuación podemos observar la frase del ejemplo etiquetada por el etiquetador de castellano integrado en LanguageTool:

```
<S> Se [se/P0000000,se/P03CN000,se/PP3CN000,] a[a/NCFS000,a/SPS00,]
hecho[hecho/NCMS000,hacer/VMP00SM,] un[uno/DIOMS0,] gran[gran/AQ0CS0,]
esfuerzo[esfuerzo/NCMS000,esforzar/VMIP1S0,] en[en/SPS00,] el[e1/DA0MS0,]
diseño[diseño/NCMS000,diseñar/VMIP1S0,] de[de/SPS00,] este[este/DD0MS0,]
corrector[corrector/AQ0MS0,corrector/NCMS000,</S>,]
```

El corrector gramatical es capaz de etiquetar el texto que analiza, es decir, darle a las palabras las etiquetas que expresan sus categorías gramaticales y algunas subcategorizaciones, aunque no es capaz de desambiguar y asigna todas las etiquetas posibles.

LanguageTool (<https://www.languagetool.org/>) es un corrector gramatical de código abierto que está disponible para más de 20 lenguas, entre ellas el castellano, catalán e inglés. Los usuarios avanzados pueden modificar las reglas o crear nuevas reglas de forma que el corrector va mejorando con el tiempo.

LanguageTool puede funcionar como aplicación independiente y además se integra perfectamente en LibreOffice/OpenOffice y también se puede instalar como extensión de Firefox. También es posible instalar LanguageTool a la herramienta de traducción asistida OmegaT.

## Bibliografía

Lippmann E. O. (1971) *An approach to Computer - Aided Translation*. IEE Transactions on Engineering Writing and Speech, Vol. EWS-14, No. 1, February 1971

## Licencia de este documento



Traducción y Tecnología by Antoni Oliver

is licensed under a [Creative Commons Atribución - ShareAlike 3.0 Unported License](https://creativecommons.org/licenses/by-sa/3.0/) .