Análisis de la concordancia paralela en la traducción de verbos en gerundio *(-ing)* en un corpus *ad hoc* de textos técnicos de mercadotecnia de salud de mascotas

Parallel concordance in the translation of gerund verbs (-ing) *in an* ad hoc *corpus of technical pet health marketing texts*

Claudia Elizabeth Leyra Parrilla
Universidad Autónoma del Estado de México, México / Universidad Autónoma de Baja California, México
celeyrap@uaemex.mx

José Cortez Godínez
Universidad Autónoma de Baja California, México
jose_cortez@uabc.edu.mx

Original recibido: 28/09/23
Dictamen enviado: 06/03/2024
Aceptado: 10/07/2024

Resumen

Esta investigación plantea abordar el fenómeno lingüístico de la ambigüedad en la traducción de verbos en gerundio (*-ing*) que se encuentran en textos técnicos de marketing de salud de mascotas del inglés al español. Se formará un corpus *ad hoc* paralelo bilingüe (que podría considerarse una muestra de un corpus más grande) mediante el uso de la herramienta *Sketch Engine* con la finalidad de activar la función de concordancia paralela para cumplir con el objetivo de contrastar con esta función la traducción de verbos en gerundio (*-ing*) del inglés al español. La pregunta de investigación de este trabajo es: ¿es posible contrastar aspectos que tienden al error en el proceso de posedición de la traducción de textos técnicos de salud de mascotas del inglés al español para verificar el grado de calidad de sus traducciones? La presente investigación es mixta. En primera instancia, es de corte cualitativo porque se contrastan las propuestas de traducción de verbos en gerundio (*-ing*) del inglés al español en un corpus *ad hoc* que fue construido para verificar la calidad de traducción de esta tipología de textos. Por otro lado, también es cuantitativa, porque se extraen tanto el número de errores, así como, el porcentaje del grado de calidad que se encontró en el corpus *ad hoc*. Debido a esto, la lingüística de corpus puede fungir como una metodología y ser aplicada al análisis de fenómenos lingüísticos, en este caso, para la posedición de la traducción de textos. Los datos indican que el grado de calidad sí puede ser verificable

a través de la construcción de un corpus *ad hoc* y esto puede fungir como una manera de revisar varios aspectos que tienden al error en la traducción de textos técnicos del inglés al español.
Palabras clave: concordancia paralela en *Sketch Engine*, posedición de la traducción, traducción de textos técnicos del inglés al español

Abstract

This research takes into consideration the linguistic phenomena of ambiguity in the translation of -ing verbs that appear in technical pet health marketing texts from English to Spanish through an bilingual parallel ad hoc *corpora, which could be a sample of a bigger corpora, through the Sketch Engine tool and by using the parallel concordance function, and to prove the main objective that is to contrast by using such function that the translation of -ing verbs from English to Spanish. Because of that the research question is the following: is it possible to contrast aspects that tend to be mistaken during the process of translating technical texts of pet health from English to Spanish to verify the quality in such translations? This research is mixed, firstly, it is qualitative, because there is a contrast of translation of -ing verbs from English to Spanish in an* ad hoc *corpus, which was done exclusively to verify the quality of translation of such type of texts, and at the same time, it is also quantitative, because the number of mistakes are also given, as well as the percentage of the quality range that was found in the* ad hoc *corpus. Due to that, corpus linguistics could be seen as a methodology that can be applied to analyze linguistic phenomena, in this case, to postedit translated texts. Thanks to this analysis the conclusion is that the range of quality can be verified through the construction of an* ad hoc *corpus, and this can be a way to verify many aspects that tend to be mistaken in the translation of technical texts from English to Spanish.*
Keywords*: parallel concordance in Sketch Engine, posedition in translation, translation of technical texts from English to Spanish*

Introducción

De acuerdo con Fernández (2018, p. 43) el uso actual de la posedición profesional en México es del 1 %, ya que no se considera una labor que contribuya a la mejoría de la traducción de textos. Sin embargo, es importante referir que, con el paso del tiempo existe cada vez más un volumen considerable de textos y menos traductores que puedan realizar la labor de traducción desmedida que las tecnologías de la información requieren (ITC Translations, 2023).

Debido a lo anterior, el uso del corpus lingüístico paralelo bilingüe permitiría a los traductores establecer comparaciones de textos ya traducidos que contienen

errores derivados del uso de un programa de traducción automática, o bien, de una herramienta de traducción asistida por computadora con apoyo de algún motor de traducción. En primer lugar, el uso del corpus podría ayudar a conocer la tipología de textos a los que se enfrentan. Seguido de ello, pueden establecer estrategias pertinentes para mejorar el refinamiento de un texto que fue pretraducido[1] con alguna herramienta de traducción automática. Y también puede contribuir a la realización de investigación al respecto de su aplicación para mejorar y entender los procesos y técnicas de traducción y posedición de textos técnicos. Según Kenning (2010)

> Un corpus paralelo es la colección de textos electrónicos que se relacionan de manera cercana entre sí, aunque de manera distinta. […] Los corpus paralelos prototípicos consisten en un conjunto de textos en una lengua A y sus traducciones en una lengua B (ejemplo: los escritos de Dickens y sus traducciones al francés). (p. 487)

Ahora bien, un corpus *ad hoc,* de acuerdo con Castillo (2009, p. 2), es el que se crea en específico para suplir las necesidades a las que obedece. Por lo que es muy común que los traductores que trabajan con muchos textos de la misma tipología (por ejemplo, bajo la temática de salud para mascotas) tiendan a tener corpus creados únicamente para realizar la labor de traducirlos. Por esta razón, se suple solamente su uso para sus fines particulares, ya sea para revisarlos, corregirlos, tener compilada una memoria de traducción o conformar un glosario terminológico específico para el tipo de traducción que realice.

Debido a ello, a través del uso de un corpus lingüístico paralelo bilingüe unidireccional (Leiva, 2018, p. 60), el cual fue creado *ad hoc* (Castillo, 2009, p. 2) para poder realizar un análisis de la función de concordancia paralela, es posible establecer el contraste entre textos, con la finalidad de poder establecer estrategias que permitan refinar y corregir las ambigüedades en textos técnicos del inglés al español que fueron pretraducidos de manera automática.

Por tanto, el objetivo principal de esta investigación es analizar los errores de traducción de verbos en gerundio (*-ing*) que aparecen en un corpus paralelo bilingüe *ad hoc* para buscar si existen errores, contrastarlos y mejorar la calidad en el proceso de posedición de textos técnicos del inglés al español. La pregunta de investigación

[1] Se utiliza en el presente la noción de "pretraducción" como la acción de utilizar un programa de traducción automática o alguna herramienta de traducción asistida por computadora con motor de traducción para realizar la traducción previa a la traducción humana de un texto del inglés al español.

es: ¿existe una manera de extraer una muestra de corpus para analizar errores específicos para hacer eficiente el proceso de posedición de textos técnicos del inglés al español? Por lo que la hipótesis es la siguiente: a través del uso del compilador *Sketch Engine* se puede extraer un corpus *ad hoc* para poder analizar un error específico que se da al poseditar textos técnicos del inglés al español.

Corpus, corpus paralelo bilingüe y corpus *ad hoc*

En la actualidad existe un crecimiento en la cantidad de estudios de traducción automática de textos técnicos del inglés al español que hacen uso de la lingüística de corpus para poder analizar fenómenos lingüísticos específicos que puedan desvelar soluciones que nos permitan mejorar la labor de los traductores día con día. Por ejemplo, Castillo (2009) refiere cómo conformar un corpus *ad hoc* paralelo multilingüe para su aplicación en el uso de traducciones en diferentes lenguas, y se basa en la discusión al respecto de su compilación y alineación. Por otro lado, Sánchez (2010) discute sobre el uso del corpus para la desambiguación automática de palabras o unidades de traducción polisémicas de los textos, a través de la comparación de campos semánticos, para corregir las traducciones de textos. Doval (2017) también discute cómo es que se puede conformar un corpus paralelo bilingüe multifuncional, destaca que su uso es para fines de lingüística contrastiva, traducción y enseñanza de lenguas. Leiva (2018) presenta una investigación sobre las herramientas de alineación que se pueden utilizar para compilar un corpus paralelo bilingüe (*WinAlign, Sketch Engine*, etcétera) y cómo es que alguien puede utilizarlas para alinear textos paralelos.

El artículo "La construcción de un corpus paralelo bilingüe multifuncional" de Doval (2017), aborda la creación y desarrollo del corpus PaGeS (*Parallel Corpus German/Spanish*), un recurso paralelo bilingüe del español/alemán diseñado para múltiples aplicaciones, especialmente en los campos de la lingüística contrastiva, la traducción y la enseñanza de lenguas extranjeras. El principal objetivo del proyecto es superar las limitaciones de los corpus paralelos existentes que incluyen el español y el alemán, los cuales suelen estar restringidos a textos de dominios muy específicos y no ofrecen material suficiente para investigaciones contrastivas amplias. Para ello, se recopiló un corpus alineado de tamaño y variedad léxica suficiente que proporcionara una base empírica robusta para la investigación.

Sánchez (2019) en "Corpus paralelos y traducción especializada: ejemplificación de diseño, compilación y alineación de un corpus paralelo bilingüe (inglés-español) para la traducción jurídica" se centra en la creación y desarrollo de un corpus paralelo bilingüe de inglés-español destinado a la traducción jurídica. El corpus en su fase

inicial consta de 127 y 145 sentencias alineadas de tribunales ingleses y españoles, respectivamente, con un total de 29,983 pares de oraciones alineadas. El proceso de alineación se llevó a cabo utilizando el programa *InterText*. El documento detalla las etapas de diseño y los desafíos técnicos asociados con el proceso de compilación del corpus. Se enfatiza la importancia de los corpus paralelos en la investigación de la traducción especializada, particularmente en el ámbito jurídico, donde la terminología y la fraseología son fundamentales. El artículo aborda la relevancia de los corpus textuales en la enseñanza de la traducción y cómo pueden ayudar a superar las dificultades inherentes a la traducción de textos jurídicos, que no solo presentan complejidades lingüísticas sino también pragmáticas y culturales.

Por otro lado, Pérez-Carrasco y Seghiri (2020) publicaron el artículo "Motocor: compilación de un corpus técnico paralelo bilingüe (inglés-español) y su aplicación en la traducción directa de un texto sobre automoción" el cual es un estudio académico que se centra en la creación y aplicación de MOTOCOR (*Motor vehicles Corpus*), un corpus virtual, paralelo y bilingüe (inglés-español), diseñado específicamente para la traducción de textos técnicos relacionados con la automoción, en particular, manuales de usuario de motocicletas. El estudio destaca la importancia de los corpus virtuales como herramientas de traducción, ya que permiten acceder a información conceptual, terminológica, estilística y gramatical a partir de un único recurso. MOTOCOR se creó siguiendo una metodología protocolizada para garantizar su representatividad tanto cualitativa como cuantitativa. Se utilizó la herramienta de gestión de corpus *Sketch Engine* para explotar el corpus, lo que facilitó la resolución de dificultades terminológicas, fraseológicas y gramaticales en la traducción de un fragmento de texto técnico.

Además, Corpas (2003) refiere que la lingüística de corpus ha experimentado un renacimiento notable en los últimos años, en tanto ha ido ampliando su campo de aplicación, sobre todo, en lo que respecta al procesamiento del lenguaje natural, la ingeniería lingüística, la terminología y la traducción. Recientemente se ha reconocido el valor de los corpus en traducción como recursos útiles para el estudio lingüístico contrastivo de las lenguas, lo cual ha contribuido de sobremanera al rápido desarrollo de los estudios con corpus y, muy especialmente, de aquellos relacionados con la traductología.

De acuerdo con el *Expert Advisory Group on Language Engineering Standards* ([EAGLES], 1996) donde al corpus se le define como "a collection of pieces of language that are selected and ordered according to explicit linguistic criteria in order to be used as a sample of the language" (Leiva, 2018, p. 2), se destaca que, el corpus debe obedecer a los criterios que se buscan en este para que resulte de

utilidad. Ahora bien, según Baker "a translation corpus consists of original, source language-texts in language A and their translated versions in language B" (1995, p. 230). Entonces, el corpus alineado es, a su vez, el corpus paralelo que resulta de un proceso denominado alineación, y que puede definirse como: "finding correspondences, in bilingual parallel corpora, between textual segments that are translation equivalents" (Kraif, 2002, p. 273).

El tipo de corpus que se trabaja en esta investigación es un corpus paralelo bilingüe *ad hoc*. En palabras de Harris (1988), un corpus paralelo bilingüe o bitexto considera que un texto original y su traducción no son en realidad dos textos, sino que conforman "a single text in two dimensions, each of which is a language" (p. 8).

Los corpus paralelos alineados no son, con frecuencia, objeto de estudio de investigaciones de la lingüística de corpus, lo cual se debe al hecho de que las herramientas disponibles tienen funciones limitadas o se desarrollaron para proyectos o investigaciones específicas, por lo que su acceso es restringido (Leiva, 2018, p. 60). Es por esta razón que en la presente propuesta solamente analizamos la concordancia paralela de un fenómeno lingüístico, que son los verbos con terminación *-ing* o gerundio en inglés, y propuestas de traducción automática que arroja el corpus construido para la misma.

Los corpus *ad hoc* (denominación acuñada por Aston, 1999) han recibido distintas denominaciones. A saber, *virtual corpus* (Ahmad, Holmes-Higgin y Raza-Abidi, 1994); *special purpose corpus* (Pearson, 1998); *corpus especials* (Sánchez-Gijón, 2003); *customized corpus* (Austermülh, 2001); *disposable* (Varantola, 2000); *do-it-yourself* (DIY) *corpora* (Zanettin, 2002) (en Leiva, 2018, p. 2).

Si bien el término "corpus paralelo" puede tener distintas acepciones dependiendo del ámbito de estudio, nosotros nos acercamos a este tipo de corpus desde el punto de vista de la traducción, es decir, corpus formados por originales en una lengua y sus correspondientes traducciones (Olohan, 2004). Este tipo de corpus será de gran utilidad para estudiar, por ejemplo, las distintas estrategias o técnicas de traducción –equivalencia, modulación, transposición, etc.– que se recopilan en López y Minett (2006); la traducción de la fraseología empleada, o la naturaleza de la lengua traducida, entre otros (Sánchez, 2019, p. 273); así como para realizar investigación en traducción y en este caso concreto solo el fenómeno lingüístico de las propuestas de traducción de los verbos con terminación *-ing* o gerundio en inglés.

Características de un corpus

De acuerdo con Reppen (2010, p. 31), para construir un corpus no existen parámetros específicos predeterminados, no obstante, refiere que se tiene que

corroborar que no exista ya un corpus como el que se quisiera construir, y que va a depender del tipo de preguntas en la investigación para poder definir los elementos que han de conformar el corpus que se desea. Menciona también que es importante la representatividad, aunque no necesariamente se refiere a una cantidad muy grande de datos, pues por eso el diseño recae sobre las necesidades de uso del corpus. Doval (2017, pp. 720-723), por su parte, establece los siguientes parámetros para la compilación de un corpus paralelo bilingüe multifuncional:

1. Identificación de necesidades y objetivos: se deben definir los propósitos del corpus, que pueden incluir investigación en lingüística contrastiva, traducción y enseñanza de lenguas extranjeras; e identificar las necesidades específicas de los diferentes usuarios potenciales
2. Diseño y planificación: se debe establecer una tipología de los corpus bilingües, diferenciando entre comparables y paralelos. Se debe decidir sobre la bidireccionalidad de las traducciones y la variedad de textos a incluir y planificar la distribución aproximada de los textos originales y las traducciones en el corpus.
3. Compilación de textos: se deben seleccionar textos narrativos escritos después de 1960 con un predominio de obras de las dos últimas décadas. Además, asegurar la calidad del material mediante la selección de textos publicados por editoriales reconocidas.
4. Preprocesado de los datos: se deben digitalizar los textos seleccionados y prepararlos para la alineación. Además, se deben eliminar textos no correspondientes, caracteres erróneos, imágenes y revisar los textos para corregir errores de digitalización.
5. Alineación oracional: se deben segmentar los textos en unidades de alineación, generalmente oraciones y utilizar herramientas de alineación, como *LF-Aligner*, para emparejar oraciones en los textos fuente y meta. Además, se debe realizar una revisión manual para validar la alineación y asegurar una tasa de error inferior al 0.5 %.
6. Anotación Lingüística: se deben de separar los textos por lenguas, para proceder a su anotación lingüística, implementar la lematización y el etiquetado en clases de palabras (*PoS-tagging*) utilizando herramientas como *FreeLing* para el español.
7. Metadatos y marcado textual: se deben adjuntar los metadatos individuales a los archivos de texto que incluyan información bibliográfica y estadística básicas, al igual que marcar las divisiones internas de los textos como partes de capítulos, para facilitar la localización de cadenas de texto.
8. Presentación y explotación: se deben indexar y gestionar los textos anotados con plataformas de búsqueda simples y avanzadas, así como la exportación de resultados.

9. Trabajo futuro y mejoras: se debe de continuar añadiendo nuevas obras para ampliar la variedad de géneros textuales e implementar búsquedas por etiquetas y alineación de palabras, además, se debe explorar el uso de otras herramientas de búsqueda y consulta específicas para corpus con anotaciones lingüísticas.
10. Respeto a los derechos de autor: se deben adoptar medidas y aplicar restricciones de uso, tales como: limitar la visualización de textos a contextos específicos y restringir el uso a fines académicos y científicos.

Estos pasos reflejan un proceso complejo y detallado que requiere una planificación cuidadosa y un trabajo meticuloso, para asegurar que el corpus resultante sea un recurso valioso y multifuncional para la investigación y la práctica en diversos campos relacionados con el lenguaje. Es muy importante referir que, lo anterior es un ejemplo de cómo compilar un corpus, no obstante, como ya se mencionó, lo que se requiera para establecer el diseño del corpus dependerá de las necesidades particulares de uso o investigación.

Concordancia paralela y ambigüedades en traducciones automáticas

Sketch Engine es una herramienta de análisis textual de pago que permite consultar y crear corpus (monolingües y paralelos) (Leiva, 2018, p. 68). Dicha herramienta presenta varias funciones, como lo son: relaciones gramaticales, diferencia de uso entre palabras, creación de un tesauro automático, ejemplos del uso en contexto, listas de palabras, n-gramas, palabras clave, creación de un diccionario automático, concordancia paralela y tendencias (Arias, Fernández Pampillón Cesteros, Samy y Arús Hita, 2019, pp. 14-45) que pueden ser de mucha utilidad para un traductor, sobre todo al establecer comparativos en el bitexto, para poder verificar las propuestas de traducción de palabras específicas, términos, frases, etcétera. Dentro de *Sketch Engine* existe la función de concordancia paralela, la cual permite ver en el bitexto la propuesta de traducción de las categorías gramaticales que sean de interés, o bien, las unidades de traducción que se quieran comparar, para analizar la concordancia del texto y tratar de buscar ambigüedades.

Guzmán (2008, p. 1) en lugar de referirse a errores, los denomina ambigüedades, pues se trata de pretraducciones realizadas por programas de traducción automática, dichas ambigüedades son: "palabras, frases y oraciones que pierden o tienen más de un significado y que necesitan ser esclarecidas, o bien, desambiguadas". Según este mismo autor (2008, pp. 3-7), existen tres tendencias al respecto de las ambigüedades en textos que se traducen de manera automática del inglés al español y que clasificó de la siguiente manera:

1. Traducción errónea de palabras que terminan en *-ing* (eng-sp):
   a. En inglés funge como sustantivo, mientras que en español tiene una construcción de: *ser*/*estar* verbo *-ando*/-iendo
      - Gerundios en títulos o encabezados

2. Traducción errónea de cláusulas subordinadas (*that, then, while* tienen más de un significado o la conjunción se pierde en la oración, lo cual resulta en ambigüedad).
   a. Cláusulas temporales:
      There was an error downloading the update.
      Hubo un error ***que descarbaga.***
      Hubo un error mientras ***se*** *descarg****aba.*** *(*pronombre *se)*
   b. Cláusulas relativas subordinadas (*that*):
      Installing computers that run windows.
      Instalación de equipos ***ese ejecutados*** *Windows.*
      Instalación de equipos ***que ejecutan*** *Windows.*
   c. Cláusulas subordinadas condicionales:
      Then '*luego*/*después*/*a continuación*', *entonces* '*in that case*'
      If you want to print a file, then you need to click…
      Si usted quiere imprimir un documento ***después es necesario*** *hacer clic…*
      Si usted quiere imprimir un documento ***entonces es necesario*** *hacer clic…*

3. Errores de traducción de verbos polisémicos
   Check that *your computer…*
   Active que *su equipo…*
   Compruebe **que**…

Para la presente investigación solo se analizó la primera ambigüedad, que corresponde a los verbos en gerundio (*-ing*) del inglés. Posteriormente nos centramos en analizar aquellos que se presentan como sustantivos, pues resulta interesante su propuesta de traducción.

Metodología

Los autores Teubert (2005) y Hardie y McEnery (2010) realizaron un desglose histórico de lo que hoy se ha conformado como la lingüística de corpus, en el cual se aprecia que existe un debate en torno a si esta se considera una metodología, o bien, una teoría. Se destaca que los lingüistas de corpus se consideran primero

metodólogos y la lingüística de corpus es primordialmente una metodología. Sin embargo, existe la ICAME (International Computer Archive of Modern and Medieval English), la cual es una organización que representa a la lingüística de corpus y que tiene representantes de ambas perspectivas, tanto la metodológica, como quienes la perciben como una teoría.

Para la lingüística de corpus, la lengua es el centro de una cultura, esta implica una descripción de su sistema de reglas, uso y significado. Por lo que Teubert (2005, p. 13) la define desde su perspectiva personal como "un esquema conceptual para estudiar la transmisión de contenido de una comunidad discursiva como se evidencia en la intertextualidad del discurso". En consecuencia, para esta investigación se adopta la postura de Teubert para poder tomar a la lingüística de corpus como una metodología, esto debido a que la aplicación se realiza a la traducción de textos técnicos del inglés, el cual es un contexto particular y que nos sirve de análisis para explorar la concordancia paralela.

A su vez, con base en la naturaleza de los datos y su indagación, la presente es una investigación mixta. Por un lado, es cualitativa, ya que el proceso de indagación es flexible y se mueve entre los eventos y su interpretación, las respuestas y el desarrollo de la teoría. Su propósito consiste en "reconstruir" la realidad tal como la observan los actores de un sistema social definido previamente. A menudo se llama "holístico" porque consideraa el todo, sin reducirlo al estudio de sus partes (Hernández, Fernández y Baptista, 2014, p. 15). Por otro lado, también se trata de una investigación cuantitativa, pues se extrae la cantidad de verbos (*-ing*) traducidos al español y se marca el porcentaje de tendencia y se define así ya que se examina la causa y cómo esta influye en los resultados (Rojo, 2013, p. 127).

Construcción del corpus

En primera instancia, se decidió trabajar con un corpus *ad hoc* conformado solamente por ocho textos técnicos en inglés>español, del sector del marketin. Su temática versa sobre la salud de mascotas. La extensión del corpus es aparentemente corta (4,909 palabras), pues solamente se quería explorar el bitexto con la función de concordancia paralela.

Para la recopilación de los textos, se requirió que pertenecieran al campo de la mercadotecnia y sobre el tema salud de mascotas. Esto debido a que se desarrolla en el ámbito de los textos técnicos de baja a mediana complejidad para su traducción (Gamero, 2001, como se citó en Terán Rueda, 2016, pp. 18-19). Se extrajeron cuatro textos en inglés de la página de *barksandwhiskers.com*, sobre la temática establecida, así como sus versiones -cuatro textos más- traducidas al

español y que aparecen en la página con su versión hispana. Todos los textos se publicaron en enero de 2023.

Para la construcción del corpus se extrajo el texto en formato plano .txt, tanto en inglés como en español, dicho proceso se realizó manualmente. Después, se conformó un archivo en formato Excel, esto para conformar un corpus paralelo bilingüe alineado, de igual manera, la alineación se llevó a cabo manualmente, para evitar posibles errores al respecto de las concordancias de las unidades de traducción en los textos.

Para la extracción de la información se recurrió a la técnica de *web scrapping*, la cual es la extracción de las partes específicas de la información que resultan trascendentales para esta investigación (Nieves y Molina, 2023, p. 47). Debido a esto, se obtuvo el título, el resumen, el desarrollo y las conclusiones de cada artículo o texto que conformó el corpus.

Procesamiento del corpus

Una vez procesados los textos en formato Excel, se procedió a utilizar la herramienta de *Sketch Engine*, para el análisis en esta propuesta. Se recurrió a cargar el corpus paralelo bilingüe y obtener sus características (Figura 1).

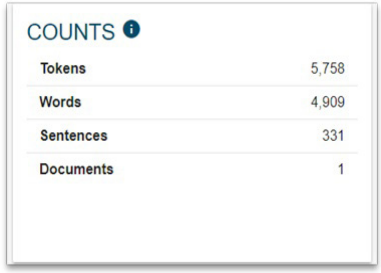


Figura 1. Resumen de tokens, words y sentences de Sketch Engine.

Debido a que con *Sketch Engine* no se permite, en el caso de corpus paralelos, cargar archivos individuales (Leiva, 2018, p. 68), en documentos solamente aparece uno, aunque en realidad se trata de ocho textos, cuatro en inglés y cuatro en español.

Para poder buscar la traducción que nos interesa en los textos del corpus, se optó por utilizar la función de *Parallel Concordance* (Figura 2), a su vez, en la parte de búsqueda se utilizó el regex[2] *-ing* en inglés, lo que arrojó la herramienta

[2] RegEx es la abreviatura de *Regular Expressions*, y esta es una cadena de texto genérica, que se usa a modo de patrón y que sirve para localizar trozos de texto dentro de otro texto mayor.

fue lo siguiente, no se acotó en un inicio a buscar solamente verbos, sino palabras, por eso se usa "lemma" en la herramienta.

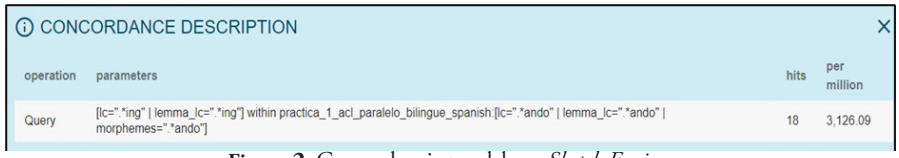


Figura 2. Concordancia paralela en *Sketch Engine*.

Análisis de resultados

Para poder interpretar el análisis del corpus, es importante resaltar que existen diferentes funciones que se puede activar dentro de *Sketch Engine* y que bien pueden apoyar a analizar un corpus desde diferentes ángulos y con base en las necesidades de investigación que se tengan. Por lo que, otra manera de exportar el resultado de concordancias paralelas se presenta en la Figura 3.

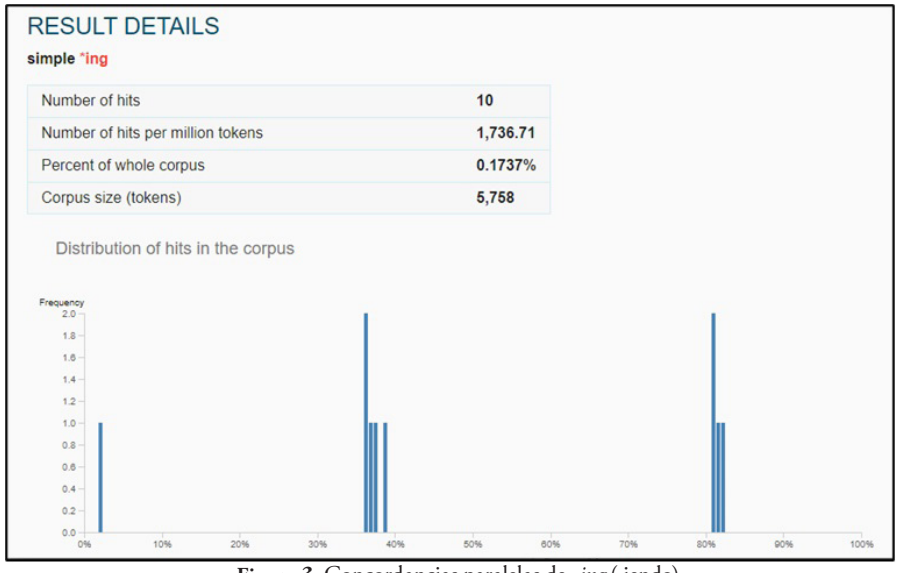


Figura 3. Concordancias paralelas de *-ing* (-iendo).

Esta figura a manera de gráfica representa el resultado del estado del corpus *ad hoc*, en el cual se detalla que se buscaron las palabras que terminan en *-ing* y que se localizaron 10 ejemplos (*hints*) dentro del corpus. La herramienta compara los ejemplos utilizados por un millón de tokens, el porcentaje de ejemplos totales encontrados dentro de este corpus, el cual es de .17 %, además del tamaño total del corpus *ad hoc*, el cual es de 5,758 tokens. Lo anterior, para el manejo convencional

de un corpus representa una cantidad pequeña, no obstante, referimos que se utilizó esta cantidad para poder hacer el ejercicio de la función de concordancia paralela. Es importante señalar que los 10 ejemplos se destilaron pues reunían la característica de terminar en *-ing*, ser verbos en gerundio y se corroboraron sus propuestas de traducción al español.

Por otro lado, en la Figura 3, en donde aparece una gráfica de barras se tiene que la frecuencia de las palabras utilizadas *-ing* es de 1.0 y 2.0 (se marca en el eje X de la gráfica). Además, se especifica el porcentaje de uso de dichas frecuencias (aparece en el eje Y de la gráfica) pues la herramienta encontró acepciones del gerundio en español con terminación: -ndo, aunque no todos los ejemplos encontrados se tradujeron con esa tendencia en la versión final de los textos al español. Aunado a esto, este tipo de visualización nos permite ver el porcentaje del fenómeno que buscamos (.17 %) que, si bien es poco, resulta funcional para el objetivo de la presente investigación debido a la extensión del corpus *ad hoc* conformado.

Para exportar las concordancias paralelas, la herramienta ofrecía los formatos en .txt o PDF. Se optó por exportarlos en .txt, para poder darle forma a las figuras, depurar los datos que nos interesan y, así, realizar el análisis detallado; se exportaron en PDF para observar las concordancias sin que se muevan o se altere el bitexto.

Se puede observar en la Figura 4 que la herramienta de concordancia paralela ha extraído el regex: [1c= ". *ing | lemma_1c= ".*ing], por lo que se ha acotado a la herramienta a buscar esa expresión en los ocho textos que conforman el corpus. Se puede observar entonces que, de manera subsecuente, se procesaron las oraciones en el texto que contienen esas expresiones, por lo que *Sketch Engine* procede a realizar el cálculo de la concordancia paralela.

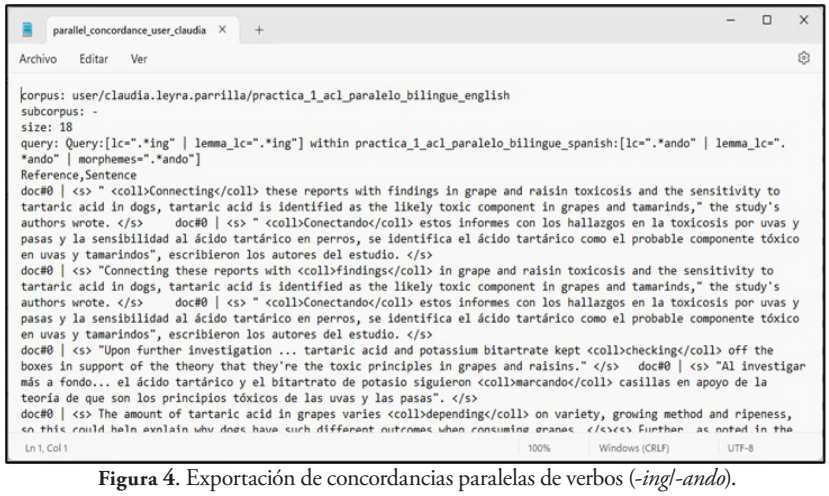
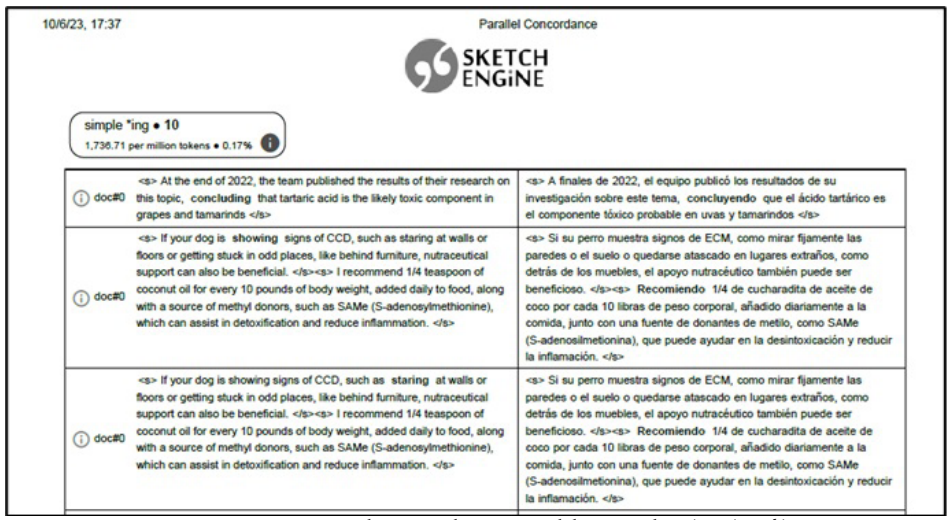


Figura 4. Exportación de concordancias paralelas de verbos (*-ing*/*-ando*).

La interfaz nos indica el nombre del corpus *ad hoc*: práctica_1_ac1_paralelo_bilingue_english, no existe un subcorpus que se haya ligado a este, por lo que el espacio aparece vacío. En el tamaño indica que encontró 18 coincidencias del regex buscado lemma_1c=".ing en coincidencia con [morphemes=".ando .endo] dentro del corpus indicado. Posteriormente, la herramienta enlista las oraciones en donde se encuentran los ejemplos (*hints*) localizados dentro del corpus, por lo que enlista en número de documento (aparece: 0, pues es un solo bitexto compilado ya que juntamos los 4 textos en inglés y los cuatro en español en la herramienta) y arroja el texto codificado a manera de lenguaje natural programado.

Ahora bien, en la Figura 5 se tiene ya activada la función de concordancia paralela, que nos brinda un documento a manera de bitexto en formato PDF, para no mover las columnas. Se puede observar nuevamente en la Figura 5 que la herramienta extrajo el .17 % del total de los tokens que concuerdan con el regex buscado; por lo que extrae los fragmentos del corpus en donde se localizan las expresiones con *-ing* en formato de bitexto, es decir, en inglés y español.


10/6/23, 17:37 Parallel Concordance

SKETCH ENGINE

simple *ing • 10
1,736.71 per million tokens • 0.17%

doc#0	<s> At the end of 2022, the team published the results of their research on this topic, **concluding** that tartaric acid is the likely toxic component in grapes and tamarinds </s>	<s> A finales de 2022, el equipo publicó los resultados de su investigación sobre este tema, **concluyendo** que el ácido tartárico es el componente tóxico probable en uvas y tamarindos </s>
doc#0	<s> If your dog is **showing** signs of CCD, such as staring at walls or floors or getting stuck in odd places, like behind furniture, nutraceutical support can also be beneficial. </s><s> I recommend 1/4 teaspoon of coconut oil for every 10 pounds of body weight, added daily to food, along with a source of methyl donors, such as SAMe (S-adenosylmethionine), which can assist in detoxification and reduce inflammation. </s>	<s> Si su perro muestra signos de ECM, como mirar fijamente las paredes o el suelo o quedarse atascado en lugares extraños, como detrás de los muebles, el apoyo nutracéutico también puede ser beneficioso. </s><s> **Recomiendo** 1/4 de cucharadita de aceite de coco por cada 10 libras de peso corporal, añadido diariamente a la comida, junto con una fuente de donantes de metilo, como SAMe (S-adenosilmetionina), que puede ayudar en la desintoxicación y reducir la inflamación. </s>
doc#0	<s> If your dog is showing signs of CCD, such as **staring** at walls or floors or getting stuck in odd places, like behind furniture, nutraceutical support can also be beneficial. </s><s> I recommend 1/4 teaspoon of coconut oil for every 10 pounds of body weight, added daily to food, along with a source of methyl donors, such as SAMe (S-adenosylmethionine), which can assist in detoxification and reduce inflammation. </s>	<s> Si su perro muestra signos de ECM, como mirar fijamente las paredes o el suelo o quedarse atascado en lugares extraños, como detrás de los muebles, el apoyo nutracéutico también puede ser beneficioso. </s><s> **Recomiendo** 1/4 de cucharadita de aceite de coco por cada 10 libras de peso corporal, añadido diariamente a la comida, junto con una fuente de donantes de metilo, como SAMe (S-adenosilmetionina), que puede ayudar en la desintoxicación y reducir la inflamación. </s>

Figura 5. Exportación de concordancias paralelas en verbos (*-ing*/*-endo*).

Al analizar esta figura resultó muy interesante el uso de esta herramienta, pues, por ejemplo, en la primera línea aparece lo siguiente (Cuadro 1):

Cuadro 1. Cuadro de análisis del texto origen y el texto traducido con verbo en inglés y español.

TO	TT
…, conclud**ing** that tartaric acid is the likely toxic component in grapes and tamarinds.	, concluy**endo** que el ácido tartárico es el componente tóxico posible en uvas y tamarindos.

En este caso, la traducción de *concluding* se realizó como 'concluyendo'. Quien revisó la versión final del texto optó por utilizar el gerundio, inclinación interesante, puesto que pudo haberse adoptado utilizar: 'se concluye' a manera de reflexivo, para evitar calcar la estructura literal del inglés; sugerencia de procedimiento de traducción que brindan López y Minett (2006). Por consiguiente, en la siguiente línea se observa (Cuadro 2):

Cuadro 2. Análisis de verbos *staring*/*getting* y su traducción al español.

TO	TT
…such as star**ing** at walls or floors or gett**ing** stuck in odd places…	…como **mirar** fijamente las paredes o el suelo o **quedarse** atascado en lugares extraños…

Aquí se puede ver que se tienen dos verbos en inglés que terminan en *-ing* y la propuesta de traducción al español tiende en este caso a no calcar de manera literal a dichos verbos, por lo que, *staring* se traduce como 'mirar', mientras que *getting* como 'quedarse'. Es debido a esta observación que se adoptó como postura de la presente investigación centrarnos en analizar la concordancia paralela solamente de los verbos *-ing* del inglés, para poder entender y contrastar la inclinación al realizar su traducción. Los datos se limpiaron manualmente.

De acuerdo con la herramienta, se obtuvieron siete propuestas de traducción de verbos que terminan con *-ing* (*-ando*), dado que la instrucción que se le dio a *Sketch Engine* se centró en buscar el lemma, la herramienta marca como "*18 hints*" en total, aunque se resalta que se depuraron y al final solamente quedaron 14 (Cuadro 3). Para *-ing*/*-ando* se tienen los siguientes:

Cuadro 3. Concordancias paralelas exportadas de *Sketch Engine* (*-ing*/ando).

No.	English	Spanish
1	Connetc**ing** with these reports	Conect**ando** estos informes con
2	Kept check**ing** off the boxes	Siguieron marc**ando** casillas
3	Depend**ing** on variety, grow**ing** method, and ripeness	**Varía** en función de la variedad, el método de cultivo y madurez,
4	So have fun experiment**ing** to see which one your pet likes best	Así que diviértete experiment**ando** para ver cuál le gusta más a tu mascota
5	Avoid**ing** the skin and core.	Evit**ando** la piel y el corazón.
6	so many websites incorrectly list foods as be**ing** "toxic" that aren't.	Porque muchos de los sitios web **presentan** incorrectamente como "tóxicos" alimentos que no lo son.
7	the odds of develop**ing** CCD dropped to 52% for each additional year of life.	Los problemas de salud y los niveles de actividad, las probabilidades de **desarrollar** CCD descendieron al 52 % por cada año adicional de vida.

Como se observa en el Cuadro 3, para la primera línea en donde el verbo aparece en posición inicial, existe un error de traducción, de acuerdo con las sugerencias de errores de traducción a evitar de Guzmán (2007, pp. 3-7). El autor indica que si el verbo se ubica en posición inicial puede traducirse en infinitivo o con reflexivo -se, en lugar de *connecting* por 'conectando', puede ser: 'conectar' o 'se conectan'. Otra sugerencia del autor puede ser utilizar una transposición a un sustantivo 'la conexión', o bien, utilizar la palabra *cómo* más el verbo, para mejorar la estructura sintáctica del español. De tal manera que, la propuesta para la línea 1 sería: 'Cómo relacionar estos informes con'.

Para la segunda línea se tiene nuevamente que la propuesta de traducción es literal (*checking*, 'marcando'). Si seguimos las recomendaciones, podría optarse por traducirlo como: 'se marcan' con reflexivo, o si nos apegamos a Guzmán (2007) podría quedar: 'Cómo marcar las casillas'. Por lo que la tendencia es calcar la estructura también aparece en esta línea.

En la línea tres se observa que se realiza una buena propuesta para omitir *-ando* en español. Se optó por utilizar un verbo transitivo "varía", el cual hace más natural la traducción al español. Para la línea cuatro, nuevamente se tiene la inclinación a la tendencia de uso del gerundio en español. Se traduce *experimenting* por 'experimentando', que bien podría ser: 'al experimentar' y así dar naturalidad en español. Ocurre exactamente lo mismo con la línea cinco (*avoiding*, 'evitando'), en la que también se adopta la tendencia en el uso del gerundio. En este ejemplo, se podría adoptar otra forma en español como 'evitar', 'evite'.

Para la línea seis, la propuesta es un cambio de la forma del verbo del inglés al español *being* por 'presentan'. En este caso hubo una adaptación al español del verbo en presente. Para la línea siete la propuesta de traducción se apega a la sugerencia de uso del infinitivo, pues *developping* queda como 'desarrollar'.

Ahora, para la búsqueda *-ing*/*-endo* se obtuvieron ocho ejemplos (*hints*). Por lo que el Cuadro 4 queda de la siguiente manera:

Cuadro 4. Concordancias paralelas exportadas de Sketch Engine (*ing/-iendo y otras propuestas arrojadas).

No.	English	Spanish
1	**Concluding** that tartaric acid is the likely toxic component in grapes and tamarinds	**concluyendo** que el ácido tartárico es el componente tóxico probable en uvas y tamarindos
2	If your dog is showing signs of CCD, such as **staring** at walls	Si su perro muestra signos de ECM, como **mirar** fijamente las paredes o el suelo
3	or floors or **getting** stuck in odd places,	o **quedarse** atascado en lugares extraños,
4	such as **staring** at walls or floors	como **mirar** fijamente las paredes o el suelo

No.	English	Spanish
5	or **getting** stuck in odd places, like behind furniture,	o **quedarse** atascado en lugares extraños, como detrás de los muebles,
6	it's possible you can stave off dementia altogether just by **staying** active together.	es posible que puedas evitar la demencia simplemente **manteniéndote** activo.
7	**Walking** is great lifelong exercise for both you and your older canine companion.	**Pasear** es un ejercicio estupendo para toda la vida, tanto para usted como para su compañero canino mayor.
8	and while seniors might not be able to do much power **walking**,	aunque las personas mayores no sean capaces de **caminar** mucho,

En el primer ejemplo se vuelve a observar la tendencia de utilizar el gerundio en la traducción al español. *Concluding* se traduce como 'concluyendo', que bien podría haberse traducido por: 'se concluye'. La línea dos es interesante, se tiene que para el verbo *staring* la traducción es 'mirar' en infinitivo, mientras que para *getting* es 'quedarse', por lo que ambas propuestas se apegan a darle naturalidad al texto en español. A su vez, la línea tres presenta un ejemplo similar al de la línea dos y se inclina por traducir *staring* por 'mirar'.

Por otro lado, en la línea cuatro se observa que la traducción también es parecida a la de la línea dos. En esta se opta por traducir *getting* como 'quedarse' utilizando el verbo con un reflexivo. Para la línea cinco, se mantiene la inclinación de conservar la estructura en gerundio en español; se optó por traducir *staying* por 'manteniéndote'. Parecería que no tiene la estructura de gerundio, pero sí, aunque incorpora el pronombre *te* aglutinado en el verbo.

A partir de los datos anteriores es posible observar cuatro patrones de traducción en los verbos que terminan en *-ing* en el corpus *ad hoc* que se analizó, los cuales se presentan a continuación en el Cuadro 5.

Cuadro 5. Resumen de patrones de traducción en los verbos que terminan en *-ing*.

Patrón de traducción	English	Spanish	Desambiguación
Verbo en posición inicial	**Concluding** that tartaric acid is the likely toxic component in grapes and tamarinds	**concluyendo** que el ácido tartárico es el componente tóxico probable en uvas y tamarindos	**Se concluye** que el ácido tartárico es el componente tóxico…
Verbo con preposición	the odds **of developing** CCD dropped to 52% for each additional year of life.	las probabilidades **de desarrollar** CCD descendieron al 52 % por cada año adicional de vida.	
Dos verbos juntos	or **getting stuck** in odd places, like behind furniture,	o **quedarse atascado** en lugares extraños, como detrás de los muebles,	O **atascarse** en lugares extraños, como detrás de los muebles,
Perifrasis verbal (*to be*)	If your dog is **showing** signs of CCD,	Si su perro **muestra** signos de ECM,	Si tu perro **no deja de mostrar** signos de ECM,

Se observa, además, una columna que puede fungir como estrategia de posedición para la desambiguación de las propuestas de traducción observadas. Solamente en la categoría Verbo con preposición no se presenta, porque la tendencia es siempre la de utilizar la preposición. En el caso de la categoría "dos verbos juntos", López y Minett (2006) sugieren utilizar el pronombre reflexivo *-se*, mientras sea posible. La propuesta para las perífrasis verbales que ofrece López y Minett (2006) es utilizar "no parar de+ infinitivo del verbo" y de esa manera se apega más a la sintaxis del español.

Se puede establecer que la tendencia de traducción de los verbos *-ing* del inglés al español en esta tipología de textos y en este corpus *ad hoc*, mostrados en los Cuadros 3 y 4 queda de la siguiente manera: verbos en inglés que terminan en *-ing* y que tienden a traducirse con terminación *-ando* (ver Cuadro 6) y verbos en inglés que terminan en *-ing* y que tienden a traducirse con terminación *-endo* (ver Cuadro 7):

Cuadro 6. Verbos *-ING -ANDO*.

No.	Verbo *-ING*	Propuesta de traducción
1	connecting	conectando
2	checking	marcando
3	depending	**varía**
4	experimenting	experimentando
5	avoiding	evitando
6	being	**presentan**
7	developing	desarrollar

Cuadro 7. Verbos *-ING -ENDO*.

No.	Verbo *-ING*	Propuesta de traducción
1	concluding	concluyendo
2	staring	**mirar**
3	getting	**quedarse**
4	staring	**mirar**
5	getting	**quedarse**
6	staying	manteniéndote
7	walking	**pasear**
8	walking	**caminar**

En el Cuadro 6 se observa que, de siete ejemplos, cuatro se tradujeron con terminación *-ando*, mientras que tres presentan mejores propuestas al español, de acuerdo con lo que sugieren López y Minett (2006) y Guzmán (2008); además de que se apegan a mejorar la calidad del texto poseditado. Por su parte, en el Cuadro 7, dos se tradujeron utilizando gerundio y el resto tuvo una mejor propuesta.

Luego de todo lo anterior, se intuye entonces que, al realizar un conteo de la tendencia en el contraste de la traducción de los verbos *-ing* del inglés, se tiene una mayor desambiguación de estos al español, por lo que en esta tipología de textos se podría deber a la labor de revisión a nivel posedición de los textos que se traducen (ver Tabla 1.)

Tabla 1. Porcentajes de traducción de verbos *-ing*.

Cuadro	Traducción de verbos con gerundio	Propuestas de traducción desambiguadas
6	4 (*-ando*)	3
7	2 (*-endo*)	6
Total:	6 (40 %)	9 (60 %)

Conclusiones

La lingüística de corpus es una metodología que, en conjunto con la traducción, puede resultar de gran utilidad para la investigación en traducción, ya que puede contribuir a refinar procedimientos de traducción, entender terminología, fraseología y la construcción sintáctica de los textos paralelos para mejorar sus versiones traducidas y aumentar su calidad de manera automática.

La construcción de un corpus *ad hoc* paralelo bilingüe permite destilar textos para necesidades específicas, que si bien pueden no tener una extensión demasiado amplia, al menos alcanzan para comprender la tipología textual del texto técnico que se trabaje en ese momento, hacer un análisis exprés, y así, poder comprender el comportamiento del texto y trabajar de manera específica sobre los puntos corales de este, por lo que beneficia la labor de posedición del texto pretraducido con traducción automática y su revisión.

Valdría la pena en una investigación futura ampliar la búsqueda para indagar todas las categorías que propone Guzmán (2008), y así ver su comportamiento en esta tipología de texto que se utilizó en la presente investigación.

La construcción manual del archivo plano .txt es laboriosa, ya que se tienen que limpiar los textos de etiquetas e imágenes al ser estos provenientes de páginas web, por ello, la estrategia de *web scrapping* propuesta por Nives y Molina (2023) es ideal para trabajar este tipo de textos. Además, la realización de la alineación manual permite ver aspectos en el bitexto que posiblemente un procedimiento automático no permitiría. No obstante, esta técnica de trabajo manual no es recomendable para un corpus de mayor extensión, pues no resultaría útil y viable, si se busca contrastar el corpus para trabajar en la traducción. El análisis de concordancias paralelas permite corroborar a manera de tablas, las cuales se extrajeron, las propuestas de traducción de los verbos *-ing* que se presentaron en inglés y que se

tradujeron al español, no siempre con la terminación -ndo, la cual sería y fue una ambigüedad, de acuerdo con lo propuesto por López y Minett (2006, p. 129-140), quienes recopilan los procesos de traducción de: equivalencia, modulación, transposición y demás, para respetar la naturalidad del texto en español.

A través del presente análisis de datos se cumplió el objetivo principal de la investigación, que fue analizar los errores de traducción en verbos en gerundio (*-ing*) en inglés que se buscaron específicamente a través del corpus *ad hoc* paralelo bilingüe construido para ello. Se contrastaron los errores y se pudo determinar a través del porcentaje del conteo de propuestas de traducción que el texto presentaba una desambiguación (véase figura 2) de 9 de los 15 verbos poseditados (60 %), lo cual podría no representar una calidad alta para la posedición de un texto, pero sí representa una marca del trabajo de los traductores sobre la posedición de los textos y seguramente ese porcentaje se debe al tiempo dedicado a la posedición de los textos designados para ello.

Al respecto de la pregunta de investigación, la manera de extraer una muestra de corpus es que posiblemente quien sea el encargado de revisar la calidad de los textos poseditados puede extraer muestras de los textos ya publicados al español y analizar un aspecto en específico para ir tomando muestras del conteo de la posedición que se realiza e ir midiendo si mejora o no la propuesta de traducción. Por otro lado, la hipótesis de la presente aludía a que, a través de la construcción de un corpus *ad hoc*, es posible contrastar aspectos en los textos en inglés que tienen tendencia al error en español en textos técnicos, lo cual, en efecto, ayuda al poseditor a verificar la calidad final del texto traducido.

Por último, la aplicación de la metodología mixta para revisar la calidad de la traducción a través de un análisis contrastivo de segmentos específicos (en este caso de verbos en gerundio y su traducción) permite establecer un sondeo para verificar aspectos específicos que se pueden mejorar en el proceso de posedición de textos técnicos del inglés al español.

Referencias

Ahmad, K., Holmes-Higgin, P. y Raza Abidi, S. (1994). "A description of texts in a corpus: 'Virtual' and 'real' corpora". En W. Martin, W. Mejis, M. Moerland; E. ten Pas, P. van Sterkenburg y P. Vossen (Eds.), E*URALEX 1994: Proceedings. Papers submitted to the 6th Euralex International Congress on Lexicography* (pp. 390-402). Ámsterdam: Vrije Universiteit.

Aston, G. (1999). "Corpus use and learning to translate". Textus, *12,* pp. 289-314. https://godzilla.sslmit.unibo.it/~guy/textus.htm

Arias, I., Fernández Pampillón Cesteros, A., Samy, D. y Arús Hita, J. (2019). *Taller sobre herramientas de análisis textual: La herramienta Sketch Engine*. Madrid: Universidad Complutense de Madrid. https://biblioteca.ulpgc.es/files/repositorio_de_docum152/noticias/taller_de_sketch_engine_18_02_2019_0.pdf

Baker, M. (1993). "Corpus linguistics and translation studies: implications and applications". En M. Barker, G. Fracis y E. Tognini-Bonelli (Eds.), *Text and technology: In honour of John Sinclair (*pp. 233-250). Amsterdam: John Benjamins.

Baker, M. (1995). "Corpora in translation studies: An overview and some suggestions for future research". *Target. International Journal of Translation Studies, 7*(2), pp. 223-243. https://doi.org/10.1075/target.7.2

Bowker, L. y Pearson, J. (2002). *Working with specialized language: A practical guide to using corpora.* London: Routledge.

Castillo, C. (2009). "La elaboración de un corpus ad hoc paralelo multilingüe". *Revista tradumática. Traducció i Tecnologies de la información i la Comunicació,* (7), pp. 1-11. http://www.fti.uab.cat/tradumatica/revista

Corpas, G. (2003). "Turicor: Compilación de un corpus de contratos turísticos (alemán, español, inglés, italiano) para la generación textual multilingüe y la traducción jurídica". *Panorama actual de la investigación en traducción e interpretació*n, *2*, pp. 373-384.

Doval, I. (2017). "La construcción de un corpus paralelo bilingüe multifuncional". *Moenia, 23*, pp. 717-734. https://revistas.usc.gal/index.php/moenia/article/view/4071/5556

Expert Advisory Group on Language Engineering Standards (1996). *Text corpora Working Group reading Guide. EAGLES Document EAG-TCWG-FR-2*. https://www.ilc.cnr.it/EAGLES96/corpintr/corpintr.html

Fernández, L. (2018). *La profesión del traductor en México*. Ciudad de México: Universidad Intercontinental. https://www.uic.mx/wp-content/blogs.dir/8/files/2020/09/la_profesion_del_traductor_en_mexico.pdf

Gaya, C. (2023). "¿Qué son las expresiones regulares? (RegEx)". *ViewNext. An IBM Subsidiary*. https://www.viewnext.com/que-son-las-expresiones-regulares-regex/

Guzmán, R. (2007). "Automatic MT post-editing using regular expressions". *Multilingual*, *90*(18), pp. 49-52.

Guzmán, R. (2008). "Advanced automatic MT post-editing". *Multilingual computing*, *19*(95), pp. 52-57. https://www.rafaelguzman.ie/wp-content/uploads/2022/04/advancedautomaticMTpost-editing.pdf

Hardie, A. y McEnery, T. (2010). "On two traditions in corpus linguistics, and what they have in common". *International Journal of Corpus Linguistics, 15*(3), pp. 384-394. https://doi.org/10.1075/ijcl.15.3.09har

Harris, B. (1988). "Bi-text, a new concept in translation theory". *Working papers on Bilingualism o language monthly*, *54*, pp. 8-10. https://www.academia.edu/34281085/Bitext_the_original_article_of_1988

Hernández, R., Fernández, C. y Baptista, M. (2010). *Metodología de la investigación*. McGraw Hill. https://www.icmujeres.gob.mx/wp-content/uploads/2020/05/Sampieri.Met.Inv.pdf

ITC Global (2023). La demanda de traductores está aumentando a través de ITC Translations. https://www.itcglobaltranslations.com/es/blog/la-demanda-de-traductores-esta-aumentando

Kennedy, G. (1998). *An Introduction to corpus linguistics.* London: Longman.

Kenning, M. (2010): "What Are Parallel and Comparable Corpora and How Can We Use Them". En A. O'Keeffe, M. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics* (pp. 487-500). London/New York: Routledge.

Kraif, O. (2002). "Translation alignment and lexical correspondences: A methodological reflection". En B. Altenberg y S. Granger (Eds.), *Lexis in contrast: Corpus-based approaches* (pp. 271-289). Amsterdam: John Benjamins. https://doi.org.10.1075/scl.7.19kra

Leiva, J. (2018). "Diseño y compilación de corpus paralelos alineados: dificultades y (algunas) soluciones en el ejemplo de un corpus de textos museísticos traducidos (inglés-español). *Revista de Lingüística y Lengua Aplicada*s, *13*, pp. 59-73. http://doi.org/10.4995/rlyla.2018.7912

López, J. y Minett, J. (2006). *Manual de traducción. Inglés*|*Castellano*. Barcelona: Gedisa Editorial.

Nieves, S. y Molina, J. (2022). "Hacia una extracción de metáforas conceptuales en un corpus de economía a partir del procesamiento de lenguaje natural". *Estudios de Lingüística Aplicada*, *41*(76), pp. 39-67. https://ela.enallt.unam.mx/index.php/ela/article/view/1027

O'Keeffe, A. y McCarthy, M. (Eds.). (2010). *The Routledge Handbook of Corpus Linguistics.* New York: Routledge Handbooks.

Olohan, M. (2004). *Introducing corpora in translation studies*. New York: Routledge.

Pearson, J. (1998). *Terms in Context, Studies in Corpus Linguistics*. Ámsterdam y Filadelfia: John Benjamins.

Pérez-Carrasco, M. y Seghiri, M. (2020). "Motocor: compilación de un corpus técnico paralelo bilingüe (inglés-español) y su aplicación en la traducción

directa de un texto sobre automoción". *Anales de Lingüística, 5*, pp. 123-148. http://revistas.uncu.edu.ar/ojs/index.php/analeslinguistica

Reppen, R. (2010). "Building a corpus. What are the key considerations?" En O'Keeffe, A. y McCarthy, M. (Eds.), *The Routledge handbook of corpus linguistics.* (2a ed., pp. 31-37). New York: Routledge Handbooks.

Rojo, A. (2018). *La investigación en traducción. Una revisión metodológica de la disciplina.* Barcelona: Anthropos.

Sánchez, A. (2010). "Traducción automática, corpus lingüísticos y desambiguación automática de los significados de las palabras". *Lengua, traducción, recepción*, *1*, pp. 555-587. http://hdl.handle.net/10612/4728

Sánchez-Gijón, P. (2003). "És la web pública la nova biblioteca del traductor?". Tradumàtica: *Traducció i tecnologies de la informació i la comunicación*, *2*, pp. 1-7. https://revistes.uab.cat/tradumatica/article/download/n2-sanchez-gijon/162-pdf-ca

Sánchez, M. (2019). "Corpus paralelos y traducción especializada: ejemplificación de diseño, compilación y alineación de un corpus paralelo bilingüe (inglés-español) para la traducción jurídica". *De Gruyter, 64*(2), pp. 269-285. https://doi.org/10.1515/les-2019-0015

Terán, P. (2016). *La traducción técnica y las herramientas de traducción* [Tesis de licenciatura]. Universidad de Valladolid. https://uvadoc.uva.es/bitstream/handle/10324/18869/TFG-O%20820.pdf?sequence=1&isAllowed=y

Teubert, W. (2005). "My version of corpus linguistics". *International Journal of Corpus Linguistics, 10*(1), pp. 1-13.

Varantola, K. (2000). "Translators, dictionaries and text corpora". En S. Bernardini y F. Zanettin (Eds.), *I corpora nella didattica della traduzione* (pp. 117-133). Bolonia: CLUEB.

Zanettin, F. (2002). "DIY Corpora: The WWW and the Translator". En B. Maia, J. Haller y M. Urlrych (Eds.), *Training the Language Services Provider for the New Millennium* (239-248). Oporto: Facultade de Letras, Universidade do Porto.