

**Aplicación de recursos TAO de software libre
para la traducción de textos especializados
en el campo de la Demografía Histórica
(par lingüístico: del español al inglés).
El caso de OmegaT.**

Mary Patricia Zacher

Tutor: BARTOLOMÉ MESA LAO

TRABAJO FIN DE MÁSTER

Traducción especializada



20 de enero 2020

Índice

RESUMEN.....	3
ABSTRACT.....	4
1. INTRODUCCIÓN.....	5
1.1. Motivación.....	5
1.2. Objetivos.....	6
1.3. Aplicaciones.....	7
2. METODOLOGÍA.....	7
2.1. ¿Por qué OmegaT?.....	7
2.2. Creación de recursos.....	8
2.2.1. Compilación de corpus paralelos bilingües.....	9
2.2.2. Conversión de formatos y pos-edición.....	13
2.2.3. Alineación de bitextos.....	13
2.2.4. Extracción de recursos.....	13
2.2.5. Depuración de glosarios.....	14
2.3. Test de funcionamiento.....	15
2.3.1. Memorias de traducción.....	15
2.3.2. Bases de datos terminológicos.....	16
3. RESULTADOS.....	17
3.1. Memorias de traducción.....	17
3.1.1. Resultados cualitativos.....	17
3.1.2. Resultados cuantitativos.....	18
3.2. Bases de datos terminológicos.....	19
3.2.1. Resultados cualitativos.....	19
3.2.2. Resultados cuantitativos.....	19
4. CONCLUSIONES.....	21
4.1. Memorias de traducción.....	21
4.2. Bases de datos terminológicos.....	22
4.3. Líneas de investigación futura.....	22
5. BIBLIOGRAFÍA.....	24
6. ANEXOS.....	26
1. Textos en INGLÉS obtenidos en Encyclopedia.com.....	26
2. Textos en ESPAÑOL-INGLÉS (resúmenes y abstracts) obtenidos de los artículos de Revista de Demografía Histórica.....	26

Tablas y figuras

Tabla 1. Textos de artículos de enciclopedia	11
Tabla 2. Textos de resúmenes de artículos de revista	12
Tabla 3. Textos para proyectos de traducción	16
Tabla 4. Relaciones de términos especializados	16
Tabla 5. Descripción cuantitativa de las MT	18
Tabla 6. Estadísticas de coincidencias con MT en los textos a traducir	18
Tabla 7. Descripción cuantitativa de las BDT	19
Tabla 8. Estadísticas de coincidencias de términos especializados (en %).	19
Figura 1. Coincidencias totales de los glosarios	20

RESUMEN

La finalidad de este trabajo es evaluar la aplicación de una herramienta de Traducción Asistida por Ordenador (TAO), en concreto OmegaT, y su utilidad en la traducción del español al inglés de textos especializados en el ámbito temático de la Demografía Histórica. Las ventajas de esta herramienta son su gratuidad, versatilidad y fácil manejo. Cuenta con los cuatro componentes básicos: memorias de traducción, gestores terminológicos, herramientas de alineación y editor.

Se ha observado que en el campo de la Demografía Histórica no hay disponibles memorias de traducción (MT) ni Bases de datos terminológicos (BDT), dos componentes necesarios que se insertarían en OmegaT para asegurar su eficacia. Debido a ello, se busca solventar este déficit en dos fases distintas. En la primera, se diseñan dos estrategias para crear ambos recursos por medio del alineación de textos. En la segunda, cada uno de esos recursos es analizado en su funcionamiento (eficacia y eficiencia) en casos reales de traducción especializada.

Para la elaboración del corpus paralelo bilingüe, en primer lugar, se recopilan textos en inglés de Demografía Histórica de la web de una enciclopedia en línea de conocida solvencia científica: Encyclopedia.com. Los 8 textos seleccionados son traducidos al español usando una herramienta de Traducción Automática (TA): Google Translate. En segundo lugar, se toman los resúmenes en español y sus correspondientes *abstracts* traducidos al inglés de 16 artículos. Estos artículos fueron escritos por especialistas y publicados recientemente en la *Revista de Demografía Histórica*, una publicación conocida por su prestigio en ese campo.

En la segunda fase, las MT y las BDT obtenidas mediante cada estrategia son sometidas a prueba utilizando OmegaT y varios textos especializados que no han sido traducidos hasta ahora. De este modo, se miden tanto la cantidad de las sugerencias automáticas como el uso correcto de términos técnicos. El principal indicador del análisis es la proporción del número de términos y expresiones especializadas que el texto fuente contiene y cuántos de ellos son sugeridos por la herramienta TAO tanto parcial como completamente.

De los resultados obtenidos se puede deducir: la eficacia de ambas estrategias para la creación de MT y BDT; la ineficiencia de las MT creadas, dado que no se obtienen ventajas significativas en los dos casos; la utilidad de las BDT, que prueban ser mucho más útiles en ambos casos, ya que incrementan la productividad y aportan coherencia a las traducciones proporcionando con precisión los equivalentes lingüísticos de los términos especializados.

En conclusión, el trabajo avala la idoneidad de la implantación de OmegaT en el campo señalado. Los resultados son muy esperanzadores, sobre todo en cuanto a la terminología, y mejorarán considerablemente cada vez que se vayan añadiendo a esa herramienta TAO las MT y las BDT originadas por las traducciones que se vayan realizando.

Palabras clave: OmegaT, Demografía Histórica, alineación, corpus paralelo bilingüe, traducción asistida por ordenador, traducción automática, Google Translate.

ABSTRACT

The objective of this work is to evaluate the computer-assisted translation tool (CAT tool), OmegaT, and its usefulness in translating specialized texts in the field of historical demography from Spanish to English. The advantages of this tool are that it is free, versatile, and easy to manage. It comes with four basic components: translation memory, terminology management tools, alignment tools and an editor.

It has been observed that in the field of historical demography, there is a general lack of available translation memories (TM) or terminology databases (TD), two necessary components which would be inserted into the OmegaT program to ensure its efficacy. Due to this lack, I have set out to offset this deficit in two separate phases. In the first phase, two strategies were designed to create both resources from the beginning through the use of bilingual alignment texts. In the second, each of the obtained resources was analyzed for its performance (efficacy and efficiency) in a real translation situation.

To create the necessary bilingual parallel corpus, first, texts in English were compiled from the web in the field of historical demography from a trustworthy online scientific encyclopedia: Encyclopedia.com. The 8 selected texts were then translated to Spanish using a free and open machine translation tool (MT): Google Translate. Second, the Spanish summary and their corresponding abstracts translated to English were taken from 16 articles. The articles were written by specialists and were published in recent years in the *Revista de Demografía Histórica*, a magazine known for its excellence in the field.

In the second phase, the TM and the TD obtained through each strategy were put to the test using OmegaT and a specialized texts which had not been previously translated. In this way, the quantity of automatic suggestions could be measured as well as their appropriate use as technical terms. The main indicator of the analysis was the proportion of the number of specialized terms and expressions that the source text contained, and how many of them are suggested by the CAT, either partially or totally.

From the results obtained, the following can be deduced: the efficacy of both strategies in the creation of the TM and TD; the inefficiency of the TM created, as it obtained no significant advantage in either case; the usefulness of the TD, as they proved to be much more useful in both cases. The TD increased the productivity and brought about a coherency in the given translations through precise and unambiguous linguistic equivalents of specialized terms.

In conclusion, this work supports the use of OmegaT in the said field. The results are very encouraging, especially in regard to the terminology, and will be significantly improved with more translation memories and terminology databases each time a finished translation is added to this CAT tool.

Key words: OmegaT, Historical Demography, Alignment, Bilingual Parallel Corpus, Computer-assisted Translation, Machine Translation, Google Translate.

1. INTRODUCCIÓN

1.1. Motivación

La autora de este Trabajo Fin de Máster (TFM) desea dedicarse profesionalmente a la traducción de textos especializados de Demografía Histórica desde el español al inglés, su lengua materna. Atendería así la creciente demanda de especialistas en ese campo que desean publicar sus estudios más allá del espacio hispanohablante y que deben recurrir a la lengua inglesa, que se ha convertido en un medio para la difusión científica.

La Demografía Histórica es una disciplina que estudia la población humana en el pasado. Cuenta con un lenguaje especializado muy preciso, con terminología y fraseología propias, que proviene tanto de la Demografía como de la Historia. Hoy en día, para la divulgación internacional de la producción científica producida en español ha de recurrirse, como en casi todos los campos, a su traducción a la lengua inglesa. Para ello, ha de tenerse en cuenta que en los lenguajes de especialidad las normas retóricas y convenciones textuales tienen un impacto directo en el proceso de traducción a una segunda lengua (Bermúdez Bausela, 2014).

Por otra parte, hoy en día en el desarrollo profesional de la traducción especializada es prácticamente ineludible el recurso a las herramientas de Traducción Asistida por Ordenador (TAO). De hecho, es casi imposible encontrar un traductor que no haga uso de ellas. Según una encuesta realizada en 2013 entre tres millares de traductores profesionales de todo el mundo, casi nueve de cada diez de ellos manifestaron servirse en su labor de alguna TAO. El campo de especialización mayoritario era el de la Tecnología (34%). A considerable distancia le seguían el de los Negocios-Finanzas, el Marketing, la Medicina-Farmacia y el Derecho, con porcentajes de entre el 15% y 10%. Llamaba la atención, asimismo, su escaso uso en el campo humanístico y más concretamente en el de las Ciencias Sociales. A este área se dedicaban menos de un 5% de los traductores que hacían uso de esas herramientas (Tabor, 2013). La misma encuesta se ha realizado en 2019, aunque los resultados no están aún disponibles. No es aventurado intuir que no variarán sustancialmente salvo para confirmar el incremento del uso de las TAO.¹

Son varios los trabajos que han destacado los incrementos de productividad que el uso de la TAO proporciona al traductor profesional. Vallianatou (2005)², por ejemplo, aseguraba que la utilización de cualquiera de las tres TAO por ella analizadas aumentó considerablemente el promedio estándar de palabras/hora. Reconocía que la productividad real dependía del tipo de trabajo y de las MT y BDT utilizadas, y que, cuanto más específicas estas son, mejores resultados se obtienen. Encuestas realizadas entre otros usuarios (profesionales y estudiantes) han destacado las ventajas adicionales que este recurso aporta a la labor traductora: mejora de

1 Efectivamente, desde enero de 2019 se viene realizando una nueva encuesta sobre el uso de las TAO (<https://prozcomblog.com/2019/01>).

2 A pesar de que la investigación de esta traductora profesional tiene ya casi tres lustros, creemos que los resultados son ilustrativos. Durante 18 meses, comparó los rendimientos de TAO comerciales (DejaVu, Trados y una anónima proporcionada por el cliente) y la traducción sin herramientas para el par de lenguas inglés-griego.

la coherencia, eliminación del trabajo repetitivo, reaprovechamiento de recursos e incluso su función pedagógica (Paulsen y Schjoldager, 2011; LeBlanc, 2013).³

Por tanto, esta investigación pretende explorar las posibilidades de dotarse en esta labor de un recurso TAO para incrementar tanto la productividad como la calidad las traducciones porque, aunque la del traductor es “una profesión donde se trabaja a contrarreloj”, el incremento de la productividad no puede ir en detrimento de la calidad final. Así pues, esta investigación se enmarca y apoya en los estudios previos realizados por distintos autores tanto en el campo de las herramientas TAO como en el de la obtención de los recursos que estas necesitan (Corpas, 2001; Castillo Rodríguez, 2009; Filatova, 2010; Ramírez Polo, 2011; Caetano, 2016; Vigier 2016; Rankine, 2017).

Como herramienta TAO se ha elegido el programa OmegaT. Se trata de un programa gratuito con el que nos hemos familiarizado a lo largo del máster y queremos evaluar los rendimientos que esta herramienta podría reportar en la labor profesional. Como en otros recursos similares, sus componentes esenciales son las Memorias de Traducción (MT) y las Bases de Datos Terminológicos (BDT) (Oliver, 2013).

Tal y como se ha podido comprobar a lo largo de este trabajo, en el campo de la Demografía Histórica apenas existen ese tipo de recursos ya compilados y disponibles.⁴ Es obligada, por tanto, su elaboración “ex novo” valiéndose de otro de los componentes que OmegaT incorpora: la alineación de textos (originales y traducciones).

Las investigaciones sobre la obtención de este tipo de recursos se han realizado en distintas áreas de conocimiento. En general, el uso de MT y BDT ha mostrado su efectividad en aquellas que se sirven de un lenguaje formal, estereotipado y, en consecuencia, repetitivo. Nos referimos a áreas como las de Derecho, Economía-Finanzas o Medicina-Farmacia (Vallianatou, 2005). De hecho, como se comprueba en la citada encuesta de Tabor (2013), esos son los ámbitos donde más se recurre a herramientas TAO. En áreas de conocimiento más creativas y sujetas a mucha más variabilidad del lenguaje –la creación literaria, comercial, periodística, por ejemplo– o donde se da mayor cantidad de textos “pesados” (tablas, gráficos, etc.) la misma encuesta y varias investigaciones recientes (Rankine, 2017:13-14) han revelado unos rendimientos mucho menores e incluso insignificantes. En este sentido es significativo que casi una décima parte de los encuestados por Tabor en 2013 mostraban sus dudas respecto a la eficiencia de las TAO en la labor traductora.

1.2. Objetivos

Así pues, esta investigación tiene un enfoque descriptivo en cuanto a los procedimientos y un enfoque empírico en cuanto al análisis de los resultados. Sus objetivos principales se establecen a dos niveles:

1. Diseño y evaluación de procedimientos para la elaboración de MT y BDT que alimenten el funcionamiento de OmegaT. Se valorarán aquí la disponibilidad de textos paralelos

3 Es cierto que estos autores también han puesto sobre la mesa las desventajas señaladas por sus usuarios: cambio de la relación traductor-texto, obstáculo para la creatividad, incremento del acriticismo, propagación de errores, etc.

4 Se han consultado las páginas institucionales de la “International Commission for Historical Demography” en búsqueda de BD (<https://www.ru.nl/ichd/>) o DocuTradSo: Fuentes de información para la actividad traductora (<http://www3.uva.es/docutradso/#1.10>)

bilingües, los recursos técnicos y el tiempo necesario para la obtención de esos recursos, así como los rendimientos obtenidos.

2. Evaluación del rendimiento de la propia herramienta OmegaT en la traducción de textos especializados de Demografía Histórica. Esta evaluación se realizará considerando dos aspectos:

a. Productividad. Incremento en la rapidez de traducción mediante la sugerencia de equivalentes pertinentes para expresiones o términos especializados. De producirse, ayudaría en la reducción de tareas repetitivas, en una menor carga de trabajo, en el acortamiento de los plazos de entrega y en la posibilidad de ofertar precios más competitivos.

b. Coherencia. Grado de coherencia de que dota al texto al proporcionar equivalencias exactas, precisas y unívocas de los términos especializados. Ello repercutiría en una mejora de la corrección lingüística, es decir, en una mayor calidad de la traducción realizada y en un enriquecimiento progresivo de las MT y BDT.

En ambos casos, se realizará una evaluación tanto desde el punto de vista de los recursos necesarios para crearlos (eficiencia) como de los resultados obtenidos en la práctica (eficacia). Para ello se recurrirá a datos cualitativos y cuantitativos.

1.3. Aplicaciones

Como queda dicho, la principal aplicación sería la de proporcionar un conjunto de herramientas de trabajo básica y en continuo perfeccionamiento que permita una mayor competitividad en el campo de la traducción profesional. En concreto, se trata de valorar la idoneidad tanto de la herramienta OmegaT como de los procedimientos para elaborar sus principales componentes: las MT y BDT. Asimismo, explorar si esa metodología de trabajo podría extenderse fácilmente a otros campos temáticos de la traducción especializada.

2. METODOLOGÍA

Esta investigación se plantea en dos fases. En la primera, se aplican dos estrategias para crear los recursos esenciales que necesita OmegaT para funcionar como herramienta TAO. En la segunda, esos recursos obtenidos se someten a una prueba empírica con textos especializados. Los resultados serán analizados en el siguiente capítulo. Previamente a todo ello, es necesario justificar la elección de OmegaT como herramienta.

2.1. ¿Por qué OmegaT?

Hoy en día el mercado ofrece muchos productos TAO. La mayoría de ellos son de pago y requieren de aprendizaje previo. Para un traductor que se inicia en la profesión, estos costes económicos y de tiempo son factores que deben considerarse desde el punto de vista de las ventajas que a cambio se obtienen con el uso de la herramienta.

Paralelamente a la mejora en la oferta y extensión de su uso, en los últimos años se han hecho diversos estudios comparativos entre diversas herramientas TAO. Según la mencionada encuesta de Tabor (2013), la herramienta más usada por los traductores profesionales era SDL Trados Studio, un programa de pago. Sin embargo, en los últimos años varias investigaciones vienen mostrando que OmegaT va ganando en consideración como alternativa a las opciones comerciales (Filatova, 2010; Anica, 2014; Caetano, 2016; Rankine, 2017). De entre las ventajas señaladas por estos autores destacan las siguientes:

- *Gratuidad* tanto en su adquisición inicial como las futuras actualizaciones.
- *Versatilidad*. Es compatible con Windows, Linux y Macintosh, y admite una gran variedad de formatos.
- *Suficiencia*. Cuenta con los componentes y las funciones esenciales: memoria de traducción, gestor de terminología, alineación de textos, herramientas de localización, propagación de textos, etc.
- *Facilidad de instalación y uso*. Tiene una interfaz bien diseñada, clara y minimalista, y un manual de instrucciones asequible. La creación y gestión de proyectos es muy sencilla.
- *Familiaridad*. Ha sido la TAO que se ha utilizado durante el máster.

En cuanto a su eficacia y eficiencia, OmegaT ha sido comparada con diversas TAO. La investigación de Filatova (2010) señalaba algunas deficiencias respecto a Trados y Wordfast. Algunas de ellas han podido ser subsanadas por versiones posteriores del programa. La de Anica, en cambio, tras someter OmegaT a un caso práctico, sostiene que es una herramienta “excelente y que merece la pena introducir en nuestra estación de trabajo” (Anica, 2014:55). De parecida opinión es Caetano (2016), quien considera que OmegaT es la TAO gratuita más versátil y completa de las que evalúa.

Para la realización de este trabajo con OmegaT nos hemos servido de la versión 4.3.1. aparecida hace unos años. Desde hace poco se encuentra disponible la versión 5.1.0 que incluye nuevas aplicaciones. No obstante, dado que carece aún de manual específico y puede plantear algunos problemas de funcionamiento, hemos preferido servirnos de la versión anterior por encontrarse suficientemente testada.⁵

2.2. Creación de recursos

Para una correcta gestión de un proyecto de traducción, OmegaT debe ser alimentado con una serie de recursos. Aquí nos centraremos en los que consideramos los principales: MT y BDT (glosarios). Es verdad que la propia herramienta es capaz de crearlos automáticamente conforme se van realizando traducciones al modo tradicional. Pero este proceder no es viable en una investigación de las dimensiones temporales de este TFM. En este caso se necesita la asignación de esos recursos previamente elaborados.

Hemos indagado en Internet sobre la existencia previa de estos recursos en el campo concreto de nuestro interés: la Demografía Histórica. Aunque, se han consultado las web institucionales y profesionales de entidades dedicadas a la investigación o docencia de esa disciplina⁶, apenas

5 Consultado en <https://sourceforge.net/p/omegat/news/2019/11/omegat-latest-version-510-released/>

6 En la búsqueda de estos recursos, entre el 15 y el 20 de noviembre de 2019 se han consultado las páginas institucionales de la “International Commission for Historical Demography” (<https://www.ru.nl/ichd/>), DocuTradSo: Fuentes de información para la actividad traductora (<http://www3.uva.es/docutradso/#1.10>),

se han obtenido resultados. Únicamente se ha podido conseguir un glosario para el término genérico “Humanidades” en el recurso terminológico *Interactive Terminology for Europe (IATE)*.

Debido a que no podemos disponer de MT y BDT previamente elaboradas, tenemos que crear un número suficiente de ellas para someter a prueba el funcionamiento de OmegaT. Para ello nos vamos a servir de una de las aplicaciones más útiles incorporadas a la herramienta: la alineación de textos. Evidentemente, por razones de tiempo, el número de recursos creados no puede ser tan elevado como para ser plenamente representativo a nivel estadístico, pero sí suficiente para que obtengamos estimaciones que puedan orientarnos sobre la conveniencia o no de implantar esta TAO en el campo concreto de la Demografía Histórica.

2.2.1. Compilación de corpus paralelos bilingües

Para proceder a la extracción de recursos mediante la alineación de textos necesitamos disponer de textos de Demografía Histórica tanto en su versión en inglés como en su versión en español.

Sin embargo, esos textos deben ofrecer ciertas garantías. Por un lado, la terminología utilizada en la lengua de origen debe ajustarse al lenguaje especializado propio del área. Por otro, la traducción de esos términos debe remitirse a los términos equivalentes también especializados utilizados en la lengua de llegada. En este sentido, no sirven tan solo las traducciones lingüísticamente correctas. La terminología especializada de la lengua de partida debe corresponderse unívocamente con la terminología especializada de la lengua de llegada.

Mostramos aquí un ejemplo. En Demografía Histórica, la palabra española *hogar* denomina a una unidad de residencia que agrupa a una o más personas que hacen provisión común de alimentos. Se distingue conceptualmente de *familia*, que agrupa sólo a personas relacionadas por sangre, adopción o matrimonio. Pues bien, el equivalente inglés del término especializado *hogar* es *household* y no otros como *family* o *home* que podrían ser traducciones correctas en un entorno no especializado.

a. Criterios

Procedemos a la localización de textos que integren un “corpus paralelo bilingüe especializado” destinado a la elaboración de una base de datos para una herramienta TAO. Asumiendo la definición de Sinclair (citado por Corpas, 2001:157), entendemos por “corpus” una colección de piezas de lenguaje seleccionadas y ordenadas de acuerdo con criterios lingüísticos explícitos para ser usados como ejemplos del lenguaje.

Así pues, establecemos una serie de criterios que garanticen el nivel de lengua deseado. Para ello, tomamos en consideración todas las observaciones y recomendaciones que realizan diversos autores que se han ocupado de esta cuestión (Corpas, 2001; Castillo Rodríguez, 2009; Bermúdez Bausela, 2014).

A) *Propósito*. Obtención de bases de datos de terminología especializada.

Base de Datos Terminológica Multilingüística de Naciones Unidas (UNTERM) <https://www.un.org/es/hq/dgacm/unterm.shtml>, Interactive Terminology for Europe (IATE) <https://iate.europa.eu/home> y Asociación de Demografía Histórica (ADEH) <http://www.adeh.org/?q=es/contenido/asociacion-de-demografia-historica>.

- B) *Cantidad*. Ajustado a las posibilidades del proyecto (unas 40.000 palabras) y considerando su representatividad y equilibrio (límites temporales y geográficos, niveles de lengua, etc.)
- C) *Medio*. Textos disponibles en línea (Internet).
- D) *Temática*. Demografía Histórica.
- E) *Nivel de especialización*. Comunicación especialista-especialista o especialista-público especializado.
- F) *Tipología textual*. Publicaciones electrónicas (formatos HTML y PDF).
- G) *Autoría*. Especialistas.
- H) *Fecha de publicación*. Posteriores a 2010.
- I) *Lenguas seleccionadas*. Inglés y español.

b. Obtención de textos especializados

Siguiendo estos criterios, la búsqueda se ha realizado en Internet por medio del buscador Google. Este es considerado el mejor motor de búsqueda en cuanto a la calidad de recuperación mediante palabras clave. En este caso, se han utilizado expresiones del tipo “Demografía Histórica”, “Historia de la población”, “familia”, “evolución demográfica”, etc.

Se han seguido dos procedimientos o estrategias:

1. Artículos de enciclopedia (8)

Se han consultado artículos de Demografía Histórica publicados en la web de Encyclopedia.com. Consideramos que esta fuente es capaz de proporcionar con rapidez información relevante y fiable. Se trata de una enciclopedia en línea en inglés iniciada en 1998 que integra información procedente de diccionarios, enciclopedias y trabajos de referencia como Oxford University Press, Columbia Encyclopedia o Gale.⁷ El acceso a artículos concretos se ha hecho a través de palabras clave y de ellos se han seleccionado 8 en formato HTML con una extensión total de casi 32.000 palabras (ver anexo 1).

Encyclopedia.com no cuenta con las versiones en español de estos artículos. Para obtenerlas hemos recurrido a la herramienta de Traducción Automática (TA) más asequible y fiable, y que además está incorporada al motor de búsqueda: Google Translate.

Google Translate se fundamenta en un sistema basado en corpus que aúna la traducción automática estadística y la traducción neuronal. En otras palabras, un sistema que aprende automáticamente a través de patrones y ejemplos encontrados en los millones de documentos traducidos que se encuentran en Internet. Para ello usa un poderoso software y la lengua inglesa como lengua intermedia, dado que es en ella en la que pueden encontrarse la mayor cantidad de textos traducidos. Basándose en esos patrones, Google Translate no trata las palabras de manera individualizada sino que se fija en las frases completas y tiene en cuenta el contexto. Su tecnología basada en el llamado “aprendizaje profundo” (*deep learning*) permite que la herramienta TA dé

⁷ “Encyclopedia.com” posee más de 200.000 entradas y 50.000 sumarios temáticos. Datos obtenidos en Wikipedia (consultada el 22/11/2019)

sentido a la frase y no traduzca de forma aislada o a pedazos (Norvig, 2011; Francés, 2018; Padilla, 2018).

Los resultados parecen ser satisfactorios. Los autores que han comparado Google Translate con otras herramientas TA han destacado la calidad de sus traducciones. Jordan lo considera el mejor sistema para la traducción del inglés al español de textos de Ciencias Sociales (Jordan *et alter*, 2017: 248). Más recientemente aún, tras comparar otras TA, Viver coincide al afirmar que Google Translate “es el motor de traducción automática más preciso y el que ofrece mejores traducciones” (Viver, 2018: 32).

No obstante, somos conscientes de que la calidad de las traducciones generadas por esta TA contienen problemas graves debidos a la herramienta en sí. Por ello nos servimos de ella tan sólo para generar los recursos necesarios a los fines señalados.

Así, mediante Google Translate se ha procedido a la traducción al español de los ocho artículos de Encyclopedia.com que originalmente estaban en inglés. Hemos observado que, como es habitual (Anica, 2014), los textos en español resultantes son más extensos que sus versiones en inglés (casi un 16% más de palabras).

Texto de artículo	Palabras ENG	Palabras ESP	KB	Segmentos	Palabras/ segmento
Historical Demography	3.688	4.019	52	120	33,5
Demography of Ancient World	2.019	2.446	40	106	23,1
The Population of Europe: The Demographic Transition	6.215	7.654	131	293	26,1
Household Composition	1.117	1.248	25	49	25,5
Peasant and Farming Villages	5.340	6.101	51	200	30,5
Family Demography	3.254	3.791	69	138	27,5
Family and Household Structure	3.551	4.174	62	163	25,6
The Population of Europe: Early Modern Patterns	6.737	7.545	145	280	27,0
TOTALES	31.921	36.978	575	1.349	27,4

Tabla 1. Textos de artículos de enciclopedia

2. Resúmenes de artículos de revista (16)

Como segundo procedimiento, se han recopilado los resúmenes y correspondientes traducciones al inglés (*abstracts*) de 16 artículos publicados por especialistas en la *Revista de Demografía Histórica* entre 2011 y 2016 (ver anexo 2). Insistimos en que se trata de los resúmenes y no del texto completo de los artículos. Esta prestigiosa publicación es editada por una entidad científica, la Asociación Española de Demografía Histórica, y se encuentra disponible en su página web.⁸ En este caso, las traducciones al inglés han sido facilitadas por los propios autores en sus *abstracts*, lo que, en principio, hace suponer la utilización de terminología especializada y sus

⁸ Los números de esta revista desde 1983 a 2018 –denominada Boletín de la Asociación Demografía Histórica hasta el 2000– están disponibles en <http://www.adeh.org/?q=es/node/92>

correspondientes equivalentes en inglés. Los resúmenes/*abstracts* cuentan además con sus correspondientes palabras clave/*key words*. Estos 16 textos se encuentran en formato PDF y suman en español aproximadamente 3.400 palabras.

Resumen de artículo	Palabras ⁹ ENG	Palabras ESP	Segmentos	Palabras/ segmento
El crecimiento demográfico...	156	177		
Hogar y familia...	202	216		
Women, family, work and welfare in Europe...	178	204		
Familia, género y migración interna en España...	234	251		
Las edades de acceso al mercado de trabajo formal...	152	175		
Un estudio de caso sobre abandono infantil...	426	454		
El impacto demográfico de las creencias...	154	180		
Grado de fiabilidad del Movimiento Natural de la Población...	255	287		
Ingresos, ocupaciones e identidades entre las clases altas...	168	184		
El acceso de los jóvenes de la nobleza a la herencia...	99	112		
Familia y estrategias de alianza en el Reino de Mallorca...	128	139		
Marginación y pobreza desde la cuna: el niño expósito...	238	273		
Ya en pleitos desde la más tierna infancia: menores...	151	174		
Consanguinidad, exogamia y estrategias colectivas...	174	172		
El descenso pionero de la mortalidad...	218	198		
Viudedad en la España del siglo xx...	175	175		
TOTALES	3.108	3.371	136	24,8

Tabla 2. Textos de resúmenes de artículos de revista

Antes de seguir adelante debemos tratar brevemente sobre la representatividad de los textos seleccionados. Esta va ligada al tamaño del corpus y este a la disponibilidad de tiempo para su recopilación. Se trata de reflexionar sobre cuántos textos o qué extensión de textos son necesarios para que el corpus refleje las características del lenguaje especializado seleccionado. Nosotros contamos en total con 24 textos de Demografía Histórica de distintos autores y referidos a distintos aspectos y épocas. En español, suman en conjunto 40.349 palabras.

El volumen óptimo del corpus de trabajo es una cuestión compleja. Hay autores que opinan que para tener una representatividad garantizada son necesarios textos que sumen millones de palabras. Sin embargo, un creciente número de investigadores defiende que con textos mucho menores y manejables pueden obtenerse buenos resultados. Hay quien, como Williams, concluye que la verdadera representatividad es imposible (citados por Corpas y Seghiri 2010: 120 y 134). Por otro lado, en opinión de varios especialistas, la representatividad del corpus estaría vinculada más a la naturaleza de los textos que a su número. Según Vigier, aquellos muy estereotipados, donde las expresiones y el vocabulario se repiten con frecuencia, requerirán apenas de entre 10 y 25 textos (Vigier, 2016). Browker y Pearson (2002: 45-59) consideran que sobre este tema no hay reglas fijas y Kennedy (2014: 70-82) que no siempre

9 Incluyendo las palabras clave/*key words*

un corpus más extenso proporciona mejores resultados. Existen también programas informáticos que realizan complejos estudios estadísticos sobre esta cuestión.¹⁰

En nuestra investigación, nos encontramos con limitaciones muy similares a las reflejadas en el trabajo sobre textos médicos de Corpas (2001: 165). Nos referimos a la limitación de tiempo disponible para compilar y analizar un gran número de textos.

2.2.2. Conversión de formatos y pos-edición

Las versiones en inglés y español de los textos obtenidos han sido convertidas a texto plano en formato [.odt] de LibreOffice, software como OmegaT también libre. Los artículos enciclopédicos y sus traducciones, a partir del formato HTML y los resúmenes/*abstracts* de los artículos de revista a partir del formato PDF. Esto se ha hecho para poder trabajar con OmegaT con más comodidad y sin problemas de compatibilidad de formatos.

Posteriormente, se ha procedido a una pos-edición eliminando elementos gráficos (guiones medios, puntos de iniciales, fotografías, tablas, publicidad, etc.) que pueden provocar errores de lectura o de reconocimiento. También, mediante la aplicación para buscar y reemplazar de LibreOffice, se han transformado los puntos y comas en puntos finales.

Por su extensión, los artículos de enciclopedia se han reproducido en dieciséis documentos separados (ocho en inglés y ocho en español), mientras que los artículos de revista se han fundido en solo dos documentos (uno en inglés y otro en español).

2.2.3. Alineación de bitextos

Para la alineación se ha utilizado la correspondiente aplicación de OmegaT. Se han creado ocho proyectos para los artículos de enciclopedia y uno para los resúmenes de los artículos de revista.

La segmentación aplicada ha sido la estándar proporcionada por la herramienta, es decir, a partir del elemento textual [.] (punto). Se han detectado algunos errores de reconocimiento originados por los puntos de iniciales y acrónimos, transformación de signos gráficos como guiones medios, puntos y comas, inserciones de espacios después de puntos, etc. Se ha observado también que algunas de las correcciones realizadas en la pre-edición de los textos en formato [.odt] no son reconocidas por OmegaT, que las transforma en signos como [`<t0/>`] y similares.

A continuación, la segmentación ha sido corregida manualmente. Los resultados finales arrojan un total de 1.349 segmentos para los artículos de enciclopedia y 136 para los resúmenes de los artículos de revista. Esto supone que los segmentos tienen una longitud media de 25 y 27 palabras por segmento respectivamente.

2.2.4. Extracción de recursos

¹⁰ ReCore, por ejemplo, determina a posteriori el tamaño mínimo del corpus o colección textual con independencia de la lengua o género textual de dicha colección.

a. Creación de proyectos de traducción

Con el alineación de textos, a partir de los artículos de enciclopedia se han creado 8 memorias de traducción. A pesar de su generación individual, en el análisis se utilizarán de forma conjunta bajo la denominación “MT_enciclopedia”.

En cambio, a partir del alineación del documento en el que figuran agrupados los resúmenes de los artículos de revista, se ha creado una única memoria de traducción que utilizaremos en el análisis con el nombre “MT_revista”.

A partir de los 8 artículos de enciclopedia se han creado otros tantos proyectos de traducción. En cada uno de ellos, se ha insertado como documento fuente a traducir el texto en español. A continuación se le ha asignado las memorias de traducción agrupadas en “MT_enciclopedia”. Somos conscientes de que, aunque hemos traducido automáticamente del inglés al español, en la parte empírica de este estudio la combinación lingüística es inversa, es decir, del español al inglés.

De igual forma se ha procedido con el único documento que agrupa los resúmenes de los artículos de revista, al que se le ha asignado “MT_revista”.

En su funcionamiento, OmegaT sugiere para cada segmento en español la traducción importada de las MT. Lógicamente, el nivel de coincidencia es sistemáticamente del 100%. No obstante, se corrigen manualmente algunos detalles y errores de las traducciones con el fin de evitar su propagación interna. Una vez validadas, quedan incorporadas automáticamente a las memorias de traducción.

b. Bases de datos terminológicos

En cada proyecto, se procede a la extracción de glosarios. A cada término en español seleccionado se le asigna su correspondencia en inglés sugerida por la MT respectiva. Los términos seleccionados incluyen expresiones de diversa naturaleza de amplio uso en la literatura especializada en Demografía Histórica.

- A) *Demográficas*: tasas, índices, ratios y proporciones de natalidad, mortalidad, fecundidad o nupcialidad.
- B) *Documentales*: libros parroquiales, registros sacramentales, etc.
- C) *Socio-económicas*: trabajo asalariado, explotación infantil, asociaciones gremiales, etc.
- D) *Cronológicas*: Alta Edad Moderna, Periodo de entreguerras, Antiguo Régimen, etc.
- E) *Geográficas*: Canal de la Mancha, Franja celta, Flandes belga, etc.
- F) *Acrónimos y su desarrollo*: PIB, encuesta nacional de familias y hogares, etc.

En el caso de los artículos de enciclopedia, los términos seleccionados y sus correspondientes traducciones engrosan un único glosario. Este se genera incorporando sucesivamente a cada uno de los ocho proyectos el glosario del proyecto anterior. De esta forma, la terminología extraída va enriqueciendo ese glosario único, denominado “BDT_enciclopedia”.

En el caso de los resúmenes de los artículos de revista, se origina un único glosario denominado “BDT_revista”.

2.2.5. Depuración de glosarios

Finalmente, los términos de ambos glosarios son exportados en texto plano a una hoja de cálculo en formato [.xml] también de LibreOffice. Allí son puestos en orden alfabético con el fin de detectar la repetición de términos, eliminar selecciones no procedentes y unificar traducciones en el caso de que se den varias para un mismo término. En concreto, las operaciones realizadas con el fin de que posteriormente se ofrezcan traducciones unívocas (un mismo término, una misma traducción) son del siguiente tipo.

- Eliminación de términos duplicados.
- Desglose de términos que aparecen agrupados: *marital and cohabiting unions* → *marital unions* y *cohabiting unions*.
- Corrección de términos “desdoblados”. Por ejemplo, en algunas traducciones se han equiparado por error los términos “familia” y “hogar”.
- Estandarización de terminología, bien cuando se ofrecen dos traducciones para el mismo término, bien cuando existen alternativas (*breadwinner-homemaker family* → “familia sostén económico-ama de casa”).
- Estandarización de formas: especialmente, el uso del guión o mayúsculas en algunas expresiones.

Todas estas correcciones se han trasladado a los glosarios. También se ha procedido a la corrección manual de traducciones en la propia interfaz de OmegaT. Con todo, pese a la corrección general de las traducciones, se han detectado algunos errores. He aquí algunos ejemplos.

En uno en uno de los resúmenes de artículos de la revista se traduce la expresión “Siglos XVII y XVIII” por “XVIIth and XVIIIth centuries”. Es un claro error, pues en inglés no se usa la numeración romana para denominar a los siglos.

También en la traducción de artículos enciclopédicos se observa algún error. Google Translate ha traducido *records series* por “series de discos”. El contexto del artículo indica, sin embargo, que en este caso la traducción pertinente sería “series documentales”.

Finalmente, el número de términos por glosario queda establecido en 330 para “BDT_encyclopedia” y 147 para “BDT_revista”.

2.3. Test de funcionamiento

Una vez elaborados los recursos necesarios para poder traducir con OmegaT, se examinan los resultados utilizando casos de traducción reales.

2.3.1. Memorias de traducción

Para el test de las MT, se crean en OmegaT tres proyectos para la traducción de otros tantos textos de Demografía Histórica. Buscando su representatividad, han sido elegidos buscando una extensión similar, y una temática, ámbito geográfico y periodos cronológicos variados.

Autor	Texto	Temática	Ámbito	Cronología	Palabras
1. MILEKARENA	“La estructura de los hogares navarros”	Estructura familiar	Navarra	S. XVIII	3.074
2. QUILODRÁN	“Los cambios de la familia”	Dinámica familiar	Hispanoamérica	S. XX	3.086
3. CACHINERO	“El modelo europeo de matrimonio”	Dinámica social	Europa	S. XVI-XX	2.933

Tabla 3 . Textos para proyectos de traducción

A los tres proyectos de traducción se les asignan, en dos fases, las memorias de traducción creadas mediante cada uno de los dos procedimientos anteriormente especificados: “MT_encyclopedia” y “MT_revista”. En ambos casos se extraen, con el fin de compararlas, las estadísticas de coincidencias proporcionadas por OmegaT y que resumimos en la tabla 6. Estas son comprobadas manualmente a la vista de las sugerencias de la herramienta para cada segmento.

2.3.2. Bases de datos terminológicos

En el caso de los glosarios, el test de funcionamiento real en la traducción de textos se amplía a 5, añadiendo a los 3 anteriores 2 libros especializados.

Autor	Texto de procedencia	N.º términos extraídos
1. MIKELARENA	“La estructura de los hogares navarros”	70
2. QUILODRÁN	“Los cambios de la familia”	60
3. CACHINERO	“El modelo europeo de matrimonio”	53
4. HENRY	“Manual de Demografía Histórica” ¹¹	60
5. VINUESA	“Demografía: análisis y proyecciones” ¹²	60

Tabla 4. Relaciones de términos especializados

Como puede verse, de todos ellos se ha extraído una lista o relación con los términos especializados que contienen. No se contemplan las repeticiones de cada término. Estos incluyen conceptos similares a los señalados en el punto 2.2.4.b de la extracción de términos de BDT. Estas relaciones de términos constituyen el documento a traducir en cada uno de los 5 proyectos de traducción generados. Se ha recurrido a este sistema ante la limitación de tiempo para traducir íntegramente tantos textos y tan extensos.

En un procedimiento similar al descrito para las MT, se les ha asignado en dos tiempos los glosarios “BDT_encyclopedia” y “BDT_revista”.

En este caso, se han querido contrastar además los resultados obtenidos con la única BDT que hemos podido obtener de forma gratuita en la web *Europa Union Terminology (IATE)*¹³. Por

11 Henry, L. (1983). *Manual de Demografía Histórica*. Barcelona: Crítica.

12 Vinuesa, J. (Editor) et al (1994). *Demografía: análisis y proyecciones*. Madrid: Síntesis.

13 Recurso obtenido en <https://iate.europa.eu/home> el 20/11/2019.

no encontrar una más específica, hemos extraído los términos propuestos para el campo de “Humanidades”. Con ella hemos conformado el glosario adicional “BDT_IATE”.

3. RESULTADOS

Recapitulemos. La conveniencia de implantar la herramienta TAO OmegaT vamos a evaluarla a partir de su funcionamiento real. Para ello, vamos a examinar el nivel de respuesta obtenido por dos MT y dos BDT creadas siguiendo dos procedimientos distintos mediante el alineación de OmegaT. El primero, basado en artículos de enciclopedia traducidos con Google Translate, ha originado un conjunto de MT que hemos denominado “MT_enciclopedia” y una BDT denominada “BDT_enciclopedia”. El segundo, basado en los resúmenes de artículos de revista que aparecen traducidos en el propio texto, ha creado respectivamente “MT_revista” y “BDT_revista”.

3.1. Memorias de traducción

3.1.1. Resultados cualitativos

El número y condición de las palabras implicadas en las frases sugeridas por las MT muestran su ineficacia. Las coincidencias se producen en segmentos muy cortos y la mayoría de las veces son componentes lingüísticos no especializados.

En cuanto a la eficiencia, hemos de considerar los costes de elaboración de estas MT. La extracción de MT de los artículos de enciclopedia ha sido relativamente rápida. Se han obtenido textos extensos en inglés, rápidamente traducidos al español mediante TA y fácilmente exportables al formato [.odt]. La segmentación con OmegaT ha dado algunos problemas (supresión, repetición o fragmentación de párrafos), lo que, unido a su longitud, ha hecho más costoso el alineación. Con todo se han obtenido memorias de traducción que, en conjunto, ocupan 575 KB.

En el caso de los artículos de revista, el procedimiento ha sido algo más complejo. Han proporcionado simultáneamente textos en castellano e inglés, pero han sido más cortos. Además, debido a que se trata de textos en formato PDF, su conversión a un documento único y en formato [.odt] ha sido algo más laboriosa. En cambio, debido a su corta extensión, el alineación de los textos ha resultado más sencillo. Sólo en uno de los artículos el resumen en español difería en cuanto a estructura con el *abstract* en inglés.

Respecto a la eficiencia, cabría suponer que, por su naturaleza (deben de ser más sintéticos), el texto de los resúmenes de los artículos de revista contendría unos conceptos más precisos y frases más cortas. Sin embargo, no ha sido así. La longitud de los segmentos (el número de palabras por segmento) es similar en ambos casos. Y, con ello, las probabilidades de encontrar coincidencias en los textos a traducir se igualan.

3.1.2. Resultados cuantitativos

Desde el punto de vista de la eficacia en la generación de MT, la estrategia basada en los artículos de enciclopedia parece más apropiada que la basada en artículos de revista. No sólo

el procedimiento de obtención es más rápido, sino que, como puede verse en la tabla siguiente, ofrece mayor cantidad total de segmentos traducidos.

Textos	Formato	Palabras	KB	Segmentos	Pal./seg.
Enciclopedia	HTML	36.978	575	1.349	27,4
Revista	PDF	3.371	41	136	24,8

Tabla 5. Descripción cuantitativa de las MT

Por otro lado, para medir el nivel de respuesta obtenida al intentar traducir los tres artículos con las MT creadas, nos hemos servido de las “estadísticas de coincidencias” que OmegaT proporciona para cada proyecto. Todas ellas se resumen en la siguiente tabla.

Texto →	1-Mikelarena		2-Quilodrán		3-Cachinero	
Memorias de traducción →	Enciclopedia	Revista	Enciclopedia	Revista	Enciclopedia	Revista
Repeticiones	0	0	0	0	0	0
Coincidencia exacta	2	0	0	0	2	0
95% - 100%	0	0	0	0	0	0
85% - 94%	0	0	0	0	0	0
75% - 84%	0	0	0	0	0	0
50% - 74%	0	0	4	0	1	0
Menos 50%	1	0	1	0	4	0
Sin coincidencia (+ 50%)	72	74	98	102	97	100
Total segmentos	74	74	102	102	100	100
Total palabras	3.088	3.088	3.087	3.087	2.935	2.935

Tabla 6. Estadísticas de coincidencias con MT en los textos a traducir

Como puede observarse, la utilidad de las MT ha sido prácticamente nula. En los textos seleccionados, no hay repeticiones internas. Entre las coincidencias de segmentos, tan sólo el conjunto que hemos denominado “MT_enciclopedia” arroja dos correspondencias exactas. Hay también once parciales, pero con porcentajes muy poco representativos que raramente superan la mitad de las palabras de segmento. La “MT_revista” no muestra ningún nivel de respuesta.

3.2. Bases de datos terminológicos

3.2.2. Resultados cualitativos

Repetimos aquí lo dicho para el proceso de alineación de los textos obtenidos en ambas estrategias. Los artículos de enciclopedia proporcionan de forma más sencilla unos textos de mayor tamaño que, sin embargo, cuesta más alinear. Los resúmenes de los artículos de revista

proporcionan textos más breves que además exigen mayor control en su conversión, aunque su alineación es más fácil.

Textos	Formato	Palabras	KB	Términos	Pal./Término
Enciclopedia	HTML	36.978	16	330	112
Revista	PDF	3.371	6	147	23

Tabla 7. Descripción cuantitativa de las BDT

Tal y como se aprecia en la tabla, en términos relativos los textos procedentes de la revista proporcionan cuatro veces más terminología que los basados en los artículos de enciclopedia. Esto se debe sin duda a la mayor densidad terminológica de los resúmenes y sobre todo a la presencia de las “palabras clave” que dichos textos contienen.

3.2.2. Resultados cuantitativos

La proporción de respuestas obtenidas tras introducir sucesivamente las cinco relaciones de términos especializados y las tres BDT que se desean contrastar queda resumida en la tabla 8. Se distinguen las coincidencias exactas (cuando OmegaT sugiere el término idéntico), parciales (cuando sugiere parte del término) y las erróneas (cuando sugiere coincidencias no procedentes).

BDT →	Relación	Términos	IATE			Enciclopedia			Revista		
			exacta	parcial	error	exacta	parcial	error	exacta	parcial	error
	1. MILEKARENA	70	4,3	17,1	2,9	14,3	21,4		10,0	30,0	
	2. QUILODRÁN	60	1,9	9,4	3,8	9,4	11,3		11,3	20,8	
	3. CACHINERO	53	5,0	8,3	5,0	13,3	13,3		10,0	21,7	
	4. HENRY	60		1,7	3,3	33,3	38,3		8,3	23,3	
	5. VINUESA	60	10,0	25,0		15,0	40,0		3,3	28,3	
	Totales	303	4,3	12,5	3,0	17,2	25,1		8,6	25,1	

Tabla 8. Estadística de coincidencias de términos especializados (en %)

En todos los casos, el nivel de respuesta no ha superado la mitad de los términos.

Los mejores resultados los proporciona el glosario “BDT_enciclopedia” con una considerable coincidencia de términos exactos (17,2%) y parciales (25,1%). Los resultados son llamativos en la relación número 4 (basada en el manual clásico de Demografía Histórica de Henry) donde proporciona más del 70% de los términos especializados, aproximadamente la mitad de forma exacta y la mitad de forma parcial. Las coincidencias exactas también son destacables en las relaciones número 5, 1 y 3.

El glosario “BDT_revista” arroja resultados más modestos en cuanto a la coincidencia exacta (8,6%), pero equiparables al anterior en cuanto a las coincidencias parciales (25,1%). Los valores de eficacia más altos se producen en las relaciones 1, 2 y 3, aunque a un nivel modesto. Las coincidencias parciales son más regulares en todas las relaciones.

A pesar de su mayor tamaño, el glosario externo “BDT_IATE” arroja los peores resultados globales. La coincidencia exacta es casi irrelevante (4,3%) y la parcial, muy modesta (12,5%). Además, y esto es algo a considerar, es el único glosario que proporciona equivalentes erróneos (3%). Los mejores resultados los obtiene con la relación 5, un manual de Demografía contemporánea no histórica. Esto es coherente con la naturaleza no-histórica de ambos elementos.

Los resultados totales pueden apreciarse de forma visual en este gráfico.

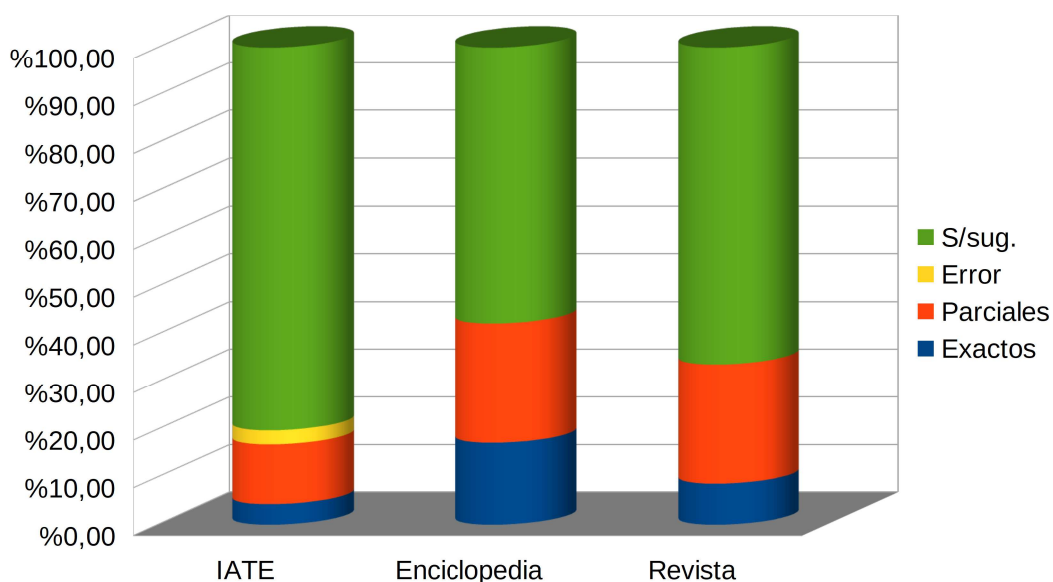


Figura 1. Coincidencias totales de los glosarios

4. CONCLUSIONES

En este TFM se ha intentado explorar empíricamente la posibilidad de recurrir a una herramienta TAO para realizar la traducción de textos especializados en el ámbito de la Demografía Histórica. De entre las existentes, se ha utilizado OmegaT, porque es gratuita, completa, versátil y fácil de usar.

Ante la falta de recursos disponibles para alimentar dos elementos básicos en el funcionamiento de OmegaT –las MT y las BDT–, estos han tenido que ser creados utilizando otra de sus aplicaciones, la alineación de textos. Para la recopilación del necesario corpus de textos paralelos bilingües se han diseñado dos procedimientos. En el primero, se ha recurrido a artículos de enciclopedia en inglés traducidos automáticamente al español mediante una herramienta de TA. En el segundo, a los resúmenes en español y sus correspondientes *abstracts* en inglés de artículos de una revista especializada.

Se ha evaluado el rendimiento de OmegaT observando su funcionamiento práctico en la traducción de varios textos. La eficiencia y eficacia tanto en la elaboración de recursos como en su rendimiento en OmegaT se ha evaluado tanto mediante indicadores cualitativos como cuantitativos. Debido a las posibilidades que ofrece una investigación de este alcance, las conclusiones obtenidas tienen una representatividad limitada y un valor meramente orientativo. Sin embargo, pueden señalarse las siguientes conclusiones que coinciden en lo esencial con las obtenidas por otros autores en cuanto al funcionamiento de la herramienta OmegaT (Filatova, 2010) y la calidad de las traducciones obtenidas (Ramírez Polo, 2011; Rankine, 2017; Viver, 2018).

OmegaT se confirma como una herramienta TAO adecuada y útil para la traducción de textos especializados de Demografía Histórica. Se respaldan en este estudio las virtudes señaladas en otras investigaciones respecto a su carácter gratuito, versátil, completo, ágil y de fácil uso. Del mismo modo, la extracción de MT y BDT resulta sencilla recurriendo a la alineación de textos.

Respecto a la eficacia y eficiencia de los tipos de recursos creados, no obstante, la valoración debe ser matizada.

4.1. Memorias de traducción

Las pruebas realizadas muestran la utilidad muy limitada de las MT en el campo de la Demografía Histórica. Ninguna de las creadas por OmegaT, sea cual fuese el procedimiento de obtención, ha ofrecido resultados que incrementen la productividad o la coherencia de los textos a traducir. Esto puede haberse debido a varias razones:

1. *Reducida representatividad cuantitativa del corpus de textos paralelos utilizados.* Esto nos sitúa ante la ya comentada cuestión relativa al número de textos que debe integrar ese corpus para que los resultados sean representativos.
2. *Especificidad del lenguaje de la disciplina.* Sin duda alguna, los textos de Demografía Histórica, además de una terminología científica muy específica, poseen un alto grado

de creatividad y, por lo tanto, de variabilidad. Es decir, apenas tienen fórmulas estereotipadas (como el Derecho, la Farmacia, la Medicina o la Economía) y el grado de repetición externa e interna suele ser muy bajo. Esto, al igual que Rankine (2017) concluye para los textos literarios, supone unas bajas probabilidades de que las MT proporcionen segmentos traducidos con un aprovechable nivel de coincidencia.

3. *Excesiva longitud de los segmentos.* Derivado de lo anterior, las frases son largas y, en consecuencia, si se utiliza como criterio de delimitación de las unidades de traducción el punto, se originan segmentos también largos. Como bien señala Oliver, “la probabilidad de encontrar fragmentos iguales o similares en una memoria de traducción es menor si el fragmento es largo” y si son demasiado cortos, “la unidad y la coherencia de texto se romperían y harían imposible la traducción” (Oliver, 2013: 32).

Aunque sin ninguna validez estadística, hemos explorado esta posibilidad con un texto extenso segmentado recurriendo a la coma como delimitador de las unidades de traducción. La alineación ha planteado algunos problemas por el diferente uso de la coma en inglés y en español. Tras ello, se han obtenido segmentos notablemente más cortos (una media de 16 palabras por segmento). El porcentaje de coincidencias ha sido mayor, aunque se han centrado más en elementos gramaticales del tipo “sin embargo”, “en cuanto al primer punto”, “en primer lugar” o similares. Consideramos que estas coincidencias incrementan muy poco la productividad y no aportan casi nada a la coherencia de texto. Por otra parte, los segmentos que sí contienen terminología especializada coinciden casi siempre con los términos incluidos en los glosarios ya obtenidos.

No obstante, somos conscientes de que queda pendiente llevar a cabo un estudio más profundo y sistemático para obtener conclusiones más sólidas.

4.2. Bases de datos terminológicos

A diferencia de las MT, los glosarios elaborados sí han mostrado su utilidad. Respecto a los incrementos de productividad, esta ha aumentado, especialmente cuando los términos especializados sugeridos se repiten internamente. Su utilidad es aún mayor en el aumento de la coherencia (y consiguiente calidad) de los textos al proporcionar de forma constante equivalentes únicos y correctos en inglés para los términos especializados utilizados en castellano.

Respecto a los procedimientos utilizados en la elaboración de glosarios, cabe destacar que ambos se han mostrado muy útiles. La extracción de terminología a partir de artículos de enciclopedia traducidos ha proporcionado de forma rápida un gran número de términos, pero son los artículos de revista los que lo han hecho con mayor densidad debido, sobre todo, a la inclusión en ellos de palabras clave. En todo caso, cualquiera de los procedimientos se ha mostrado eficaz.

4.3. Líneas de investigación futura

Así, pues, la investigación contenida en este TFM, a pesar de su modesta representatividad, respalda la conveniencia de adoptar OmegaT como herramienta para la traducción de textos del ámbito de la Demografía Histórica del español al inglés. Sugiere, igualmente, que la metodología aplicada en este TFM podría extenderse al campo de la Historia en general.

Es cierto que en un estado inicial debería invertirse tiempo en la elaboración simultánea de MT y, sobre todo, de BDT. Para ello, la alineación de textos obtenidos por ambos procedimientos se ha mostrado muy útil. La disponibilidad en la red de estos textos parece también amplia. Además de ello, todos estos recursos se pueden ir enriqueciendo progresivamente mediante propagación (incorporación automática de los textos traducidos) con el propio uso profesional de la herramienta TAO.

Finalmente, la eficiencia de implantar la herramienta OmegaT y elaborar los correspondientes recursos para la traducción en estos campos debería valorarse también a la vista del volumen y frecuencia con que se van a abordar encargos de traducción de textos especializados en el ámbito de la Demografía Histórica.

5. BIBLIOGRAFÍA

- Anica, V. N. (2014). *Free CAT tools as an alternative to commercial software: OmegaT* (Trabajo Fin de Grado). Universidad de Valladolid.
- Browker, L. y Pearson, J. (2002). *Working with Specialized Language. A Practical Guide to Using Corpora*. London/New York: Routledge.
- Caetano, A. L. (2016). Software livre e de código aberto no ensino de programas de apoio à tradução: OmegaT, uma alternativa viável. *Belo Horizonte*, 9-2, 219-230.
- Castillo, C. (2009). La elaboración de un corpus ad hoc paralelo multilingüe. *Revista tradumática*, Diciembre, 1-11. Recuperado de <http://www.fti.uab.cat/tradumatica/revista/num7/articles/03/03art.htm>
- Corpas, G. (2001), Compilación de un corpus ad hoc para la enseñanza de la traducción inversa especializada. *TRANS. Revista de traductología*, 5, 155-184.
- Corpas, G. y Seghiri, M. (2010). Size Matters: A Quantitative Approach To Corpus Representativeness. *Lengua, traducción, recepción. En honor de Julio César Santoyo*, 112-146.
- Filatova, I. (2010). *Evaluación de herramientas y recursos informáticos (TAO y Ofimática) para la traducción profesional. Hacia la configuración de un entorno óptimo de trabajo para el traductor autónomo* (Tesis Doctoral). Universidad de Málaga.
- Forcada, M. L., Sánchez-Martínez, F. y Pérez-Ortiz, J. A. (2016), *Manual de informática y de tecnologías para la traducción*. Alicante.
- Francés Oltra, N. (2018). *Evaluación de la traducción automática para la asimilación* (Trabajo Fin de Grado). Universidad de Alicante.
- Jordan-Núñez, K., Forcada, M. L. y Chia, E. (2017). Usefulness of MT output for comprehension -an analysis from the point of view of linguistic intercomprehension. *Proceeding of MT Summit XVI*, 1, 241- 253.
- Kennedy, G. (2014), *An Introduction to Corpus Linguistics*. Londres; New York: Routledge.
- LeBlanc, M. (2013). Translators on translation memory TM. Results of an ethnographic study in three translation services and agencies. *The International Journal for Translation and Interpreting Research*, 5-2, 1-13.
- Norvig, P. (2011). *Una mirada dentro de la tecnología de Google Translate*. (20 de diciembre de 2019). Recuperado de <https://latam.googleblog.com/2011/11/una-mirada-dentro-de-la-tecnologia-de.html>
- Oliver, A. (2013). *Herramientas tecnológicas para traductores*. Barcelona: Universitat Oberta de Catalunya.

Ortiz, C. (2016). *Implementing Machine Translation and Post-Editing to the Translation of Wildlife Documentaries through Voice-over and Off-screen Dubbing* (Tesis Doctoral). Universidad Autónoma de Barcelona.

Padilla, Ó. (2018). *¿Cómo funciona el traductor de Google?*. (20 de noviembre de 2019). Recuperado de <https://elandroidelibre.lespanol.com/2018/12/como-funciona-traductor-de-google-translate.html>

Paulsen, T. y Schjoldager, A. (2011), The Impact of Translation-Memory TM. Technology on Cognitive Processes: Student-Translator's Retrospective Comments in an Online Questionnaire. *Proceedings of the 8th International NLPSC Workshop: Human-Machine Interaction in Translation*, 119-130.

Ramírez Polo, L. (2011). Managing the translation workflow with a Computer Assisted Translation Tool: SLD Trados 2011. *Language Value*, 5 (1),161-174.

Rankine, H. E. (2017). *El uso de la Traducción Asistida por Ordenador en la traducción de textos literarios. Un estudio de caso: traducción de un fragmento de la novela 'Catch 22' con OmegaT* (Trabajo Fin de Postgrado). Universitat Oberta de Catalunya.

Rodríguez Bausela, M. (2014). Aplicación de la Traducción Asistida por Ordenador a la investigación en traducción. *Estudios de Traducción*, 4, 129-143.

Seguiri, M. (2011). Metodología protocolizada de compilación de un corpus de seguros de viajes: aspectos de diseño y representatividad. *Revista de Lingüística Teórica y Aplicada*, 49 (2), 13-30.

Tabor, J. (2013). *CAT tool use by translator: who is using?*. (20 de noviembre de 2019). Recuperado de <https://prozcomblog.com/2013/03/22/cat-tool-use-by-translator-who-is-using>.

Vallianatou, F. (2005). CAT tools and Productivity. Tracking Words and Hours. *Translation Journal*, 9 (4). (23 de noviembre de 2019). Recuperado de <http://translationjournal.net/journal/34CAT.htm>

Vigier, F. J. (2016). Los corpus *ad hoc* en la traducción inversa de los textos jurídicos: ejemplos de su utilización como fuentes de información fraseológica y terminológica. *Revista Académica LiLETRAd*, 2, 867-878.

Viver, P. (2018). *La evaluación de las herramientas de traducción automática (TA) desde la perspectiva del traductor: Google Translate, Bing, Babylon y Systran* (Trabajo Fin de Grado). Universidad de Valladolid.

6. ANEXOS

1. Textos en INGLÉS obtenidos en Encyclopedia.com

1. "Historical Demography ." Encyclopedia of Population . Consultado el 28/11/2019 en Encyclopedia.com: <https://www.encyclopedia.com/social-sciences/encyclopedias-almanacs-transcripts-and-maps/historical-demography>
2. "Ancient World, Demography of ." Encyclopedia of Population . Consultado el 28/11/2019 en Encyclopedia.com: <https://www.encyclopedia.com/social-sciences/encyclopedias-almanacs-transcripts-and-maps/ancient-world-demography>
3. "The Population of Europe: The Demographic Transition and After ." Encyclopedia of European Social History . Consultado el 28/11/2019 en Encyclopedia.com: <https://www.encyclopedia.com/international/encyclopedias-almanacs-transcripts-and-maps/population-europe-demographic-transition-and-after>
4. "The Population of Europe: The Demographic Transition and After ." Encyclopedia of European Social History . Consultado el 28/11/2019 en Encyclopedia.com: <https://www.encyclopedia.com/international/encyclopedias-almanacs-transcripts-and-maps/population-europe-demographic-transition-and-after>
5. "Household Composition ." Encyclopedia of Population . Consultado el 28/11/2019 en Encyclopedia.com: <https://www.encyclopedia.com/social-sciences/encyclopedias-almanacs-transcripts-and-maps/household-composition>
6. "Peasant and Farming Villages ." Encyclopedia of European Social History . Consultado el 28/11/2019 en Encyclopedia.com: <https://www.encyclopedia.com/international/encyclopedias-almanacs-transcripts-and-maps/peasant-and-farming-villages>
7. "Family Demography." Encyclopedia of Population . Consultado el 28/11/2019 en Encyclopedia.com: <https://www.encyclopedia.com/social-sciences/encyclopedias-almanacs-transcripts-and-maps/family-demography>
8. "Family and Household Structure ." Encyclopedia of Sociology . Consultado el 28/11/2019 en Encyclopedia.com: <https://www.encyclopedia.com/social-sciences/encyclopedias-almanacs-transcripts-and-maps/family-and-household-structure>

2. Textos en ESPAÑOL-INGLÉS (resúmenes y abstracts) obtenidos de los artículos de Revista de Demografía Histórica

1. Ferrer. L. (2016). El crecimiento demográfico de las comarcas de Girona en el siglo XVIII y los cambios hacia la transición demográfica en la primera mitad del siglo XIX. *Revista de Demografía Histórica*, 35 (1), 17-52.

2. Chacón, F. y Chacón, A. (2015). Hogar y Familia: corrientes interpretativas y realidades sociales. Los ejemplos de movilidad de la población (1771) y movilidad social (1797) en Lorca. *Revista de Demografía Histórica*, 33 (2), 29-64.
3. Moring, B. (2015). Women, family, work and welfare in Europe in the long 19th century. Budget studies, the nuclear family and the male breadwinner. *Revista de Demografía Histórica*, 33 (2), 119-154.
4. Recaño, J. (2015). Familia, género y migración interna en España. *Revista de Demografía Histórica*, 33 (2), 153-184.
5. Iturralde, M. (2015). Las edades de acceso al mercado de trabajo formal: de los oficios tradicionales a la industria algodonera moderna, Barcelona, 1784-1856. *Revista de Demografía Histórica*, 33 (1), 65-98.
6. León, M. (2015). Un estudio de caso sobre abandono infantil en la Andalucía Moderna: los expósitos de la inclusa antequerana. *Revista de Demografía Histórica*, 33 (1), 99-133.
7. Dopico, F. (2014). El impacto demográfico de las creencias. Una evaluación desde el siglo XVIII español. *Revista de Demografía Histórica*, 32 (2), 51-76.
8. Durán, A. (2014). Grado de fiabilidad del Movimiento Natural de la Población en los municipios de la provincia de Badajoz. 1900-1935. *Revista de Demografía Histórica*, 32 (2), 77-115.
9. Artola, M. (2014). Ingresos, ocupaciones e identidades entre las clases altas: Madrid, 1930. *Revista de Demografía Histórica*, 32 (2), 21-50.
10. Bartolomé, J. M. (2013). El acceso de los jóvenes de la nobleza a la herencia: el ejemplo de las familias de regidores leoneses (1700-1850). *Revista de Demografía Histórica*, 31 (2), 25-46.
11. Picazo, A. (2014). Familia y estrategias de alianza en el Reino de Mallorca en el siglo XVII: esbozo de un modelo analítico. *Revista de Demografía Histórica*, 31 (2), 113-129.
12. Sánchez Fernández, L. V. (2014). Marginación y pobreza desde la cuna: el niño expósito en el Concejo de Siero, Asturias (1800-1936). *Revista de Demografía Histórica*, 31 (2), 131-165.
13. García Fernández, M. (2014). Ya en pleitos desde la más tierna infancia: menores, tutores, litigios. *Revista de Demografía Histórica*, 31 (2), 87-112.
14. Blanco, J. P. (2013). Consanguinidad, exogamia y estrategias colectivas en la sociedad moderna. Reflexiones en torno a un ejemplo rural (1700-1820). *Revista de Demografía Histórica*, 30 (2), 25-54.
15. Gonzalvo-Cirac, M. (2013). El descenso pionero de la mortalidad en la provincia de Tarragona, 1900-1960: análisis epidemiológico. *Revista de Demografía Histórica*, 30 (2), 85-125.

16. Spijker, J. J. A. (2011). Viudedad en la España del siglo xx: La evolución histórica de la población viuda y sus determinantes demográficos. *Revista de Demografía Histórica*, 29 (2), 119-150.