

UNIVERSIDAD DE SALAMANCA  
FACULTAD DE TRADUCCIÓN Y DOCUMENTACIÓN  
GRADO EN TRADUCCIÓN E INTERPRETACIÓN  
Trabajo de Fin de Grado

LA RECUPERACIÓN DE  
SUBSEGMENTOS EN SDL TRADOS  
STUDIO 2017

**Análisis del mecanismo upLIFT y evaluación de  
su aprovechamiento**

Carmen Gragero Carmona  
Tutor: Emilio Rodríguez Vázquez de Aldana

Salamanca, 2018

# ÍNDICE

1	INTRODUCCIÓN .....	1
1.1	JUSTIFICACIÓN, MOTIVACIÓN Y CONTEXTUALIZACIÓN DEL TRABAJO 1	
1.2	OBJETIVOS.....	1
1.3	ESTRUCTURA.....	2
1.4	METODOLOGÍA .....	3
2	MARCO TEÓRICO .....	6
2.1	CORPUS .....	6
2.2	MEMORIAS DE TRADUCCIÓN .....	7
2.2.1	Historia de memoria de traducción .....	8
2.2.2	Tipos de memorias de traducción .....	9
2.3	RECUPERACIÓN DE SEGMENTOS .....	9
2.4	RECUPERACIÓN DE SUBSEGMENTOS .....	10
2.4.1	Aspectos generales sobre la recuperación de subsegmentos .....	10
2.4.2	La traducción automática y la recuperación de subsegmentos en SDL Trados Studio 2017.....	12
3	MARCO EXPERIMENTAL.....	16
3.1	SDL TRADOS 2017 Y LA RECUPERACIÓN DE SUBSEGMENTOS.....	16
3.1.1	¿Qué es?.....	16
3.1.2	¿Cómo funciona?.....	16
3.1.3	¿Cómo se activa upLIFT en SDL Trados 2017? .....	21
3.2	PREPARACIÓN DEL EXPERIMENTO .....	21
3.3	DESARROLLO DEL EXPERIMENTO: COMPROBACIÓN DEL MECANISMO UPLIFT.....	24
3.3.1	CASO 1: Solo upLIFT TU Match .....	24

3.3.2	CASO 2: También upLIFT Fragment Match .....	26
3.3.3	CASO 3: reduciendo el tamaño de los fragmentos en upLIFT TU Match .....	27
3.3.4	CASO 4: reduciendo el tamaño de los fragmentos en upLIFT Fragment Match.....	28
3.3.5	CASO 5: El diccionario de Autosugerencias.....	30
3.3.6	CASO 6: upLIFT Fragment Recall versus Autosugerencias.....	32
3.4	RESULTADOS DEL EXPERIMENTO .....	39
4	CONCLUSIONES .....	40
5	BIBLIOGRAFÍA.....	42

## TABLA DE ILUSTRACIONES

Figura 1: Paso 1 en la traducción automática estadística. ....	14
Figura 2: Paso 2 en la traducción automática estadística. ....	15
Figura 3: Mecanismo upLIFT TU Match activado.....	17
Figura 4: Recuperación de upLIFT TU Match. ....	18
Figura 5: Mecanismo upLIFT Fragment Match activado.....	18
Figura 6: Recuperación de upLIFT Fragment Match. ....	19
Figura 7: Recuperación de upLIFT Fuzzy Repair. ....	20
Figura 8: Recuperación de dos unidades de traducción en upLIFT TU Match.....	26
Figura 9: Aumento de propuestas al bajar el nivel de coincidencia en <i>upLIFT TU Match</i> . ....	28
Figura 10: Recuperación de <i>upLIFT Fragment Match</i> en el segmento 364.....	29
Figura 11: Recuperación de <i>upLIFT Fragment Match</i> en el segmento 365.....	29
Figura 12: Lista de propuestas erróneas en <i>upLIFT Fragment Match</i> . ....	30
Figura 13: Recuperación de subsegmentos por parte del Diccionario de Autosugerencias en el segmento 826.....	32
Figura 14: Recuperación de subsegmentos de <i>upLIFT Fragment Recall</i> funcionando como Diccionario de Autosugerencias en el segmento 826. ....	34
Figura 15: Recuperación de subsegmentos de <i>upLIFT Fragment Recall</i> funcionando como Diccionario de Autosugerencias en el segmento 151. ....	36
Figura 16: Recuperación de subsegmentos de <i>upLIFT Fragment Match</i> en el segmento 8.....	37
Figura 17: Recuperación de subsegmentos de Diccionario de Autosugerencias y <i>upLIFT Fragment Match</i> en el segmento 191. ....	38

## Resumen

En su deseo de aprovechar al máximo la información que contienen las memorias de traducción, las empresas están desarrollando continuamente nuevos mecanismos para que dicha información no quede infrautilizada. La herramienta de traducción asistida por ordenador SDL Trados Studio ha incluido, en su última versión (2017), un complemento de recuperación de subsegmentos que pretende ayudar al traductor en su tarea.

Este trabajo está constituido por dos partes diferenciadas. Primeramente, el marco teórico, en el que, a partir de una estructura expositiva, se realiza una contextualización de la recuperación de subsegmentos en el marco de la recuperación de traducciones. En segundo lugar, el experimento práctico con la herramienta SDL Trados Studio 2017 con el fin de comprobar cuán útil es el mecanismo *upLIFT Fragment Recall*, que se encarga de recuperar subsegmentos a partir de la memoria de traducción que se use en cada proyecto. Para ello, utiliza unas técnicas similares a las de la traducción automática estadística.

**Palabras clave:** SDL Trados Studio 2017, memorias de traducción, recuperación de subsegmentos, herramientas de traducción asistida por ordenador, recursos tecnológicos para la traducción.

## Abstract

As one of the main objectives for companies is to squeeze the information stored in translation memories, they are developing new devices in order to avoid its underutilization. In its last version (2017), the CAT tool SDL Trados Studio has included a subsegment recall mechanism in order to help the translator with his task.

This project consists of two different parts. First of all, a theoretical framework that, from an exposition structure, offers a theoretical approach in order to become familiar with translation technologies. Secondly, a practical experiment with the tool SDL Trados Studio 2017, whose objective is to evaluate the utility of *upLIFT Fragment Recall*. Using techniques based on statistical-based machine translation, it tries to recall subsegment coincidences from the translation memories.

**Keywords:** SDL Trados Studios 2017, translation memories, subsegment recall, CAT tools, translation technologies.

# 1 INTRODUCCIÓN

## 1.1 JUSTIFICACIÓN, MOTIVACIÓN Y CONTEXTUALIZACIÓN DEL TRABAJO

El presente trabajo pretende centrarse en el ámbito de los recursos tecnológicos para la traducción, que está cobrando gran importancia tanto en el mundo empresarial como en el académico desde hace varios años y sobre el que se está investigando en profundidad. Dado que este campo es muy amplio, nos centraremos en el aprovechamiento de la información que las memorias de traducción contienen. Concretamente, en la recuperación de subsegmentos, que implicaría la máxima utilización de los datos de que disponemos por parte de una herramienta dedicada a este fin. Para analizar cómo estos programas organizan la información para sacar el máximo partido de ella, emplearemos la nueva versión de SDL Trados Studio (2017), que utiliza bases de datos terminológicas y memorias de traducción con el fin de facilitar y agilizar la tarea del traductor.

Este proyecto surge a principios del curso 2017/18, dado que comenzamos a utilizar en clase la versión de 2017 de SDL Trados Studio. Teniendo en cuenta que la versión de esta herramienta con la que habíamos trabajado anteriormente era la del año 2015, consideramos que sería interesante probar el funcionamiento de las novedades que presentaba esta última versión. Por tanto, centramos el objetivo de este proyecto en contrastar cómo ahora SDL Trados Studio 2017, gracias a un nuevo mecanismo, es capaz de recuperar subsegmentos y ofrecérselos al usuario traductor para que los utilice mientras trabaja.

## 1.2 OBJETIVOS

El objetivo del presente trabajo consiste en demostrar cuán efectivo y útil es para el traductor la nueva herramienta implementada en la última versión de SDL Trados Studio, que recupera fragmentos de oraciones y las propone al traductor. A partir de esta idea principal, podemos establecer los siguientes puntos:

1. Explicar qué es la recuperación de subsegmentos y revisar el estado del arte de esta cuestión.

2. Explicar las técnicas de recuperación de subsegmentos que ha incluido SDL Trados Studio en su última versión (2017):
3. Buscar una manera de evaluar el grado de utilidad de estos nuevos mecanismos de recuperación de subsegmentos incorporados en esta nueva versión a través de una serie de experimentos que detallaremos a continuación.
4. Comparar la utilidad de estas nuevas técnicas con las clásicas que ya tenía incorporadas. Por ejemplo, el Diccionario de Autosugerencia.

Además, analizaremos en qué circunstancias esta herramienta ofrece más resultados y si estos son de calidad para que el traductor conozca cómo sacarle el máximo partido, no solo a este mecanismo en particular, sino también al programa completo.

No obstante, el proyecto no se centrará únicamente en evaluar cómo de eficaz es dicho mecanismo de aprovechamiento de memoria de traducción (TM)<sup>1</sup> de manera aislada, sino que se realizará un estudio comparativo con respecto al Diccionario de Autosugerencias. Este análisis nos ayudará a determinar cuál de las dos herramientas es más útil según qué circunstancias.

### 1.3 ESTRUCTURA

Con el fin de cumplir los objetivos anteriormente mencionados, este proyecto se basa en un enfoque de tipo teórico y práctico al mismo tiempo. Por ello, comenzaremos ofreciendo un marco teórico-conceptual en el que se explicarán conceptos clave para este trabajo y se proporcionará información sobre las memorias de traducción, los corpus y los tipos de recuperación de información en la que se incluyen la recuperación de segmentos y subsegmentos. Este último está estrechamente relacionado con la traducción automática, dado que algunas de las

---

<sup>1</sup> La abreviatura utilizada para “memoria de traducción” es “TM”, que proviene del inglés “Translation Memory”. Esto se debe a que “MT” podría confundirse con el término también en inglés “Machine Translation”, que corresponde a la traducción automática.



técnicas utilizadas en traducción automática estadística se emplean en la recuperación de subsegmentos.

En el siguiente punto, nos centraremos en el enfoque práctico, por lo que realizaremos un experimento con el fin de explicar y comprender cómo funciona la recuperación de subsegmentos en los mecanismos que SDL Trados Studio 2017 ha incluido. Este experimento consta de seis casos diferentes en los que probaremos cada uno de los mecanismos *upLIFT Fragment Recall*. Para conseguirlo, emplearemos diferentes textos en francés que traduciremos con ayuda de un corpus bilingüe en francés y en español. El corpus con el que trabajaremos estará relacionado con el tema sobre el que tratan dos de los textos con los que vamos a trabajar con el fin de que la información que se recupere sea lo más relevante posible para nuestro producto final. No obstante, como hemos mencionado anteriormente, también se incluirán otros dos textos de temática diferente para evaluar el nivel de recuperación en los casos en que el tema no coincida.

El experimento se llevará a cabo introduciendo tanto los archivos del texto de partida como el del corpus en la herramienta de traducción y analizaremos los resultados con y sin los diferentes mecanismos que SDL Trados Studio 2017 ha incluido para la recuperación de subsegmentos. Así, lograremos comprobar cuál de ellos funciona mejor o si, por el contrario, no son de gran utilidad. Además, realizaremos diferentes pruebas modificando el nivel de coincidencia para observar cómo afecta este dato a la recuperación de información y, por consiguiente, al proceso de traducción. Finalmente, se incluirán los resultados de cada uno de los casos con los que extraeremos unas conclusiones.

#### 1.4 METODOLOGÍA

Con el fin de establecer una estructura coherente y de aportar información sobre el tema elegido, partimos de la base de los conocimientos adquiridos durante el grado de Traducción e Interpretación. En primer lugar, en la asignatura Informática Básica (2º curso), que proporciona al alumno conocimientos elementales para maquetar proyectos, impartida por Emilio Rodríguez Vázquez de Aldana. También,

Recursos Tecnológicos para la Traducción (3º curso) ha sido de crucial importancia, ya que pone en contacto al alumno con el ámbito de las tecnologías para la traducción, impartida por Emilio Rodríguez Vázquez de Aldana y Jesús Torres del Rey. Por último, Gestión Terminológica y de Proyectos (4º curso), impartida por Joaquín García Palacios y Jesús Torres del Rey, que ofrece la oportunidad de trabajar con un proyecto real de traducción para el que se necesitan herramientas TAO (Traducción Asistida por Ordenador) y resulta de gran interés para el presente análisis, ya que fue necesario utilizar SDL Trados 2017. También, la bibliografía e información propuestas por los profesores mencionados anteriormente, constituyeron un punto de partida en la búsqueda temática sobre este ámbito especializado. Finalmente, y al cabo de un exhaustivo periodo de búsqueda, lectura, selección y revisión sobre la recuperación de subsegmentos, confeccionamos una estructura expositiva que ofreciera la aproximación teórica necesaria al tema de este trabajo.

Por otro lado, para cumplir los objetivos de la parte experimental, hemos estudiado, en primer lugar, la documentación de SDL Trados Studio para descubrir y ser capaces de explicar en el presente trabajo las diferentes técnicas que este utiliza para llevar a cabo la recuperación de subsegmentos. También, dado que no existe una metodología establecida para la valoración de su utilidad, hemos optado por escoger corpus creados por la Dirección General de Traductores (DGT), que utilizaremos como memorias de traducción. En concreto, trabajaremos con dos, pues así nos aseguraremos de que el corpus sea lo suficientemente extenso como para que se obtengan resultados interesantes. Con el fin de que el rango de cobertura de dichas pruebas sea lo más amplio posible, seleccionaremos textos de diversa índole. Así, con documentos distintos (dos originales de la Comisión Europea, uno de Naciones Unidas de carácter administrativo y, finalmente, uno de carácter científico sobre el sida), nuestra intención es comprobar si este mecanismo resulta eficiente o no sea cual sea el campo en que se trabaje.

El experimento práctico contará con cinco casos diferentes, todos ellos orientados a comprobar dos de los mecanismos de recuperación de subsegmentos: *upLIFT TU Match* y *upLIFT Fragment Match* incorporados por la empresa SDL y

cuyos funcionamientos detallaremos más adelante. En el primer caso, activaremos el mecanismo de recuperación de unidades completas de traducción. En el segundo, haremos lo propio con el mecanismo de recuperación de fragmentos. En los dos siguientes, modificaremos los umbrales de coincidencia para comprobar si aumenta el número de resultados. Para terminar, en el último caso compararemos la recuperación de fragmentos con la que ofrece el Diccionario de Autosugerencias.

## 2 MARCO TEÓRICO

### 2.1 CORPUS

Según Corpas Pastor, un corpus es una colección de textos en soporte digital que ha sido compilada conforme a unos criterios de diseño apropiados que le confieren representatividad en una lengua determinada. (Corpas Pastor, 2012: 2).

Existen diferentes tipos de corpus según del criterio que siga a la hora de seleccionar y recopilar los textos a partir de los que se va a crear. Pueden regirse en función del idioma en que estén escritos, su formato o el fin con el que se construyen, esto es, el uso que se le dará en un futuro. En primer lugar, si tomamos como referencia la lengua del corpus, es posible encontrar corpus monolingües, bilingües o multilingües. Estos son recopilaciones de un idioma, dos o varios, respectivamente. Los primeros suelen emplearse para obtener más detalles sobre el uso de una lengua en un contexto determinado, extraer terminología específica, etc. Los segundos y terceros se constituyen por textos en dos o más lenguas que han sido alineados, de manera que cada expresión cuenta con su equivalente en el resto de idiomas.

No obstante, también pueden organizarse en función de si se trata de un corpus escrito u oral. Entre los escritos, es posible encontrar los corpus de referencia, los monitores y los paralelos. Los primeros se utilizan para confeccionar gramáticas, diccionarios o cualquier material que sirva de referencia. Por otra parte, los corpus monitores son aquellos que se mantienen actualizados permanentemente y se emplean para estudiar la lengua desde un enfoque diacrónico. Finalmente, los paralelos son recopilaciones de textos traducidos a una o más lenguas (Otero Martínez, 2015:9). Este último grupo es el que resulta más interesante para el presente trabajo, pues los corpus procedentes de la Dirección General de Traductores (DGT)<sup>2</sup> que utilizaremos para llevar a cabo el experimento, que detallaremos más adelante, pertenecen a este tipo de corpus. De hecho, la siguiente explicación de Koehn se ajusta en gran medida a los corpus con los que hemos trabajado. “A parallel corpus is a collection of

---

<sup>2</sup><https://ec.europa.eu/jrc/en/language-technologies/dgt-translation-memory#dgt-memory>

text, paired with translations into another language. Several parallel corpora have recently become available. These are often in the order of tens to hundreds of millions of words”<sup>3</sup>. (Koehn, 2010: 54)

Los corpus tienen gran importancia en el mundo de las tecnologías de la traducción, ya que tanto los sistemas de memorias de traducción como la traducción automática tienen su base en ellos. Ambas herramientas constituyen una ayuda al traductor y tienen como fin reducir su trabajo. Dichos sistemas de memorias de traducción utilizan bases de datos donde se almacenan los corpus alineados para, una vez rastreada esta base de datos, recuperar segmentos previamente traducidos y que se puedan volver a utilizar en el caso de que estos segmentos coincidan en un porcentaje mínimo previamente establecido. Además, las memorias de traducción no solo se valen del corpus, sino que son en sí mismas son corpus bilingües o multilingües alineados (Corpas Pastor, 2012: 4).

## 2.2 MEMORIAS DE TRADUCCIÓN

Teniendo en cuenta la idea sobre la que se basan las memorias de traducción, podríamos definir las como un archivo de traducción o base de datos que almacena segmentos de un conjunto de textos en lengua origen alineados con su traducción correspondiente. Los segmentos normalmente son frases completas, pero también pueden estar constituidos por fragmentos más cortos. (SDL Trados Studio, 2015)

Con el fin de que tanto el proceso de almacenamiento como la posterior recuperación de propuestas de traducción sean más sencillos, la herramienta divide el texto en segmentos de manera automática. Para ello, se basa en unas reglas de segmentación que se centran en signos ortotipográficos como los puntos finales de oración y de párrafo, los signos de interrogación y exclamación finales, etc. Cada segmento en lengua origen junto con su equivalente en lengua de destino recibe el nombre de unidad de traducción (UT). Dichas unidades se encuentran almacenadas en el archivo y pueden recuperarse en futuras traducciones en que aparezcan en segmentos iguales o similares.

---

3 Un corpus paralelo es una colección de textos, alineados con traducciones en otro idioma. Recientemente, se han puesto a disposición de los usuarios diferentes corpus paralelos. Estos suelen contener desde decenas hasta cientos de millones de palabras.

Hoy en día, los sistemas de memorias de traducción son la principal herramienta para la ayuda a la traducción, dado que ninguna otra tecnología ha cambiado de la misma manera las condiciones generales de la traducción. Quizá, esto puede estar relacionado con que las memorias de traducción sirven, principalmente, para ayudar a los profesionales en su trabajo sin influir de manera radical en los procesos cognitivos de la traducción que requieren la creatividad y el conocimiento de un traductor humano. (Reinke, 2013:27)

### 2.2.1 Historia de memoria de traducción

Las memorias de traducción surgieron en los años sesenta, cuando la Comunidad Europea del Carbón y del Acero desarrolló y utilizó un sistema informático para recuperar términos y sus contextos de traducciones almacenadas. Dichas traducciones no habían sido realizadas por un ordenador, sino por un traductor humano (Reinke, 2013 :2).

De hecho, en los años setenta, el servicio de traducción del Ejército Federal de Alemania desarrolló un modelo que consistía en reutilizar traducciones ya existentes realizadas por un humano a través de un ordenador. Su principal objetivo era facilitar y acelerar el proceso de revisión identificando los pasajes que coincidían. Además, esto ayudaría en la revisión de nuevas ediciones de traducciones en las que solo sería necesario reescribir aquellos fragmentos que fueran diferentes a los encontrados en la memoria (Krollmann, 1971: 123).

A finales de los años setenta, Peter Arthern, traductor de la Comisión Europea, propuso un apoyo más avanzado para el traductor. El modelo que el proponía era diferente al de Krollmann, pues no solo incluía la reutilización de segmentos idénticos, sino también la recuperación de segmentos similares a los que se están traduciendo. Hoy en día, estas coincidencias se conocen como correspondencia total o *exact match* y correspondencia parcial o aproximada o *fuzzy matching*, respectivamente.

En los años ochenta, Kay ya hablaba de un hipotético sistema en el que el ordenador traduciría automáticamente la unidad seleccionada. El usuario tendría el poder de intervenir en el proceso cuando lo creyese necesario, editando la traducción que propusiera la máquina. Además, la herramienta consultaría al traductor cuando tuviese que hacer frente a alguna

decisión o cuando se topase con ambigüedades en el texto origen. (Kay, 1980: 18 en Otero Martínez, 2015: 14).

### 2.2.2 Tipos de memorias de traducción

Es posible encontrar dos tipos de sistemas de memorias de traducción, principalmente: los basados en frases y los basados en cadenas de caracteres en bitextos (*character string in bitext*). Los primeros, con los que trabajaremos en este trabajo, parten de la división del texto origen y el texto meta en unidades de traducción equivalentes a oraciones o frases, el almacenamiento de dichas UT en una base de datos y la identificación de frases idénticas o similares a estas en el texto con el que se está trabajando.

Los corpus basados en cadenas de caracteres se basan en almacenar bitextos enteros en una base de datos y buscar cadenas de caracteres idénticas de cualquier longitud (Gow, 2003: 2). Los bitextos son textos de origen asociados a sus correspondientes traducciones. La diferencia con el modelo anterior es que, en este caso, la memoria se construye a medida que se van encontrando las correspondencias.

## 2.3 RECUPERACIÓN DE SEGMENTOS

En primer lugar, cabe destacar que los sistemas de memorias de traducción se sirven de tres tipos de archivos: el documento con el que vamos a trabajar y su traducción final, el archivo bilingüe que la propia herramienta crea en un formato propio y las bases de datos correspondientes, ya sean memorias de traducción, bases de datos terminológicas, etc. El archivo de origen no se modifica durante la traducción, sino que es el archivo bilingüe el que se va completando conforme se va avanzando en el trabajo. Por último, una vez terminada la traducción, se crea el documento de texto en destino.

De esta manera, el traductor trabaja sobre el archivo bilingüe en el editor, en que aparece el texto dividido en segmentos en la parte izquierda de la pantalla y otra columna a la derecha con sus equivalencias correspondientes. Estas últimas las puede haber escrito el traductor manualmente o haber sido recuperadas por la herramienta a partir de la memoria de traducción, pues, durante todo el proceso de trabajo, el programa busca constantemente en la

memoria segmento que coincidan total o parcialmente con los del documento. Esto viene determinado por un umbral de coincidencia mínimo que fijamos nosotros.

Para que la herramienta sea capaz de determinar si el segmento en cuestión se ajusta o no al porcentaje que hemos establecido (principalmente, en el caso de las coincidencias parciales), es necesario que tenga en cuenta el número de modificaciones, eliminaciones, inserciones o sustituciones que deben hacerse para transformar una cadena de caracteres determinada en la que se pide. Así, la base de esto reside en buscar las diferencias entre la cadena de caracteres que se va a traducir y la que se encuentra la memoria. (Recursos Tecnológicos Aplicados a la Traducción).

## 2.4 RECUPERACIÓN DE SUBSEGMENTOS

### 2.4.1 Aspectos generales sobre la recuperación de subsegmentos

Dado que la mayoría de los usuarios de las herramientas TAO consideran que las memorias de traducción quedan infrutilizadas con solo la recuperación de segmentos completos, las empresas que distribuyen sistemas de memorias de traducción están intentando desarrollar mecanismos que permitan hacer un uso más completo de ellas con el fin de sacarles el máximo partido. Por ello, SDL Trados Studio ha incluido un sistema que analiza y recupera fragmentos más pequeños sin necesidad de que sean frases completas únicamente. Estos fragmentos reciben el nombre de subsegmentos. Esta idea se basa en que es bastante frecuente que se repitan unidades de menor longitud que los segmentos a lo largo de un documento. Estas unidades pueden estar constituidas por una o más palabras, pero que con los sistemas de recuperación a niveles mayores que teníamos hasta ahora quedaban ocultas. El fin de estos nuevos mecanismos es el de recuperar estas unidades menores y aumentar la productividad y reusabilidad de las memorias de traducción. De esta forma, el traductor necesitaría menos tiempo y esfuerzo para realizar su trabajo.

No obstante, no existe una única manera de llevar a cabo esta tarea, pues cada herramienta de traducción emplea una técnica diferente. Además, la situación actual es que la recuperación de subsegmentos existe, pero aún se puede mejorar para cumplir las expectativas de los clientes, puesto que, hoy en día, los traductores esperan ciertas



características de estas herramientas. Según Flanagan (2015), entre las expectativas de los clientes, se encuentra la recuperación a nivel de subsegmentos (*Subsegment Recall Typology*) en que se buscan, principalmente, unidades de traducción a partir de una Base de Datos Terminológica o una memoria de traducción. De esta manera, cuando se traduce, la herramienta comprueba si hay coincidencias en alguno de estos recursos y lo ofrece. La ventaja que esto presenta es que no es necesario un trabajo previo con respecto a la memoria o a la base de datos. También contempla la búsqueda de concordancias automática (*Automatic Concordance Search, ACS*) que permite hacer una selección de los subsegmentos que puedan ser relevantes para el traductor. Funciona con un regulador de longitud del fragmento. El traductor tiene que analizar y evaluar la propuesta para situarla. Por último, resulta relevante para el presente trabajo la extracción de fragmentos bilingüe (*Bilingual Fragment Extraction*) en el que se incluye el Diccionario de Autogerencias que SDL Trados Studio incluyó en 2009. Este mecanismo reutiliza contenido previamente traducido a modo de diccionario.

Por otra parte, y según Flanagan, la recuperación de subsegmentos puede tener consecuencias negativas. Principalmente, la descontextualización, ya que los subsegmentos aparecen aislados y sin contexto y la pérdida de variación, ya que, si se usa una memoria de traducción, es posible ver diferentes traducciones para un mismo segmento o una traducción para diferentes segmentos. Esto es positivo para el traductor, pero cuando se trabaja con una base de datos, suele aparecer solo un resultado por segmento. En cualquier caso, también existen alternativas para enriquecer la codificación de las traducciones en TMX con información sobre la equivalencia de los subsegmentos. De esta manera, se obtendrán mejores resultados en la recuperación de subsegmentos y, por lo tanto, un mejor aprovechamiento de la memoria. Dos de ellas consisten en hacer anotaciones aparte antes de crear la memoria de traducción o bien crearlas en una memoria ya existente y con segmentos guardados. Estas reciben el nombre de Stand-Off Annotation y Embedded Annotation, respectivamente<sup>4</sup>. (Forcada, 2014)

---

<sup>4</sup> Para más información sobre anotaciones en el formato TMX de las memorias de traducción, leer <https://pdfs.semanticscholar.org/9eee/122860041d5d769b687f6e22699a2a305516.pdf> Forcada, 2014.

Reinke también establece tres estrategias diferentes para llevar a cabo este tipo de recuperación. Para empezar, la más sencilla sería buscar unidades completas que constituyan partes del segmento en lengua origen en la memoria o la base de datos terminológica con la que estemos trabajando e insertar las equivalencias en lengua de destino de manera automática. Teniendo en cuenta que el presente trabajo probará el nuevo mecanismo que se ha introducido en la última versión de SDL Trados Studio, la de 2017, esta primera estrategia podría compararse con *upLIFT TU Match*, con la diferencia de que en nuestra interfaz no se incluye la equivalencia en lengua de destino, pues es el traductor quien tiene que hacerlo manualmente. En segundo lugar, se encontraría otra basada en encontrar cadenas de caracteres que se repitan en la memoria, lo que podríamos equiparar con *upLIFT Fragment Match*. Finalmente, la última manera consiste en que la herramienta propone diferentes equivalencias para los fragmentos que se recuperen a partir del segmento origen en una pantalla que se despliega al comenzar a escribir la traducción. Esto se conoce como Diccionario de Autosugerencias en SDL Trados Studio, pero, además, también se podría comparar con *upLIFT Fragment Match* de nuevo, ya que actúa de esta manera de forma automática (Reinke, 2013: 33)

### 2.4.2 La traducción automática y la recuperación de subsegmentos en SDL Trados Studio 2017

En general, la traducción automática tiene como objetivo construir sistemas artificiales que puedan simular la capacidad humana para procesar el lenguaje natural. Su origen se encuentra en los años cincuenta cuando se creó el primer programa de traducción automática. Después de varias décadas de evolución, en 1990 los sistemas de traducción automática basada en corpus comenzaron a cobrar importancia. Dentro de esos sistemas, se encuentran la traducción automática estadística (SMT, Statistical Machine Translation) y la traducción automática basada en ejemplos (EBMT, example-based machine translation). En estos últimos años, hemos asistido a la entrada en escena de un nuevo paradigma, la traducción automática neuronal (NMT, Neural Machine Translation), también basada en Corpus (Forcada, 2017). No obstante, la traducción automática estadística es la que nos interesa en este trabajo, pues, con el paso del tiempo, ha ido adquiriendo importancia y ha

pasado a ser una de las bases de los motores de traducción, ya que es relativamente sencillo disponer de colecciones extensas de textos en diferentes idiomas.

En SDL Trados Studio 2017, la recuperación de subsegmentos recoge técnicas utilizadas en la traducción automática estadística. En concreto, las empleadas para la conformación de lo que en dichos sistemas se conoce como “tabla de frases” (phrase table) (Koehn, 2010:136).

Al igual que la traducción automática estadística, SDL Trados Studio crea unos modelos de traducción, que según Hearne y Way (2011), está constituido por una serie de frases en el idioma de destino asociadas con una lista de traducciones de frases en lengua origen y con las probabilidades correspondientes. Dichas probabilidades se basan en cálculos relativos o teniendo en cuenta las probabilidades de traducción léxicas (Koehn, 2010: 8). Recordemos que una memoria de traducción en SDL Trados no es más que un conjunto de N pares de oraciones origen y destino pareadas. Pero este alineamiento solo a nivel de oración necesita completarse con un alineamiento a nivel de fragmentos (palabras o secuencias de palabras).

Hay softwares libres, como por ejemplo, Giza ++<sup>5</sup>, que realizan este tipo de alineaciones. Para ello, es necesario un corpus a partir del cual se crea un modelo de traducción. En este proceso, tiene lugar una primera fase de alineación de palabras y, una vez que este proceso se termina, se produce una segunda fase de alineación de frases a partir de las palabras alineadas en la primera. Este último consiste en extraer todas las parejas de frases que sean consistentes con respecto a la alineación de palabras (Koehn, 2010:8). En los sistemas de traducción automática estadística basada en frases, una frase es una secuencia de palabras de tamaño variable. No tiene por qué tratarse de sintagmas, pues los signos de puntuación también pueden formar parte de ella (Recursos Tecnológicos para la Traducción).

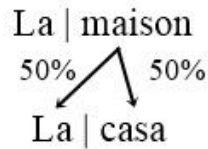
---

<sup>5</sup> [www.statmt.org](http://www.statmt.org)

Para llevar esto a cabo, intentan ser como un humano a la hora de deducir qué palabra corresponde con cuál. Por ejemplo, si encontramos “La maison” y “La casa”, la herramienta considera que “la” tiene el mismo porcentaje de posibilidad para corresponderse con “la” o con “casa”. No obstante, si, además de esta oración, tenemos “la maison rouge” y “la casa roja”, la herramienta acabará deduciendo que “la” tienes más probabilidades de corresponderse con “la” y “maison” con “casa”. Esta explicación se ilustra en las Figuras 1 y 2.



Existe la misma probabilidad para que se alinee "la" con "la" o "casa".



Existe la misma probabilidad para que se alinee "casa" con "la" o "casa".

**Figura 1: Paso 1 en la traducción automática estadística.**

Teniendo esto en cuenta:



La herramienta deduce que "rouge" corresponde a "roja", pero sigue sin saber si "la" corresponde a "la" o a "casa".

Figura 2: Paso 2 en la traducción automática estadística.

## 3 MARCO EXPERIMENTAL

### 3.1 SDL TRADOS 2017 Y LA RECUPERACIÓN DE SUBSEGMENTOS

#### 3.1.1 ¿Qué es?

En la última versión de la herramienta de ayuda a la traducción de la empresa SDL Trados, SDL Trados Studio 2017<sup>6</sup>, se ha incluido el mecanismo de recuperación de subsegmentos que recibe el nombre de *upLIFT Fragment Recall*. Esta novedad permite recuperar palabras o frases de nivel inferior al segmento entre los ya traducidos que se encuentran almacenados en nuestra TM. Lo que pretende SDL con esta implementación es aprovechar al máximo la memoria de traducción, ya que solo con la recuperación de segmentos completos, queda infrutilizada. Esto, a su vez, podría reducir el tiempo de búsqueda a los traductores, pues, en caso de no encontrar un segmento completo, el programa recupera automáticamente partes de oraciones y las propone como posibles traducciones al segmento con el que se está trabajando. Además, con este mecanismo es posible añadir los resultados propuestos a una base de datos terminológica en caso de que estos fueran acertados.

#### 3.1.2 ¿Cómo funciona?

Como se puede comprobar en la página web de SDL Trados Studio 2017<sup>7</sup>, es posible encontrar dos tipos de resultados cuando se hace uso del mecanismo upLIFT. El primero recibe el nombre de *Fragment Recall* y busca fragmentos cuando no hay ningún otro tipo de coincidencia. Esto es, en el caso de que no se encuentren equivalencias a nivel de segmentos en la TM, comienza la búsqueda de unidades más pequeñas relacionadas con el segmento que se está traduciendo. Estos fragmentos varían en longitud y pueden ser, a su vez, de dos tipos: *TU Match* (unidades de traducción completas anteriormente traducidas y almacenadas en la

---

<sup>6</sup><https://www.sdl.com/es/software-and-services/translation-software/sdl-trados-studio/uplift-translation-memory/>

<sup>7</sup><https://www.sdl.com/es/software-and-services/translation-software/sdl-trados-studio/uplift-translation-memory/>

memoria) y *TU Fragment Match* (partes de segmentos que aparecen como fragmentos en un nuevo contexto y contenido). Ambos mecanismos presentan dos opciones a la hora de marcar los niveles de coincidencias: una primera casilla para decidir el “número mínimo de palabras para una coincidencia” y otra para el “número mínimo de palabras significativas para una coincidencia”. La diferencia entre ambas reside en que la herramienta intenta deducir cuáles son las palabras significativas en una secuencia. Estas serían los adjetivos, sustantivos o verbos, pero, dado que no cuenta con conocimientos lingüísticos, se basa, aparentemente, en la longitud de las palabras, ya que, en la mayoría de ocasiones, los nombres o verbos son más largos que los artículos o pronombres.

En la Figura 3, puede comprobarse cómo se pone en marcha el mecanismo de recuperación upLIFT TU Match, en que se buscan coincidencias 100% con segmentos ya traducidos. Por ello, aparece el mensaje de “coincidencia de unidad de traducción completa” y se muestra con un icono azul.

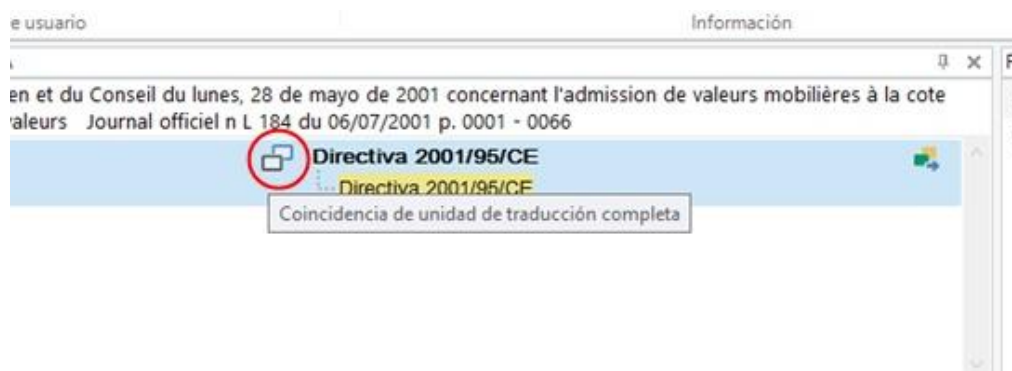


Figura 3: Mecanismo upLIFT TU Match activado.

Para TU Match, la herramienta SDL Trados no necesita ningún tipo de alineación porque los subsegmentos ya se encuentran divididos así en la memoria. En este caso, el mecanismo encuentra coincidencias para fragmentos del documento que son unidades completas de traducción en la TM. Este procedimiento se ilustra en la Figura 4.

### upLIFT TU Match

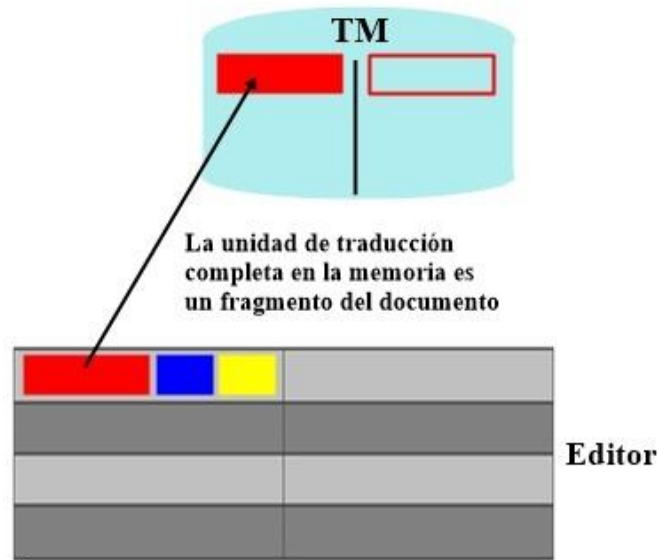


Figura 4: Recuperación de upLIFT TU Match.

Por otra parte, en la Figura 5 podemos comprobar cómo se ha encontrado una “coincidencias de fragmento de unidad de traducción”, representada con un icono rojo. Esto nos muestra el mecanismo *upLIFT Fragment Match* en marcha, a través del cual la herramienta busca pequeñas partes de subsegmentos ya traducidos.



Figura 5: Mecanismo upLIFT Fragment Match activado.



En este caso, el fragmento para el que se busca una correspondencia figura también como parte de un subsegmento en la TM. De esta manera, varios fragmentos de un mismo segmento en la memoria pueden utilizarse como equivalencia de fragmentos pertenecientes a segmentos diferentes del documento sin importar, además, el orden en que aparezcan en el segmento original. Esto es, un fragmento que se encuentre al inicio de una oración podrá ocupar cualquier lugar de la frase para la que se recupera. Este procedimiento se muestra en la Figura 6.

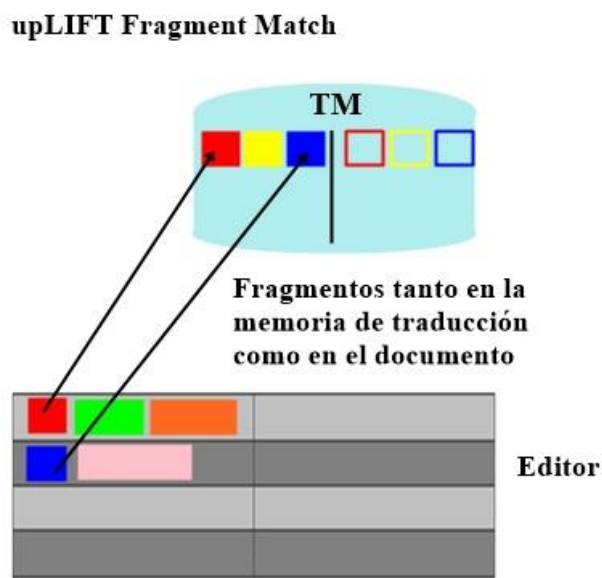


Figura 6: Recuperación de upLIFT Fragment Match.

El otro tipo de mecanismo del que dispone upLIFT recibe el nombre de *upLIFT Fuzzy Repair*, en el cual la información se recuperará en aquellas situaciones en que se produzca “fuzzy matching”. En la página web upLIFT Trados Studio 2017<sup>8</sup>, se explica cómo la herramienta recurrirá a nuestras TM, bases de datos o a la traducción automática para tratar de mejorar la coincidencia entre las oraciones. De esta manera, el programa recupera posibles propuestas de traducción en las que se puedan realizar algunas modificaciones con el fin de que el resultado sea el correcto. Es la propia herramienta la que se encarga de realizar los

<sup>8</sup> [https://blog.sdltrados.com/uplift-fuzzy-repair-sdl-trados-studio-2017-transformational-tm-technolog](https://blog.sdltrados.com/uplift-fuzzy-repair-sdl-trados-studio-2017-transformational-tm-technology/)  
y/

cambios que sean necesarios para arreglar el “fuzzy matching”. Esto se realiza de manera automática, por lo que el traductor no necesita hacer nada más. De manera general, *upLIFT Fuzzy Repair* utiliza los resultados de *upLIFT TU Fragment Match*.

En este sistema, la tecnología *upLIFT Fuzzy Repair* es capaz de deducir qué palabra necesita ser modificada en el segmento que recupera como propuesta de traducción. En el caso de que la memoria de traducción o la base de datos con la que estemos trabajando contenga esta palabra traducida, el programa automáticamente la reemplazará. En el caso de que las palabras en cuestión sean del mismo género o tengan el mismo número, no habrá necesidad de que el traductor intervenga en absoluto. No obstante, si estos no coinciden, se requerirá un proceso de posesición, pero que, en la mayoría de las ocasiones, será mucho más rápido y ligero de trabajo que en situaciones normales. Este mecanismo no tiene un icono propio para indicar el tipo de recuperación, sino que usa el que se asigna para “Fuzzy Matching”. Este procedimiento se muestra en la Figura 7.

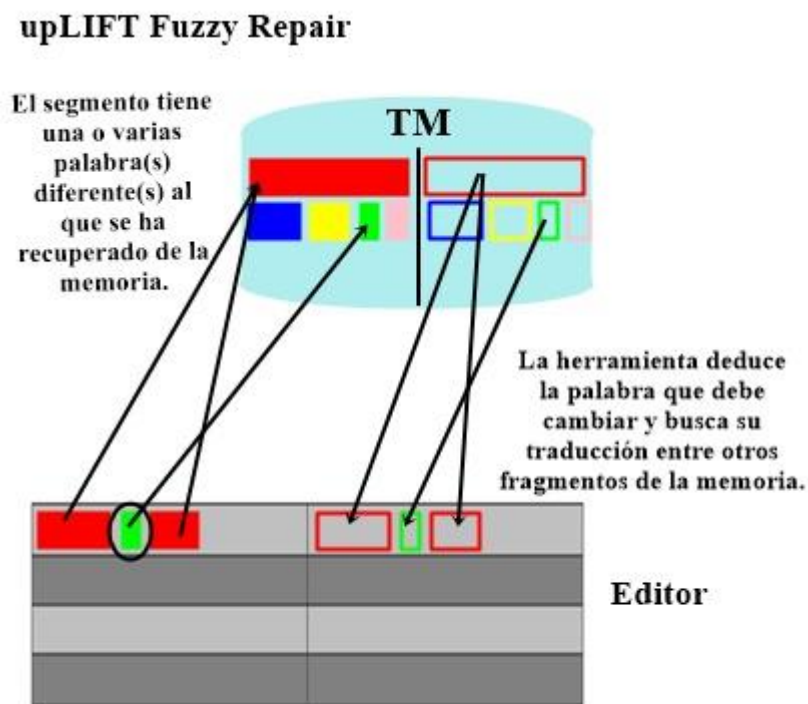


Figura 7: Recuperación de upLIFT Fuzzy Repair.

### 3.1.3 ¿Cómo se activa upLIFT en SDL Trados 2017?

Para activar la utilidad upLIFT<sup>9</sup>, es necesario abrir una memoria de traducción, y, en lugar de abrir la “Configuración del proyecto”, hay que ir a la configuración de la propia memoria y clicar en “Alineación de fragmentos”. Después, seleccionar “activar” en el primer recuadro para que los segmentos se alineen. Este proceso puede tardar un tiempo considerable si trabajamos con corpus de gran tamaño. Al mismo tiempo, también es recomendable marcar “alinear el nuevo contenido automáticamente” para que lo que se traduzca pueda utilizarse en otras situaciones. Una vez hecho esto, para visualizar las pantallas en las que la herramienta muestra los resultados obtenidos, hay que clicar en la pestaña “Ver” situada en la parte superior de la pantalla. Posteriormente, es necesario activar el botón “Ventana de coincidencia de fragmentos”, que está situado en el panel superior de la esquina derecha.

## 3.2 PREPARACIÓN DEL EXPERIMENTO

Para realizar este experimento, se escogieron dos TM de la Dirección General de Traductores de la Comisión Europea de diferentes fechas. En su página web<sup>10</sup>, es posible encontrar en formato ZIP sus propios corpus creados a partir de documentos oficiales en numerosos idiomas. Dichos archivos contenían un número de unidades de traducción de entre 70 000 y 90 000. Para pruebas de este tipo, es necesario que los corpus sean, aproximadamente, de este tamaño, ya que la herramienta debe contar con suficientes unidades de traducción para ofrecer resultados tanto a nivel de unidad de traducción completa como a niveles más inferiores. Como esta prueba requería únicamente los idiomas francés y español, se realizó una extracción en TMX de las partes correspondientes a los mismos con un programa Java<sup>11</sup> que ellos mismos ofrecen en su sitio web. Este proceso no sobrepasó los cinco minutos a pesar de estar trabajando con memorias de tamaño

---

9 Para más información, ver [https://www.youtube.com/watch?time\\_continue=93&v=d6Pic4t--mM](https://www.youtube.com/watch?time_continue=93&v=d6Pic4t--mM)

10 <https://ec.europa.eu/jrc/en/language-technologies/dgt-translation-memory#download>

11 <https://ec.europa.eu/jrc/en/language-technologies/dgt-translation-memory#how-to>

considerable. Uno de los corpus tiene fecha del 2010 y aparece nombrado en la página web como “Vol\_2010\_3” y que cuenta con más de 81 000 segmentos. El otro, data de 2016, contiene un total de 30 435 unidades de traducción y su denominación es “Vol\_2016\_9”.

Por otra parte, los documentos que empleamos para traducir tratan temas muy diferentes. Primeramente, escogeremos dos documentos de la Comisión Europea. Ambos son directivas publicadas en el Journal Officiel, por lo tanto comparten tipología. El motivo para la elección de estos textos reside en que se trata de documentos muy repetitivos, lo cual podría ayudarnos a aumentar la recuperación de subsegmentos y conseguir resultados más numerosos para comentar. De hecho, estas directivas se utilizan para crear los corpus que ofrece la Dirección General de Traductores. No obstante, para evitar que los segmentos guardados en la memoria de traducción fueran exactamente los mismos que los del texto elegido, descargamos documentos publicados en fechas diferentes a la creación de dichos corpus.

En primer lugar, una de las directivas de la Comisión Europea se publicó en 2001<sup>12</sup>. Para facilitar el trabajo con SDL Trados Studio, y dado que la página web ofrece la posibilidad de descargarlo en diferentes formatos (PDF o HTML), escogimos este último, pues el formato PDF, como sabemos, requiere de una conversión a un formato editable que en SDL Trados es el formato DOCX. Esta conversión puede, en ciertos casos, plantear problemas. Una vez en el editor, el texto había sido dividido en 1 505 segmentos.

Por otra parte, el otro texto procedente de la Comisión Europea<sup>13</sup> es mucho más reciente, pues se publicó en el Journal Officiel de la misma institución en marzo de 2018. Una vez descargado en formato HTML y abierto en el editor, el documento cuenta con 1 862 segmentos.

En segundo lugar, añadiremos un documento de tipología parecida. No obstante, en este caso, se trata de un informe de Naciones Unidas<sup>14</sup> publicado en 2018. El texto original

---

12 <https://eur-lex.europa.eu/legal-content/FR/TXT/HTML/?uri=CELEX:32001L0034&from=FR>

13 <http://eur-lex.europa.eu/legal-content/FR/TXT/HTML/?uri=OJ:L:2018:067:FULL&from=ES>

14 [https://www.unjiu.org/sites/www.unjiu.org/files/a\\_72\\_34\\_jiu\\_annual\\_report\\_fr.pdf](https://www.unjiu.org/sites/www.unjiu.org/files/a_72_34_jiu_annual_report_fr.pdf)

solo puede encontrarse en formato PDF, por lo que, como se ha mencionado anteriormente, es posible que plantee problemas al añadirlo a SDL Trados. En efecto, la segmentación era errónea al abrirlo en el editor, ya que había numerosos segmentos completamente vacíos. Para intentar mejorar esta división, convertimos el texto a formato TXT. Así, el número de segmentos se redujo radicalmente a 1 875 y ya no figuraba ninguno en blanco.

Por último, seleccionaremos un texto de ONUSIDA<sup>15</sup>, con fecha del año 2017, que versa sobre una campaña contra el sida dirigida a los adolescentes para informar sobre la infección por VIH, la enfermedad del sida y su contagio entre los jóvenes de África. A primera vista, es un documento cuya temática no coincide con la del corpus con el que estamos trabajando. No obstante, deseamos exprimir al máximo el mecanismo upLIFT para estudiar todas sus posibilidades. Con este texto, pretendemos analizar el tipo de resultados obtenidos con el fin de definir el grado de utilidad de esta herramienta incluso en ocasiones en que no dispongamos de una memoria de traducción relacionada con el mismo tema que el del documento a traducir. El documento ha sido dividido en 1873 segmentos.

Estos dos últimos documentos se han elegido con el fin de comprobar si, a pesar de tener una temática diferente, la herramienta sigue recuperando resultados. Además, queremos evaluar, por supuesto, su calidad, pertinencia y analizar los contextos en los que se producen o tienen más posibilidades de aparecer

Todas las partes del siguiente experimento han sido realizadas en un ordenador portátil SONY Vaio con procesador Intel® Core™ i5-4200U/1,60 GHz, tarjeta gráfica HD Intel® 4400/NVIDIA® GeForce® GT 740M con tecnología NVIDIA® Optimus™: 1GB, con 8GB de memoria total y con un almacenamiento de disco duro de 750 GB aproximadamente.

---

<sup>15</sup> [http://www.unaids.org/sites/default/files/media\\_asset/20150217\\_ALL\\_IN\\_brochure.pdf](http://www.unaids.org/sites/default/files/media_asset/20150217_ALL_IN_brochure.pdf)

### 3.3 DESARROLLO DEL EXPERIMENTO: COMPROBACIÓN DEL MECANISMO UPLIFT

Con el fin de aportar un apoyo práctico al objetivo fijado para este trabajo, procedemos a experimentar con la herramienta upLIFT para evaluar su utilidad. El experimento consiste en introducir los textos seleccionados en Trados con la recuperación de subsegmentos activada con el fin de comprobar si, realmente, aumenta la productividad del especialista durante el proceso de traducción. Así, podremos comprobar en qué documentos hay más recuperaciones.

Este experimento cuenta con cinco casos, clasificados en función de lo que se pondrá en práctica en cada uno de ellos. En el caso 1, trabajaremos con el mecanismo *upLIFT TU Match* para comprobar su eficiencia. Este caso contará con dos partes, ya que realizamos intentos con dos memorias de traducción diferentes. En el caso 2, activaremos el *upLIFT Fragment Match* y mantendremos *upLIFT TU Match* activado. En el caso 3, procederemos a modificar los umbrales de coincidencia de *upLIFT TU Match* con el fin de evaluar si las recuperaciones varían. Posteriormente, en el caso 4, cambiaremos dichos umbrales para upLIFT Fragment Recall manteniendo los modificados para el caso anterior. Por último, en el caso 5 realizaremos una comparación con un diccionario de AutoSugerencias que crearemos a partir del corpus con el que trabajamos.

#### 3.3.1 CASO 1: Solo upLIFT TU Match

En este primer caso se estudian los resultados obtenidos con el mecanismo *upLIFT TU Match* activado, sin comprobar, por el momento, la búsqueda de fragmentos (*upLIFT Fragment Match*). El umbral elegido para realizar las búsquedas es 2, ya que es el número que marca el propio programa por defecto.

Cabe destacar que, en algunas ocasiones, el buscador no recupera las propuestas de manera instantánea, sino que tarda alrededor de 15 o 20 segundos hasta mostrar los resultados. Tras pasar por gran parte de los segmentos de todos los textos, hemos podido comprobar que, en efecto, se recuperan algunos fragmentos, denominados por la herramienta como “unidades de traducción completas”. En la mayoría de casos son nombres de directivas, nombres de instituciones, tratados u otras colocaciones recurrentes acuñadas que

aparecen exactamente de la misma manera en la memoria de traducción. Esto ocurre, principalmente, en los dos primeros textos, ya que comparten temática con el corpus. Además, lo más frecuente es que solo haya una propuesta de traducción en la pantalla de “coincidencia de fragmentos”. No obstante, otras veces, el programa ha recuperado más de un subsegmento y, por tanto, los ofrece en una lista ordenados según su pertinencia basándose, en nuestra opinión, en la coincidencia de contextos. Según la página web de SDL Trados Studio 2015<sup>16</sup>, las coincidencias de contextos se dan cuando se produce una coincidencia 100%, pero que, además, comparten contextos, pues el mismo segmento les precede. En cualquier caso, los resultados no son tan frecuentes como cabría esperar, ya que buena parte de los segmentos no presentan ningún tipo de recuperación de la memoria.

De entre los cuatro textos, cabría destacar que el que más recuperaciones presenta es el texto número dos (directiva de 2018), pues, aunque la frecuencia con la que aparecen las propuestas de traducción es similar a la del resto de documentos, sí que ofrece varias opciones en la mayoría de los segmentos para los que ha encontrado resultados.

Otro aspecto interesante es que, a veces, aparecen resultados de traducción, que en realidad son simplemente copias de secuencias de números. Como la herramienta ha emparejado esas secuencias como equivalentes, aparecen como propuestas, pero, evidentemente, no son traducciones reales.

Puesto que los resultados no han sido muy fructuosos en este primer intento con una única memoria (“Vol\_2010\_3), procedemos a fusionar las dos TM con las que estamos trabajando por separado con el fin de comprobar si la herramienta recupera más subsegmentos. De esta manera, a partir de ahora contaremos con una herramienta de trabajo que contiene 90 812 unidades de traducción. Puesto que el tamaño de la memoria puede ser determinante, realizaremos el resto de pruebas con la memoria resultante de la fusión de las dos originales, denominada FR\_ES\_FUSIONADA .

---

<sup>16</sup>[http://producthelp.sdl.com/SDL\\_Trados\\_Studio\\_2015/client\\_es/About/TM/EVWorkingwithTMsAbout\\_Translation\\_Memory\\_Matches.htm](http://producthelp.sdl.com/SDL_Trados_Studio_2015/client_es/About/TM/EVWorkingwithTMsAbout_Translation_Memory_Matches.htm)

Ahora, usando el mismo mecanismo (*upLIFT TU Match*), hemos mantenido el umbral de 2 palabras que aparece por defecto con el fin de no reducir demasiado el margen de coincidencia. No obstante, y a pesar de contar con una memoria de traducción más grande, no se observan cambios notables en los resultados. La herramienta no presenta problemas, de nuevo, a la hora de recuperar las unidades completas, pero de forma general, no se observa una gran diferencia entre los resultados que obtuvimos con el corpus previo y los recuperados con este nuevo corpus fusionado. En efecto, el número de unidades de traducción recuperadas con un nivel de coincidencia 100% es aproximadamente el mismo que en la recuperación de la fase anterior.

Como caso particular, mencionaremos que, para el documento de carácter administrativo extraído de Naciones Unidas, sí se ofrecen resultados solo en los primeros segmentos aunque no son demasiado frecuentes. No obstante, dichos resultados son, sorprendentemente, coincidencias de segmentos completos, por lo que todas se presentan en la ventana de “coincidencias” y no en la de “fragmentos”. Teniendo esto en cuenta, podríamos decir que no obtenemos, en principio, unos resultados fructuosos en relación con lo que nosotros estamos buscando. No es hasta el segmento número 200<sup>17</sup> que encontramos dos propuestas para esta misma oración. Se trata de una unidad de traducción de 3 palabras, “échange d’informations”. Esto se ilustra en la Figura 8.

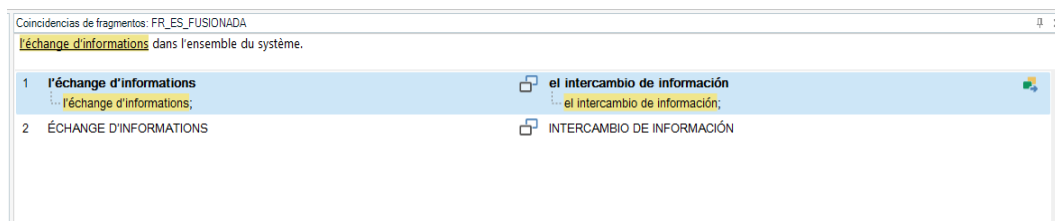


Figura 8: Recuperación de dos unidades de traducción en upLIFT TU Match.

### 3.3.2 CASO 2: También upLIFT Fragment Match

Una vez que hemos analizado los resultados con *upLIFT TU Match*, activamos el segundo mecanismo (sin desactivar el primero, pues si lo hiciéramos, se desactivaría *upLIFT* por completo) para evaluar si hay diferencias en la recuperación de subsegmentos. En primer

<sup>17</sup> Texto del segmento: “l’échange d’informations dans l’ensemble du système”.



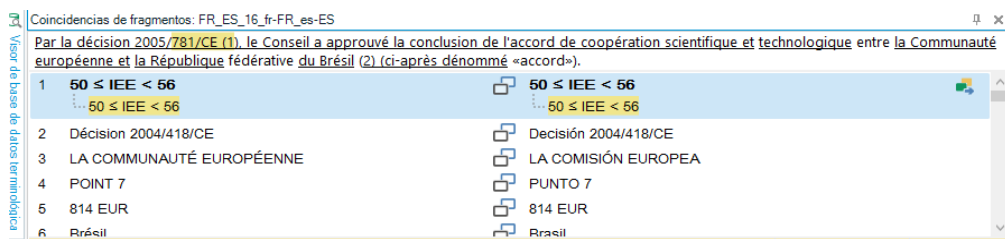
lugar, respetamos el umbral 3 que nos ofrece la interfaz de manera automática. Con este límite, obtenemos resultados muy parecidos a los de la primera fase. Recuperamos, principalmente, unidades de traducción completas y, apenas se encuentran fragmentos de traducción. Solo en algunos segmentos se produce algún cambio: se añade un fragmento como propuesta de traducción. No obstante, son, de nuevo, nombres acuñados en lugar de expresiones. Es decir, lo mismo que ocurría en el Caso 1 con los resultados de *upLIFT TU Match*, pero, esta vez, se recuperan como fragmentos.

Los resultados son, en general, los mismos para todos los textos sin importar demasiado su temática. Así, para los dos últimos también se recuperan simplemente algunos fragmentos, pero no con demasiada frecuencia. Esto se debe, probablemente, a que el umbral de coincidencia de *upLIFT Fragment Match* (coincidencias de fragmentos) es demasiado alto y hay pocas secuencias de tres palabras que coincidan con lo que estamos buscando nosotros.

### 3.3.3 CASO 3: reduciendo el tamaño de los fragmentos en *upLIFT TU Match*

Como el umbral de coincidencia con el que hemos estado trabajando para ha sido de 2 palabras para *upLIFT TU Match*, lo reduciremos a 1 para tratar de aumentar el número de propuestas de traducción. Para que esto funcione correctamente, debemos desactivar la recuperación de fragmentos (*upLIFT Fragment Match*), pues, si no lo hacemos, no será posible evaluar las recuperaciones de unidades completas de traducción.

En la Figura 9, es posible observar que, al modificar este umbral, lo que conseguimos es que aparezcan numerosas unidades de traducción formadas por tan solo una sola palabra en la lista de sugerencias, pues así están almacenadas en nuestra TM. Al reducir el mínimo de coincidencias, las unidades de traducción recuperadas en su forma completa aparecen formando una lista, en ocasiones, de hasta nueve o diez elementos; mientras que, en las fases anteriores, este número era considerablemente menor y oscilaba entre una y dos propuestas en los mejores casos.



**Figura 9: Aumento de propuestas al bajar el nivel de coincidencia en *upLIFT TU Match*.**

### 3.3.4 CASO 4: reduciendo el tamaño de los fragmentos en *upLIFT Fragment Match*

Para continuar con el análisis, procedemos a bajar el número de palabras coincidentes a 2 para *upLIFT Fragment Match*, una vez más, sin desactivar la recuperación de unidades completas de traducción. Con esta opción activada, observamos que empiezan a aparecer más resultados para algunos segmentos. Ahora, hay una gran diferencia entre la longitud de las listas de propuestas, pues oscilan entre los 1 o 2 elementos, como mínimo, y más de 20, en los casos en que se han recuperado muchas coincidencias.

En general, se observa un aumento notable en las recuperaciones en todos los textos con los que trabajamos. No obstante, el tercer texto, de Naciones Unidas con una temática diferente a la del corpus, vuelve a presentar más resultados de los esperados. Por ejemplo, en el segmento número 186<sup>18</sup>, se producen 22 recuperaciones de diferentes fragmentos de unidades de traducción. También, en el número 364<sup>19</sup> encontramos una lista de propuestas con 34 opciones. En el siguiente segmento, 365<sup>20</sup>, a pesar de tener una secuencia de palabras en común con el segmento anterior “procédures internes”, la lista de propuestas de traducción es mucho más larga, pues cuenta con 54 elementos en total. La diferencia reside en que, en este caso, el programa no solo ha reconocido “ses procédures internes”, sino también las palabras que le preceden: “d’accroître l’efficacité des procédures internes”. Además, dichas propuestas no son equivalencias solamente para esta secuencia, sino para

18 Texto del segmento 186: “fonctions de contrôle de la manière dont les secrétariats gèrent les ressources”.

19 Texto del segmento 364: “Le Corps commun a achevé un examen de ses procédures internes, qui a permis”.

20 Texto del segmento 365: “d’accroître l’efficacité des procédures internes et de renforcer l’application des”.

otros fragmentos de la oración, por lo que es normal que haya una diferencia de 20 elementos en la lista. Esta comparación entre ambos segmentos se ilustra en las figuras 8 y 9.

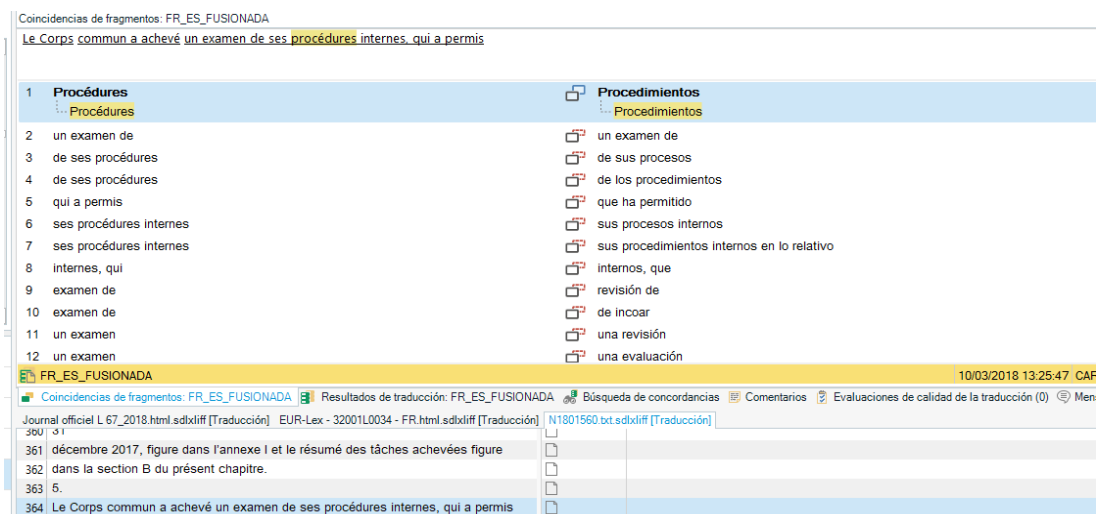


Figura 10: Recuperación de *upLIFT Fragment Match* en el segmento 364.

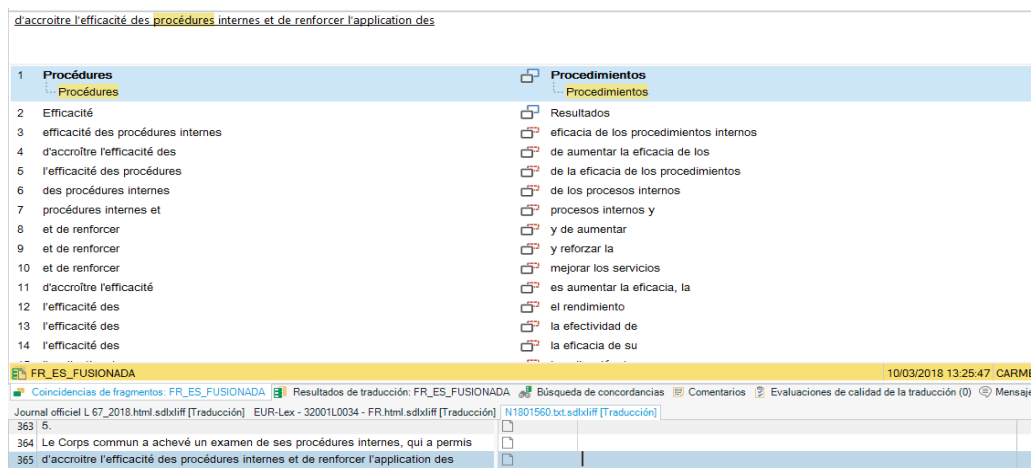
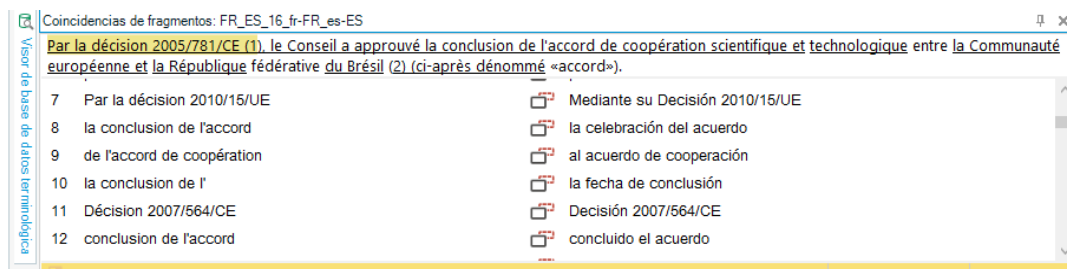


Figura 11: Recuperación de *upLIFT Fragment Match* en el segmento 365.

En el resto de segmentos, también se encuentran opciones aunque todas son similares a las que se acaban de exponer. Suelen ser unidades o fragmentos muy pequeños, de unas dos, tres o cuatro palabras como máximo y que cuentan con numerosas propuestas en cada segmento. Todas estas listas presentan, en primer lugar, las coincidencias de unidades de traducción completas recuperadas por *upLIFT TU Match* en el caso de que las haya y, posteriormente, el resto de opciones buscadas con *upLIFT Fragment Match*.

Decidimos bajar aún más el nivel de coincidencia para este mecanismo (*upLIFT Fragment Match*), por lo que establecemos el mínimo, 1, al igual que hicimos con *upLIFT TU Match* en el Caso 3. De esta manera, obtendremos más cantidad de propuestas con total seguridad, si bien corremos el riesgo de que dichas propuestas sean de menor calidad y, por ello, el ruido aumente; es decir, el grado de precisión a la hora de recuperar posibles traducciones es bajo (Recursos Tecnológicos aplicados a la Traducción).

Como era de esperar, los resultados son más numerosos y, además, más variados. Esto es, unos son recuperaciones de la unidad de traducción completa (mecanismo *upLIFT TU Match*) y los otros constituyen fragmentos de unidades de traducción mayor (*upLIFT Fragment Match*). De hecho, se presentan resultados en casi todos los segmentos. De esta manera, podemos comprobar que el programa, en efecto, está recuperando elementos que antes quedaban ocultos por realizar búsquedas con un nivel de coincidencia más elevado. No obstante, la cantidad no garantiza la calidad y, en efecto, con este nivel de coincidencia tan bajo, muchas de las propuestas que encontramos no son, en absoluto, acertadas. Esto se muestra la Figura 12.



**Figura 12: Lista de propuestas erróneas en *upLIFT Fragment Match*.**

### 3.3.5 CASO 5: El diccionario de Autosugerencias

En este caso del experimento, procederemos a probar el funcionamiento del Diccionario de Autosugerencias con el fin de compararlo con *upLIFT* en el caso siguiente. Esto se debe a que, como se ha mencionado anteriormente, el mecanismo de recuperación *upLIFT Fragment Recall* puede actuar de la misma manera que el diccionario de Autosugerencias que SDL Trados introdujo hace años. Este diccionario propone una lista de posibles términos cuando la memoria no encuentra ninguna coincidencia en la TM.

Para estudiar este comportamiento en profundidad, creamos un diccionario de AutoSugerencias con la memoria fusionada a partir de los dos corpus; es decir, la misma memoria que hemos empleado para estudiar el resto de casos, por lo que tendremos 90 812 unidades de traducción para exportar. Este proceso nos lleva aproximadamente unos 30 minutos. Lo que pretendemos con este análisis, en concreto, es realizar una comparación con el fin de determinar cuál de las dos herramientas es más útil, más rápida o más eficiente. Este funcionamiento se probará solamente con los textos dos y cuatro (Directiva 2018 y ONUSIDA, respectivamente).

Posteriormente, añadimos el diccionario al proyecto y desactivamos los mecanismos upLIFT TU Match y upLIFT Fragment Recall. De esta manera, podremos analizar íntegramente los resultados con el diccionario sin que la recuperación a partir de la memoria de traducción nos resulte molesta.

Cuando comenzamos a escribir la traducción en la parte derecha del editor, el diccionario comienza a actuar, pues aparece una nueva ventana debajo de la frase que estamos escribiendo con la lista de opciones que nos ofrece. En algunas ocasiones, nos ofrece más cantidad de resultados que el mecanismo *upLIFT Fragment Match* en este texto durante las pruebas de los casos anteriores. No obstante, cuando se presentan tantos resultados, la mayoría de ellos no coinciden con la palabra para la que estemos buscando una traducción. Este es, principalmente, el caso de palabras con poco significado como los artículos. En cambio, suele acertar las equivalencias de los adjetivos y sustantivos. El segmento 826<sup>21</sup> del texto procedente de la directiva de 2018 ilustra este caso en particular, por lo que lo detallamos a continuación:

Una vez que empezamos a teclear la traducción del texto, no se presentan equivalencias para palabras como “vu” o “le”. No obstante, en cuanto continuamos tecleando la siguiente palabra, “réglement”, el diccionario despliega una nueva lista de resultados, entre

---

<sup>21</sup> Texto del segmento: “vu le règlement (CE) no 881/2002 du Conseil du 27 mai 2002 instituant certaines mesures restrictives spécifiques à l'encontre de certaines personnes et entités liées aux organisations EIL (Daech) et Al-Qaida (1), et notamment son article 7, paragraphe 1, point a), et son article 7 bis, paragraphe 1,”.

los que es posible observar “reglamento”. Además, no solo ha recuperado la traducción correcta para el término, sino que también ha encontrado “no.” a continuación, pues ya tenía esta secuencia almacenada en la memoria. E, incluso, el diccionario marcó automáticamente la opción más conveniente, esto es: “reglamento no.”, algo que, con la recuperación de subsegmentos, no sucede, pues, en el caso encontrar una traducción apropiada en la lista de sugerencias, es el traductor quién tiene que seleccionar el término correcto de entre todos los ofrecidos. Este mismo proceso, recogido en la Figura 13, ocurre en las siguientes partes del segmento, por lo que se traduce de manera muy rápida.

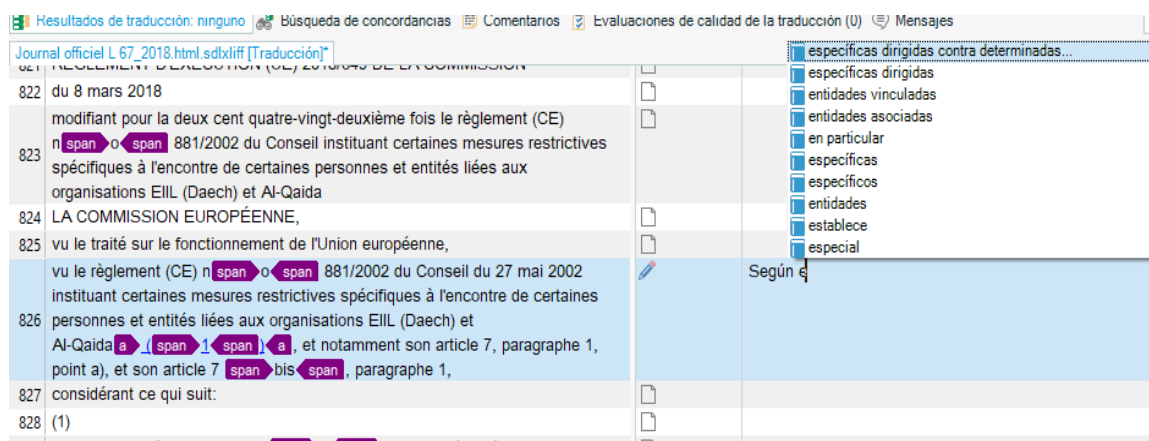


Figura 13: Recuperación de subsegmentos por parte del Diccionario de Autosugerencias en el segmento 826.

Sin embargo, el ejemplo del segmento anterior ilustra una de las mejores situaciones que pueden encontrarse en el documento. Para la mayoría de los segmentos del texto, no se encuentran resultados interesantes pues la lista de opciones no aparece siempre y, cuando lo hace, no ofrece las propuestas correctas. Por lo tanto, lo normal es que solo se recuperen propuestas de traducción para casos puntuales.

### 3.3.6 CASO 6: upLIFT Fragment Recall versus Autosugerencias

En este último caso, compararemos los resultados obtenidos en el caso precedente con los obtenidos a partir del mecanismo de recuperación *upLIFT Fragment Recall*. Como hemos mencionado anteriormente, *upLIFT* puede funcionar a modo de Diccionario de Autosugerencias. Por lo tanto, recupera fragmentos de segmentos ya traducidos y los muestra en forma de lista con solo teclear las primeras letras de la palabra que estemos traduciendo.

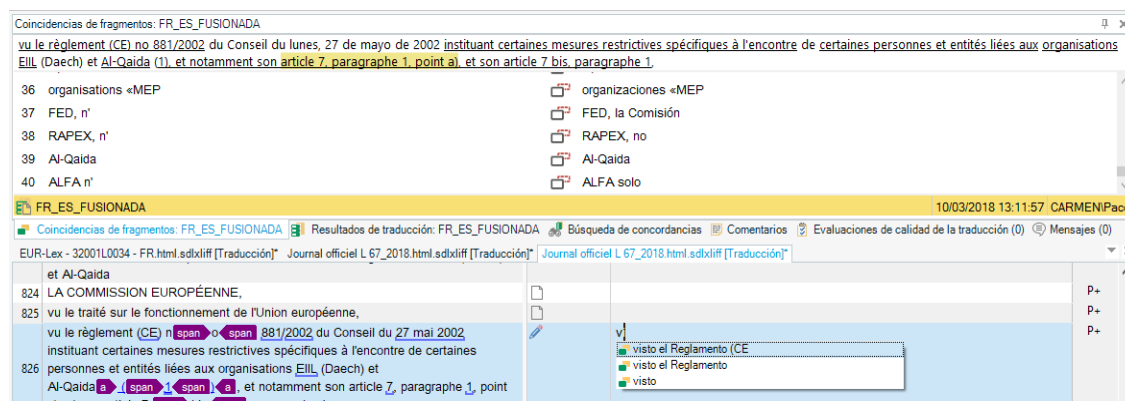
Este uso se activa de manera automática solo cuando se está llevando a cabo la búsqueda por *upLIFT Fragment Recall*, algo que pretende facilitar al traductor su tarea al ofrecerle una lista de posibilidades con solo escribir la o las primeras letras de una palabra. Para ello, tiene que estar activada la recuperación de subsegmentos (*upLIFT Fragment Match*), ya que, aunque esté llevándose a cabo la búsqueda de unidades de traducción completas (*upLIFT TU Match*), no se pone en marcha.

Lo que pretendemos con esta prueba es averiguar cuál de las dos herramientas ofrece una respuesta con mayor rapidez o más ajustada a la hora de escribir la traducción del segmento en el editor. La búsqueda se realiza con un umbral 2 tanto en recuperación de *upLIFT TU Match* como en *upLIFT Fragment Match*. Para este caso trabajaremos con el mismo segmento (826), que pertenece a la Directiva de 2018.

Cabe destacar que, cuando nos disponemos a traducir el segmento con el que estemos trabajando en el editor, el propio mecanismo de recuperación de subsegmentos (*upLIFT Fragment Match*) funcionando como *Diccionario de Autosugerencias* propone más opciones que las que aparecen en la ventana de coincidencias de fragmentos. Estas propuestas suelen ser palabras sueltas o combinaciones más pequeñas de las que se han marcado en el umbral mínima de coincidencia. No obstante, hemos observado que, en el caso de que el mecanismo *upLIFT TU Match* recupere unidades de traducción completas no se activa la lista de opciones a modo de diccionario si escribimos la primera palabra de la traducción en el editor.

Para demostrar lo que acabamos de explicar, vamos a comentar de manera detallada dos casos concretos que hemos escogido durante el proceso. Uno de ellos es el segmento 826, con el que trabajamos anteriormente y en el que se encuentra una combinación de recuperaciones a partir de ambos mecanismos (*upLIFT Fragment Recall* y *Diccionario de Autosugerencias*); en el otro, solo se han recuperado fragmentos de unidades de traducción. A pesar de que tanto el *Diccionario de Autosugerencias* y *upLIFT* pueden trabajar al mismo tiempo, no es habitual que aparezcan juntas las propuestas de ambos. Al menos, en este experimento, parece que el mecanismo de recuperación de subsegmentos .sustituye al diccionario en la mayoría de las ocasiones.

Empezaremos por el segmento 826<sup>22</sup>, para el que la herramienta recupera la unidad de traducción « son article 7, paragraphe 1, point a) » gracias a *upLIFT TU Match* y coloca estos resultados al comienzo de la lista en la ventana de coincidencias de fragmentos. En total, encontramos dos coincidencias de unidad de traducción completa y, luego, 28 resultados más de coincidencia de fragmento de unidad de traducción. Esto no quiere decir que el programa haya reconocido 30 partes diferentes de la oración, sino que ha encontrado varias propuestas para un mismo fragmento. Además, no sigue ningún orden, ni siquiera el de la propia frase, pues aparecen antes traducciones para “notamment son article 7” que para “Al-Qaida” que está situado justo delante. Al escribir nuestra traducción en el editor, se despliega la ventana con las propuestas a modo de diccionario. El orden de dichas propuestas difiere del que se presenta en la ventana de recuperación de fragmentos, por lo que es, en nuestra opinión, aleatorio. Esto se muestra en la Figura 14.



**Figura 14: Recuperación de subsegmentos de *upLIFT Fragment Recall* funcionando como Diccionario de Autosugerencias en el segmento 826.**

El segundo segmento elegido es el número 151<sup>23</sup> y presenta únicamente propuestas recuperadas a través del mecanismo *upLIFT Fragment Recall*. En la ventana de

<sup>22</sup> Texto del segmento: “vu le règlement (CE) no 881/2002 du Conseil du 27 mai 2002 instituant certaines mesures restrictives spécifiques à l'encontre de certaines personnes et entités liées aux organisations EIL (Daech) et Al-Qaida (1), et notamment son article 7, paragraphe 1, point a), et son article 7 bis, paragraphe 1,”.

<sup>23</sup> Texto del segmento: “le traitement dont auraient bénéficié les actionnaires et créanciers à l'égard desquels des mesures de résolution ont été exécutées, ou le système de garantie des dépôts concerné, si l'entité

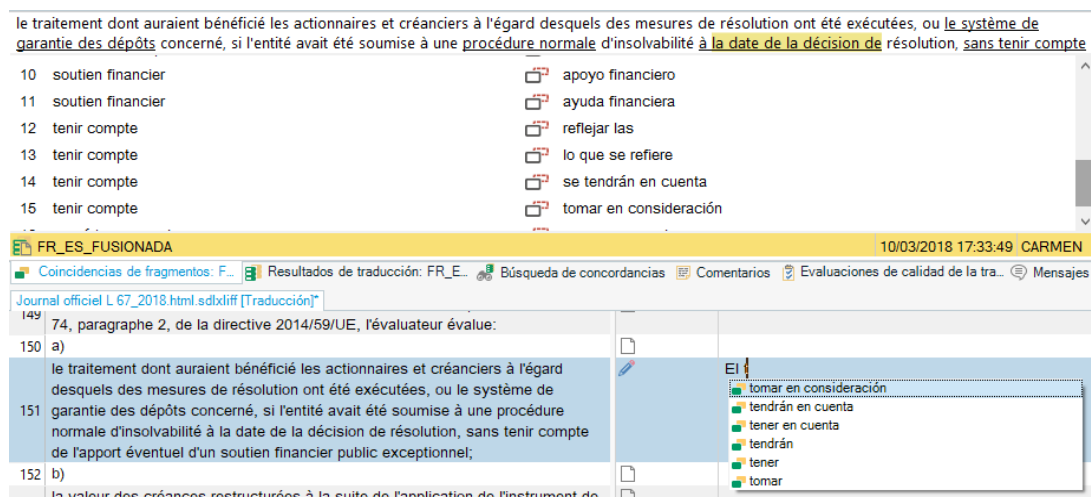


coincidencias de fragmentos, podemos observar una lista con un total de 23 resultados. No obstante, y como se ha mencionado anteriormente, la lista que aparece funcionando como Diccionario de Autosugerencias no sigue el mismo orden que el de la ventana de recuperación de fragmentos. Esto se debe a que, en este caso, las propuestas de la ventana de fragmentos no se han realizado para esta palabra en concreto, por lo que el programa trata de revisar toda la memoria para recuperar el máximo de resultados posibles. Esto demuestra que el hecho de que salgan ordenadas de una manera determinada para un fragmento de la oración, no implica que vayan a respetar el mismo orden para el resto. Por ejemplo, al escribir la primera “t” de la palabra “tratamiento” en el editor, aparece, en primer lugar, “tomar en consideración” seguido de “tendrán en cuenta”, mientras que figuran en el orden opuesto en la ventana de propuestas de traducción. Además, siguiendo con las opciones que nos ofrece este mecanismo como diccionario, aparecen las siguientes: “tener en cuenta”, “tendrán”, “tener” y “tomar”. Como se puede comprobar, no aparece en ningún momento la traducción correcta para esta palabra.

Como ejemplo de un caso en que sí coincida el orden de las propuestas, mencionamos “la date de decision”, que se encuentra en el mismo segmento 151. Cuando comenzamos a escribir “fecha”, nos aparecen, de nuevo, las opciones debajo de la frase. En esta ocasión, estas propuestas siguen el mismo orden que en la lista de arriba, con la única diferencia de que, aquí, hay más variabilidad en las palabras. Esto es, el programa es capaz de ofrecer también fragmentos aún más cortos hasta llegar a las palabras individuales. De esta manera, encontramos como primera opción (y más acertada) “la fecha de la decisión del”; como segunda, “la fecha de la decisión” en la que se ha eliminado “del”; “fecha de la” y, por último, “fecha”. Esto, además, nos permite afirmar que, en estas situaciones, lo que hace el programa es ir descartando palabras y ordenando las opciones en función de su pertinencia y de sus posibilidades de encajar con el contexto. Esto se recoge en la siguiente ilustración (Figura 15):

---

avait été soumise à une procédure normale d'insolvabilité à la date de la décision de résolution, sans tenir compte de l'apport éventuel d'un soutien financier public exceptionnel ;».



**Figura 15: Recuperación de subsegmentos de *upLIFT Fragment Recall* funcionando como Diccionario de Autosugerencias en el segmento 151.**

Para terminar con este caso, estudiaremos los resultados obtenidos en el texto procedente de ONUSIDA. En cuanto se activa la recuperación de fragmentos, aparecen tres para el primer fragmento cuyo texto es “ALL IN”, nombre de una plataforma que ha creado ONUSIDA para concienciar al público sobre la cantidad de jóvenes que padecen sida en el mundo. A pesar de que el título está en inglés, la memoria parece mostrar propuestas como “in”, “de” y “del día siguiente”. Como podemos comprobar, únicamente la primera sería válida, ya que, al ser el nombre de la campaña, no debe traducirse. En cualquier caso, las otras dos correspondencias no presentan ninguna relación evidente entre la palabra de origen y el resultado que se ha mostrado.

Ahora, probaremos con otros segmentos cuyo texto esté íntegramente en francés. Por ejemplo, el número 8<sup>24</sup>, para el que aparecen 21 resultados en la ventana de coincidencias de fragmentos. Solo uno de ellos se ha recuperado mediante el mecanismo *upLIFT TU Match*, mientras que los otros han sido encontrados por *upLIFT Fragment Match*. Por lo tanto, el resto de resultados estarán disponibles para utilizarse como diccionario. En este segmento concreto, es observable, que la mayoría de los resultados obtenidos no coinciden con la

<sup>24</sup> Texto del segmento: “Dans le monde, les deux tiers des adolescents récemment contaminés par le VIH sont des filles”.

necesidad del traductor a pesar de que se recuperan numerosas propuestas. Esto se muestra en la Figura 16.

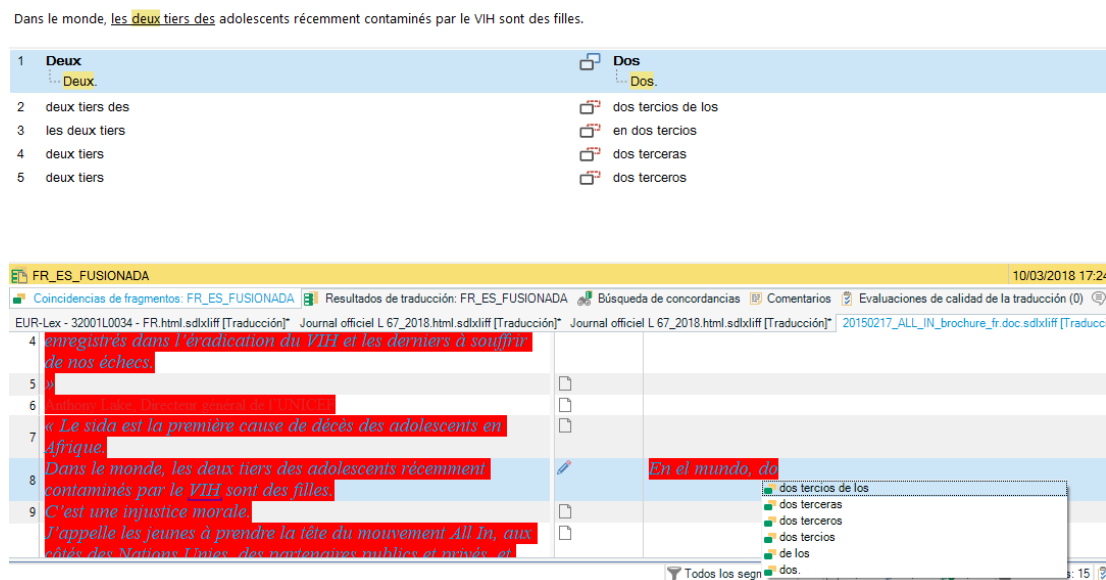


Figura 16: Recuperación de subsegmentos de *upLIFT Fragment Match* en el segmento 8.

Cabría destacar que, con respecto a “le VIH”, solo encontramos un resultado: “el STC” (traducido desde “le STC”). Es evidente que el “STC” no tiene relación alguna con el “VIH”, pero es posible entender la razón por la que lo ha recuperado. Es una secuencia similar, formada por un artículo y unas siglas. Este tipo de resultados se da a lo largo de todos los segmentos. Por lo tanto, podríamos decir que, en esta situación, el Diccionario de Autosugerencias sería más efectivo. No obstante, no siempre está tan claro cuál es el más eficiente. Para demostrarlo, trabajaremos con el segmento 191<sup>25</sup> del mismo documento.

En esta ocasión, la ventana de coincidencia de fragmentos, encontramos 47 resultados. Al teclear las primeras palabras de la traducción, se despliega una lista con lo recuperado a partir de la memoria. Para algunas secuencias, el diccionario ofrece respuestas totalmente equivocadas, mientras que el mecanismo *upLIFT TU Match* propone soluciones más acertadas. Para otras, es *upLIFT* el único que recupera algún resultado. No obstante,

<sup>25</sup> Texto del segmento: “Les données actuellement disponibles révèlent que l’intensification du dépistage et de la prise en charge chez les enfants et les adolescents vivant avec le VIH est encore beaucoup trop lente”.

también se encuentran palabras que no tienen ningún tipo de resultado o bien ninguno de los mostrados por ambos mecanismos es correcto, algo comprensible, pues ni el corpus ni el diccionario tratan el tema del texto de manera específica. Esto se muestra en la Figura 17.

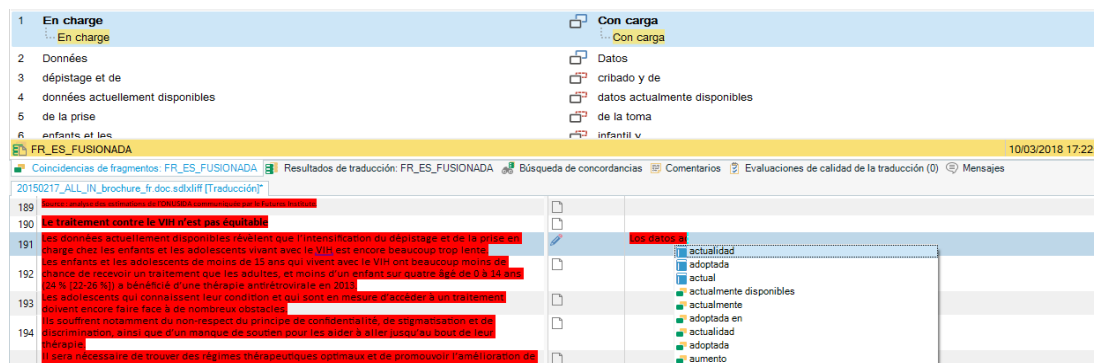


Figura 17: Recuperación de subsegmentos de Diccionario de Autosugerencias y *upLIFT Fragment Match* en el segmento 191.

En cualquier caso, resulta interesante comentar que, en todas las recuperaciones que se han ofrecido para este segmento en concreto, la memoria y el diccionario han coincidido tan solo una vez en sus traducciones. Este fenómeno es curioso, pues, aunque ambos mecanismos parten de la misma fuente de información (una memoria de traducción resultante de la fusión de dos corpus de la Comisión Europea), recuperan subsegmentos de manera diferente y, por tanto, algunos resultados no coinciden. Teniendo esto en cuenta, a veces, las propuestas de uno son más acertadas que las otras y viceversa. *UpLIFT Fragment Recall* parece tener más tendencia a recuperar parejas de palabras y el diccionario de autosugerencias propone, la mayoría de veces, palabras individuales. Esto no elimina, no obstante, la posibilidad de que *upLIFT* recupere tan solo una palabra y que el diccionario recupere subsegmentos más largos.

Según hemos podido comprobar, la diferencia real entre el Diccionario Autosuggest y *upLIFT* reside en que el diccionario es mucho más preciso al tener un número limitado de palabras a las que recurrir mientras que el mecanismo de recuperación propone, en numerosas ocasiones, resultados que no tienen nada que ver con la traducción, pues aunque siempre recupera palabras contiguas, el emparejamiento que ha realizado es incorrecto. Por lo tanto, las traducciones que ofrece no se corresponden con la palabra con la que se está trabajando en algunos casos. Por ello, normalmente, ofrece listas de palabras o combinaciones de las mismas demasiado largas en las que ni siquiera aparece la traducción

correcta del subsegmento. Esto viene a confirmar, al fin y al cabo, que el uso del diccionario reduce la cantidad de ruido a la hora de trabajar con el editor.

Quizá si hubiéramos empleado un diccionario confeccionado con vocabulario propio del campo de cada uno de los textos, los resultados hubieran sido totalmente diferentes. Así, probablemente, hubiera habido una gran diferencia entre la recuperación de subsegmentos en la memoria y el diccionario, pues este último sería mucho más preciso.

### 3.4 RESULTADOS DEL EXPERIMENTO

Teniendo en cuenta los resultados de la mayoría de las pruebas que se han llevado a cabo y previamente detallados en los apartados anteriores, hemos concluido que los mecanismos *upLIFT TU Match* y *upLIFT Fragment Match* pueden llegar a ser útiles únicamente si se mantiene un nivel mínimo de coincidencia mayor que 1. En el caso de tener una memoria de traducción de tamaño adecuado y de la misma temática que el texto con el que trabajemos, el mecanismo de recuperación de unidades de traducción completas (*upLIFT TU Match*) puede ser mucho más provechoso que el otro, pues, en muchas ocasiones, la equivalencia que propone es la correcta.

Sería una muy buena idea que los resultados de *upLIFT* se mostrasen por orden de aparición en la frase y, dentro de este requisito, por orden de pertinencia. Siempre aparecen mezclados y supone una pérdida de tiempo para el traductor, ya que tiene que ir leyendo todos y cada uno de las opciones en un orden totalmente aleatorio. En muchas ocasiones, resulta más sencillo teclear directamente la traducción en el editor. Es cierto que se producen listas de propuestas muy largas cuando el número de coincidencia mínima es muy bajo, pero, en cualquier caso, ayudaría a que el traductor ordenase las opciones en caso de que una expresión o término tenga más de una traducción posible.

Por lo que se ha demostrado en este experimento, el Diccionario de Autosugerencias del propio programa, permite obtener mejores resultados que los del mecanismo *upLIFT* usado de una manera similar. Esta conclusión se basa en el hecho de que el diccionario tiene las correspondencias hechas de una manera más directa, por lo que, en mucho menos tiempo,

SDL Trados Studio 2017 es capaz de recuperar y ofrecer unas propuestas más ajustadas en relación a la palabra que estemos traduciendo. Además, en todos los casos, dichas opciones comienzan por la letra que tecleamos y están ordenadas según su pertinencia. No obstante, cuando el mecanismo *upLIFT* funciona automáticamente de forma parecida al diccionario, los resultados que se obtienen son más numerosos, pero con una coincidencia menor. La herramienta recupera absolutamente todas las opciones que podrían tener una relación con el segmento a traducir, por mínima que esta sea, y la ofrece. Algo que parece no ocurrir con *Autosuggest*, pues solo muestra aquellas que tienen muchas posibilidades de coincidir. De hecho, en el mecanismo *upLIFT* las propuestas no siempre comienzan por la letra inicial de la palabra en cuestión y no siguen ningún criterio para aparecer ordenadas.

Asimismo, no hemos encontrado casos de *upLIFT Fuzzy Repair* en los segmentos con los que hemos trabajado. Por ello, creemos que no es un mecanismo que se ponga demasiado en acción, pues los otros dos, pertenecientes a *upLIFT Fragment Recall*, son los principales.

## 4 CONCLUSIONES

La recuperación de subsegmentos a partir de una memoria de traducción o de una base de datos terminológica puede ser muy interesante para el traductor si se emplea correctamente. Esto sería contar con una memoria de traducción del tema exacto del texto con el que se va a trabajar y de un tamaño lo suficientemente extenso como para que se produzcan recuperaciones de manera frecuente. De hecho, como hemos podido demostrar con este experimento, los textos con más resultados son aquellos que trataban una temática similar a la del corpus. Los otros dos presentaban recuperaciones numerosas, pero no eran pertinentes. Además, sería importante no bajar demasiado el umbral de coincidencias para obtener subsegmentos de calidad.

A partir de lo que hemos comprobado con este proyecto, el mecanismo *upLIFT TU Match*, que recupera unidades de traducción completas y las recupera para un nuevo contexto, sería, en un principio el más útil de cara a agilizar el trabajo del traductor, pues los resultados suelen ser precisos. También, la idea de que *upLIFT Fragment Match* funcione como un Diccionario de Autosugerencias puede ser interesante, pero, creemos que aún la

alineación de subsegmentos que la herramienta lleva a cabo no es pertinente en numerosas ocasiones, lo que hace que el propio Diccionario de Autosugerencias siga siendo la mejor opción en cuanto a rapidez se refiere.

En relación con *upLIFT Fuzzy Repair*, no hemos podido comprobar hasta qué punto es efectivo, dado que no hemos obtenido ningún resultado relevante para el presente trabajo. No obstante, creemos que también necesitaría una alineación más ajustada y que, en realidad, no sería un avance tan grande para el traductor, ya que, en ocasiones, tendría que modificar la oración que se le propone. Teniendo esto en cuenta, las conclusiones para los tres mecanismos coinciden en que, a veces, es más rápido que el traductor introduzca manualmente su traducción antes que mirar cuál de entre todas las propuestas es la adecuada, pues no aparecen ordenadas por pertinencia. Esto no plantea ningún problema cuando se trata de listas relativamente cortas, con cuatro o cinco opciones como máximo, ya que se ven todas en la pantalla de fragmentos. No obstante, cuando este número asciende a diez, veinte o, incluso, treinta como se ha dado el caso a lo largo del experimento, resulta bastante incómodo tener que leer absolutamente todas las propuestas y encontrar las respuestas. Además, el traductor se ve obligado a teclear la traducción, pues, una vez seleccionada, no se incluye automáticamente en la columna de la derecha del editor. Por estas razones, seguimos pensando que estos mecanismos deben mejorarse para conseguir no solo que la memoria se aproveche al máximo posible, sino que también sea rápido para el traductor y le ahorre tiempo y esfuerzo.

En nuestra opinión, consideramos que el tema tratado en este trabajo sienta las bases de una posible futura investigación, ya que se trata de un asunto de actualidad y gran extensión. En tanto que traductores, debemos estar informados de cada avance que se produzca en herramientas de traducción como SDL Trados Studio y en sus futuras versiones. De esta manera, estar pendiente de esta evolución en el campo de las tecnologías de la traducción es crucial con el fin de sacar el máximo partido al apoyo que pueden suponer estas herramientas para el traductor en su labor y, por lo tanto, obtener el mejor resultado posible.

## 5 BIBLIOGRAFÍA

- Corpas Pastor, Gloria. 2012. “Corpus, tecnología y traducción”. Universidad de Málaga.
- Flanagan, Kevin. 2014. “Filling in the gaps: what we need from TM subsegment recall”. *Conference Translating and the Computer* 36: 12–23. <http://www.kftrans.co.uk/lift/FillingInTheGaps.pdf>
- Flanagan, Kevin. 2015. “Subsegment recall in Translation Memory — perceptions, expectations and reality”. *The Journal of Specialised Translation* 23: 64–88.
- Forcada, Mikel L. 2014. “On the annotation of TMX translation memories for advanced leveraging in computer-aided translation”. *Ninth International Conference on Language Resources and Evaluation* , 4374–4378. 1 de junio de 2018. [http://www.lrec-conf.org/proceedings/lrec2014/pdf/373\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2014/pdf/373_Paper.pdf)
- Forcada, Mikel L. 2017. “Making sense of neural machine translation”. *Translation Spaces* 6:2 (2017): 291–309,
- Goldsmith, Emma. 2015. “An introduction to translation memory”, *SDL Trados Studio 2015*. 1 de junio de 2018. <https://www.sdltrados.com/download/an-introduction-to-translation-memory/10190/2/>
- Gow, Francie. 2003. *Metrics for Evaluating Translation Memory Software*. Tesis doctoral. Universidad de Ottawa.
- Hearne, Mary y Andy Way. 2011. “Statistical Machine Translation: A Guide for Linguists and Translators”. *Linguistics and Language Compass* 5: 205–226. 7 de junio de 2018. <https://doi.org/10.1111/j.1749-818X.2011.00274.x>
- *How to use upLIFT Fragment Recall in SDL Trados Studio 2017*. 9 de marzo de 2018. [https://www.youtube.com/watch?time\\_continue=93&v=d6Pic4t--mM](https://www.youtube.com/watch?time_continue=93&v=d6Pic4t--mM)
- Koehn, Philipp. 2010. “Words, Sentences, Corpora”. En *Statistical Machine Translation*. Cambridge: Cambridge University Press, 33-62.
- Krollmann, Friedrich. 1971. “Linguistic data banks and the technical translator”. *Meta: Translators Journal* 16: 117–124



- Otero Martínez, Lucía. 2015. “La recuperación de subsegmentos en Déjà Vu x3”. Trabajo fin de grado. Universidad de Salamanca.
- Reinke, Uwe. 2013. “State of the Art in Translation Memory Technology”. *Translation: Computation, Corpora, Cognition* 3: 27–48. Retrieved from <http://t-c3.org/index.php/t-c3/article/view/25>
- Rodríguez Vázquez de Aldana, Emilio. 2016. “La recuperación de traducciones en los sistemas de memoria de traducción”. Asignatura Recursos Tecnológicos para la Traducción. Universidad de Salamanca.
- Rodríguez Vázquez de Aldana, Emilio. 2016. “Introducción a la traducción automática. Asignatura Recursos Tecnológicos para la Traducción. Universidad de Salamanca.
- *SDL Trados Studio 2015*. 9 de junio de 2018. [http://producthelp.sdl.com/SDL\\_Trados\\_Studio\\_2015/client\\_es/About/TM/EVWorkingwithTMsAbout\\_Translation\\_Memory\\_Matches.htm](http://producthelp.sdl.com/SDL_Trados_Studio_2015/client_es/About/TM/EVWorkingwithTMsAbout_Translation_Memory_Matches.htm)
- SDL Trados Studio 2017. “upLIFT Translation Memory Technology”. 2 de junio de 2018. <https://www.sdltrados.com/download/uplift-translation-memory-technology-beginners-guide/119151/>
- *UpLIFT: Next-generation Translation Memory Technology*. 2 de junio de 2018. <https://www.sdl.com/software-and-services/translation-software/sdl-trados-studio/uplift-translation-memory/>
- Webb, Lynn E. 1992. *Advantages and Disadvantages of Translation Memory: A Cost/Benefit Analysis*. Tesis doctoral. San Francisco State University.