



---

**Universidad de Valladolid**

**FACULTAD DE TRADUCCIÓN E INTERPRETACIÓN**

**Grado en Traducción e Interpretación**

**TRABAJO DE FIN DE GRADO**

**Las herramientas TAO: creación de diccionarios  
de autosugerencias y modificación de una  
memoria de traducción a partir de la herramienta  
SDL Trados Studio 2009**

**Presentado por Eva López Sainz-Aja**

**Tutelado por Dra. María Teresa Sánchez Nieto**

**Soria, 2015**

# ÍNDICE

ÍNDICE DE FIGURAS .....	4
RESUMEN .....	5
ABSTRACT .....	5
INTRODUCCIÓN .....	6
1. Estructura .....	6
2. Justificación .....	6
3. Motivación .....	7
4. Vinculación del trabajo con las competencias del Grado en Traducción e Interpretación ..	7
OBJETIVOS .....	9
METODOLOGÍA Y PLAN DE TRABAJO .....	10
Metodología .....	10
Plan de trabajo .....	10
CAPÍTULO 1: LAS HERRAMIENTAS DE TRADUCCIÓN ASISTIDA POR ORDENADOR (TAO) .....	12
1. ¿Qué son las herramientas TAO? ¿En qué se diferencian de las herramientas de TA? .....	12
2. ¿Cuándo y por qué surgen las herramientas TAO? .....	13
3. Programas que forman la herramienta de entorno de traducción (TEnt) .....	14
3.1. Memorias de traducción: funcionamiento. ....	14
3.1.1. Segmentación o alineado .....	16
3.1.2. Coincidencias .....	16
3.2. Alineador .....	18
3.3. Programas de estadísticas .....	19
3.3.1. Aproximación o enfoque lingüístico .....	19
3.3.2. Aproximación estadística .....	20
3.4. Gestor terminológico .....	20
3.4.1. Almacenamiento .....	21
3.4.2. Recuperación .....	22
3.4.3. Reconocimiento activo de terminología: .....	22
3.4.4. Extracción de términos: .....	22
3.5. Sistemas de localización .....	23
3.6. Corpus electrónicos y generadores de concordancias .....	25
3.6.1. Tipos de corpus electrónicos .....	26
3.6.2. Herramientas de análisis de corpus .....	26

3.7.	Editor y correctores ortográficos .....	27
3.8.	Diccionario de autosugerencias .....	28
4.	Clasificación de las herramientas TAO .....	30
4.1.	Clasificación según su ubicación .....	30
4.2.	Clasificación según licencia y código .....	31
CAPÍTULO 2: CREACIÓN DE UN DICCIONARIO DE AUTOSUGERENCIAS A PARTIR DE LA MEMORIA DE TRADUCCIÓN DE LA DGT.....		
1.	La Dirección General de Traducción y su memoria de traducción.....	36
1.1.	Qué es el <i>Acquis communautaire</i> ? .....	36
1.2.	La Dirección General de Traducción (DGT): estructura y organización .....	36
1.3.	La memoria de traducción de la DGT .....	37
1.4.	Política de multilingüismo .....	37
1.5.	Utilidades de la memoria de traducción de la DGT.....	38
2.	Descarga e instalación de la memoria de traducción .....	39
2.1.	Descarga de la memoria de traducción y del programa de extracción de pares de lenguas. ....	39
2.2.	Extracción del par de lenguas DE>ES .....	40
2.3.	Creación de la memoria de traducción en SDL Trados 2009 .....	41
3.	Generar el diccionario de autosugerencias.....	44
CAPÍTULO 3: APROVECHAMIENTO DE UNA MEMORIA DE TRADUCCIÓN MEDIANTE EL USO COMBINADO DE CAMPOS Y FILTROS: ETIQUETADO TEMÁTICO DE LAS UNIDADES DE LA MEMORIA DE TRADUCCIÓN DE LA DGT.....		
1.	Descarga y recopilación términos clasificados por campo temático .....	48
2.	Personalización de los campos de la memoria de traducción .....	48
3.	Filtrado de términos por campo temático .....	51
3.1.	Creación de filtros en Trados Studio .....	51
3.2.	Edición de unidades de traducción por lotes con ayuda de filtros .....	53
3.3.	Exportación de los filtros.....	54
4.	Empleo del sistema de penalizaciones.....	55
CONCLUSIONES .....		58
REFERENCIAS BIBLIOGRÁFICAS .....		61
ANEXOS .....		64

## ÍNDICE DE FIGURAS

Figura 1: Consulta de las coincidencias parciales localizadas en la MT .....	29
Figura 2: Selección de carpetas de origen y destino .....	40
Figura 3: Selección del par de lenguas (DE y ES) .....	41
Figura 4: Creación de la memoria de traducción en SDL Trados .....	42
Figura 5: Lista de los archivos TMX extraídos .....	42
Figura 6: Importación de los archivos TMX a la memoria de traducción.....	43
Figura 7: Descripción de la memoria de traducción resultante .....	44
Figura 8: Selección de las unidades de traducción para el diccionario de autosugerencias .....	45
Figura 9: Creación del diccionario de autosugerencias.....	45
Figura 10: Configuración de los campos de la memoria de traducción .....	51
Figura 11: Configuración de los campos de la memoria de traducción .....	51
Figura 12: Inserción de condiciones en un filtro .....	52
Figura 13: Edición de los filtros por lotes .....	53
Figura 14: Actualización de valores de campo para los filtros .....	54
Figura 15: Exportación de filtros .....	55
Figura 16: Adición de filtros al sistema de penalizaciones de las memorias de traducción .....	56

## RESUMEN

Las herramientas de traducción asistida por ordenador (TAO) han cambiado completamente el método de trabajo de los traductores en apenas unas décadas. Una de estas herramientas son los gestores de memorias de traducción, que ofrecen múltiples posibilidades con el fin de mejorar la calidad de las traducciones.

En el presente trabajo nos serviremos de las posibilidades de una de estas herramientas TAO, SDL Trados 2009, para sacar más provecho a la memoria de traducción de la Dirección General de Traducción creando un diccionario de autosugerencias y empleando el sistema de filtrado del gestor para modificar dicha memoria.

**Palabras clave:** herramientas TAO, memoria de traducción de la DGT, Dirección General de Traducción, filtros, SDL Trados Studio, diccionario de autosugerencias.

## ABSTRACT

Computer-assisted translation tools (CAT Tools) have completely changed translators working method in scarcely a few decades. One of these tools is the translation memory manager, which provides translators with a lot of possibilities to improve the quality of their work.

In this project we will make use of a CAT Tool like SDL Trados 2009 in order to exploit more DGT Translation Memory by creating an AutoSuggest dictionary. Also we will use the tool filtering system to modify this memory and make it more profitable.

**Key words:** CAT Tools, DGT Translation Memory, Directorate-General for Translation, filters, SDL Trados Studio, AutoSuggest dictionary.

# INTRODUCCIÓN

Nos encontramos en la era de las tecnologías en la que, de una forma u otra, todo pasa por algún medio electrónico. Con la globalización, la información generada a la que tenemos acceso se ha multiplicado y nos hemos encontrado con la necesidad de adaptarla al mundo entero. Esto ha dado lugar a la expansión de la labor de traducción, encargada de reproducir en todas las lenguas posibles esta información. A partir de entonces, un sector de la informática ha fijado sus objetivos en desarrollar herramientas que faciliten la labor de difusión de la información generada. Tanto traduciendo nuevos documentos como reutilizando otros antiguos, las denominadas herramientas TAO (CAT *tools* en inglés) funcionan como complemento al traductor, siendo este el factor principal de la ecuación.

## 1. Estructura

El presente trabajo está estructurado en tres capítulos principales que derivan en unas conclusiones. En el primer capítulo realizaremos un recorrido por las herramientas de entorno de traducción<sup>1</sup> (historia, programas que forman una herramienta TAO, clasificación de las herramientas TAO atendiendo a diferentes criterios, etc.). El segundo apartado lo dedicaremos a la Dirección General de Traducción y su gran memoria de traducción. Tras explicar la naturaleza de esta institución, detallamos cómo extraer la parte de la memoria que nos interesa e importarla al gestor de memorias SDL Trados con el fin de generar un diccionario de autosugerencias. Y ya en el tercer capítulo profundizaremos más en el gestor mencionado y emplearemos algunas de sus aplicaciones para modificar y sacar partido a la memoria, ampliando así nuestros conocimientos sobre esta herramienta de traducción y sobre la memoria en sí. Por último, extraeremos unas conclusiones de todo lo realizado a lo largo del proyecto, tanto de las ventajas de estas herramientas y recursos como de las desventajas.

## 2. Justificación

El presente trabajo tiene como finalidad, además de hacer un breve recorrido por la gran variedad de herramientas TAO de las que disponemos hoy en día, demostrar que estas herramientas tienen más posibilidades de las que podemos ver a simple vista, pues es posible modificar una memoria de traducción para sacarle más partido del que ya tiene de por sí para facilitar la tarea del traductor.

---

<sup>1</sup> Por sus siglas en inglés TEnT, las herramientas de entorno de traducción son todas aquellas que contribuyen a crear un entorno diseñado para facilitar el trabajo del traductor. Se trata de un término acuñado por Jost Zetzsche en 2007. (Balcells, 2013)

La Unión Europea pone a nuestra disposición un recurso tan importante como es la memoria de traducción de la DGT. Utilizando algunas de las múltiples posibilidades que nos ofrece una de estas herramientas TAO de las que hablamos, SDL Trados Studio, podemos sacar provecho a esta herramienta de una forma ~~fácil y sencilla~~. Relativamente sencilla, con una programación previa de ciertas operaciones.

Los traductores se sirven de las herramientas TAO para agilizar y mejorar la calidad de sus traducciones de forma instantánea. Sin embargo, dedicando un tiempo a mejorar y personalizar los recursos que nos ofrecen, los objetivos de estos profesionales pueden verse superados con creces. Una forma de conseguir esto es ahondar en las utilidades de estas herramientas, en las aplicaciones que muchas veces escapan a la vista de los traductores porque precisan un dominio más avanzado de la herramienta. Con este trabajo demostraremos que es posible modificar una memoria tan grande y compleja como la que nos ofrece la DGT sirviéndonos de las múltiples funciones de una herramienta tan compleja como es SDL Trados.

### **3. Motivación**

La motivación para realizar este trabajo surgió de la idea de hacernos con un recurso tan valioso como es la memoria de traducción de la DGT y tratar de exprimir al máximo sus posibilidades gracias a una de las herramientas TAO con la que hemos trabajado a lo largo de todo el grado, SDL Trados (en su versión de 2009). Al mismo tiempo, pretendíamos aprender a utilizar algunas de las aplicaciones que nos ofrece esta herramienta a un nivel más avanzado que el hasta ahora conocíamos. Entre estas aplicaciones se encontraban la de creación de diccionarios de autosugerencias, una aplicación que no nos ofrece ninguna otra herramienta comercial de entorno de traducción, y el sistema de filtrado, que ofrece muchas más posibilidades de las que podemos ver a simple vista.

### **4. Vinculación del trabajo con las competencias del Grado en Traducción e Interpretación**

Entre las competencias específicas de nuestro Grado en Traducción e Interpretación, vinculamos a nuestro proyecto las siguientes:

E8. “Conocer y gestionar las fuentes y los recursos de información y documentación en Lengua A/B/C necesarios para el ejercicio de la traducción general/especializada B/C”.

E9. “Reconocer la diversidad y multiculturalidad de la Lengua A/B/C/D”.

E11. “Comprender las relaciones internacionales en el contexto europeo y mundial y su organización interna”.

E12. “Conocer la evolución social, política y cultural para comprender la diversidad y la multiculturalidad”.

E15. “Familiarizarse con los distintos tipos de centros de información y con los servicios que ofrecen al usuario”.

E16. “Manejar las últimas tecnologías documentales aplicadas a la traducción: sistemas de gestión y recuperación de información electrónica”.

E17. “Dominar conceptos básicos sobre el funcionamiento de herramientas informáticas que faciliten su utilización y su integración en la labor del traductor”.

E18. “Utilizar las herramientas informáticas básicas como instrumento específico de ayuda a la traducción en las diferentes fases del proceso traductológico”.

E19. “Desarrollar un método de trabajo organizado y optimizado gracias al empleo de herramientas informáticas”.

E20. “Conocer las herramientas más apropiadas de localización”.

E21. “Familiarizarse con los programas y con las técnicas de pre y post-edición desarrolladas en el campo de la traducción automática”.

E34. “Aplicar a proyectos de traducción general/especializada las herramientas profesionales de traducción asistida por ordenador (TAO) y de gestión terminológica”.

E35. “Conocer el trabajo terminológico en cada una de sus fases y aplicarlo a su labor traductora”.

E36. “Adquirir conocimientos básicos de terminología”.

## OBJETIVOS

Antes de explicar la metodología y el plan de trabajo seguido en este proyecto, es necesario establecer los objetivos que tenemos intención de conseguir al final de este.

El objetivo principal del trabajo se basa en conseguir modificar y sacar partido a una memoria de traducción con el fin de aprender a utilizar las funciones avanzadas que nos ofrecen los gestores de memorias de traducción, sirviéndonos para ello del gestor SDL Trados Studio. Este objetivo principal se estructura a su vez en otros objetivos que derivan de él:

- Contextualizar el proyecto haciendo un recorrido por las herramientas TAO, sus componentes, características y funciones desde que surgieron hasta la actualidad, así como los tipos de herramientas TAO con los que podemos contar para facilitar nuestra tarea como traductores.
- Realizar una clasificación de las herramientas TAO disponibles en el mercado actual.
- Descargar la memoria de traducción de la Dirección General de Traducción (DGT), concretamente el par de lenguas DE-ES e importarla a los gestores de memorias de traducción.
- Generar un diccionario de autosugerencias a partir de esa memoria de traducción.
- Personalizar esta memoria marcando todas sus unidades de traducción según el dominio al que pertenezcan, aprendiendo a utilizar los sistemas de filtros, campos personalizados y penalizaciones de dicha herramienta. Esto conlleva aprender a manejar Trados a un nivel avanzado.

Abordaremos todos estos objetivos a lo largo de los tres capítulos del proyecto, uno principalmente teórico y los dos siguientes de carácter más práctico.

# METODOLOGÍA Y PLAN DE TRABAJO

## Metodología

Este trabajo es una compilación de elementos relacionados entre sí. En principio, la metodología que hemos tratado de seguir ha sido deductiva, pues comenzamos familiarizándonos e informándonos sobre las herramientas TAO y cómo funcionan, para luego adentrarnos de lleno en una de estas herramientas y hacer un uso avanzado de sus aplicaciones para conseguir nuestros objetivos.

Por otro lado, la metodología también tiene, en parte, carácter inductivo, pues, en cierto modo pretendíamos describir las herramientas TAO y tratar de mostrar algunas funciones útiles que requieren cierto conocimiento básico de estas herramientas. Sin embargo, a lo largo del proceso hemos descubierto otras funciones que no conocíamos y que también resultan muy útiles. De esta forma, hemos adquirido más conocimientos sobre esta herramienta y hemos tratado de proyectarlos en el trabajo.

## Plan de trabajo

Antes de comenzar a realizar el proyecto, hubo un trabajo de documentación sobre la situación actual y la historia de las herramientas TAO, así como una compilación de información (reflejada en la bibliografía) que nos sirvió para contextualizar el resto del trabajo técnico.

En cuanto a la parte teórica, que nos servirá como base para adquirir los conocimientos necesarios para entender la parte práctica, establecemos el estado de la cuestión en relación con las herramientas TAO, realizando un recorrido por su historia, sus características, sus ventajas y desventajas y la situación actual de este tipo de herramientas. Así mismo, realizaremos una clasificación de las herramientas TAO atendiendo a ciertos criterios de gran importancia en la situación actual de la traducción como proceso ligado a internet.

En segundo lugar, nos centramos en la parte práctica, que también requería una gran labor de documentación, tanto a nivel informático como a la hora de entrar en contacto con los recursos que la Unión Europea pone a nuestra disposición y con la política de multilingüismo que lleva a cabo esta institución. Una vez establecidas las bases de esta parte, comenzó el desarrollo de la práctica, que consiste en la descarga e incorporación a Trados Studio de una memoria de traducción DE>ES con un mínimo de 30.000 unidades a partir de la memoria de traducción de la DGT. El resultado de este proceso supuso uno de los productos principales con los que vamos a trabajar el resto de la fase práctica: describir el proceso de creación de un diccionario de autosugerencias a partir de esa memoria y una

forma de aprovechamiento del sistema de filtros de SDL Trados para etiquetar las unidades de traducción de la memoria de traducción e conjunción con la función de campos personalizados que también nos ofrece esta herramienta.

Por último, tras analizar los resultados de todos los procesos, proyectaremos las conclusiones extraídas de todo el proyecto.

# CAPÍTULO 1: LAS HERRAMIENTAS DE TRADUCCIÓN ASISTIDA POR ORDENADOR (TAO)

Este apartado se centra en analizar las herramientas de “Traducción Asistida por Ordenador” disponibles actualmente en el mundo profesional. Día a día, las herramientas TAO se están convirtiendo en el gran aliado de los traductores profesionales puesto que facilitan y agilizan el trabajo de estos. En los últimos años se han producido importantes avances en el desarrollo de las herramientas TAO; tanto es así que, actualmente, contamos con una gran variedad de herramientas de Traducción Asistida por Ordenador.

Tradicionalmente, la mayoría de los traductores tiene formación principalmente lingüística, es decir, “de letras”. Sin embargo, tarde o temprano todo profesional de la traducción debe enfrentarse al método actual de trabajo ligado estrechamente a la informática. La traducción asistida, la gestión de proyectos o los métodos de envío de archivos son aspectos técnicos cuyo manejo es indispensable para todo traductor moderno, pues esto puede influir en su trabajo de tal forma que de ello dependa su competencia y su éxito a la hora de adquirir una cartera de clientes (Ortiz Jiménez, 2011).

Actualmente el ordenador es la herramienta más utilizada por los traductores. Por ello, además de las herramientas TAO, es importante que estos cuenten con un amplio rango de herramientas de software genérico: procesadores rápidos, lectores y editores de PDF, gestores de correo, antivirus y antiespías potentes, compresores versátiles y todas las herramientas que hagan el trabajo más fácil (López Sánchez, 2014:8).

Por norma general, los traductores son usuarios hábiles de las tecnologías. Todo traductor, desde un estudiante hasta un traductor profesional o docente, debe conocer y aprender a manejar estas herramientas. El uso y conocimiento de estas da lugar a un importante campo en la Traducción Asistida por Ordenador: la investigación para mejorar y perfeccionar estas herramientas (Bowker y Fisher, 2002 en Gambier et al., 2010: 60).

## 1. ¿Qué son las herramientas TAO? ¿En qué se diferencian de las herramientas de TA?

Antes de sumergirnos en el mundo de las herramientas de Traducción Asistida por Ordenador, es necesario aclarar en qué consiste la traducción asistida y que la diferencia de la traducción automática.

La Traducción Asistida por Ordenador es un tipo de traducción basada en el uso de programas que acompañan al traductor en el proceso de reproducción de un texto meta a partir de un texto origen. Esta traducción atribuye la responsabilidad principal del texto al traductor; el software sirve como herramienta de ayuda para facilitar el trabajo del

profesional. La diferencia principal de este tipo de traducción con la traducción automática es que esta última la realiza una máquina sin ser imprescindible la actuación humana, aunque suele requerir la intervención de profesionales para editar y corregir (Bowker y Fisher, 2002 en Gambier et al., 2010: 60).

Las herramientas TAO más extendidas a nivel profesional y académico son las herramientas de entorno de traducción o TEnT (término acuñado por Jost Zetzsche, 2007), que consisten en un paquete de herramientas diseñadas para crear un entorno que facilite el trabajo del traductor y así no necesite complementos externos. Estas herramientas están diseñadas especialmente para el traductor, van más allá de las herramientas generales como procesadores de textos o gestores de correos.

Hutchins (1998) realizó una distinción entre las herramientas de traducción humana asistida por una máquina y la traducción de máquina asistida por humanos, para aclarar las diferencias entre las herramientas de gestión de memorias de traducción y la traducción automática, en la que la tarea del traductor se limita a la postedición (Balcells, 2013: 99).

## **2. ¿Cuándo y por qué surgen las herramientas TAO?**

Todas las herramientas de entorno de traducción han evolucionado mucho desde que comenzaron a desarrollarse en la década de 1970 y 1980, después de que, en 1966 el gobierno de Estados Unidos publicara el informe ALPAC (*Automatic Language Processing Advisory*). Este informe, realizado tras una investigación fruto de la preocupación por el poco progreso que se estaba realizando en la traducción automática, reveló que esta traducción resultaba mucho más lenta, cara y menos precisa que la traducción humana. Tras esto, los investigadores rechazaron la viabilidad de la investigación en TA y comenzaron a desarrollar sistemas de ayuda al traductor (Hutchins, 1998: 3).

Las herramientas TAO nacieron en el entorno informático con el comienzo del desarrollo de los ordenadores personales en la década de 1970. Las cantidades de información se multiplicaron de forma muy rápida, por lo que se volvió necesario desarrollar métodos de gestión de esta información de forma automática y rápida. Con este propósito surgieron las memorias de traducción, que no solo almacenaban cantidades ingentes de información, sino que también servían para hacer búsquedas (Bowker y Fisher, 2002 en Gambier et al., 2010: 60).

Una serie de investigadores introdujeron importantes novedades que consolidaron el núcleo de las TAO: las memorias de traducción. Fue Arthern (1979) quien creó un sistema que permitía el almacenamiento de bitextos (texto origen y texto meta) y la inserción fragmentos en nuevas traducciones. Esto permitía comparar textos originales con

textos traducidos. Más tarde, Martin Kay (1980) introduciría procesador de textos con la pantalla dividida en dos, un glosario en el que los términos se buscarían automáticamente y un sistema de sugerencias de posibles traducciones que se pudieran aceptar o descartar.

Pero fue Melby (1981) quien desarrolló definitivamente el entorno de herramientas para el traductor: definió un procesador de textos para traducir el texto directamente mientras se permitía el acceso a una base de datos terminológica. También se podría sustituir los términos que aparecieran en el banco de bitextos de forma automática y se sugerirían traducciones automáticas que el traductor pudiera elegir si aceptarla o no.

Estos investigadores son considerados unos visionarios de la traducción, pues la informática de aquella época distaba mucho de lo que actualmente tenemos, pero ellos imaginaron un futuro donde los ordenadores son la herramienta más utilizada por traductores profesionales.

A partir de entonces se desarrollaron varios prototipos de paquetes de herramientas TAO para los primeros ordenadores de mediados de los ochenta. A medida que la informática avanzaba, también lo hacían estos proyectos. Los diseñadores comenzaron una guerra para situarse a la cabeza del sector, luchando por distribuir sus productos entre grandes compañías e instituciones públicas como la Unión Europea. A día de hoy la guerra continúa, con nuevos rivales cada vez más avanzados y que incluyen más herramientas útiles para facilitar la labor del traductor.

El futuro de la traducción, al igual que el de todas estas herramientas, se encuentra en la nube, dado que esta es la forma más sencilla de adquisición de memorias de traducción y colaboración con otros traductores (Balcells, 2013: 100, 107).

### **3. Programas que forman la herramienta de entorno de traducción (TEnt)**

En los últimos años ha habido un incremento del uso de la tecnología en la traducción. Esto ha provocado el nacimiento de herramientas de asistencia al traductor, corpus electrónicos y memorias de traducción entre otros, que aportan mayor eficiencia al profesional. Puesto que la cantidad y calidad de herramientas de traducción no hace sino aumentar, el traductor debe ser capaz de conocer y dominar gran parte de ellas para determinar cuáles pueden ayudarle más a realizar su trabajo (Bowker, 2002: 12).

#### **3.1. Memorias de traducción: funcionamiento.**

El concepto de memoria de traducción existe desde hace tiempo. La idea original se atribuye a Martin Kay que en 1980 publicó "The Proper Place of Men and Machines in

Language Translation", en el que propuso un borrador para lo que ahora llamamos "herramientas de entorno de traducción". De hecho, en su obra no se refiere directamente a las memorias de traducción (Somers, 2003: 32):

*...the translator might start by issuing a command causing the system to display anything in the store that might be relevant to [the text to be translated]... Before going on, he can examine past and future fragments of text that contain similar material (Kay, 1980: 19).*

Las ideas de Kay se basaban en las de Peter Arthern, que sugerían que los traductores perdían demasiado tiempo traduciendo textos que ya habían sido traducidos previamente. Propuso el almacenamiento de textos fuente y sus traducciones de forma que cualquier parte de estos pudiera ser fácil de recuperar e insertar en nuevos documentos. Se refirió a esto como *translation by text-retrieval* o «traducción mediante recuperación de textos». Sus propuestas describían claramente lo que hoy llamamos «memorias de traducción» (Hutchins, 1998: 7):

*The pre-requisite for implementing my proposal is that the text-processing system should have a large enough central memory store. If this is available, the proposal is simply that the organization in question should store all the texts it produces in the system's memory, together with their translations into however many languages are required. This information would have to be stored in such a way that any given portion of text in any of the languages involved can be located immediately...together with its translation (Arthern 1979: 94f).*

El primer sistema de memorias de traducción accesible a nivel comercial fue *Alps*, en 1981, que llevaba el nombre de *Repetitions Processing* (Procesamiento con Repeticiones). El término "memorias de traducción" aparecería poco después. A nivel comercial, los sistemas de memorias de traducción se expandieron a partir de la mitad de la década de 1990 (Somers, 2003: 33).

Una memoria de traducción es un banco de datos, vacío la primera vez que se utiliza, que va almacenando las unidades que traducimos por pares (original y meta). Esta estructura se llama corpus paralelo o bitexto. Las traducciones están divididas en segmentos o unidades de traducción (López Ciruelos, 2013: 182).

Estas unidades de traducción se almacenan en la base de datos de la memoria de traducción. El objetivo de estos sistemas es que el traductor pueda consultar una base de datos repleta de traducciones realizadas previamente con el fin de localizar frases similares a la que está traduciendo y utilizarlas como modelo para su nueva traducción.

Hasta la primera mitad de los años 90, los traductores no acostumbraban a guardar las traducciones hechas y, quienes sí lo hacían, las coleccionaban de una manera poco sistemática y localizar posibles segmentos útiles era una tarea muy costosa (localizar el archivo, encontrar el segmento apropiado, abrir la traducción, localizar el segmento aproximado, encontrar el equivalente adecuado, copiar y pegar el segmento, etc.) que requería demasiado tiempo. Sin embargo, desde entonces, los progresos realizados han dado como resultado estas memorias de traducción que permiten a los traductores beneficiarse de su propio trabajo realizado anteriormente (Somers, 2003: 31).

Las memorias de traducción funcionan comparando automáticamente el texto fuente con la base de datos de textos traducidos anteriormente. Cuando el traductor ha de traducir un nuevo segmento, el sistema consulta la base de datos para ver si este corresponde con algún segmento ya traducido. Si se encuentra una coincidencia, el propio sistema ofrece al traductor la traducción previa. Así, el profesional puede consultarla y decidir si quiere o no incorporarla a la nueva traducción.

A continuación, y siguiendo las ideas de Bowker (2002: 94-96), explicaremos cómo funcionan las memorias de traducción que utilizamos actualmente:

### **3.1.1. Segmentación o alineado**

Normalmente, la unidad de segmentación de un sistema de memorias de traducción es una oración, pero también pueden aparecer otros tipos de división, pues listas, tablas y encabezados no se consideran oraciones. Muchos sistemas de memorias de traducción permiten definir unidades de segmentación particulares, pudiendo incluir en un mismo segmento fragmentos de frases o párrafos completos.

Aun así, la segmentación de un texto no es tan sencilla como parece. En principio, una frase está determinada por un signo de puntuación, pero abreviaturas como Sr., Dr., EE.UU, etc. incluyen también esos signos de puntuación sin poder considerarse oraciones. La solución a este tipo de problemas pasa por incluir listas de parada o "stoplist" en la memoria de traducción, que contienen abreviaturas que no indican el final de una oración. En el caso de que aparezcan otros problemas de segmentación errónea, es el traductor quien debe adaptar y corregir a su gusto.

### **3.1.2. Coincidencias**

Aunque un segmento traducido no sea idéntico al que se compara con él, según Somers (2003: 37-41) y Bowker (2002: 96-107) existen diversos tipos de coincidencias dependiendo del grado de semejanza de los segmentos:

- Coincidencias exactas: esto ocurre cuando los dos segmentos son exactos al cien por cien, incluyendo puntuación, inflexión, números y formato (cursiva, negrita, etc.). Incluso cuando el sistema recupera una coincidencia exacta, el traductor no está obligado a aceptar esta propuesta. Además, aunque haya el segmento sea exacto, es posible que no concuerde con el contexto en el que se encuentra el nuevo segmento a traducir, por eso es importante comprobar si esta coincidencia es adecuada para el nuevo texto y que se transmite el mensaje correcto.
- Coincidencias completas: se producen cuando un segmento fuente solo se diferencia del localizado en la memoria en términos llamados "elementos variables", que incluyen números, fechas, horas, monedas, medidas e incluso nombres propios. Todos estos elementos deben tratarse de forma especial en cada texto dado que, aunque los nombres no suele traducirse, si lo hacen las fechas y horas, que en cada idioma emplean un formato diferente.
- Coincidencias parciales o fuzzy: los segmentos similares también pueden resultar útiles a la hora de traducir nuevos textos. La mayoría de los sistemas de memorias de traducción son capaces de localizar estas coincidencias "parciales". Algunos sistemas señalan las diferencias entre el segmento localizado y el nuevo para que el traductor pueda editar las partes que no coinciden y después incorporar el segmento al texto meta. El grado de semejanza de una coincidencia parcial puede variar desde un 1% hasta un 99% y existe la posibilidad de que el sistema localice más de un segmento similar. En ese caso, el propio sistema ofrecerá todos los segmentos por orden de semejanza.
- Coincidencias de términos: los sistemas que operan junto a un gestor terminológico dentro de un paquete suelen emplear sus bases de datos bilingües para ofrecer los términos al traductor en el caso de que coincidan con los términos de los nuevos segmentos que se quieren traducir. La base de datos terminológica actúa como un diccionario automático que sugiere equivalentes de traducción a todos los términos que estén registrados en esta.
- Coincidencias de «subsegmentos»: este tipo de coincidencias solo se incluyen en los sistemas de memorias de traducción más desarrollados. Lo que se compara en este caso son pequeños fragmentos de los segmentos, siempre que estos fragmentos tengan un alto grado de semejanza con

respecto al que se va a traducir. Esta coincidencia es similar a la coincidencia de términos, pero no se limita solo a localizar términos registrados, sino fragmentos que pueden no ser especializados.

- No coincidencias: como no hay ningún texto idéntico, puede ocurrir que haya segmentos para los que no se encuentre ninguna coincidencia. En estos casos, el traductor deberá traducir el segmento completo. Una vez traducido el segmento, este puede ser añadido a la memoria para ser recuperado en futuras ocasiones.

### 3.2. Alineador

Somers (2003: 33) analiza los posibles métodos para hacerse con memorias de traducción eficientes. Uno de ellos consiste en desarrollarlas a medida que se van realizando traducciones, insertándolas en una base de datos a la que se puede recurrir posteriormente para realizar otras traducciones del mismo tema. Este método puede resultar muy lento, pues sería necesario realizar una gran cantidad de traducciones para llegar a componer una memoria consistente.

Con el desarrollo de internet, compartir recursos es cada vez más común. Cada vez resulta más fácil intercambiar memorias de traducción entre usuarios. El hecho de que la mayoría los desarrolladores de *software* acepten el mismo formato de memorias de traducción (*Translation Memory Exchange* o *TMX*) facilita aún más las cosas.

Sin embargo, hay un tercer método, más complejo a nivel técnico, que consiste en construir la base de datos partiendo de un recopilatorio de textos originales con sus traducciones. Se trata del proceso de alineación, que consiste en hacer coincidir el texto origen y su traducción segmento a segmento. Estos segmentos paralelos se convierten en pares de traducción. Si la traducción es sencilla y fiel al original, este proceso no resulta complejo.

Sin embargo, hay tres factores que pueden hacer del alineado un proceso complicado: el primero es la dificultad que reside en definir cuándo comienza y termina una frase. Otra dificultad es que, dependiendo del par de lenguas, una frase puede no corresponder con su traducción a otra lengua. El tercer factor es el posible hecho de que una traducción tenga diferente distribución que su texto original.

Hay muchos textos cuyas traducciones difieren de los originales por su longitud, por lo que resulta útil poder ajustar el alineado. También es frecuente encontrar frases ordenadas de forma diferente en las dos lenguas; estos sistemas permiten cruzar

segmentos de forma manual. Debido a estos problemas, muchos sistemas de herramientas de traducción ofrecen herramientas de alineado que realizan una primera alineación pero permiten al traductor corregir los segmentos alineados propuestos (Somers, 2003: 34).

### **3.3. Programas de estadísticas**

Algunos sistemas de memorias de traducción incluyen una herramienta de extracción o identificación terminológica. La mayoría de herramientas de extracción terminológica son monolingües e intentan analizar los textos de origen con el fin de identificar los términos candidatos. Sin embargo, desde hace algunos años se están desarrollando algunas herramientas bilingües que analizan los textos origen y meta tratando de identificar los términos potenciales con sus equivalentes. Este método ayudaría al traductor a construir una base de datos terminológica de forma muy rápida (a pesar de que la lista creada automáticamente debería ser verificada).

La labor de las herramientas de extracción de términos consiste en identificar las unidades multipalabra mediante dos enfoques diferentes: el lingüístico y el estadístico.

#### **3.3.1. Aproximación o enfoque lingüístico**

Este enfoque utiliza la aproximación para identificar combinaciones de palabras que concuerdan con patrones hablados. Con este enfoque, cada palabra del texto original es etiquetada con su parte apropiada de la oración. Cuando el texto completo es etiquetado, la herramienta de extracción de términos identifica todos los patrones emparejados (por ejemplo, sustantivo + sustantivo o sustantivo + adjetivo) como términos potenciales.

Sin embargo, no todos los patrones pueden localizarse perfectamente, pues no todas las combinaciones que cumplen los patrones pueden calificarse como términos. Otro problema surge a la hora de localizar términos cuyos patrones no son comunes por su complejidad.

Además, el enfoque lingüístico depende en gran medida de las particularidades del lenguaje, es decir, de la formación de términos en cada idioma. En inglés resulta común encontrar términos “adjetivo + sustantivo” o “sustantivo + sustantivo”, mientras que en francés y español es más común encontrar el patrón “sustantivo + adjetivo”. Por eso la mayoría de herramientas de extracción de términos suelen estar diseñadas para trabajar en inglés o en otras lenguas similares desde el punto de vista lingüístico.

### 3.3.2. Aproximación estadística

Este es el enfoque más sencillo y extendido de la extracción de términos. El trabajo de esta herramienta consiste en buscar series de elementos léxicos repetidos. El usuario puede determinar el intervalo de frecuencia (el número de veces que una serie de elementos debe ser repetida). Un término debe aparecer al menos dos veces en el mismo orden para ser reconocido como término.

Esto puede, sin embargo, conducir a problemas, pues el lenguaje contiene muchas repeticiones y no todas ellas pueden calificarse como términos. Además, en muchas situaciones se pueden identificar términos que contienen “ruido”, es decir, palabras que nunca podrían considerarse términos. Para evitar esto, las listas de parada o *stoplist* resultan muy útiles.

Otro inconveniente del enfoque estadístico es que puede haber términos que no se repitan y no cumplan con el umbral de frecuencia mínimo y, sin embargo, deberían ser considerados términos.

Otro método estadístico para identificar posibles términos es calcular la información mutua<sup>2</sup>: si dos elementos léxicos aparecen juntos con más frecuencia que separados, la unidad multipalabra podría ser un candidato a término que el usuario se encargaría de corroborar o desechar.

La mayor ventaja del enfoque estadístico en las herramientas de extracción de términos es que estas no dependen del idioma. Así, una herramienta estadística de extracción de términos puede utilizarse para procesar textos en varias lenguas (Bowker, 2002: 82-86).

### 3.4. Gestor terminológico

Un trabajo muy importante en el proceso de traducción consiste en identificar equivalentes para términos especializados. Campos especializados como informática, comercio, derecho o medicina tienen mucha terminología específica. Encontrar los términos adecuados en cada traducción consume mucho tiempo y, con las presiones y la velocidad a la que se trabaja actualmente, los traductores profesionales no quieren repetir esta tarea cada vez que comienzan una nueva traducción. Un sistema de gestión terminológica ayuda al profesional dándole la posibilidad de almacenar, recuperar y actualizar términos. Esto hace mucho más rápido y menos costoso el trabajo del traductor.

---

<sup>2</sup> En teoría de la probabilidad y en teoría de la información, la información mutua de dos variables aleatorias es la cantidad que mide la dependencia mutua de estas dos variables. Se mide en bits. (Docsetools, 2015)

El resultado son traducciones con mayor consistencia terminológica y calidad lingüística, sin errores de comunicación.

Los inicios de los sistemas de gestión terminológica se remontan a la década de 1960, cuando se desarrollaron grandes bancos de términos como *Eurodicautom*, *Termium* y *The Banque de terminologie du Québec*.

En 1980, con el desarrollo de los ordenadores personales, esta herramienta fue una de las primeras disponibles para traductores. Sin embargo, estos gestores solo funcionaban en ordenadores individuales y no se podían compartir fácilmente.

Actualmente, el software es mucho más flexible y potente. Las funciones más importantes de las bases de datos terminológicas son las siguientes:

#### **3.4.1. Almacenamiento**

La función más importante de los gestores de terminología es que estos actúan como repositorios donde la información se almacena de forma onomasiológica, lo que permite localizar términos en direcciones múltiples. También son flexibles en el tipo y cantidad de información que permiten almacenar en cada entrada terminológica. Pudiendo crear campos libremente en una base de datos terminológica, los campos más comunes en cada entrada son:

- Término (DE)
- Término (ES)
- Campo temático
- Definición
- Contexto
- Sinónimos
- Fuente
- Comentarios
- Información administrativa adicional (fecha, autor, etc.)

Para crear gestores terminológicos efectivos es necesario establecer una nomenclatura preliminar que resulte fácil de visualizar y haga las búsquedas más sistemáticas. También es recomendable definir los valores de cada campo y tener claros los criterios que se van a seguir. Las fichas terminológicas contienen información conceptual (campos y subcampos, definiciones y contexto), lingüística (término,

abreviaturas, etc.), pragmática (observaciones y ejemplos) e información de gestión (número de término, fecha de creación, autor, fuentes, etc)

### **3.4.2. Recuperación**

Una vez que la terminología ha sido almacenada, la labor del traductor es recuperar esta información cuando sea necesaria. Para ello se utilizan mecanismos de búsqueda y recuperación. Uno de ellos consiste en tratar de encontrar coincidencias exactas mediante patrones de búsqueda. Un ejemplo sería utilizar "*comput\**" para localizar términos para "*computer*" o "*computing*". También se pueden utilizar métodos para localizar coincidencias parciales, que permiten encontrar registros morfológicamente variables, por ejemplo, palabras con prefijos o sufijos, palabras compuestas, etc.

### **3.4.3. Reconocimiento activo de terminología:**

Esta característica es propia de las bases de datos que funcionan dentro de un paquete de herramientas de entorno de traducción. Se trata de una característica que hace que el gestor terminológico funcione como un diccionario automático. Mientras el traductor se desliza a través del texto, la función de reconocimiento terminológico compara los términos con los que se encuentra con los términos de la base de datos; si encuentra una coincidencia, la herramienta lo notifica para que el traductor lo consulte y decida si es adecuado o no.

### **3.4.4. Extracción de términos:**

También llamado sistema de identificación de términos. Esta herramienta intenta identificar las unidades lingüísticas de más de una palabra mediante aproximación lingüística (combinaciones de palabras que coinciden con patrones establecidos o hablados) o aproximación estadística (series repetidas de elementos léxicos). Los inconvenientes de estos métodos de aproximación residen en que presentan errores, bien por ruido o por silencios, por lo que el factor humano es imprescindible en esta tarea. (Ver apartado 3.3 Programas de estadísticas)

Como afirma Bowker, el avance de la tecnología ha sido y es el responsable del cambio que se ha producido en el desarrollo de los sistemas de gestión terminológica. Pero también influye la evolución de la profesión. Los actuales sistemas de gestión terminológica permiten insertar información detallada en cada entrada del glosario; sin embargo, los plazos de la industria de esta profesión son tan cortos que, a menudo, la preparación de glosarios detallados resulta imposible.

El hecho de que muchas herramientas de este tipo vengan integradas en paquetes de más herramientas facilita el trabajo de traductor de forma que, al estar vinculada una herramienta a otra, es posible insertar la terminología de las bases de datos automáticamente en el texto de destino. Además, estos sistemas de gestión terminológica permiten el intercambio de información con otros usuarios, al estar conectados a la red. Así, diferentes usuarios pueden acceder y contribuir a una base a largo plazo.

Actualmente hay numerosos sistemas de gestión terminológica en el mercado. El costo de estos productos varía dependiendo del paquete en el que van incluidos, pero también existen programas de «*software libre*» que funcionan de forma muy similar a los no gratuitos aunque presentan limitaciones de memoria y potencia de procesamiento (Bowker, 2002: 77-91).

### 3.5. Sistemas de localización

Las memorias de traducción también son un componente importante de otras herramientas de traducción asistida: las herramientas de localización de software. La localización, también conocida como L10n, es el proceso de adaptación de un producto a la situación del mercado de llegada. Esto incluye no solo traducir los textos y tablas o gráficos que pudiera haber, sino adaptar el producto a las normas culturales del mercado local. Aunque la localización no se limita solo a adaptación de *software*, es en esta área donde más demanda existe (López Sánchez, 2014: 21).

Las herramientas utilizadas en el proceso de localización de *software* combinan diferentes aplicaciones de traducción asistida en una sola. La mayoría de ellas ofrecen funciones de extracción terminológica y sistemas de gestión de glosarios, También suelen ofrecer un cuadro de diálogo en modo WYSIWYG<sup>3</sup> que permite a los traductores revisar los elementos localizados al instante, como si estuvieran utilizando el *software*.

Una característica común de todas las herramientas de localización es que todas ellas ofrecen la posibilidad de trabajar directamente en los archivos mientras el código fuente permanece protegido. Además, la mayoría de estas herramientas contiene rutinas predefinidas que permiten al usuario detectar errores comunes de combinaciones o posiciones de símbolos no válidas (Austermühl, 2001: 146).

Observando el mercado actual de herramientas de localización podemos encontrarnos diversos programas para la localización de imágenes, tarea desempeñada

---

<sup>3</sup> WYSIWYG: acrónimo de “*What You See Is What You Get*”. Sistema que emplean las aplicaciones o interfaces gráficas cuyo principal objetivo consiste en mostrar en las pantallas de los ordenadores un reflejo fiel al producto final; se presenta al usuario la página Web tal y como se verá al final del trabajo. (López y Ponce, 2014 )

hasta ahora por los diseñadores gráficos y para la que los traductores cada vez tienen más competencias.

Para desarrollar productos de *software* de manera multilingüe y que estos tengan salida nivel global, la localización incluye actividades de internacionalización (I18N) y globalización (G11N) (López Sánchez, 2014: 21).

La Asociación de Estándares de Localización (LISA) define *internacionalización* como:

*The process of designing an application so that the feature design and code design don't make assumptions that are based on a single language or locale. Internationalization simplifies the creation of different language editions of a program. One goal of internationalization is to ensure that international conventions (including rules for sorting strings and for formatting dates, times, numbers, and currencies) are supported. Another goal is to design the product in such a way that users will experience consistent appearance and functionality across different language editions of a product. Internationalization is often abbreviated as I18N. The abbreviation is formed using the first and last letters of the word (I, N) and the number 18, which specifies the number of letters between the I and the N (Alonso y de la Cova, 2013: 35).*

Esta misma asociación se refiere al término *globalización* como:

*The process of making all the necessary technical, financial, managerial, personnel, marketing and other enterprise decisions necessary to facilitate international business (Alonso y de la Cova, 2013: 35).*

La localización de páginas web requiere conocimientos sobre las herramientas de localización, pero también sobre el diseño web y los lenguajes más utilizados en este ámbito (HTML y XHTML). En localización de *software*, la integración de herramientas tecnológicas de traducción siempre ha sido mayor que en traducción tradicional. Debido a la naturaleza repetitiva de las páginas web y productos de *software*, la reutilización de traducciones existentes resulta una gran ventaja, por eso el uso de memorias de traducción está muy extendido en este campo. También es muy común el uso de herramientas TAO de extracción terminológica y sistemas de traducción automática (Esselink, 2000 en Somers, 2003: 71).

### 3.6. Corpus electrónicos y generadores de concordancias

En el sentido más amplio de la palabra, un corpus es «una colección de textos del lenguaje natural, seleccionados para caracterizar el estado de una lengua o una variedad de la misma» (Sinclair, 1991: 171 en Vargas, 2002).

Hasta la llegada de las tecnologías, todos los corpus que se generaban estaban en formato impreso. Su recopilación requería un gran trabajo y su extensión no era comparable a la actual. Pero desde entonces han surgido nuevos formatos de corpus entre los que destaca el electrónico. Este tipo de corpus presenta muchas ventajas frente al impreso pues, dado que la mayoría de los textos generados actualmente están en formato digital, resulta mucho más fácil recopilar corpus de este tipo. Además, los actuales corpus electrónicos se pueden modificar fácilmente gracias a los gestores de corpus y también pueden distribuirse a varios usuarios mediante internet. Incluso hay corpus en plataforma online y que pueden ser consultados por todos sus usuarios al mismo tiempo.

A pesar de la gran cantidad de información valiosa que contienen, los corpus impresos tienen también muchas limitaciones. Trabajar con corpus impresos implica un gasto de tiempo que se vería reflejado en la traducción a la hora de cumplir los plazos de entrega.

Los textos almacenados de formato electrónico resultan mucho más fáciles de recopilar, procesar y analizar. La principal ventaja de un corpus electrónico es que este puede ser manipulado desde un ordenador. Otra ventaja es que se pueden compartir fácilmente a través de internet o mediante soportes de almacenamiento. También es posible digitalizar documentos en papel para insertarlos al corpus mediante programas de reconocimiento de voz o de reconocimiento óptico de caracteres (OCR). Una vez que se haya creado el corpus electrónico, entran en juego las diferentes herramientas de *software*, conocidas como herramientas de análisis de corpus, que ayudan al traductor a acceder, analizar y manipular los datos.

Es necesario aclarar que un corpus no es una colección de textos aleatorios. Los textos escogidos para realizarlo son seleccionados a través de criterios específicos para representar a subconjuntos del lenguaje. Así, los corpus tienden distinguirse por una temática concreta y los traductores pueden elaborar diferentes corpus en función de los proyectos en los que trabajen (Bowker, 2002: 43).

Si bien es cierto que los programas de análisis de corpus no se consideran herramientas TAO porque no asisten al traductor en su tarea, sí se consideran herramientas de entorno de traducción, pues el trabajo que realizan es previo a la traducción y sirve de apoyo a la labor del traductor.

Los generadores de concordancias son una aplicación incluida en los gestores de corpus electrónicos que nos permiten localizar y extraer la terminología y fraseología de los textos que conforman el corpus. Como hemos dicho, es cierto que no se consideran una herramienta de entorno de traducción, pero muchas herramientas de TAO las incluyen entre sus aplicaciones puesto que facilitan la creación de corpus y agilizan el trabajo de sus usuarios.

### 3.6.1. Tipos de corpus electrónicos

Sara Laviosa (1998: 106) en Somers (2003) realiza una clasificación de los tipos de corpus electrónicos en los que se basa un estudio traductológico: un *corpus paralelo bilingüe unidireccional* consiste en uno o más textos en una lengua A y sus traducciones en una lengua B, un *corpus paralelo bilingüe bidireccional* contiene textos originales y traducciones en ambas lenguas. Por su parte, un *corpus comparable bilingüe* consiste en dos colecciones de textos originales en lenguas A y B, similares en género textual, función y tema. Y por último, un *corpus comparable monolingüe* consiste en dos colecciones de textos en una lengua; la primera está compuesta de traducciones de una o varias lenguas originales, mientras que la segunda estaría formada por textos originales con una estructura y composición similar a los de la primera colección.

### 3.6.2. Herramientas de análisis de corpus

La principal diferencia entre un corpus electrónico y uno impreso es que el primero ofrece muchas más posibilidades con respecto al segundo. Para sacar partido a todas estas propiedades, los traductores deben disponer de una serie de herramientas que les ayuden a manipular e investigar los conceptos del corpus.

La mayoría de herramientas de análisis de corpus contienen una serie de características que permiten generar y manipular listas de frecuencia de palabras, concordancias y colocaciones

- Listas de frecuencias de palabras: Se trata de la cualidad más básica ofrecida por una herramienta de análisis de texto. Permite a los usuarios saber cuántas palabras diferentes hay en un corpus y con qué frecuencia se repiten. Además, se pueden manipular según las preferencias del usuario. Hay algunas herramientas que permiten crear listas lematizadas, en las que se pueden agrupar las palabras relacionadas como, por ejemplo, los diferentes tiempos de un mismo verbo. También existe la posibilidad incluir listas de parada o *stoplist*, que contienen todos los elementos que

pretenden ignorarse por falta de contenido semántico, como artículos, conjunciones o preposiciones.

- Concordancias: una concordancia es un listado de todas las ocurrencias de una palabra dada en un corpus acompañada de su contexto (Sinclair, 1991 en Holtsberg, 2001:135). Sirve para recuperar todos los casos aclarar el mejor uso de ciertos términos viéndolos uno o más contextos. Algunos programas de análisis de corpus solo generan concordancias de forma monolingüe, pero la mayoría son capaces de trabajar con textos de varias lenguas, como se describe previamente, alineando los textos de una lengua A con los de la lengua B. Como ya hemos mencionado, hay diversas herramientas de entorno de traducción que ya incorporan generadores de concordancias entre sus herramientas, pues estos resultan muy útiles para trabajar sobre las memorias de traducción que, al fin y al cabo, son corpus paralelos bilingües.
- Colocaciones: muchas herramientas de análisis de corpus son capaces de detectar colocaciones, que son patrones de palabras característicos y que aparecen con una mayor frecuencia que la que tendrían de no ser colocaciones. Para calcular la probabilidad de que dos palabras sean colocaciones, se emplea la fórmula de la información mutua (Ver 3.3. Programas de estadísticas), que compara la probabilidad de que dos palabras aparezcan unidas frente a la probabilidad de que aparezcan separadas (Bowker, 2002: 46-67).

### 3.7. Editor y correctores ortográficos

Todas las herramientas de entorno de traducción actuales cuentan con un editor de texto que resulta el espacio de trabajo más importante del traductor, pues es aquí donde se reproduce el resultado de todo el trabajo realizado con todas las herramientas descritas anteriormente, es decir, la traducción. Esta herramienta permite al traductor escribir y modificar su traducción constantemente.

En las actuales herramientas de traducción asistida, los editores de texto están segmentados en casillas, alineadas con el texto origen, de forma que por cada segmento del documento original, el editor ofrece un espacio para que el usuario escriba su traducción. Es en esas casillas donde la memoria de traducción inserta las coincidencias en caso de que las haya.

La mayoría de herramientas electrónicas de procesamiento de textos cuentan con correctores ortográficos integrados disponibles para varias lenguas, que son diccionarios personalizados que detectan los errores de ortografía del texto en el que se trabaja (y algunos errores gramaticales) y sugieren las opciones correctas de ortografía para que el usuario pueda rectificar.

El corrector ortográfico requiere como punto de partida dos bases de datos: un diccionario principal y un diccionario secundario o del usuario. El primero es una versión abreviada de un diccionario académico de la lengua, mientras que el segundo es el que forma el usuario con las palabras que va introduciendo en función de sus necesidades (Ariza y Tapia, 1998: 375).

Muchas herramientas de traducción emplean el corrector ortográfico del procesador de texto Microsoft Word, pues es el más compatible de todos los correctores. Sin embargo, hay ciertas herramientas TAO que utilizan otro tipo de correctores o incluso permiten combinar más de un corrector para obtener mejores resultados. Es el caso de Trados Studio 2014 o la versión Pro de Wordfast (ver apartado 4.2 Clasificación de las herramientas TAO según licencia y código), que llevan integrado el corrector Hunspell, pero permiten hacer uso del corrector MS Word (Office) y combinarlos dependiendo de las características de los textos que se van a traducir. OmegaT, sin embargo, se basa únicamente en el corrector ortográfico Hunspell (el corrector que utilizan Firefox u Openoffice entre otros).

### **3.8. Diccionario de autosugerencias**

Los diccionarios de autosugerencias son alternativas a las memorias de traducción o a las bases terminológicas que se utilizan en los proyectos de traducción. Por el momento, se trata de diccionarios propios de la herramienta de entorno de traducción Trados Studio, pues no consta información de que alguna otra herramienta TAO cuente con una opción similar.

Estos diccionarios son también bases de datos que contienen unidades de traducción, pero en lugar de sugerir unidades de traducción, sugieren palabras o grupos de palabras de estas unidades que pueden ser el equivalente que buscamos o no. De esta forma, los traductores podemos aceptar o rechazar estas sugerencias.

Este tipo de diccionarios agilizan el proceso de la traducción, pues ya no es necesario escribir todas las palabras, basta con escribir el principio. Además, ayudan a evitar los errores de ortografía y la repetición u omisión de las letras.

Se crean a partir de memorias de traducción que pueden estar tanto en formato de Trados Studio (SDLTM) como en TMX, que es un formato más común. Para crear un diccionario de autosugerencias es necesario contar con una memoria extensa (un mínimo de 25.000 unidades o segmentos) y, a diferencia de las memorias de traducción, no se actualizan automáticamente (Arranz, 2013).

Uno de los objetivos de la parte práctica de este proyecto es la producción de un diccionario de autosugerencias. Durante este proceso hemos tenido la oportunidad de valorar las ventajas y los inconvenientes de este tipo de diccionarios. La principal ventaja de estos diccionarios es que agilizan notablemente el proceso de traducción y, al sugerirte las palabras adecuadas, ayudan a evitar los errores ortográficos. Además, la mayoría de los diccionarios se crean a partir de memorias de traducción de formatos comunes como TMX o, como en este caso, SDLTM (memorias de Trados Studio).

Sin embargo, también hemos podido observar que los diccionarios no se actualizan a medida que ampliamos la memoria de traducción, sino que hay que generar uno diferente cada vez que consideremos que la memoria está más completa. Además, para crear un diccionario consistente es necesario que la memoria de traducción sea lo bastante grande para obtener un número de sugerencias que resulte efectivo. Se necesitan más de 20.000 unidades para que esto ocurra, por lo que cualquier a traductor que genere su propia memoria le llevará un tiempo recopilar todas esas unidades y poder crear un buen diccionario.

Una vez aclarado que es necesario actualizar estos diccionarios cada cierto tiempo, la cuestión estaría ahora en determinar ese tiempo. Para ello, podría utilizarse una norma periódica (cada tres o seis meses dependiendo del trabajo del traductor). Otra opción sería basarse en las unidades de traducción agregadas a la memoria desde la última actualización del diccionario. Esto lo podemos comprobar en la configuración de la memoria de traducción, donde se refleja cuantas unidades de traducción incluye nuestra memoria. También podemos basarnos en las coincidencias encontradas, pues cuantas más unidades de traducción hayan sido reutilizadas, la memoria será menos rica en nuevos términos, por tanto, actualizar el diccionario no será tan necesario.

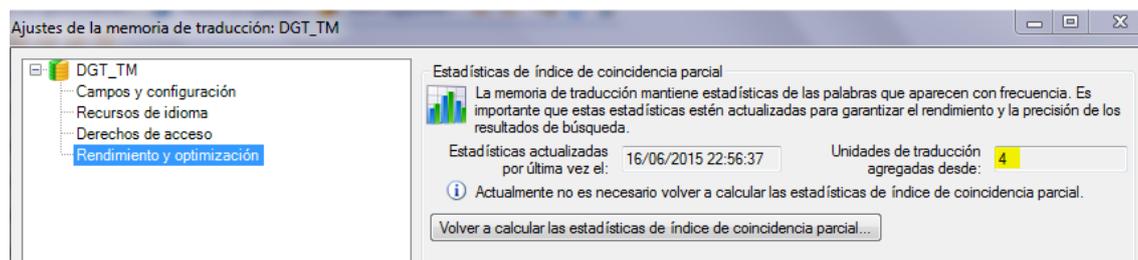


Figura 1: Consulta de las coincidencias parciales localizadas en la MT

También es cierto que no todas las sugerencias ofrecidas por este tipo de diccionarios son efectivas; los diccionarios de autosugerencias no distinguen el género ni el número, así como las mayúsculas y minúsculas, por lo que es necesario revisar cada sugerencia antes de aceptarla. Un ejemplo es el caso del alemán, una lengua con muchas declinaciones, géneros y números. En este caso, un diccionario de autosugerencias puede ofrecer una gran cantidad de propuestas de la misma palabra declinada en varias formas para un único significado. Sin embargo, si hablamos del inglés (sin géneros), las sugerencias serán mucho más concretas.

Pero a pesar de los inconvenientes que tienen, es cierto que un buen diccionario generado a partir de una memoria de traducción extensa puede resultar muy efectivo a la hora de agilizar la tarea de un traductor.

## **4. Clasificación de las herramientas TAO**

Actualmente nos encontramos inmersos en la era de las tecnologías en la que día a día surgen nuevas aplicaciones y herramientas dirigidas a ayudar al usuario en el campo de trabajo en el que se encuentre. En la traducción también se ha desarrollado una gran cantidad de programas o *software* que facilitan y agilizan el trabajo del traductor. Hoy en día, resulta casi imposible pensar en un traductor que no haya invertido en un ordenador y en una serie de herramientas que le permitan trabajar con solvencia. Por tanto, resulta casi imprescindible estar al tanto de las novedades tecnológicas que se presentan y conocer todas las herramientas y programas de traducción disponibles en el mercado.

A continuación, realizaremos una clasificación de las herramientas de traducción asistida por ordenador atendiendo a diferentes factores.

### **4.1. Clasificación según su ubicación**

Con la evolución de internet, las herramientas de *software* en “la nube” están cogiendo impulso frente a las tradicionales que se almacenan en el disco duro del equipo del usuario. Tanto es así, que ya se están desarrollando varias herramientas TAO que funcionan con este sistema. Estas herramientas permiten trabajar de forma colaborativa dentro de un mismo documento, acceder a ellas desde cualquier dispositivo con conexión a internet y supone un gran ahorro en cuanto a inversión, pues la mayoría de ellas son gratuitas o más asequibles que las herramientas tradicionales. Además, estas herramientas se mantienen actualizadas y son compatibles con cualquier sistema operativo.

Sin embargo, la realidad actual prueba que los traductores continúan decantándose por instalar en su equipo las herramientas de traducción asistida por ordenador, puesto que estas novedosas herramientas requieren una conexión a internet continua y, como los archivos no se almacenan en el equipo del usuario, este pierde el pleno control sobre ellas. Además, estas aplicaciones tienden a ser menos potentes que las tradicionales, por lo que el rendimiento puede no resultar óptimo y el riesgo de saturación es mayor.

Aún así, la mayoría de traductores sí utilizan internet para sacar partido al resto de utilidades *online*, como las herramientas de almacenamiento, los procesadores de texto, buscadores, ciertos antivirus gratuitos y gestores de correo y mensajería.

En cuanto a las herramientas TAO que se almacenan en el equipo, encontramos los paquetes de herramientas lanzadas por SDL International (SDL Trados Studio 2009, 2011 o 2014) que incluyen gestores terminológicos (Multiterm) y herramientas de localización (Passolo). También destacan Wordfast Classic, Transit o Déjà Vu entre otros, así como la herramienta de entorno de traducción MemoQ, más reciente y sencilla. Como herramientas gratuitas destacan OmegaT, Across y Anaphraseus.

Sin embargo, por ahora no son tantas las herramientas desarrolladas online. Dejando a un lado todas las utilidades online imprescindibles para el traductor actual, como herramientas de traducción asistida por ordenador cabe mencionar Google Translator Kit, de carácter gratuito, desarrollado por el gigante Google. La versión gratuita y online de Wordfast Classic, Wordfast Anywhere, también resulta muy eficiente, pues se puede acceder a ella desde cualquier equipo (López Sánchez, 2014).

#### **4.2. Clasificación según licencia y código**

A pesar de la gran cantidad de herramientas de traducción asistida por ordenador disponible en el mercado, con el paso del tiempo, el *software* libre ha ido ganando peso poco a poco entre los usuarios de equipos informáticos en comparación con los programas comerciales. De hecho, es ya una realidad que muchas herramientas de software libre y gratuito han llegado a la altura de programas de pago.

Antes de realizar una clasificación siguiendo este criterio, es necesario aclarar lo que significa que un *software* es libre y que lo diferencia de un *software* propietario.

Un software libre es cualquier programa que permite a los usuarios ejecutarlo, copiarlo, distribuir copias a terceros, estudiar cómo funciona, modificarlo a su antojo y mejorarlo. Al adquirir un programa libre (tanto gratuito como de pago), se adquiere también el código fuente de este, de manera que el programa se puede adaptar a las necesidades

del usuario. Por lo general, las herramientas de carácter libre son de código abierto, mientras que las herramientas «no libres» utilizan un código cerrado y la divulgación de esto podría ser constituyente de delito en algunos países

El *software* propietario en cambio, es un programa acompañado de una licencia que permite al comprador ejecutarlo en un número determinado de ordenadores. Esto significa que nadie puede adquirir el programa en sí, sino que se paga por el derecho de poder ejecutarlo. Mientras, el programa continúa siendo propiedad de la empresa y solo esta puede realizar cambios en él, pues nadie más dispone del código fuente (González Piñero, 2004; 12).

Un programa propietario suele ofrecer un servicio técnico destinado a solventar problemas de instalación, configuración o uso, aunque no se hacen responsables de los errores o daños que el programa pudiera causar en el equipo del usuario, ni atienden las necesidades personales de estos de adaptar alguna parte de ese *software*. Sin embargo, aunque las empresas distribuidoras de *software* libre no se responsabilizan de ningún error que el programa pudiera tener, existen empresas que se dedican a proporcionar soporte técnico a compradores de *software* libre, modificándolo para adaptarlo a las necesidades del usuario de forma personalizada y eficiente.

Es indudable que, en los últimos años, la popularidad del *software* libre ha ido aumentando hasta convertirse en una alternativa en muchos sectores que, hasta hace pocos años, no conocían de su existencia. Este tipo de *software* se mantiene constantemente en desarrollo y cada día más empresas apuestan por el cambio puesto que las ventajas trascienden lo tecnológico; un *software* libre requiere mucho menos presupuesto debido a la inexistencia de licencias y la libertad de distribución permite conseguirlos gratuitamente de forma legal (Stallman, 2004).

Los programas de gestión de memorias de traducción de carácter libre más destacados son Omega T, Google Translator Toolkit, Wordfast Anywhere y Anaphraseus. Como programas comerciales o *software* propietario cabe mencionar a los gigantes SDL Trados y Wordfast Classic, MemoQ y Transit entre otros.

En cuanto a las herramientas de gestión terminológica libres, la más destacable es TermWiki, pues en este campo sigue habiendo una tendencia a utilizar *software* propietario, como TermStar, Multiterm o Xbench.

Sin embargo, a la hora de hacer referencia de programas de gestión de corpus y concordancias, la tendencia sigue la dirección opuesta; en este caso hay un mayor desarrollo programas de gestión de corpus como AntConc o System Quirk, frente a los reconocidos gestores comerciales como Word Smith Tools o MonoConc.

Por último, en el sector de la localización, los programas de traducción de *software* tienden a ser mayoritariamente comerciales: Passolo, Language Studio, Lingobit Localizer, RC-Win Trans o Visual son solo algunos de ellos, aunque también podemos encontrar herramientas libres o a precios muy bajos, como CatsCradle, Rainbow o Visual Translate (Tradumàtica, s.f.)

SOFTWARE DE GESTIÓN DE MEMORIAS DE TRADUCCIÓN	
LIBRE	PROPIETARIO
Google Translator Toolkit	Memsources
OmegaT	Déjà Vu
EsperantiloTM	Across
Open TM2	Wordfast Classic
Wordfast Anywhere	SDL Trados
Anaphraseus	Star Transit
Heartsome	Swordfish Translation Editor
Tumatxa	MemoQ
	Metatexis
	Similis
	Cafetran

SOFTWARE DE GESTIÓN TERMINOLÓGICA	
LIBRE	PROPIETARIO
TermWiki	TermStar
	Multiterm

Xbench
--------

SOFTWARE DE GESTIÓN DE CORPUS	
LIBRE	PROPIETARIO
AntConc	MonoConc
System Quirk	Word Smith Tools

SOFTWARE DE LOCALIZACIÓN	
LIBRE	PROPIETARIO
CatsCradle	Alchemy Catalyst
Rainbow	RapidTranslation
Visual Translate	Lingobit localizar
	Crowdin
	RC-WinTrans
	Language Studio
	NLocalize
	Sisulizer
	Mutilizer
	Crowdin
	StringDB
	Passolo
	App Translator

Globalization Image Assistant (GIA)
Web Budget

En definitiva, las herramientas de entorno de traducción son un instrumento muy necesario para los traductores actuales, cuya mayoría de encargos está en formato digital. Por ello es casi imprescindible conocer y mantenerse al tanto de las herramientas disponibles en el mercado y de las novedades que se presentan cada día, así como dominar al menos una herramienta de cada tipo. Hay que tener en cuenta que, debido a las necesidades de mercado, siempre hay algunas herramientas más solicitadas que otras. También hay herramientas a precios muy bajos y otras que requieren una inversión considerable. Lo que es cierto es que cuantas más herramientas se dominen, más amplio será el rango de trabajos que un traductor puede aceptar. En resumen, el conocimiento de las herramientas TAO supone un mayor éxito en el mundo profesional y, en la era de internet, cada vez resulta más fácil disponer de todos esos conocimientos y llevarlos a la práctica diaria.

## **CAPÍTULO 2: CREACIÓN DE UN DICCIONARIO DE AUTOSUGERENCIAS A PARTIR DE LA MEMORIA DE TRADUCCIÓN DE LA DGT**

La Dirección General de Traducción (DGT) es el servicio de traducción de la Comisión Europea, que trabaja con todas las lenguas oficiales de la Unión Europea. De forma pública y gratuita, este organismo pone a disposición de los usuarios la memoria de traducción que recoge todos los documentos del acervo comunitario traducidos en 24 lenguas.

En este apartado explicaremos qué estructura tiene esta memoria de traducción y cómo podemos obtenerla y modificarla para sacarle el máximo partido como traductores del par de lenguas DE>ES. Además, seguiremos los pasos necesarios para generar un diccionario de autosugerencias a partir de esta memoria, empleando un gestor de memorias de traducción (en este caso SDL Trados 2009).

### **1. La Dirección General de Traducción y su memoria de traducción**

#### **1.1. Qué es el *Acquis communautaire*?**

En español «acervo comunitario», es el conjunto de legislaciones y jurisprudencia de la Unión Europea. Además de las actualizaciones de declaraciones, resoluciones y tratados y objetivos políticos, este acervo contiene todas las normas originarias de la fundación de la Unión.

Todos los países que presentan su candidatura para adherirse a la Unión Europea tienen que aceptar este conjunto de normas. De esta forma, cada vez que un país entra en la UE, este integra toda la normativa comunitaria en su legislación nacional (Miller, 2011).

#### **1.2. La Dirección General de Traducción (DGT): estructura y organización**

Geográficamente, la DGT está situada en dos puntos clave, en Bruselas, Luxemburgo Y Estrasburgo. Cada una de las 23 lenguas cuenta con un departamento lingüístico propio. Así, hay departamentos de alemán, búlgaro, húngaro, inglés, irlandés, italiano, letón, lituano, maltés, neerlandés, polaco, portugués, rumano y sueco.

Además, existen otros departamentos como la Dirección de Servicios Lingüísticos Transversales, que se encarga de áreas especializadas como la traducción web, la edición, la documentación y las relaciones con los Estados miembros; la Dirección de Recursos, que se ocupa del personal, la informática, la financiación y la infraestructura física; y la Dirección de Estrategia de Traducción y Multilingüismo, encargada del flujo de trabajo y las cuestiones políticas.

Los aspectos de los que se encargan estos departamentos lingüísticos abordan desde la administración; la agricultura y desarrollo rural; competencia; fiscalidad y unión aduanera; educación y cultura; empleo y asuntos sociales; energía y transporte; empresa e industria; medio ambiente; relaciones exteriores; pesca y asuntos marítimos; sanidad y protección de los consumidores; sociedad de la información y medios de comunicación; mercado interior y servicios; asuntos económicos y financieros; asuntos jurídicos; política regional; investigación; estadísticas y comercio.

La DGT cuenta con 1750 traductores trabajando a tiempo completo. Sin embargo, para afrontar el alto nivel de demanda, la Dirección General de Traducción suele recurrir a la traducción externa. Los últimos datos (2008) revelan que un 26% de la producción total proviene de traductores externos (DGT, 2009).

### **1.3. La memoria de traducción de la DGT**

La Comisión Europea de la DGT creó esta memoria multilingüe en noviembre de 2007 con el fin de apoyar y fomentar el multilingüismo. La memoria comprende todos los documentos del acervo comunitario traducidos en 24 lenguas, dando lugar a 253 pares de lenguas y a 506 combinaciones de idiomas. En realidad se trata de varias memorias multilingües de libre acceso y descarga, pero cada usuario puede descargar una única memoria de traducción que incluya su par de lenguas de trabajo, de forma que resulte más práctico para él ya que, debido a su formato (.tmx), se puede modificar y adaptar a cualquier tipo de gestor de memorias de traducción (Steinberger et al. 2013).

### **1.4. Política de multilingüismo**

Uno de los principios de la Unión Europea es establecer una comunidad donde el multilingüismo sea fundamental y se garantice la diversidad cultural y lingüística. El objetivo propuesto por la política lingüística de la UE es que todos los ciudadanos sean capaces de dominar dos idiomas además de su lengua materna.

El respeto por la diversidad lingüística es, por tanto, fundamental y necesario para la apertura a otras culturas. De hecho, según la Carta de los Derechos Fundamentales de la UE (2000), «todo ciudadano tiene derecho a dirigirse por escrito a cualquiera de las instituciones u organismos contemplados en el presente artículo [24 del TFUE] o en el artículo 13 del TUE en una de las lenguas mencionadas en el apartado 1 del artículo 55 de dicho Tratado y recibir una contestación en esa misma lengua» (TFUE, 2012).

A pesar de que todas las lenguas oficiales tienen el mismo rango, no todos los textos se traducen a todas las lenguas oficiales. Por ejemplo, una carta a un individuo o un documento de carácter interno se enviará en una única lengua. Si un comité decide

trabajar en un número limitado de lenguas, deberá comunicarlo en todas las lenguas oficiales.

La Dirección General de Traducción (DGT) tiene la función de responder a las necesidades de la Comisión Europea en asesoramiento lingüístico y traducción para reforzar el multilingüismo. Por eso opera en todas las lenguas oficiales de la Unión e incorpora las nuevas lenguas utilizadas en los países que van incorporándose a esta.

Otro de los objetivos que se propone la DGT al crear esta memoria de traducción es acercar las políticas desarrolladas en la Unión Europea a todos sus ciudadanos. De esta forma, pretende fomentar la transparencia, la democracia y la igualdad (DGT, 2009).

### **1.5. Utilidades de la memoria de traducción de la DGT**

Aunque la función de los textos y traducciones que contiene la DGT es originalmente jurídica, esta colección de textos tiene muchos otros usos fuera de dicho dominio.

Las colecciones de textos paralelos han tenido una gran acogida, pues se ha hecho un gran esfuerzo por conseguir que estas sean públicas. Aún así, solo algunas de estas colecciones son accesibles vía internet (*EuroParl*, DGT-TM y JRC-Acquis entre otros).

El uso de colecciones paralelas es importante en la creación de modelos para Traducción Automática Estadística (SMT por sus siglas en inglés), que es una estrategia una versión de traducción automática mediante la cual se generan traducciones basadas en modelos estadísticos a partir del análisis de corpus de textos bilingües.

El hecho de que los corpus, diccionarios y demás herramientas en línea sean multilingües en lugar de bilingües hace que estas herramientas sean más útiles para los traductores que, normalmente, trabajan con más de una lengua de origen. Las colecciones de textos paralelos sirven, entre otras, para producir recursos multilingües léxicos y semánticos, como diccionarios u ontologías, para detectar plagios entre diferentes lenguas, para realizar clasificaciones de colocaciones y estructuras idiomáticas en diferentes lenguas, para comprobar la consistencia de nuevas traducciones, etc. (Steinberger, 2013).

En este capítulo aprovecharemos una de estas opciones para crear un diccionario de autosugerencias o *autosuggest* que nos permita sacar partido a toda la terminología y vocabulario general contenidos en la memoria de traducción puesto que, a diferencia de una memoria de traducción, el diccionario de autosugerencias nos da la opción de emplear palabras o grupos de palabras que coincidan con nuestros textos.

No obstante, es necesario aclarar que ni la memoria de la DGT ni el corpus del que extraeremos los términos más adelante (ver apartado creación de la lista de términos

(insertar referencia cruzada)) son direccionales, pues están formados por textos en varias lenguas y traducidos en todas las lenguas de la UE. Por lo tanto, resulta evidente que los textos que en esta memoria denominamos «texto origen DE» no lo son en realidad, sino que probablemente sean traducciones del inglés o de cualquier otra lengua de la Unión Europea (Cartoni y Meyer, 2133).

Por este motivo, esta memoria no resulta válida para estudios lingüísticos, pero sí para las labores prácticas de traducción, pues los traductores necesitan soluciones rápidas a los problemas sin importar demasiado si el origen de estas soluciones es original o no.

## **2. Descarga e instalación de la memoria de traducción**

### **2.1. Descarga de la memoria de traducción y del programa de extracción de pares de lenguas.**

La memoria de traducción de la DGT se encuentra disponible de forma gratuita en la página web de la Comisión Europea <<https://ec.europa.eu/jrc/en/language-technologies/dgt-translation-memory#c2744>>. La memoria está comprimida en una colección de archivos .zip, cada uno de ellos menor de 100 MB. No es necesario descomprimir los archivos, pues la herramienta de extracción de estos accede directamente al archivo comprimido.

En nuestro caso, nos limitaremos a descargar los archivos de 2013, puesto que tan solo con estos alcanzaremos 340.000 unidades de traducción. Hay que tener en cuenta qué gestor de memorias de traducción queremos utilizar, pues algunas herramientas solo permiten un número limitado de unidades de traducción. Es cierto que, cuantas más unidades de traducción tenga la memoria con la que trabajamos, más útil y eficaz será. Pero también es cierto que ocupará más espacio y podría ralentizar la herramienta y el equipo. En nuestro caso, tras realizar varios intentos tratando de descargar la memoria completa, es decir, todos los años desde su creación, concluimos en que era imposible descargarla de una sola vez, pues el ordenador (Intel Core I5, 500GB de memoria RAM) se saturaba y el proceso se ralentizaba hasta bloquearse por completo. Al realizar el proceso con solo un año (escogimos 2013 porque el tamaño del 2014 era muy superior y el año 2015 aún no estaba completo), este se completó en un máximo de 1 hora.

La herramienta de extracción de archivos que utilizaremos será TMXtract, cuya versión gratuita está disponible haciendo click en este link <<http://optima.jrc.it/Resources/TMXtract.jar>>. Este es el programa que nos servirá para

extraer los pares de lenguas que necesitemos, pues en los archivos originales se encuentran los documentos traducidos en todas las lenguas oficiales de la Unión Europea.

## 2.2. Extracción del par de lenguas DE>ES

Para extraer el par de lenguas con el que queremos trabajar, debemos ejecutar el programa TMXtract y seleccionar los archivos comprimidos uno a uno (cada archivo ocupa 100 MB y puede contener hasta 500 archivos TMX) cada archivo en la opción “*Select Input Files*”. También hay que especificar la carpeta de destino en “*Specify an output file*”.

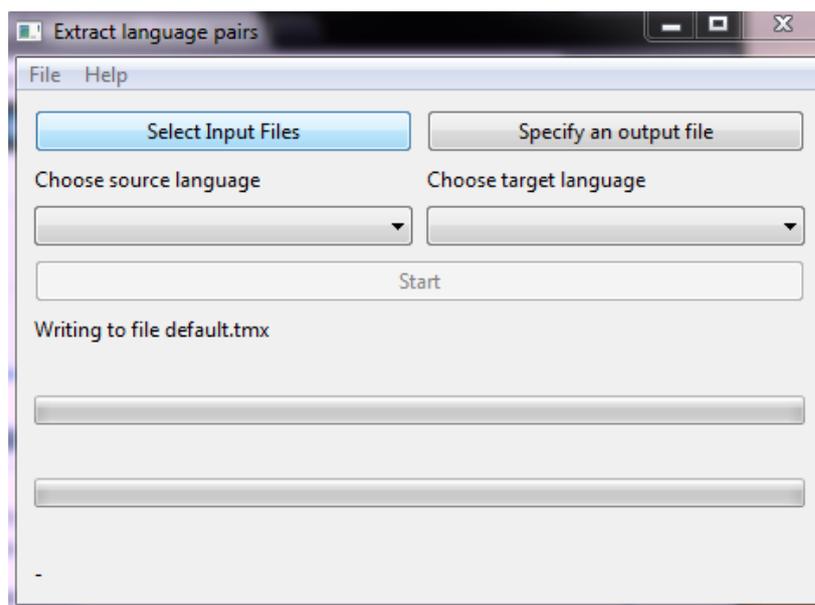


Figura 2: Selección de carpetas de origen y destino

A continuación, elegiremos las dos lenguas que vamos a extraer del documento, en este caso español y alemán, e iniciaremos el proceso. El resultado de este proceso es un archivo TMX que contendrá las unidades de traducción en alemán y español. Hay que repetir el proceso con todos los archivos comprimidos que hayamos descargado. Gómez, (2001: 2) define el formato TMX (*Translation Memory Exchange*) como:

*“TMX es un lenguaje que cumple las especificaciones XML, y cuyo propósito es proporcionar un estándar para el intercambio de las memorias de traducción. Cuando se esté trabajando con una utilidad y se desee pasar a trabajar con otra manteniendo la TM que se había ido recopilando, bastará con exportarla a formato TMX, e importarla en la nueva utilidad. Para ello es necesario que todas las utilidades soporten dicho formato: a 2001, puede decirse que ya hemos llegado a esta situación, puesto que actualmente las herramientas más importantes del mercado admiten la importación y exportación de memorias en TMX, si bien en distintos grados”* (Gómez, 2001: 2).

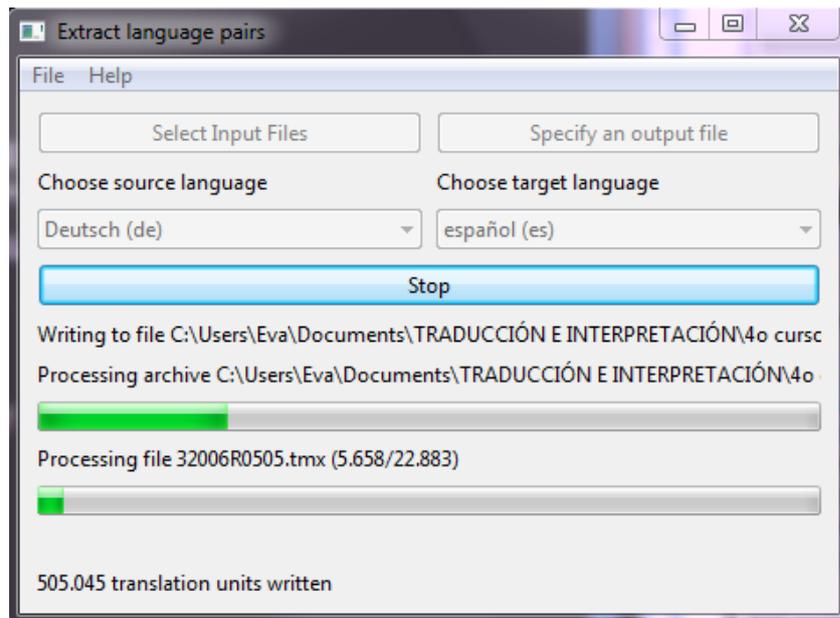


Figura 3: Selección del par de lenguas (DE y ES)

### 2.3. Creación de la memoria de traducción en SDL Trados 2009

En primer lugar, es necesario aclarar que utilizamos SDL Trados Studio (2009) y no otra herramienta TAO u otra versión de Trados Studio porque esta es la herramienta con la que trabajamos en toda la carrera y es, además, una de las más solicitadas por los clientes en el mercado de la traducción.

El siguiente paso será crear una memoria de traducción *.sdltm* que en un principio estará vacía y a la que iremos importando los archivos TMX que hemos generado previamente con la herramienta de extracción de pares de lenguas.

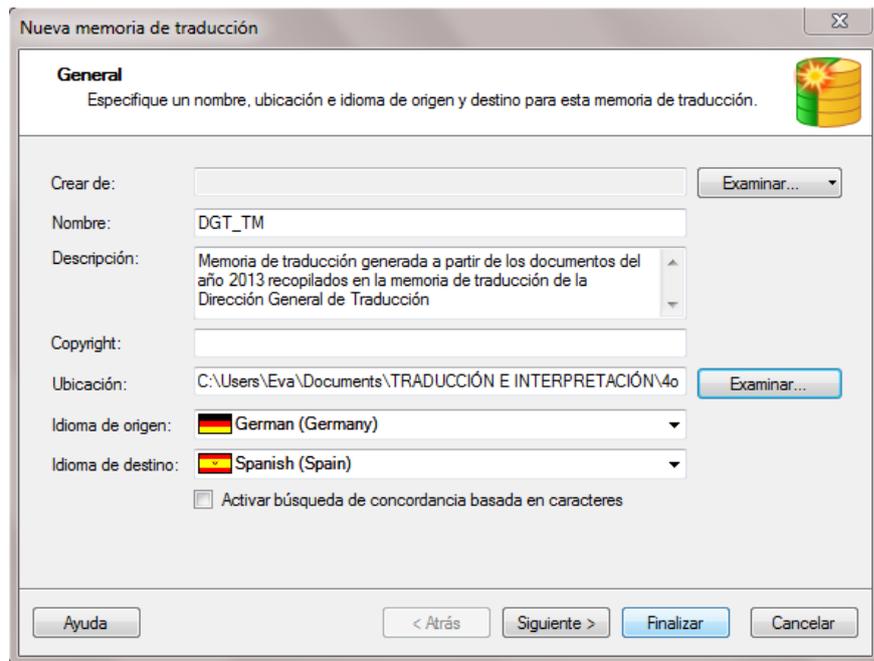


Figura 4: Creación de la memoria de traducción en SDL Trados

Una vez creada la memoria de traducción, importaremos todos los archivos descargados. Este proceso se realiza con el gestor de bases de datos de memorias de traducción o «Vista de memorias de traducción». El proceso puede durar varios minutos.

DE_ES_2013_05.tmx	16/04/2015 11:50	Archivo TMX	51.069 KB
DE-ES_2013_01.tmx	16/04/2015 19:45	Archivo TMX	64.085 KB
DE-ES_2013_02.tmx	16/04/2015 19:19	Archivo TMX	52.796 KB
DE-ES_2013_03.tmx	16/04/2015 12:24	Archivo TMX	57.882 KB
DE-ES_2013_04.tmx	16/04/2015 12:04	Archivo TMX	61.240 KB
DE-ES_2013_06.tmx	16/04/2015 11:38	Archivo TMX	54.878 KB
DE-ES_2013_07.tmx	16/04/2015 10:55	Archivo TMX	32.544 KB

Figura 5: Lista de los archivos TMX extraídos

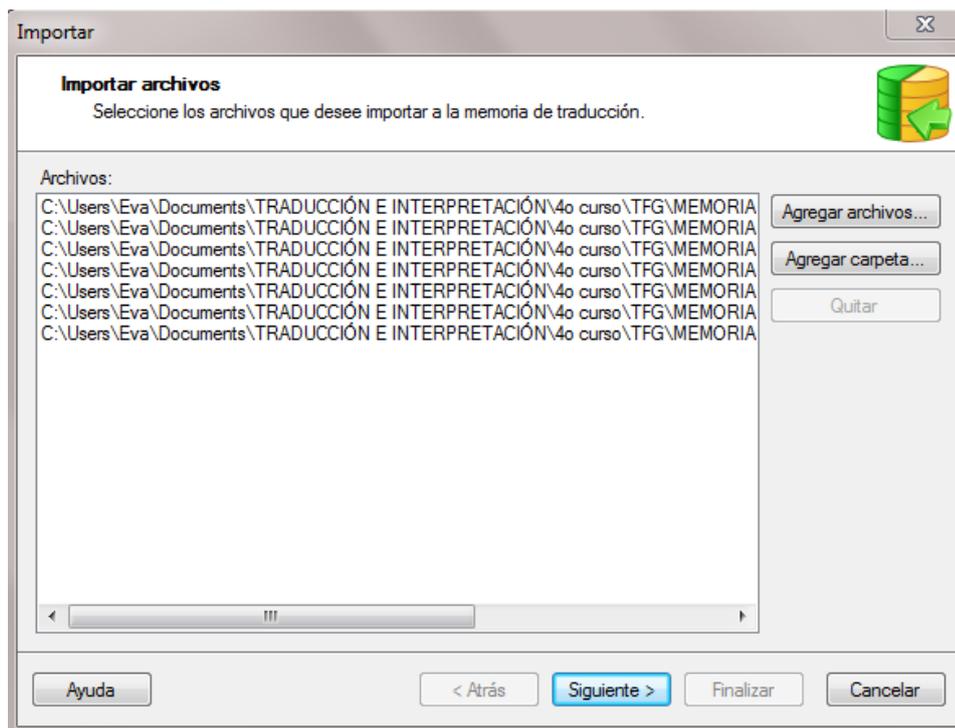


Figura 6: Importación de los archivos TMX a la memoria de traducción

El resultado, en este caso, es una memoria de traducción en formato *.sdltm* con los pares de lenguas DE>ES y consta de 342.298 unidades de traducción. Si quisiéramos ampliarla, simplemente habría que repetir este proceso, descargando desde la página de la Comisión Europea los nuevos archivos que queramos añadir a la memoria. También es posible editar la descripción de la memoria, en el caso de que la temática o el contenido de esta varíen, en la pestaña de configuración (botón derecho sobre la memoria de traducción y opción «configuración»).

Nombre:	DGT_TM
Descripción:	Memoria de traducción generada a partir de los documentos del año 2013 recopilados en la memoria de traducción de la Dirección General de Traducción
Copyright:	
Ubicación:	C:\Users\Eva\Documents\TRADUCCIÓN E INTERPRETACIÓN\4o curso\TFG
Idioma de origen:	Alemán (Alemania)
Idioma de destino:	Español (España)
	<input type="checkbox"/> Activar búsqueda de concordancia basada en caracteres
Creado por:	Eva-PC\Eva
Creado:	16/04/2015
Unidades de traducción:	342298

Figura 7: Descripción de la memoria de traducción resultante

### 3. Generar el diccionario de autosugerencias

El paso para generar el diccionario de autosugerencias es muy sencillo. A partir de la memoria de traducción creada en Trados, en la pestaña de “herramientas” de la barra de herramientas general existe la opción de “generar diccionario Autosuggest”. Cuantas más unidades queramos incluir en el diccionario, más frases estarán disponibles para ser sugeridas en el futuro, pero el proceso de generación del diccionario ocupará más memoria y tardará más tiempo en finalizar. En este caso hemos limitado las sugerencias del diccionario a unas 300.000 unidades (30 minutos con un Intel Core I5 de 500 GB de RAM). El resultado debe ser un archivo con formato BPM.

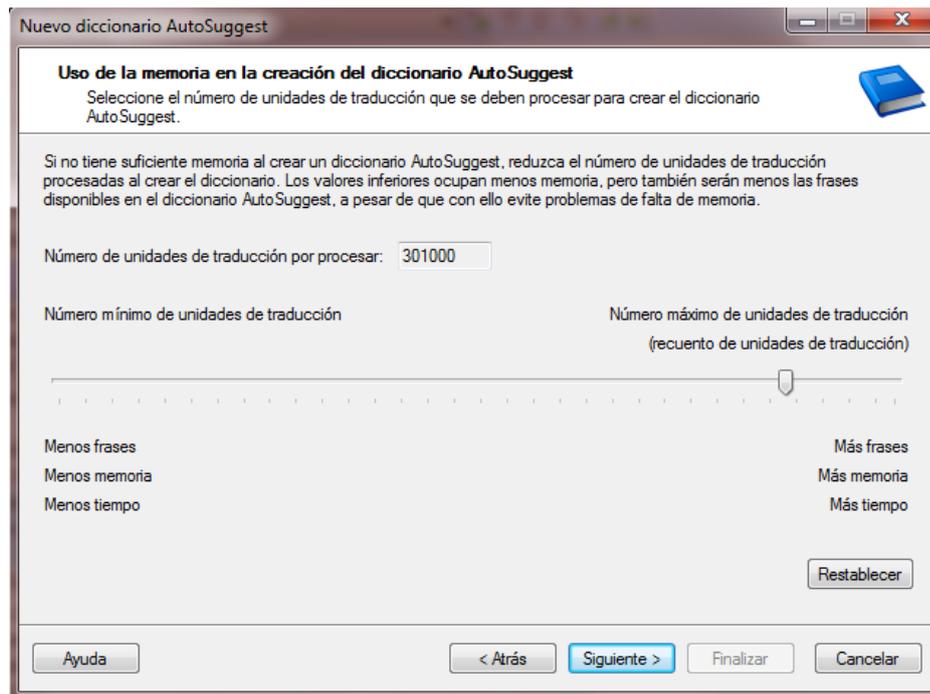


Figura 8: Selección de las unidades de traducción para el diccionario de autosugerencias

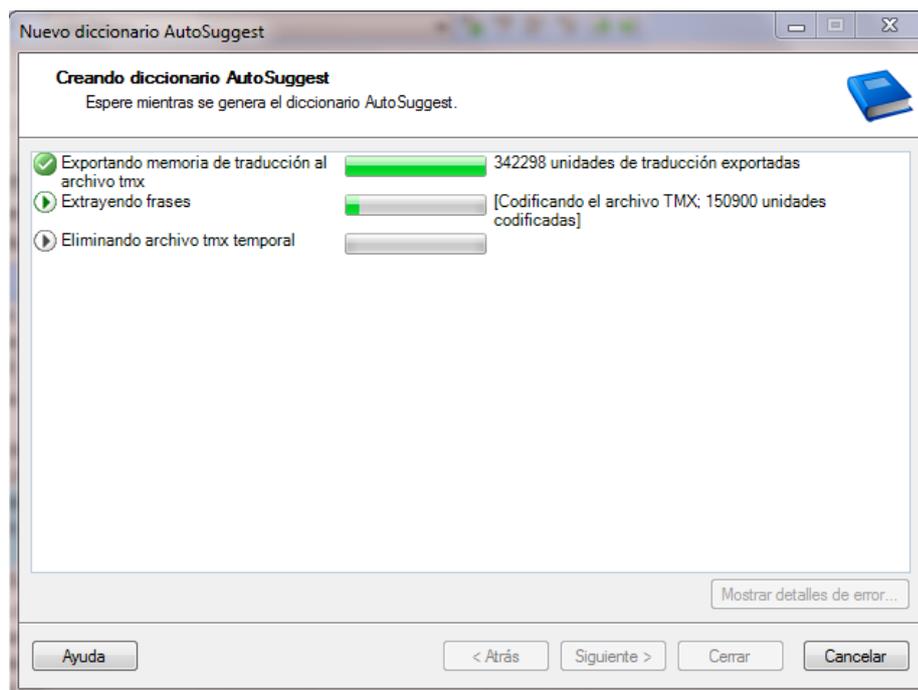


Figura 9: Creación del diccionario de autosugerencias

Al final de todo este proceso, el resultado es que contamos con un gran recurso traductológico, como es la memoria de traducción de la Dirección General de Traducción, que contiene gran cantidad de información útil para un traductor (342.302 unidades de traducción) y que se puede ampliar a medida que la Comisión Europea vaya publicando nuevos documentos.

También hemos obtenido un diccionario de autosugerencias para SDL Trados que nos facilitará futuras tareas de traducción aportándonos sugerencias de términos que estén registrados en la memoria de traducción.

Estos dos recursos resultan muy útiles a la hora de trabajar con las herramientas TAO, pues facilitan aún más la labor del traductor, incrementando la calidad de sus trabajos y mejorando la productividad.

### **CAPÍTULO 3: APROVECHAMIENTO DE UNA MEMORIA DE TRADUCCIÓN MEDIANTE EL USO COMBINADO DE CAMPOS Y FILTROS: ETIQUETADO TEMÁTICO DE LAS UNIDADES DE LA MEMORIA DE TRADUCCIÓN DE LA DGT**

Como hemos mencionado anteriormente, la memoria de traducción de la DGT contiene todos los documentos publicados por la Comisión Europea a lo largo de los años (en nuestro caso del año 2013). Por lo tanto, los documentos publicados abarcan una gran variedad de campos.

El objetivo de este apartado pasa por modificar el contenido de la memoria de la DGT de forma que podamos sacarle más partido. Una memoria de traducción con 30.000 unidades de traducción es ya de por sí muy útil, pero si aprendemos a clasificar estas unidades de traducción y agruparlas según el dominio que traten, la memoria será mucho más completa y estará mejor estructurada, de forma que se podrán realizar búsquedas de unidades en función de la temática, de la fecha de actualización y de otros muchos aspectos. También aprenderemos a manejar el sistema de filtros de la herramienta de entorno de traducción Trados Studio 2009, puesto que es una de las herramientas más empleadas en el ámbito profesional. Por último, nos centraremos también en aprender a manejar el sistema de penalizaciones de la búsqueda de concordancias de la misma herramienta, con la que podremos descartar unidades de traducción en función de su dominio cuando este no nos interese. Esta función agiliza el proceso de traducción cuando trabajamos con memorias tan grandes como la que tenemos entre manos.

En este caso, modificaremos el contenido de la memoria clasificando el mayor número posible de unidades de traducción puesto que, al estar mezcladas, resulta complicado identificar. Para conseguir esto, nos basaremos en la idea de que se puede clasificar cada unidad de traducción en función del dominio identificando la terminología que estas unidades contengan. Sin embargo, es importante tener en cuenta que es muy probable ciertas unidades de traducción pertenezcan a varios dominios y que es posible que aparezcan etiquetados con ruido, aunque la mayoría deberían ser correctos.

Como posibilidad para realizar la clasificación de campos temáticos y una recopilación de la terminología específica de cada campo, se ofrece el trabajo con un tesoro que contenga una clasificación de términos por dominio. En este caso hemos utilizado el tesoro de la Unión Europea: EuroVoc.

EuroVoc es un tesoro multilingüe de la Unión Europea que abarca la terminología de todos los ámbitos de esta institución. Tiene su origen en la voluntad del Parlamento Europeo de dotarse de un lenguaje documental común en todas las lenguas oficiales de la Unión Europea. Su primera edición fue publicada en 1984 en las siete lenguas oficiales

comunitarias de aquel momento y se editó en dos versiones: una alfabética y otra temática, pero siempre con la estructura jerárquica de un tesoro. Desde entonces se han ido publicando versiones mejoradas y más completas. A día de hoy EuroVoc es un tesoro multidisciplinar (comprende política, derecho, relaciones internacionales, economía, agricultura, información y documentación, transportes, etc.) (Maciá, 1995: 266).

A continuación comenzaremos el proceso de recopilación de los términos necesarios para crear campos temáticos y personalizaremos la memoria de traducción de la DGT para insertar estos campos y, posteriormente, crear los filtros y las penalizaciones con las que, a la hora de traducir, podremos modificar la memoria según nuestras necesidades

## **1. Descarga y recopilación términos clasificados por campo temático**

El tesoro de terminología EuroVoc está disponible de forma gratuita en su página web <<http://eurovoc.europa.eu/drupal/?q=es>>. Además de los tesauros con estructura jerárquica, EuroVoc ofrece la posibilidad de obtener una lista multilingüe de todos los términos en las lenguas que elijamos.

Como la lista de términos es inmensa, requeriría un trabajo muy grande emplear todos los términos del tesoro en clasificar las unidades de traducción. Para este proyecto hemos escogido unos 20 términos por campo temático, que sirven de modelo para posteriormente ampliar este trabajo y completar la clasificación con todos los términos del tesoro de EuroVoc.

Es necesario tener en cuenta que este proyecto tiene la intención de servir como método para proyectos de modificaciones de memorias de traducción más completos y rigurosos. Se trata, de hecho, de un modelo que sirve como ejemplo de cómo se marca y modifica una memoria de traducción., Dicho esto, el criterio que hemos seguido para la selección de términos se basa en el reconocimiento como traductores en formación, dentro de la lista multilingüe, de los términos que pertenecen a un campo o a otro, siguiendo los conocimientos que hemos adquirido en las diferentes asignaturas especializadas del grado (Terminología, traducción especializada, científica y técnica). (Dirigirse a ANEXOS)

## **2. Personalización de los campos de la memoria de traducción**

Antes de definir los campos temáticos o dominios de la memoria de traducción, es necesario aclarar un problema que puede surgir con estos términos. Existe una ambigüedad terminológica en la palabra «campo». En los proyectos terminológicos entendemos «campo temático» y «dominio» como «tema», pero no debemos confundirlo con

el término «campo» en las herramientas TAO. A continuación explicamos el significado de este término en una herramienta TAO. (SDL Trados Studio 2009, 2008)

En las herramientas TAO, los campos de una memoria de traducción son una especie de etiquetas que contienen información que se le atribuye a cada unidad de traducción de esa memoria. A cada campo se le asignan unos valores de campo y a cada unidad de traducción se le pueden atribuir uno o más valores de campo. Estos valores de campo pueden ser de varios tipos (texto, número, fecha o lista de selección de valores). Dependiendo de estos tipos, deberemos insertar una lista de valores o quedará un valor en blanco donde podremos insertar el texto que queramos, en el caso del tipo «texto»).

El sistema de memorias de traducción, además de ofrecernos la posibilidad de crear y personalizar un gran número de campos, completa sus propios campos automáticos que hacen referencia a la fecha, al autor, al número de veces que se ha reutilizado una unidad de traducción, etc.

En este proyecto aprovecharemos la posibilidad de crear campos personalizados para marcar y clasificar las unidades de traducción de la memoria con los dominios reconocidos por EuroVoc: estos constituirán los valores de campo que podemos asignar a cada unidad de traducción de nuestra propia memoria en un campo personalizado que crearemos con el nombre «dominio».

Los dominios definidos por EuroVoc son:

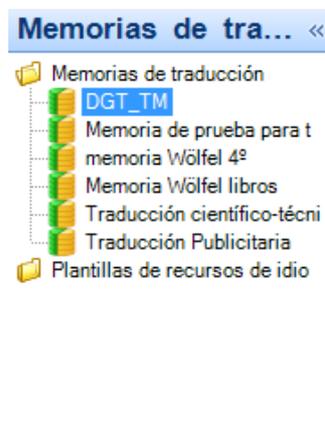
- Vida política
- Relaciones internacionales
- Unión Europea
- Derecho
- Vida económica
- Intercambios económicos y comerciales
- Asuntos financieros
- Asuntos sociales
- Educación y comunicación
- Transportes
- Medio ambiente
- Agricultura, silvicultura y pesca

- Sector agroalimentario
- Producción, tecnología e investigación
- Energía
- Industria
- Geografía
- Organizaciones internacionales
- Ciencia
- Empresa y competencia

Para crear los campos personalizados es necesario dirigirse a la ficha de configuración de la memoria, concretamente a la pestaña «campos y configuración», que solo se puede modificar con la memoria cerrada. Es conveniente realizar este proceso antes de utilizar la memoria para cualquier proyecto de traducción pues, si bien es cierto que se pueden añadir campos y valores de campo siempre que queramos, el resultado estaría alterado por las nuevas unidades de traducción que entraran en la memoria.

También incluiremos el campo «materia», donde asignaremos el valor de campo «TFG» a todas las unidades que etiquetemos. Esto nos resultará útil a la hora de hacer el recuento de las unidades de traducción marcadas durante el trabajo.

A continuación, insertaremos estos campos en la memoria de traducción desde el gestor Trados 2009. Con la memoria cerrada, pulsando con el botón derecho sobre ella en «Vista de memorias de traducción» y nos dirigimos a «configuración». Desde ahí, a la opción «campos y configuración». En este caso, el campo temático será del tipo «lista de selección» que nos permitirá predefinir los valores de campo recogidos en el punto anterior.



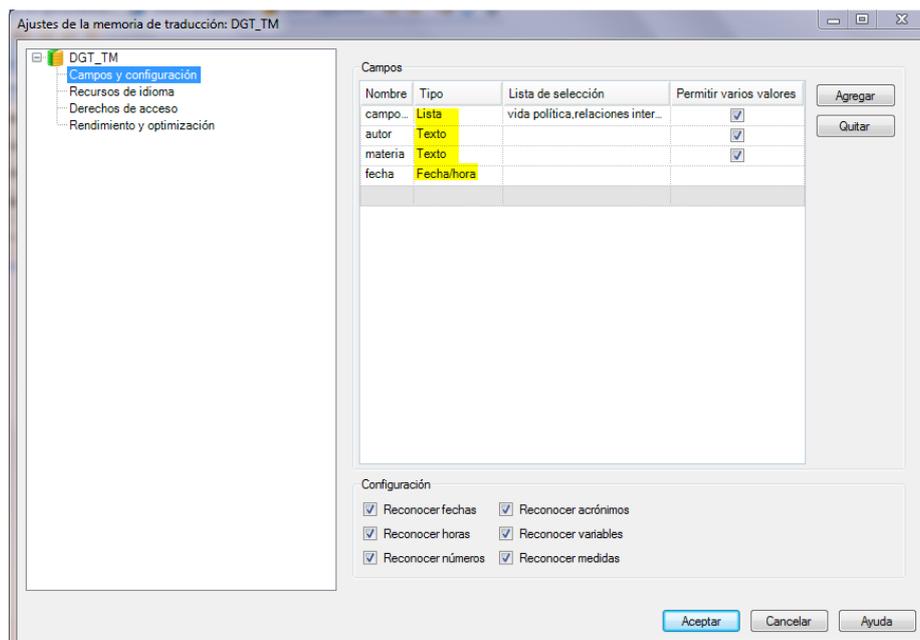


Figura 10: Configuración de los campos de la memoria de traducción

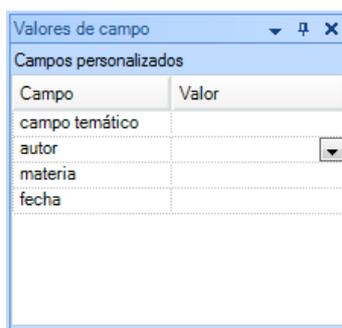


Figura 11: Configuración de los campos de la memoria de traducción

Es importante seleccionar la casilla «permitir varios valores» porque hay muchos términos que podrían pertenecer a más de un campo temático. Los campos aparecerán en un cuadro de diálogo, que es una ventana especial de la vista de «memorias de traducción» donde se pueden editar los valores de campo siempre que queramos, junto con los posibles valores que hayamos configurado.

### 3. Filtrado de términos por campo temático

#### 3.1. Creación de filtros en Trados Studio

Los filtros de Trados Studio sirven para clasificar las unidades de traducción en función de una gran variedad de criterios dependiendo de la información que contengan, por ejemplo, del texto (origen o meta) que incluyan, de la fecha de creación, el autor o el

origen de cada unidad de traducción. También se pueden filtrar unidades de traducción que no deseemos para eliminarlas, o crear un conjunto avanzado de filtros para clasificar las unidades de traducción con mayor precisión en función de nuestras necesidades. En nuestra opinión, esta herramienta resulta muy útil a la hora de mejorar los documentos de una traducción (corregir segmentos de origen erróneos), traducir y revisar las traducciones.

En este proyecto haremos uso del sistema de filtros para localizar términos en los segmentos que los contengan, tanto en los de origen como en los de destino para clasificar las unidades de traducción de la memoria de traducción de la DGT.

En cada filtro se pueden agregar varias condiciones de filtrado, por lo que no será necesario crear un filtro por término, sino que podemos insertar varios términos en un mismo filtro, procurando que el nexos entre las condiciones sea siempre “o” en lugar de “y”, que indicaría inclusión. De este modo, como veremos posteriormente, una vez guardados, podremos añadir cada filtro a un campo temático para que, a cada unidad de traducción que cumpla alguna de esas condiciones, se le asigne un campo temático correspondiente.

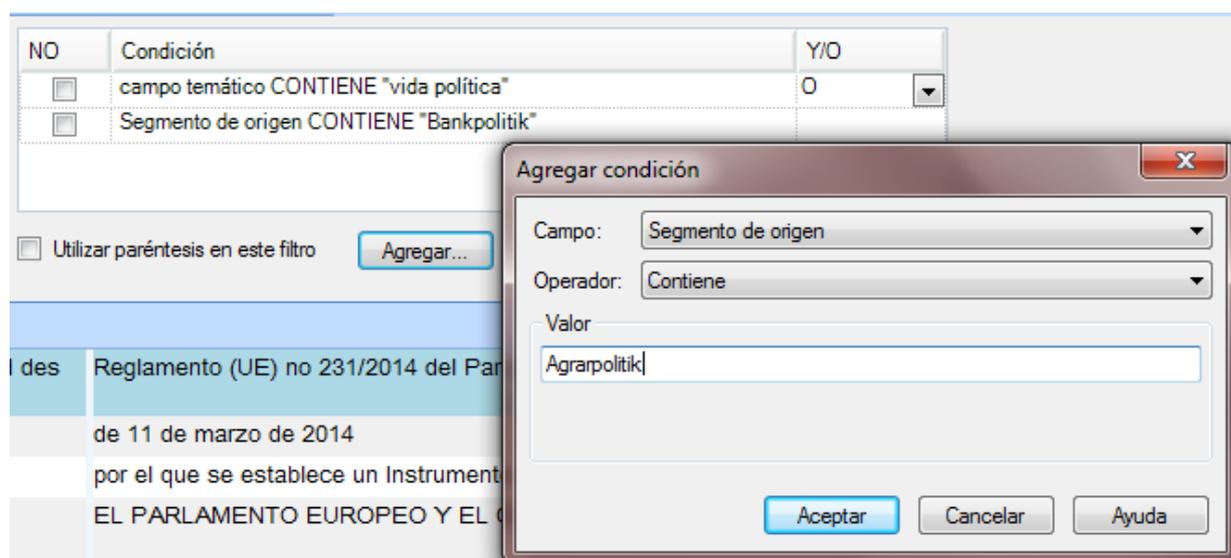
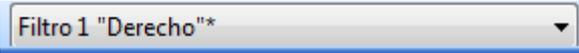


Figura 12: Inserción de condiciones en un filtro

En la vista de «filtros» haremos *click* en «agregar» e insertaremos las condiciones una a una. En este proyecto hemos recopilado los términos en alemán, pues el par de lenguas escogido es DE>ES, por lo que el segmento origen está en alemán. Por tanto, para las condiciones en las que insertaremos los términos, debemos seleccionar el campo «segmento de origen» y el operador «contiene», ya que el operador «igual que» significaría que buscamos un segmento exactamente igual al valor que escojamos. Como ya hemos explicado, es importante recordar siempre que debemos cambiar el nexos «y» que aparece automáticamente por «o».

Para evitar los errores que puedan darse y, sobre todo, que Trados se ralentice, hemos escogido insertar un máximo de 10 términos por filtro, es decir, diez condiciones por cada filtro, por lo que tendremos dos filtros por campo temático.

Para guardar un filtro una vez creado, hay que pulsar el icono «guardar» en la barra de herramientas del cuadro de filtros . Desde entonces el filtro nos aparecerá siempre en la barra de selección del mismo cuadro con el nombre que hayamos escogido . Es importante guardar los filtros para que luego nos aparezcan en esta pestaña y podamos exportarlos a un solo archivo en formato *sdltm.filter* en el caso de que queramos reutilizar estos filtros en otra memoria de traducción.

Cada vez que queramos crear un nuevo filtro, pulsaremos el icono «agregar filtro» en la misma barra de tareas .

### 3.2. Edición de unidades de traducción por lotes con ayuda de filtros

Una vez creados los filtros, queremos que a todos los segmentos que contengan las condiciones que hemos incluido, es decir, los términos, se les asigne un campo temático determinado. Para ello debemos recurrir a la función «edición por lotes», que encontraremos con el icono «edición por lotes»  en la barra de herramientas de la memoria de traducción.

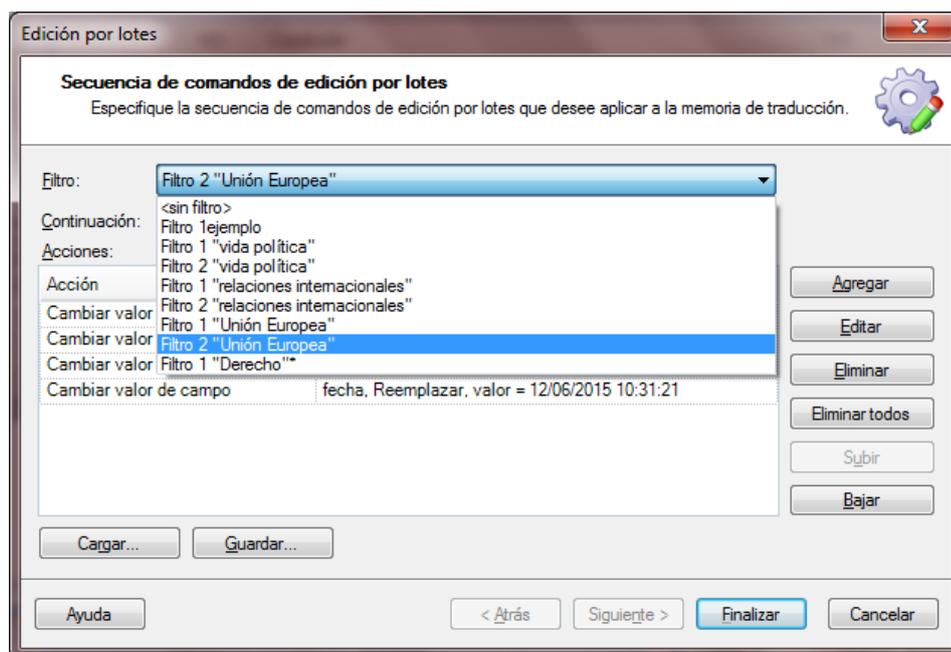


Figura 13: Edición de los filtros por lotes

Dentro de la herramienta de «edición por lotes», debemos escoger el filtro con el que queramos trabajar. Una vez escogido, en la pestaña «agregar» seleccionaremos la opción «cambiar valores de campo» (es muy importante seleccionar esta opción, pues para Trados es la única opción que nos dejará modificar las unidades de traducción. Las otras opciones son «buscar y remplazar texto», «eliminar valores de campo» y «eliminar etiquetas») y actualizaremos todos los valores de los campos que hemos personalizado previamente para la memoria. Una vez hecho esto, pulsaremos «finalizar» y esperaremos a que se actualicen los campos de todas las unidades que cumplan las condiciones de los filtros.

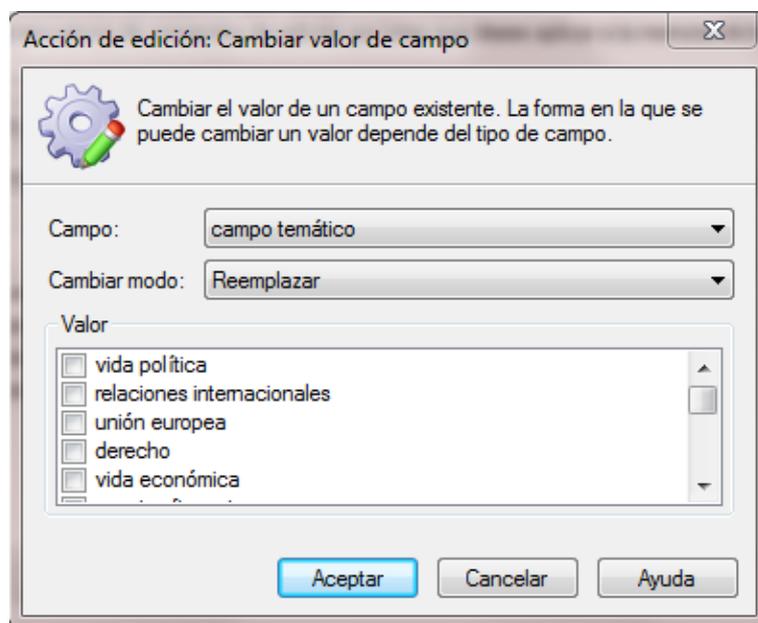


Figura 14: Actualización de valores de campo para los filtros

### 3.3. Exportación de los filtros

Una vez creados los filtros, podemos exportarlos para utilizarlos en otras memorias de traducción. Para ello, en la barra de herramientas seleccionamos la opción «exportar filtros al archivo»  y seleccionamos todos los filtros que deseamos exportar y la carpeta de destino. Los filtros se guardarán en formato *sdltm.filters*.

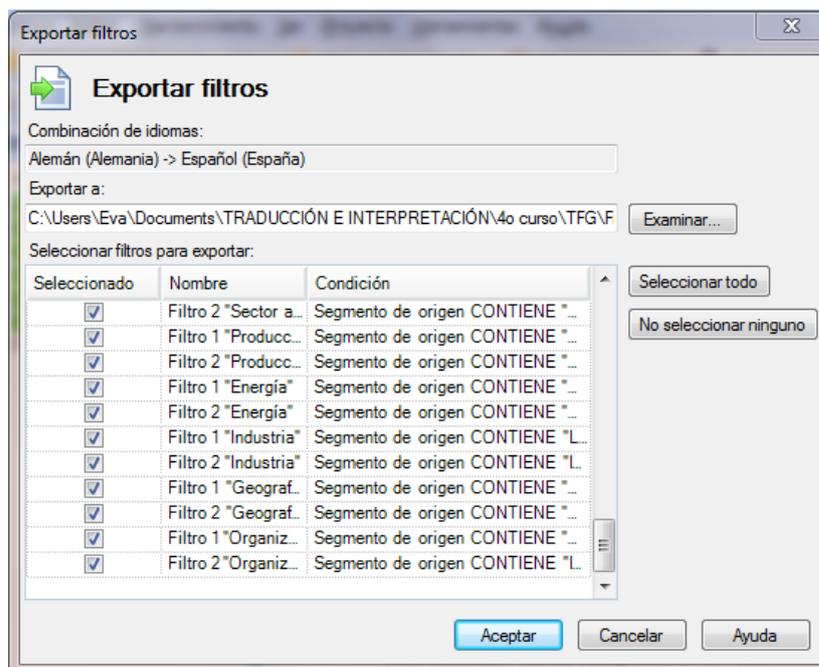


Figura 15: Exportación de filtros

#### 4. Empleo del sistema de penalizaciones

SDL Trados Studio nos da la oportunidad de excluir unidades de traducción de una memoria para proyectos determinados en los que no nos interesen estas unidades. Se trata del sistema de penalizaciones de memorias de traducción donde también se emplean los filtros y podemos hacer uso de los que ya hemos creado.

Esta herramienta está disponible al inicio de cada proyecto, en la configuración de este. Al crear un nuevo proyecto (haciendo *click* en ) , hemos de seguir los pasos correspondientes (selección de lenguas, archivo y carpeta de destino) hasta llegar a la sección de «memoria de traducción y traducción automática». Ahí, nos dirigiremos a la pestaña de filtros y agregaremos los filtros que queramos (podemos insertar filtros nuevos o reutilizar los que ya hemos hecho anteriormente).

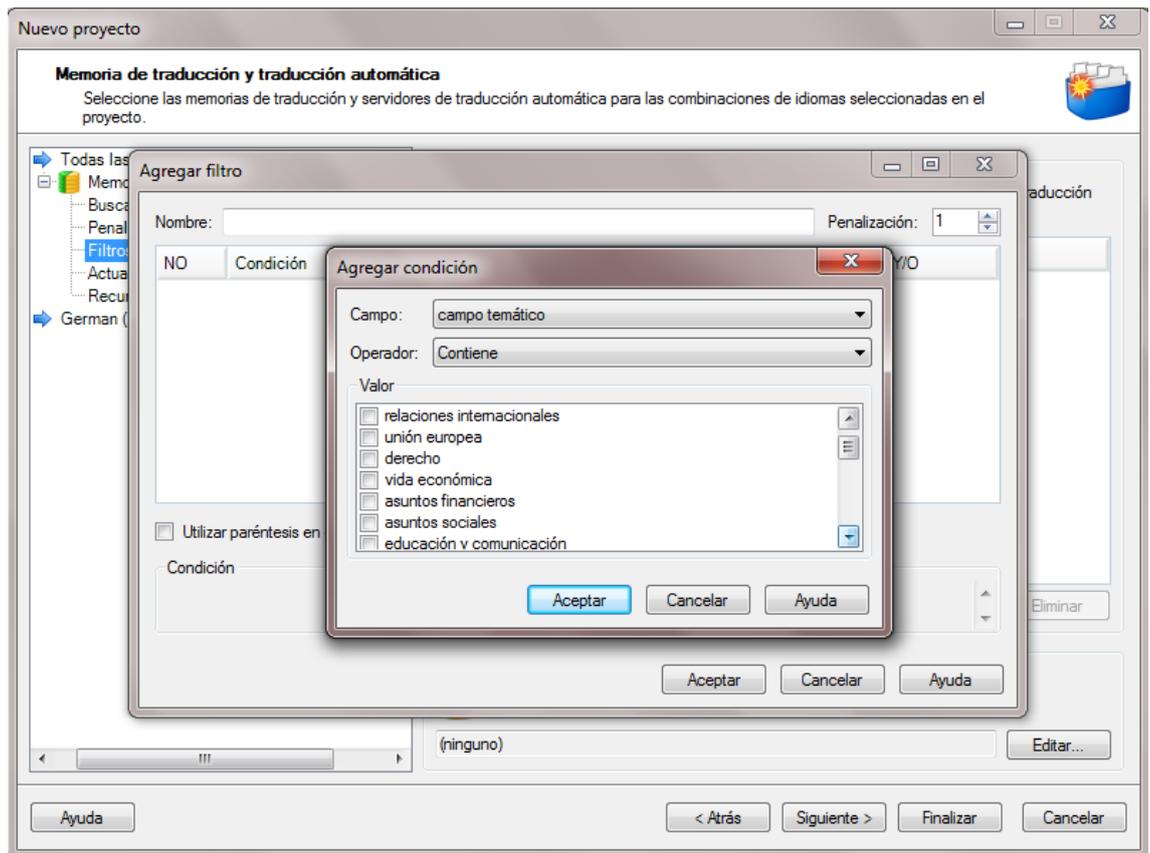


Figura 16: Adición de filtros al sistema de penalizaciones de las memorias de traducción

De esta forma, si tenemos un proyecto de carácter técnico, podemos ordenar a la memoria que, a la hora de configurar el proyecto, nos excluya las unidades de traducción clasificadas con los campos temáticos que no tengan relación con el dominio técnico (derecho, asuntos sociales, etc.). De esta forma, así, la memoria será mucho más simple para este proyecto y nos libraré de unidades de traducción inútiles que, en memorias del tamaño de la que estamos utilizando, contribuyen a ralentizar el proceso de búsqueda de concordancias. Se trata de un proceso que requiere un conocimiento avanzado de la herramienta, pues para llevarlo a cabo es necesario saber utilizar el sistema de filtros cuando estos, a su vez, dependen de la personalización de los campos de la memoria de traducción.

En general, todos los procesos con los que hemos trabajado a lo largo de este capítulo requieren conocimientos medios o avanzados de la herramienta de entorno de traducción SDL Trados Studio, pero siguiendo los pasos descritos en este método conseguiremos obtener una memoria de traducción con el mayor número posible de unidades de traducción clasificadas en base al dominio al que pertenecen y podremos

excluir las unidades de traducción que no nos interesen en futuros proyectos de traducción.

## CONCLUSIONES

Tras finalizar el proyecto, además de haber reflexionado profundamente acerca del uso de estas herramientas y de hacer un recorrido por todas ellas, hemos obtenido una serie de productos de los que podemos extraer muchas conclusiones.

El presente trabajo no se limita a un conjunto de teorías y reflexiones acerca de las herramientas de traducción asistida por ordenador y de cómo afectan a la vida profesional del traductor, sino que traspasa esos límites para enseñarnos cómo podemos aprovechar de una forma más eficaz los recursos que internet y las grandes instituciones nos ofrecen. También nos demuestra todas las posibilidades que una de estas herramientas, como es SDL Trados Studio nos ofrece y cómo podemos utilizarlas para facilitar nuestra labor profesional.

En primer lugar, hemos obtenido una memoria de traducción DE>ES con formato *.sdltm*, 342.302 unidades de traducción y 429 MB basada en la inmensa memoria de traducción de la DGT. Después de mostrar como personalizarla y modificarla, la memoria cuenta con 12.761 unidades etiquetadas según el dominio al que pertenecen. Este producto sirve como modelo para futuros proyectos, si bien puede continuar modificándose hasta etiquetar todas las unidades de traducción presentes en la memoria. Es solo una cuestión de trabajo y tiempo, pues si etiquetar todas esas unidades nos ha llevado cuatro sesiones de tres horas (12 horas), en un periodo de menos de un mes conseguiríamos etiquetar todas las unidades de traducción clasificándolas según su dominio temático.

Quizá esto suponga un gran trabajo para los profesionales que dedican su vida a la labor de traducir, pero se trata de un trabajo que podría servirles de ayuda a lo largo de gran parte de su carrera profesional. Además, tras haber realizado el proyecto hemos concluido que una de las claves para conseguir buenos resultados es la recopilación de términos especializados según los cuales determinamos el dominio de una unidad de traducción. Por cuestiones de tiempo, los términos especializados que nosotros hemos recopilado no siguen criterios profesionales, pero si nos dirigimos a los glosarios temáticos de la institución EuroVoc, podemos realizar un filtrado de términos más escrupuloso y riguroso, lo que en definitiva resultaría más eficaz a la hora de filtrar todas las unidades de traducción.

Otro de los productos obtenidos tras la realización del proyecto es un diccionario de autosugerencias generado a partir de esa memoria de traducción. El diccionario generado tiene formato *.bpm* y es compatible con SDL Trados 2009 y sus versiones posteriores. Este diccionario contiene todo el vocabulario de todas las unidades de traducción presentes en la memoria de traducción y puede utilizarse para una gran variedad de traducciones, pues

esta memoria a partir de la cual se ha generado toca muchos dominios y contiene mucho vocabulario especializado. Es cierto que, aunque estos diccionarios aún no son compatibles con otras herramientas TAO en auge, continúan siendo un recurso fantástico para quienes trabajan con la herramienta Trados, sin duda la más demandada por los clientes en el mercado actual. Tampoco podemos descartar que, en un futuro, otras herramientas incluyan esta función entre sus aplicaciones y los diccionarios de autosugerencias adquieran mayor grado de difusión. A la hora de probar el diccionario, concluimos que, efectivamente, aporta una gran cantidad de sugerencias de las cuales hay que saber escoger la más adecuada, pero supone un gran ahorro de tiempo al no necesitar realizar búsquedas en diccionarios externos para encontrar el vocabulario adecuado.

Por último, tras la finalización del proyecto también hemos un producto muy útil para los profesionales de la traducción y, posiblemente, para los estudiantes en formación: una manual de indicaciones para aprovechar de manera un gestor de memorias de traducción como es SDL Trados 2009. En él hemos tratado de explicar de la manera más clara posible cómo podemos realizar la descarga de una memoria de traducción e incorporación al sistema de Trados, cómo personalizar esta memoria añadiéndole campos en función de nuestras necesidades y cómo diferenciar estas unidades atendiendo a los diferentes campos. Además, se explican algunas de las aplicaciones que todo esto nos permite, como es la exclusión de cierto tipo de unidades que no nos interesen a la hora de traducir determinado tipo de textos mediante el sistema de filtrado, o cómo hacer un recuento de la cantidad de unidades pertenecientes a cada dominio para mantener la memoria de traducción actualizada.

Tras la conclusión de este proyecto dejamos la puerta abierta a continuar trabajando con otras aplicaciones más avanzadas de la herramienta y a descubrir si otras herramientas de este tipo tienen también propiedades similares a las que hemos descubierto en SDL Trados. Además, también queda abierta la posibilidad de continuar trabajando en el proyecto realizado que, por supuesto, no está completo y requiere más horas de trabajo para culminarlo.

En definitiva, este tipo de proyectos contribuye a acercarnos un poco más a las herramientas y cualidades que se ponen a nuestra disposición. Cada día surgen nuevas herramientas y aplicaciones que pueden aportarnos soluciones a problemas del día a día como profesionales. El mundo actual depende de la tecnología y, en nuestro caso, del multilingüismo. Podemos afirmar que este trabajo, aunque supone una mera toma de contacto con estas herramientas (a pesar de que hemos utilizado a nivel avanzado una de ellas), nos ha servido para reafirmarnos en la importancia que tiene mantenerse actualizado en lo referente a los avances de la tecnología en la traducción, pues no solo

nos generan beneficios en lo referente al tiempo, sino también en lo que se refiere a la calidad y precisión de nuestro trabajo diario.

## REFERENCIAS BIBLIOGRÁFICAS

ARIZA GARCÍA, A. y TAPIA POYATO, A.M. (1997-98) «El corrector ortográfico y la presentación del texto escrito». En: *Cauce, Revista de Filología y su Didáctica*. Sevilla. Nº20-21. 375-412 [en línea] Disponible en: [http://cvc.cervantes.es/literatura/cauce/pdf/cauce20-21/cauce20-21\\_22.pdf](http://cvc.cervantes.es/literatura/cauce/pdf/cauce20-21/cauce20-21_22.pdf) [Fecha de consulta: 01/07/2015]

ARRANZ ACEBES, A. (2013) «Ventajas e inconvenientes de usar diccionarios AutoSuggest con Trados». En: Alicia Arranz Acebes. Traductora de inglés, francés e italiano a español [En línea] Disponible en: <http://atraduccion.com/2013/10/10/ventajas-e-inconvenientes-de-usar-diccionarios-autosuggest-con-trados/> [Fecha de consulta: 23/06/2015]

ARTHERN, P.J. (1981) «Aids unlimited: the scope for machine aids in a large organization». En: *Aslib Proceedings*. Vol. 33, nº7/8, 309-319 [en línea] Disponible en: <http://www.mt-archive.info/70/Aslib-1980-Arthern.pdf> [Fecha de consulta: 01/07/2015]

BALCELLS, J. (2013) «Un cuarto de siglo de TEnT». En: *La Linterna del Traductor*. La revista multilingüe de Asetrad, nº8, 99-107 [en línea] Disponible en: [http://www.lalinternadeltraductor.org/pdf/lalinterna\\_n8.pdf](http://www.lalinternadeltraductor.org/pdf/lalinterna_n8.pdf) [Fecha de consulta: 01/07/2015]

BOWKER, L. (2002) *Computer-aided Translation Technology: A Practical Introduction*. University of Ottawa Press, Canadá. ISBN: 0-7766-0538-0

CARTONI, B. y MEYER, T. (2012) «Extracting Directional and Comparable Corpora from a Multilingual Corpus for Translation Studies». En: *Proceedings of the eighth international conference on Language Resources and Evaluation (LREC)*. Estambul. [En línea] Disponible en: <https://www.idiap.ch/~tmeyer/res/Cartoni-LREC-2012.pdf> [Fecha de consulta: 22/06/2015]

COMISIÓN EUROPEA. (2009) *Traducir para una comunidad multilingüe* [En línea] ISBN 978-92-79-09624-2. Disponible en: <http://bookshop.europa.eu/es/traducir-para-una-comunidad-multilinguee-pbHC3008600/> [Fecha de consulta: 23/06/2015]

DE LA COVA, E. y ALONSO, E. (2013) *Introducción a la localización. Materiales docentes*. Sevilla (Autoedición). ISBN: 978-84-695-9185-7 [en línea] Disponible en: <https://goo.gl/hdGUDP> [Fecha de consulta: 01/07/2015]

DOCSETOOLS, (2015) *Información mutua*. [En línea] Disponible en: [http://docsetools.com/articulos-noticias-consejos/article\\_135833.html](http://docsetools.com/articulos-noticias-consejos/article_135833.html) [Fecha de consulta: 22/06/2015]

FERNÁNDEZ SIERRA, L. (2009) «Trados como herramienta de traducción asistida por ordenador Un recorrido por su historia y su evolución». En: *Panacea*. Vol. 10, nº29, 32-37 [en línea] Disponible en: [http://www.medtrad.org/panacea/IndiceGeneral/n29\\_tribuna-Sierra.pdf](http://www.medtrad.org/panacea/IndiceGeneral/n29_tribuna-Sierra.pdf) [Fecha de consulta: 01/07/2015]

GAMBIER, Y. y VAN DOORSLAER, L. (2010) *Handbook of translation studies*. Vol. 1. John Benjamins Publishing Co., Amberes (Bélgica) ISBN: 9789027203311

GONZÁLEZ PIÑERO, D. (2004) *Software libre en los institutos*. (Disponible en el Departamento de Lenguajes y Sistemas Informáticos de la Universidad Politécnica de Cataluña) [en línea] Disponible en: [http://www.cs.upc.edu/~tonis/daniel\\_gonzalez\\_pinyero.pdf](http://www.cs.upc.edu/~tonis/daniel_gonzalez_pinyero.pdf) [Fecha de consulta: 01/07/2015]

HOLTSBERG, A y WILLNERS, C. (2001) «Statistics for sentential co-occurrence». En: *Working Papers 48*, 135-147. Lund University Department of Linguistics. [En línea] Disponible en: <http://lup.lub.lu.se/luur/download?func=downloadFile&recordId=528668&fileId=624438> [Fecha de consulta: 23/06/2015]

HUTCHINS, W. J. (1978) «Machine translation and machine-aided translation». En: *Journal of Documentation*, Vol. 34, nº2, 119-159 [en línea] Disponible en: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.97.831> [Fecha de consulta: 01/07/2015]

HUTCHINS, W. J. (1998) «The origins of the translator's workstation». En: *Machine Translation*, Vol. 13, nº4, 287-307 [en línea] Disponible en: <http://www.hutchinsweb.me.uk/MTJ-1998.pdf> [Fecha de consulta: 01/07/2015]

KAY, M. (1997) «The Proper Place of Men and Machines in Language Translation». En: *Machine Translation*. Países Bajos. 12: 3-23 [en línea] Disponible en: <http://link.springer.com/article/10.1023%2FA%3A1007911416676#page-1> [Fecha de consulta: 01/07/2015]

LÓPEZ CIRUELOS, A. (2003) «Una defensa crítica de las memorias de traducción». En: *Panacea*. Vol. 4, nº12, 180-182 [en línea] Disponible en: [http://www.tremedica.org/panacea/IndiceGeneral/n12\\_revistilo\\_LCiruelos.pdf](http://www.tremedica.org/panacea/IndiceGeneral/n12_revistilo_LCiruelos.pdf) [Fecha de consulta: 01/07/2015]

LÓPEZ SÁNCHEZ, R. (2014) Guía básica de software para traductores. [en línea] Disponible en: <https://jugandoatraducir.files.wordpress.com/2012/11/guiasoftware.pdf> [Fecha de consulta: 01/07/2015]

MACIÁ, M. (1995) «El tesoro EUROVOC». Revista General de Información y Documentación [En línea] Vol. 5, nº2. Disponible en: [revistas.ucm.es](http://revistas.ucm.es) [Fecha de consulta: 22/06/2015]

MILLER, V. (2011) «The EU's Acquis Communautaire». En: House of Commons Library [En línea] Disponible en: <http://www.isn.ethz.ch/Digital-Library/Publications/Detail/?lang=en&id=151238> [Fecha de consulta: 23/06/2015]

MIRAGLIA, M. (2013) *The influence of new technologies on translation methodology. From web corpora to cloud-based translation management*. Roma. ISBN: 978-88-548-6569-3 [en línea] Disponible en: [http://www.academia.edu/12445545/The\\_influence\\_of\\_new\\_technologies\\_on\\_translation\\_methodology\\_From\\_web\\_corpora\\_to\\_cloud-based\\_translation\\_management](http://www.academia.edu/12445545/The_influence_of_new_technologies_on_translation_methodology_From_web_corpora_to_cloud-based_translation_management) [Fecha de consulta: 01/07/2015]

ORTIZ JIMÉNEZ, M. (2011) «Traducción asistida y traductores autónomos: alternativas y avances». En: Puntoycoma. Boletín de los traductores españoles de las instituciones de la Unión Europea, Vol. 123 [en línea] Disponible en: [http://ec.europa.eu/translation/bulletins/puntoycoma/123/pyc1237\\_es.htm](http://ec.europa.eu/translation/bulletins/puntoycoma/123/pyc1237_es.htm) [Fecha de consulta: 01/07/2015]

PROGRAMARI TRADUMÀTICA (s.f.) *Llistat de programari relacionat amb la tradumàtica*. [en línea] Disponible en: <https://sites.google.com/site/programaritradumatica/> [Fecha de consulta: 01/07/2015]

SDL TRADOS STUDIO 2009 (2008) *SDL Trados Studio 2009 Intermediate for translators. Training Course*. [Manual de usuario]

SOMERS, H.L. (2003) *Computers and Translators. A translator's guide*. John Benjamins Publishing Co., Países Bajos. ISBN: 90 272 1640

STALLMAN, R. M. (2004) *Software libre para una sociedad libre*. Madrid. ISBN: 84-933555-1-8 [en línea] Disponible en: [https://www.gnu.org/philosophy/fsfs/free\\_software.es.pdf](https://www.gnu.org/philosophy/fsfs/free_software.es.pdf) [Fecha de consulta: 01/07/2015]

STEINBERGER, R. et al. (2013) *DGT-TM: A freely Available Translation Memory in 22 Languages*. [En línea] Disponible en: <http://arxiv.org/ftp/arxiv/papers/1309/1309.5226.pdf> [Fecha de consulta: 22/06/2015]

VARGAS SIERRA, Ch. (2002) «Utilización de los programas de concordancias en la traducción especializada». En: El español, lengua de traducción. Actas del I Congreso Internacional, Almagro. 522-539. Disponible en: <http://hdl.handle.net/10045/13587> [Fecha de consulta: 01/07/2015]

ZETZSCHE, J. (2007) «Tool Kit \* What Makes a Translation Environment Tool a Good TEnT?». En: The How-To Library [En línea] Disponible en: <http://www.translatorscafe.com/cafe/article79.htm> [Fecha de consulta: 23/06/2015]

## ANEXOS