

Entrenamiento y evaluación automática y humana de un motor de traducción automática neuronal para la traducción de textos literarios del francés al español en el marco del proyecto MTUOC

Jon Mikel Peñagarikano Calvo

Trabajo Final de Máster
Tutora: Dra. Silvia Rodríguez Vázquez
Máster universitario en Traducción y Tecnologías
Universitat Oberta de Catalunya
Junio de 2023

Resumen

El uso de sistemas de traducción automática para traducir textos literarios ha suscitado recientemente un gran interés dados, por una parte, el potencial cada vez mayor de estos sistemas y, por otra parte, las características propias de los textos literarios, que *a priori* no parecen aptos para la traducción automática. En este trabajo, se crea un corpus paralelo alineado mediante el algoritmo SBERT para entrenar un motor de traducción automática neuronal especializado en la traducción de textos literarios del francés al español. El entrenamiento se lleva cabo empleando herramientas propias del proyecto MTUOC de la Universitat Oberta de Catalunya y una técnica de ponderación de oraciones para dar prioridad a los segmentos que provienen del corpus especializado, lo que supone una de las principales novedades de la investigación frente a trabajos pasados. Después, se realiza una evaluación tanto automática como humana del motor y se comparan los resultados con los de un motor entrenado previamente con un corpus especializado de menor tamaño. Pese a que las traducciones ofrecidas por el motor entrenado están lejos de alcanzar la calidad de una traducción humana, los resultados obtenidos en este estudio parecen indicar que el empleo de un corpus especializado de mayor tamaño y de una técnica de ponderación de oraciones en el proceso de entrenamiento influye de manera positiva en la calidad de las traducciones del motor.

Palabras clave: traducción automática neuronal, traducción literaria, corpus paralelo, entrenamiento de sistemas de traducción automática neuronal, ponderación de oraciones, evaluación automática, evaluación humana.

Abstract

The use of machine translation systems to translate literary texts has recently aroused great interest. This has been motivated by the increasing potential of these systems and because, theoretically, literary texts do not seem suitable for machine translation. In this paper, a parallel corpus aligned by means of the SBERT algorithm is created to train a neural machine translation engine specialized in the translation of literary texts from French into Spanish. This training is performed using tools from the MTUOC project of the Universitat Oberta de Catalunya and a sentence-weighting technique so that segments from the specialized corpus are prioritized. This represents one of the main changes

when compared to previous research. Afterwards, both automatic and human evaluation of the engine are carried out and the results are compared with those of a previously trained engine with a smaller specialized corpus. Although the translations provided by the trained engine are far from reaching the quality of a human translation, the results obtained in this study seem to indicate that the use of a larger specialized corpus and a sentence-weighting technique in the training process enhances the quality of the engine's output.

Keywords: neural machine translation, literary translation, parallel corpus, neural machine translation system training, sentence-weighting, automatic evaluation, human evaluation.

Résumé

L'emploi de systèmes de traduction automatique pour traduire des textes littéraires a récemment suscité un grand intérêt étant donné, d'une part, le potentiel croissant de ces systèmes et, d'autre part, les caractéristiques des textes littéraires qui, en principe, ne semblent pas aptes pour la traduction automatique. Dans cette étude, un corpus parallèle aligné est créé à l'aide de l'algorithme SBERT pour entraîner un moteur de traduction automatique neuronale spécialisé dans la traduction de textes littéraires du français vers l'espagnol. L'entraînement est réalisé au moyen d'outils du projet MTUOC de l'Universitat Oberta de Catalunya et d'une technique de pondération de phrases pour donner la priorité aux segments issus du corpus spécialisé, ce qui constitue l'une des principales nouveautés par rapport aux études antérieures. Ensuite, une évaluation automatique et humaine du moteur entraîné est exécutée et les résultats sont comparés à ceux d'un moteur entraîné auparavant avec un corpus spécialisé plus réduit. Bien que les traductions proposées par le moteur entraîné soient loin d'atteindre la qualité d'une traduction humaine, les résultats obtenus dans cette étude semblent indiquer que l'utilisation d'un corpus spécialisé plus grand et d'une technique de pondération de phrases dans le processus d'entraînement améliore la qualité des traductions du moteur.

Mots-clés : traduction automatique neuronale, traduction littéraire, corpus parallèle, entraînement de systèmes de traduction automatique neuronale, pondération de phrases, évaluation automatique, évaluation humaine.

Índice

Resumen	2
Abstract.....	2
Résumé	3
1. Introducción.....	6
1.1. Contextualización.....	6
1.2. Motivación.....	6
1.3. Objetivos.....	8
1.4. Preguntas de investigación, hipótesis y variables	8
1.5. Estructura del trabajo	9
2. Marco teórico.....	9
2.1. Breve historia de la traducción automática	9
2.2. Evaluación de la traducción automática.....	12
2.3. La posesición	15
2.4. La traducción automática neuronal y la traducción literaria.....	16
2.5. El proyecto MTUOC	20
3. Metodología.....	21
3.1. Elaboración del corpus	21
3.1.1. Segmentación y alineación del corpus	22
3.1.2. Ampliación del corpus	26
3.2. Preprocesamiento del corpus	28
3.3. Entrenamiento del motor de TA neuronal con Marian.....	30
3.4. Puesta en marcha del servidor de MTUOC y traducción.....	33
3.5. Evaluación automática de los motores	35
3.6. Elaboración de las pruebas de evaluación humana.....	35
3.6.1. Características de las pruebas del análisis comparativo.....	36
3.6.2. Características de las personas evaluadoras	38
4. Resultados.....	38
4.1. Resultados de la evaluación automática.....	38
4.2. Resultados de la evaluación humana	40
5. Conclusiones.....	44
5.1. Limitaciones del trabajo.....	47
5.2. Futuras vías de investigación	48
Bibliografía.....	50
Anexo I: Listado de obras del corpus literario	55
Anexo II: Declaración de consentimiento	60
Anexo III: Encuesta 1 (motor 1 vs. motor 2 sin ponderación de oraciones)	62
Anexo IV: Encuesta 2 (motor 2 con ponderación de oraciones vs. motor 2 sin ponderación de oraciones).....	68

Anexo v: Muestra de la evaluación automática del motor 1	74
Anexo vi: Muestra de la evaluación automática del motor 2 con ponderación de oraciones	76
Anexo vii: Muestra de la evaluación automática del motor 2 sin ponderación de oraciones	78
Anexo viii: Resultados detallados de la encuesta 1 (motor 1 vs. motor 2 sin ponderación de oraciones).....	80
Anexo ix: Resultados detallados de la encuesta 2 (motor 2 con ponderación de oraciones vs. motor 2 sin ponderación de oraciones).....	81

Índice de tablas

Tabla 1: Escalas para la evaluación de la precisión y la fluidez (Koehn, 2011).....	13
Tabla 2: Comparación de los corpus del motor 1 y del motor literario de Moreno (2022)	21
Tabla 3: Ejemplos de segmentos bien alineados con un índice de fiabilidad alto	24
Tabla 4: Ejemplos de segmentos mal alineados con un índice de fiabilidad bajo	25
Tabla 5: Tamaño de los corpus y resultados de la evaluación manual aleatoria	26
Tabla 6: Comparación de los corpus del motor 1 y del motor 2.....	28
Tabla 7: Resumen de los valores óptimos de cada métrica (Martínez, 2023).....	39
Tabla 8: Resultados de la evaluación automática	39
Tabla 9: Ejemplos en los que el motor 2 sin ponderación de oraciones supera al motor 1.....	41
Tabla 10: Ejemplos en los que el motor 1 supera al motor 2 sin ponderación de oraciones ...	41
Tabla 11: Ejemplos en los que el motor 2 con ponderación de oraciones supera al motor 2 sin ponderación de oraciones.....	43
Tabla 12: Ejemplos en los que el motor 2 sin ponderación de oraciones supera al motor 2 con ponderación de oraciones.....	44

Índice de figuras

Figura 1: Inicio del entrenamiento del motor TAN con Marian	31
Figura 2: Inicio del entrenamiento del motor TAN con Marian.....	32
Figura 3: Archivo <i>train-weights.txt</i>	32
Figura 4: Funcionamiento del servidor MTUOC (Oliver, 2022d).....	33
Figura 5: Prueba de traducción en MTUOC-Translator.....	34
Figura 6: Resultados de la prueba 1 de la evaluación humana.....	40
Figura 7: Resultados de la prueba 2 de la evaluación humana.....	41

1. Introducción

El presente trabajo de fin de máster consiste principalmente en el entrenamiento, mediante corpus paralelo, de un motor de traducción automática neuronal (TAN) para la traducción de textos literarios del francés al español y en su posterior evaluación tanto automática como humana. Para el entrenamiento del motor, se utilizan herramientas propias del proyecto MTUOC de la Universitat Oberta de Catalunya, diseñadas por el Dr. Antoni Oliver.

1.1. Contextualización

Este trabajo se realiza como proyecto final del máster universitario en Traducción y Tecnologías, y se enmarca en el área de tecnologías del lenguaje. Sigue la línea de trabajo propuesta por trabajos previos de esta misma universidad, en concreto, los trabajos *Creación de motores de traducción automática (estadística y neuronal) inglés-español especializados en el campo de la aviación con la herramienta MTUOC* de Rodríguez (2021), *La traducción automática en la literatura: creación de un corpus paralelo inglés-español para entrenar y evaluar un sistema neuronal en el marco del proyecto MTUOC* de Moreno (2022) y *Creación de un corpus paralelo y entrenamiento de un motor de traducción neuronal para la traducción de textos turísticos del castellano al alemán en el marco del proyecto MTUOC* de Martínez (2023). Trataremos de trabajar siguiendo las líneas de investigación propuestas en estos trabajos y de suplir algunas de las carencias que mencionan las autoras en sus conclusiones. En especial, tomaremos el trabajo de Moreno como referencia, pues también está centrado en la traducción de textos literarios, y procuraremos aportar nuevo conocimiento al ámbito introduciendo los siguientes cuatro factores: una nueva combinación lingüística (francés-español), un nuevo algoritmo para la alineación de corpus (que ya incluyó Martínez como novedad frente a los otros dos trabajos), una nueva técnica de ponderación de datos (*data-weighting*) a nivel de oración y una evaluación humana (que propuso Moreno como futura vía de investigación).

1.2. Motivación

La motivación inicial para realizar este trabajo surgió de las prácticas curriculares que realizamos en estos estudios de máster durante las cuales tuvimos el primer contacto con el entrenamiento de motores de TAN mediante la creación de corpus paralelos. En estas prácticas, tuvimos la ocasión de entrenar dos motores

de TAN para la traducción de textos literarios. Decidimos optar por este tipo de textos por el hecho de que en la traducción automática suele primar mantener el significado del texto original y, por lo general, no se suelen tener en cuenta cuestiones de estilo, mientras que en los textos literarios la función estética del lenguaje, el estilo y la creatividad son de suma importancia. Por este motivo, nos pareció interesante investigar sobre la traducción automática de este tipo de textos y estudiar los resultados, puesto que, tradicionalmente, estos textos no se consideraban apropiados para la traducción automática.

De este modo, en el marco de las prácticas curriculares, entrenamos dos motores de TAN bajo la tutela del Dr. Oliver: uno para la traducción de textos literarios del inglés al español y otro del francés al español. Para el entrenamiento del motor EN-ES creamos un corpus paralelo compuesto principalmente por obras del género policíaco o de ficción detectivesca de la serie Sherlock Holmes de Arthur Conan Doyle y por algunas obras de Agatha Christie. En el caso del motor FR-ES, optamos por el mismo género y utilizamos obras de la serie de Arsène Lupin de Maurice Leblanc.

Tras entrenar los motores, los evaluamos mediante métricas automáticas y, aunque no alcanzaban la calidad de otros motores de TAN como Google Translate o DeepL, los motores de traducción entrenados presentaban, a menudo, traducciones de una calidad aceptable. No obstante, aún había un gran margen de mejora, ya que algunas de las traducciones requerirían, en términos de Oliver *et al.* (2022), una posesición «considerable» o, directamente, una «retraducción». Si bien desconocemos cuál es el motivo de la escasa calidad obtenida en algunos casos, dedujimos que podía deberse al tamaño reducido del corpus especializado empleado para el entrenamiento del motor. Este era, en especial, el caso del motor FR-ES, cuyo corpus especializado estaba compuesto solo por ocho obras. Por esta razón, hemos decidido entrenar un nuevo motor para la traducción de textos literarios del francés al español elaborando un corpus especializado mayor y aplicando una técnica de ponderación de oraciones (*sentence-weighting*), con la esperanza de que obtener resultados más satisfactorios.

1.3. Objetivos

Mediante este trabajo de fin de máster pretendemos alcanzar un objetivo principal y dos secundarios:

- **O. P. 1.** Entrenar un motor de TAN para la traducción de textos literarios del francés al español por medio de un corpus paralelo alineado con el algoritmo SBERT y una técnica de ponderación de oraciones.
- **O. S. 1.** Realizar una evaluación automática de la calidad de las traducciones del motor de traducción entrenado, y compararla con la de otro motor entrenado previamente y con la del mismo motor entrenado sin ponderación de oraciones.
- **O. S. 2.** Realizar una evaluación humana de la calidad de las traducciones del motor de traducción entrenado, y compararla con la de otro motor entrenado previamente y con la del mismo motor entrenado sin ponderación de oraciones.

1.4. Preguntas de investigación, hipótesis y variables

Asimismo, en este trabajo de fin de máster pretendemos dar respuesta a las siguientes **preguntas de investigación**:

- **P. I. 1.** ¿El uso de un corpus especializado de mayor tamaño para entrenar un motor de TAN mejora la calidad de las traducciones de dicho motor?
- **P. I. 2.** ¿El uso de la técnica de ponderación de oraciones en la fase de entrenamiento del motor de TAN mejora la calidad de las traducciones de dicho motor?

De este modo, para la consecución de este trabajo partimos de las siguientes **hipótesis**:

- **H₁:** Cuanto mayor sea el corpus especializado utilizado para entrenar un motor de TAN, mayor será la calidad de las traducciones de dicho motor.
- **H₂:** Un motor de TAN entrenado con ponderación de oraciones ofrece traducciones de mayor calidad que las del mismo motor entrenado sin ponderación de oraciones.

Por lo tanto, en este trabajo exploratorio contamos con tres **variables**: Por un lado, la calidad final de las traducciones como variable dependiente y, por otro, el tamaño del corpus especializado y el empleo de una técnica de ponderación de oraciones como variables independientes.

1.5. Estructura del trabajo

En primer lugar, presentamos en el capítulo 2 el marco teórico en el que se sitúa el trabajo. En él, hacemos un sucinto repaso de la historia de la traducción automática, de los diferentes métodos de evaluación y del concepto de posesición. También revisamos algunas de las obras que se han publicado en el ámbito de la traducción automática y los textos literarios. El capítulo finaliza con una explicación del proyecto MTUOC en el que se enmarca este trabajo.

En el siguiente capítulo, el capítulo 3, detallamos la metodología que hemos seguido para la creación de un corpus paralelo, el entrenamiento del motor de TAN, la evaluación automática y la elaboración de las pruebas de evaluación humana. Seguidamente, el capítulo 4 presenta los resultados obtenidos a través de la evaluación automática y de la evaluación humana.

Por último, el capítulo 5 ofrece las conclusiones a las que hemos llegado tras analizar los resultados obtenidos, amén de describir las limitaciones a las que ha estado sujeto este trabajo y de proponer futuras vías de investigación.

2. Marco teórico

2.1. Breve historia de la traducción automática

La historia de la traducción automática (TA) siempre ha estado muy ligada a hechos históricos, políticos, científicos y económicos, puesto que guarda una estrecha relación con el concepto de productividad. Aunque hoy por hoy la TA no reemplaza al traductor humano, este trasvase automatizado de una lengua a otra ocupa un lugar importante en el mundo de la traducción profesional, ya que los mercados y la sociedad actuales demandan una gran cantidad de información traducida en plazos de tiempo muy reducidos. La TA incluye una combinación de disciplinas, tal y como se aprecia en la definición que ya ofrecía Quah (2006) hace más de 15 años:

[MT] is an interdisciplinary enterprise that combines a number of fields of study such as lexicography, linguistics, computational linguistics, computer science and language engineering [...]. It is based on the hypothesis that natural languages can be fully described, controlled and mathematically coded [...]. (p. 57)

En general, se considera que la TA comenzó en la década de los cincuenta, tras los esfuerzos en criptografía realizados en la Segunda Guerra Mundial. Los principales avances de esta década y el éxito del experimento de Georgetown, en el que se tradujeron automáticamente unas sesenta frases del ruso al inglés, dio pie a una era de inversión en investigación sobre TA en los Estados Unidos y en la antigua Unión Soviética (Sánchez, 2017). A pesar de las altas expectativas de los científicos, que creían que serían capaces de crear un sistema de TA que pudiera traducir cualquier texto en cuestión de unos pocos años, el desarrollo de esta tecnología resultó ser más complicado de lo esperado y no progresó como habían imaginado.

Años más tarde, el informe ALPAC (*Automatic Language Processing Advisory Committee*, en inglés), publicado en 1966, afirmó que la TA era imprecisa, demasiado lenta y que no podía competir con la traducción humana. Ante estas conclusiones, el gobierno estadounidense retiró la financiación destinada a la TA y sugirió que la investigación se enfocara hacia el desarrollo de otras herramientas como diccionarios automáticos o a la Lingüística Computacional (Sánchez, 2017). El informe puso fin así a una década caracterizada por el optimismo; sin embargo, la investigación sobre TA continuó llevándose a cabo en países como Canadá, Francia y Alemania.

Los primeros sistemas de TA que se crearon eran sistemas **basados en reglas** (*Rule-Based Machine Translation*), compuestos por diccionarios monolingües y bilingües, y un conjunto de reglas. Estos sistemas presentaban sus ventajas, como el uso coherente de la terminología y la posibilidad de corregir errores modificando las reglas, pero la elaboración de dichas reglas lingüísticas para cada par de lenguas era una tarea ardua y requería una inversión de tiempo considerable. Además, estos sistemas tendían a proporcionar traducciones literales y presentaban una falta de consistencia en la fluidez sintáctica (Sánchez, 2017).

Tras varios proyectos ejecutados en los años setenta y ochenta, la década de los noventa supuso el gran impulso que la TA necesitaba. La aparición de Internet y la gran cantidad de información disponible en línea cambió la forma de entender los sistemas de TA. De esta forma, surgieron los primeros **sistemas de traducción automática estadísticos** (*Statistical-Based Machine Translation*), como el proyecto Candide, creado por la compañía IBM. Los sistemas

estadísticos ya no requerían el mismo esfuerzo humano para codificar reglas de transformación y actualizar listas terminológicas, sino que explotaban la riqueza de las traducciones existentes, que abarcaban numerosos pares de lenguas, para extraer reglas basadas en la probabilidad estadística. Es decir, los sistemas estadísticos aprendían a traducir automáticamente a partir de traducciones existentes alineadas en forma de corpus paralelo, de donde se extrae un modelo de probabilidad (Heame y Way, 2011). El proceso de entrenamiento mediante corpus era más rápido que el proceso que requería un sistema basado en reglas y el resultado presentaba una fluidez mayor, aunque también tenía ciertos inconvenientes, como las inconsistencias terminológicas, y los errores en el orden de las palabras y de concordancia.

En los años diez, aparecieron los sistemas de **traducción automática neuronal** (*Neural Machine Translation*) que, hasta hace muy poco, representaban el mayor hito en TA y trataban de solventar las deficiencias de los sistemas de TA tradicionales. Según Sánchez (2017), su principal novedad frente a los sistemas previos es que «se sirven de redes neuronales artificiales para ofrecer resultados fruto de la interpretación de frases y párrafos completos, y no solo palabra por palabra». En la TAN, las palabras y frases se representan de forma numérica mediante vectores (Casacuberta y Peris, 2017). Briva-Iglesias (2020) añade también que la TAN requiere de corpus lingüísticos de gran volumen para «estimular» las neuronas artificiales adecuadamente durante el entrenamiento del motor. Entre los motores de TAN existentes, podemos destacar DeepL¹ y algunas combinaciones de idiomas de Google Translate². Hoy por hoy, estos motores son objeto de numerosas investigaciones debido a su potencial y a la calidad de sus traducciones.

Por último, la reciente aparición de los chatbots basados en inteligencia artificial (IA) constituyen la mayor novedad en la actualidad. Estos chatbots son programas informáticos que se sirven de IA, del procesamiento del lenguaje natural y de la comprensión del lenguaje natural (NLP y NLU, respectivamente, por sus siglas en inglés) para simular conversaciones humanas y responder a las preguntas de los clientes. Se basan, además, en *machine learning* y *deep learning* para aprender de cada interacción con el usuario (Adamopoulou y

¹ <https://www.deepl.com/translator> Última consulta: 15/05/2023

² <https://translate.google.com/> Última consulta: 15/05/2023

Moussiades, 2020). ChatGPT, desarrollado por la empresa OpenAI, merece una mención especial por su eficacia a la hora de realizar cualquier tarea relacionada con el lenguaje, como la traducción, la revisión o, incluso, la generación de texto.

2.2. Evaluación de la traducción automática

Con el creciente desarrollo de los sistemas de traducción automática, llegó la necesidad de implementar técnicas de evaluación para medir la calidad de las traducciones y diseñar sistemas más eficientes.

En este punto nos centraremos en dos tipos de evaluación: la humana y la automática; que son las que emplearemos en este trabajo para evaluar el motor entrenado.

La **evaluación humana**, también conocida como manual, es aquella realizada por profesionales monolingües o bilingües. Este tipo de evaluación se suele caracterizar por su subjetividad, inversión de tiempo y alto coste (Martínez Mateo, 2014). Escribe (2019) afirma lo siguiente a este respecto:

Having human evaluators judge a MT output, either by rating it or by post-editing it according to a reference, is a difficult task because such techniques are time-consuming, rather expensive and generally not re-usable. Moreover, such studies are highly subjective, as human evaluators do not necessarily agree on the quality of the MT output. (p. 37)

Según Koehn (2010), es prácticamente imposible juzgar las traducciones de un sistema de TA como correctas o incorrectas. En su lugar, se estima la calidad de una serie de segmentos traducidos en una escala de corrección en términos de precisión o fidelidad (*adequacy*) y fluidez (*fluency*). El primer concepto comprueba si el significado del texto origen (TO) se mantiene en el texto meta (TM). El segundo, en cambio, comprueba hasta qué punto el TM ofrece una lectura similar a la del TO, con corrección gramatical y ortográfica, y coherencia terminológica. El autor (2011) presenta la siguiente escala numérica del 1 al 5 para medir cada uno de estos conceptos:

Adequacy		Fluency	
1	<i>none</i>	1	<i>incomprehensible</i>
2	<i>little meaning</i>	2	<i>disfluent English</i>
3	<i>much meaning</i>	3	<i>non-native English</i>
4	<i>most meaning</i>	4	<i>good English</i>
5	<i>all meaning</i>	5	<i>flawless English</i>

Tabla 1: Escalas para la evaluación de la precisión y la fluidez (Koehn, 2011)

TAUS (*Translation Automation User Society*), por su parte, establece unos parámetros de evaluación basados en precisión y fluidez muy similares (2014). Por un lado, la precisión, según TAUS, debe medirse en una escala del 1 al 4 de la siguiente manera:

- **1: None** – La traducción no recoge nada del significado del TO.
- **2: Little** – La traducción recoge tan solo unos pocos fragmentos del TO.
- **3: Most** – La traducción recoge prácticamente todo el significado del TO.
- **4: Everything** – La traducción recoge todo el significado del TO.

Por otro lado, TAUS propone la siguiente escala para evaluar el concepto de fluidez y valorar si la traducción puede interpretarse como si fuera un texto original (corrección gramatical y ortográfica, y uso coherente de la terminología):

- **1: Incomprehensible** – El texto está escrito de forma deficiente y no puede comprenderse.
- **2: Disfluent** – El texto muestra errores graves que dificultan su comprensión.
- **3: Good** – El texto presenta un número reducido de errores, pero puede leerse correctamente.
- **4: Flawless** – El texto no presenta errores y puede leerse perfectamente.

A pesar de los inconvenientes de la evaluación humana descritos anteriormente, Escribe (2019) destaca la importancia de la evaluación humana dado que «only human evaluators can tell whether the type of language used is adequate according to the context (register) or if a change in grammar or lexis at the post-editing stage is considerably affecting the meaning of a sentence» (p. 37).

Por otra parte, tenemos la **evaluación automática**, que se presenta como una alternativa a los altos costes y a la inversión de tiempo que suponen la evaluación humana. En la evaluación automática, se considera que la mejor

traducción es aquella que más se acerca a una traducción de referencia. Esta cercanía se calcula de manera diferente dependiendo de cada métrica. Según Escribe (2019), «many scholars claim that automatic metrics are the most efficient solution because they are objective, fast and inexpensive compared to human evaluation» (p. 37). De las diferentes métricas que existen, nos centraremos en WER, *Edit-Distance Percentage*, BLEU, NIST y TER, por ser las que incluye la herramienta MTUOC-eval³.

- **WER (*Word Error Rate*)**: Esta métrica se basa en la distancia de edición mínima. Calcula la suma de inserciones, sustituciones o eliminaciones necesarias para crear una traducción válida y la divide por el número de palabras en la traducción de referencia. Cuanto más se parezca la traducción a la de referencia, más se acercará el resultado a 0 (Goutte, 2006; Moreno, 2022; Martínez, 2023).
- ***Edit-Distance Percentage (&EdDist)* o *Levenshtein distance***: Esta métrica mide la similitud entre dos cadenas de texto calculando el número mínimo de ediciones de un solo carácter necesarias para cambiar una palabra por otra (Sánchez, 2017). Como ocurre con otras métricas automáticas, esta métrica no puede medir la calidad de una traducción por sí sola, ya que los resultados variarán si los diferentes elementos de una oración se encuentran en una posición distinta en la traducción automática y en la traducción humana de referencia. Por ejemplo, si el motor de TA ofrece la traducción «Si llego a casa antes de las ocho, te llamo por teléfono» y la traducción de referencia es «Te llamo por teléfono si llego a casa antes de las ocho», la propuesta del motor es correcta y no requiere cambios durante la fase de posesición. Sin embargo, el valor de la *Levenshtein distance* será bastante alto porque se requieren muchos cambios para convertir la primera oración en la segunda.
- **BLEU (*BiLingual Evaluation Understudy*)**: Junto con NIST, es la métrica más utilizada (Goutte, 2006). Esta métrica basada en n-gramas determina el grado de similitud entre la traducción automática y la de referencia a través de un algoritmo matemático. BLEU calcula el número de palabras o secuencias de palabras (n-gramas) que coinciden con n-gramas en los

³ <https://github.com/aoliverg/MTUOC-eval> Última consulta: 24/04/2023

textos de referencia, y otorga una puntuación mayor a las palabras o secuencias más extensas. También aplica una penalización por brevedad si la longitud de los segmentos es muy distinta (Sánchez, 2017). Cuanto más coincidan los n-gramas de la traducción automática y la de referencia, más se acercará la puntuación a 1; cuanto menos coincidan, más se acercará al 0 (Goutte, 2006; Moreno, 2022; Martínez, 2023).

- **NIST:** El nombre de esta métrica proviene del Instituto Nacional de Estándares y Tecnología de los Estados Unidos (NIST por sus siglas en inglés, *National Institute of Standards and Technology*). Se trata de una variante de la métrica BLEU cuya diferencia estriba esencialmente en el cálculo del número de referencias que se emplean para calcular la penalización por brevedad (Sánchez, 2017). NIST se centra en la poca frecuencia que algunos n-gramas presentan para otorgarles el máximo valor, por considerarlos más informativos. Igual que en BLEU, una puntuación elevada indica mejor calidad de traducción (Goutte, 2006; Moreno, 2022; Martínez, 2023).
- **TER (*Translation Error Rate*):** Esta métrica basada en correspondencia de cadenas se define como el número mínimo de ediciones que se necesitan para que la traducción propuesta sea idéntica a la de referencia (Snover *et al.*, 2006). Estos cambios incluyen inserciones, eliminaciones y sustituciones de palabras, además de cambios en la secuencia de palabras. Los valores más cercanos a 0 son los que indican mejor calidad (Moreno, 2022; Martínez, 2023).

Uno de los aspectos más cuestionables de estas métricas es el de otorgar a la traducción de referencia la categoría de «gold standard» (Saldanha y O'Brien, 2014) o «traducción de oro» con la que la traducción propuesta se debe comparar. Tal y como menciona Moreno (2022), este problema es aún mayor en el caso de la traducción literaria, en el que el estilo y la creatividad del traductor tienen un gran peso.

2.3. La posedición

La posedición es la tarea que sustituye a la traducción convencional en los proyectos de la TA. O'Brien (2011) la define como «the correction of raw machine translated output by a human translator according to specific guidelines and

quality criteria», puesto que el objetivo es que los textos traducidos por sistemas de TA alcancen un nivel de aceptabilidad establecido. La posesición cuenta con unas características propias que la diferencian de otros procesos como la traducción o la revisión (Sánchez y Rico, 2020). Al estar íntimamente ligada al concepto de productividad, la prioridad es ahorrar tiempo, por lo que no debe emplearse demasiado tiempo en las tareas de posesición (Sánchez, 2017).

Generalmente, suelen distinguirse dos tipos de posesición según las necesidades del usuario (Allen, 2003):

- **Posedición completa (*full post-editing*):** Se realizan las correcciones necesarias para asegurar que el texto alcance niveles de publicación.
- **Posedición ligera (*light post-editing*):** Se realizan los cambios mínimos para que el lector pueda hacerse una idea del texto, que, en su mayoría, son textos de uso interno y no se publican.

Por lo general, en la posesición, el significado tiene prioridad sobre la corrección gramatical y el estilo, por lo que los errores gramaticales y ortográficos solo se corrigen si afectan al significado. Se recomienda también aprovechar al máximo el texto proporcionado por el sistema de TA y realizar solo los cambios que sean estrictamente necesarios.

La norma ISO 18587:2017 establece los requisitos del proceso de posesición de los resultados de TA y las competencias de los poseedores. Asimismo, Sánchez y Rico (2020) señalan que saber poseer constituye una de las nuevas habilidades que necesitan los traductores actuales y futuros.

Tal y como señala Briva-Iglesias (2020), en los últimos años, «la posesición ha sido objeto de un gran estudio en el campo de las tecnologías de la traducción y se ha investigado desde muchos ámbitos». No obstante, y a pesar de que se trata de una etapa recomendada antes de la publicación de un texto traducido automáticamente, no constituye el objeto de estudio de este trabajo, por lo que no profundizaremos en aspectos que guarden relación con la posesición.

2.4. La traducción automática neuronal y la traducción literaria

La TAN, en su aún corta trayectoria, ha mostrado un gran potencial y hay evidencias de que puede alcanzar una calidad de traducción mayor que los sistemas estadísticos (Toral y Way, 2018). Más allá de sus resultados

generalmente positivos, la TAN reviste especial interés para los textos literarios por los siguientes motivos:

- Sus resultados en la traducción de textos con una gran riqueza léxica son especialmente prometedores (Bentivogli *et al.*, 2016). Este es, precisamente, el caso de los textos literarios.
- Hay quien afirma que los motores de TAN son capaces de encontrar los equivalentes culturales en otra lengua (Torál y Way, 2018).

Con respecto a este segundo punto, debe añadirse que, mientras que las traducciones literales pueden ser aceptables en contextos donde prima conservar el significado del texto original sin reparar en cuestiones de estilo (como en documentos técnicos), las traducciones literales no son adecuadas para textos literarios. En estos textos, las expectativas del lector son más altas, ya que se debe además preservar la experiencia de lectura del texto original (Torál y Way, 2018) a través de, por ejemplo, la fidelidad al estilo del autor. De hecho, Serrano (2020) señala que en la traducción literaria «el mensaje puede llegar a ocupar un segundo plano con relación a la forma en que se expresa dicho mensaje». La autora indica que los textos literarios se caracterizan por aspectos formales y semánticos que a las máquinas les resulta complicado interpretar, como las figuras retóricas, los juegos de palabras, el humor y la ironía. Tampoco debemos ignorar la importancia del contexto, puesto que incluso a los traductores humanos les cuesta traducir cuando el contexto no está claro, como cuando en una obra literaria se insinúan hechos que se desvelan en una etapa posterior (Serrano, 2020).

No obstante, la traducción literaria también puede aprovecharse de los avances tecnológicos como ocurre en otras especialidades de la traducción. Torál y Way (2018) citan el auge del libro electrónico y señalan la posibilidad que ofrece para crear corpus paralelos destinados a entrenar sistemas de traducción automática adaptados a la traducción de novelas.

Aun así, la creación de corpus literarios no está exenta de problemas, dado que, tal y como menciona Simard (2020), la alineación entre segmentos es a menudo una aproximación, pues la motivación del traductor para escoger una palabra o expresión sobre otra se encuentra en un contexto más amplio que el de la mera correspondencia léxica. Asimismo, los aspectos estilísticos y retóricos propios de los textos literarios son factores que fomentan la creatividad del

traductor y, en muchos casos, el texto meta puede llegar a ser irreconocible (Santoyo, 1996).

El reciente interés de la comunidad de la Lingüística Computacional por el procesamiento de textos literarios explica el gran número de investigaciones sobre la aplicación de la traducción automática en textos literarios. A continuación, se citarán algunas de las publicaciones más relevantes:

En 2013, Jonas e Irvine utilizaron sistemas de TA para traducir muestras de literatura francesa (tanto prosa como poesía) al inglés. Su principal objetivo era determinar qué era lo que volvía a la traducción literaria particularmente complicada. El año siguiente, Bernardo (2014) comparó una traducción humana al portugués brasileño de la novela *The Fortunes and Misfortunes of the Famous Moll Flanders* de Daniel Defoe con los resultados que obtuvo utilizando Google Translate. El motor, que en aquel año era un sistema estadístico mucho menos evolucionado que el motor neuronal actual, reproducía el texto origen casi palabra por palabra. Bernardo reconoció que el resultado estaba lejos de ser satisfactorio.

El mismo año, Besacier (2014) tradujo una historia corta del inglés al francés utilizando un motor de TA y el resultado fue poseído por traductores no profesionales. Esta forma de trabajar se consideró una alternativa de bajo coste frente a la traducción convencional de obras literarias, en claro detrimento de la calidad de traducción.

El siguiente estudio que merece una mención fue llevado a cabo por Toral y Way en 2015. Mientras que Besacier (2014) solo tradujo una historia corta, Toral y Way (2015a) tradujeron al catalán la novela *El prisionero del cielo* de Carlos Ruiz Zafón. Para realizar esta traducción, utilizaron un motor de TA estadística propio entrenado con textos literarios. Después, compararon su traducción con la traducción realizada por un profesional. Los autores concluyeron que aproximadamente el 20 % de las frases producidas por el motor eran exactamente iguales a las propuestas del traductor humano. También determinaron que un 60 % de los segmentos tenían una calidad similar. Estos resultados tan positivos supusieron un salto notable frente a los estudios anteriores que los autores describieron como «un prometedor primer paso». No obstante, los mismos autores obtuvieron resultados menos optimistas en otro estudio publicado el mismo año (Toral y Way, 2015b), al comparar traducciones

llevadas a cabo entre idiomas más distantes. En este caso, tradujeron de manera automática del francés al inglés *L'étranger* de Albert Camus. El motor ofrecía traducciones más literales que la traducción de referencia, aunque los autores consideraron que suponían un buen punto de partida para la posesición.

En 2018, estos autores llevaron a cabo una nueva investigación en la que entrenaron un sistema de TAN para la traducción de textos literarios del inglés al catalán, y compararon los resultados con los de un sistema estadístico basado en frases y con los de un traductor humano profesional. Para entrenar el motor, utilizaron una gran cantidad de textos (más de 100 millones de palabras) y emplearon para evaluar su motor un conjunto de 12 novelas ampliamente conocidas que datan desde la década de los veinte hasta la actualidad. Los lectores nativos que participaron en la evaluación humana determinaron que entre el 17 % y el 34 % de las traducciones del motor TAN eran de una calidad equivalente a las del traductor humano.

Aun así, en un estudio más reciente, Youdale (2020) concluyó que, si bien un motor de TA puede llegar a producir traducciones de una calidad similar a la humana en determinadas circunstancias, todavía se requieren mejoras en cuanto a la sintaxis y a factores mencionados al comienzo de este apartado, como el humor, la ironía y las figuras retóricas.

Por su parte, Guerberof y Toral (2022a, 2022b) han conducido recientemente investigaciones en torno a los conceptos de creatividad y *narrative engagement* (el grado de inmersión de los lectores durante la lectura) de textos literarios traducidos por traductores humanos y sistemas de TAN. En sus estudios, los autores concluyeron que la posesición de los resultados de TA limitaba la creatividad de los traductores.

Por último, debemos referirnos a la obra de Moreno (2022) para cerrar esta sección, ya que nos hemos basado en las futuras líneas de investigación que propone la autora para la elaboración de este trabajo. Moreno entrena un motor de TA para la traducción de textos literarios del inglés al español. Para ello, crea un corpus reuniendo la mayor cantidad de posible de materiales traducidos de contenido literario. Posteriormente, combina este corpus con otros corpus paralelos de temática variada. El motor entrenado presentó varios problemas que, según las conclusiones de la autora, se debían a un proceso de alineación mejorable y a las particularidades de la traducción literaria. De igual modo,

señaló que, además de la evaluación automática, sería conveniente realizar una evaluación humana.

2.5. El proyecto MTUOC

El proyecto MTUOC es un proyecto de investigación y de transferencia de tecnología de la Universitat Oberta de Catalunya gestionado por el Dr. Oliver. Este proyecto pretende facilitar el entrenamiento de motores de TA estadística y neuronal, así como su integración en entornos profesionales de traducción (Oliver, 2021). El proyecto proporciona *scripts* y programas para el preprocesamiento de corpus paralelos para utilizarlos en el entrenamiento de sistemas de TA, *scripts* de entrenamiento de sistemas neuronales (con Marian y OpenNMT) y estadísticos (con Moses), el servidor MTUOC, que permite poner en marcha los motores entrenados, y otras herramientas auxiliares.

Oliver (2021) señala que, aunque los proyectos Marian, OpenNMT, ModernMT, JoeyNMT y Moses se distribuyan bajo licencias libres, su instalación, configuración y puesta en marcha no siempre son sencillas y requieren conocimientos informáticos avanzados. Por consiguiente, el proyecto MTUOC tiene como objetivo poner esta tecnología a disposición de todas las personas, empresas e instituciones de manera fácil de utilizar mediante sus herramientas de licencia libre, que constan de una interfaz sencilla y se pueden utilizar en cualquier sistema operativo (Oliver, 2021).

En el seno de este proyecto se han realizado tres trabajos de fin de máster previos a este trabajo, mencionados en el apartado 1.1. Los dos primeros, llevados a cabo por Rodríguez (2021) y Moreno (2021), concluían en la importancia de un proceso de alineación fiable en la creación de corpus. Martínez (2023) introduce un nuevo algoritmo en este proceso, SBERT, que, según su investigación, logra mejores resultados que el algoritmo empleado por Rodríguez y Moreno. Asimismo, incluye una nueva combinación lingüística (alemán-español) y la particularidad de entrenar un motor para la traducción inversa en el ámbito del turismo. El presente trabajo pretende aportar nuevo conocimiento a las investigaciones conducidas en el contexto del proyecto MTUOC mediante la incorporación de la combinación lingüística francés-español, el empleo del algoritmo SBERT en la creación de un corpus de contenido literario, una técnica de ponderación de oraciones para entrenar el motor y una evaluación humana.

3. Metodología

En este capítulo describiremos los pasos seguidos para la elaboración del corpus especializado, su procesamiento para entrenar el motor de TAN y la evaluación de los resultados tanto automática como humana. En cuanto al equipo empleado, se ha utilizado un equipo de sobremesa con Windows 11 para la obtención de textos, segmentación y alineación del corpus, y la evaluación del motor; y un ordenador portátil con la distribución Ubuntu de Linux para los procesos de procesamiento de corpus y preentrenamiento del motor. El entrenamiento del motor se ha realizado en una GPU de las instalaciones de la universidad.

3.1. Elaboración del corpus

Tal y como hemos indicado en el apartado 1.2, en el contexto de las prácticas curriculares entrenamos un motor de TAN para la traducción FR-ES al que a partir de ahora llamaremos motor 1. Para entrenar este motor, utilizamos un corpus especializado combinado con un corpus general. El corpus especializado constaba de 37 541 segmentos, extraídos de ocho obras de la serie de Arsène Lupin de Maurice Leblanc. Después, lo combinamos con el corpus general CCMatrix⁴ hasta alcanzar un total de 10 000 000 de segmentos. Si comparamos estos datos con los del motor que entrenó Moreno (2022), como se puede observar en la Tabla 2, el tamaño del corpus especializado del motor 1 supone apenas un 0,38 % del corpus total, mientras que en el corpus especializado de Moreno suponía el 22,13 %.

	Corpus motor 1	Corpus motor literario EN-ES de Moreno (2022)
Tamaño del corpus especializado	37 541 segmentos (0,38 %)	1 334 526 segmentos (22,13 %)
Tamaño del corpus de combinación	9 962 459 segmentos (99,62 %)	4 696 659 segmentos (77,87 %)
Tamaño del corpus de entrenamiento	10 000 000 segmentos (100 %)	6 031 185 segmentos (100 %)

Tabla 2: Comparación de los corpus del motor 1 y del motor literario de Moreno (2022)

⁴ <https://opus.nlpl.eu/CCMatrix.php> Última consulta: 14/04/2023

Por este motivo, para entrenar el nuevo motor de TAN FR-ES, al que desde ahora denominaremos motor 2, hemos decidido utilizar un corpus especializado mayor. Para ello, amén de las ocho novelas empleadas para el entrenamiento del motor 1, hemos empleado otras 132 obras para lograr un total de 140 obras. De esta forma, hemos obtenido un corpus de un tamaño similar al que crearon Toral y Way (2018), compuesto por 133 obras literarias en inglés y catalán. La totalidad de estas obras fueron escritas originalmente en francés por autores francófonos y, tras caducar sus derechos de autor, pertenecen al dominio público. Estas obras han sido descargadas a través de los portales de Internet Ebooks libres et gratuits⁵, Atramenta⁶ y Bouquineux⁷. Para obtener las obras traducidas en español, hemos utilizado un enlace que permite la descarga de una biblioteca para la herramienta Calibre⁸, publicado en un mensaje del foro *La leyenda de Star Wars*⁹ tras comprobar que tenemos acceso a las obras traducidas de manera legal por medio de la red de bibliotecas públicas. El listado de las 140 obras que componen el corpus se encuentra en el anexo I. Estas obras han sido descargadas en formato EPUB y las hemos convertido en documentos de texto sin formato (TXT) mediante la herramienta Calibre mencionada anteriormente. Una vez convertido el formato, hemos modificado el nombre de los archivos para que tanto el original como la traducción tuvieran el mismo nombre, pero con un cambio al final indicando la lengua (p. ej., *gentleman_cambrioleur-fr.txt* y *gentleman_cambrioleur-es.txt*).

Los 140 pares de archivos de texto sin formato obtenidos en este paso contenían un total de 734 705 segmentos en francés y 690 579 segmentos en español.

3.1.1. Segmentación y alineación del corpus

Una vez terminada la obtención de los textos, hemos procedido a la segmentación de los archivos de texto y a su posterior alineación. A partir de este punto, hemos empleado las herramientas de acceso libre propias del proyecto MTUOC, diseñadas por el Dr. Oliver.

⁵ <https://www.ebooksgratuits.com/> Última consulta: 14/04/2023

⁶ <https://www.atramenta.net/> Última consulta: 14/04/2023

⁷ <http://www.bouquineux.com/> Última consulta: 14/04/2023

⁸ <https://calibre-ebook.com/> Última consulta: 14/04/2023

⁹ <https://laleyendadesw.mforos.com/1493088/11588310-biblioteca-calibre-30026-ebooks-epubs/> Última consulta: 14/04/2023

El primer paso ha consistido en la **segmentación** de los textos con el programa MTUOC-segmenter¹⁰. Este programa permite la segmentación automática de uno o varios archivos de texto sin formato por medio de un archivo SRX (*Segmentation Rules eXchange*). Esta herramienta, además de segmentar archivos individuales, ofrece las opciones de segmentar todos los archivos de un directorio y copiarlos en otro directorio, y de segmentar todos los archivos de un directorio y devolver el resultado en un único archivo de texto. El programa no solo puede utilizarse en la línea de comandos mediante *scripts* de Python, sino que cuenta también con una versión con interfaz de usuario para ejecutarse en Windows. En nuestro caso, hemos optado por utilizar la versión con interfaz de usuario en el sistema Windows y la segmentación se ha podido ejecutar sin problemas.

En lo que respecta a la **alineación** de los segmentos, cabe mencionar que Rodríguez (2021) y Moreno (2022) emplearon la herramienta MTUOC-aligner-hunalign, que funciona con el algoritmo Hunalign. Moreno concluyó que este algoritmo tiene algunas carencias y adujo la baja precisión de la alineación como uno de los motivos por los que su motor presentó algunos errores. Por consiguiente, propuso el empleo de técnicas de alineación más avanzadas como futura línea de investigación. Martínez (2023), por su parte, realizó la alineación por partida doble con los algoritmos Hunalign y SBERT, y comparó los resultados de ambas alineaciones. Martínez observó «un aumento de la fiabilidad de las alineaciones de SBERT (91-98 %) frente a las de Hunalign (84-98 %)» (p. 45), además de que con este algoritmo consiguió alinear una cantidad superior de segmentos: 52 119 segmentos frente a los 13 604 segmentos alineados con Hunalign (de un total de 270 437 segmentos sin alinear). Teniendo en cuenta las conclusiones de Moreno y los resultados de la comparativa de Martínez, hemos decidido descartar el uso del algoritmo Hunalign y utilizar únicamente SBERT para la alineación.

Para ello, hemos descargado el programa MTUOC-aligner-SBERT¹¹, que, además de presentarse como *scripts* de Python, dispone de una versión ejecutable para Windows al igual que el programa anterior. Este algoritmo «allows both the alignment of parallel document and the process of finding

¹⁰ <https://github.com/aoliverg/MTUOC-segmenter> Última consulta: 15/04/2023

¹¹ <https://github.com/aoliverg/MTUOC-aligner> Última consulta: 15/04/2023

parallel segments in comparable documents, that is, documents that are not translations of each other but talk about a similar topic» (Oliver, 2023). Este proceso también resulta útil para alinear documentos paralelos que pueden diferir en ciertas secciones.

A diferencia de MTUOC-aligner-hunalign, que es capaz de realizar la alineación de los documentos en cuestión de segundos, el proceso con MTUOC-aligner-SBERT puede demorarse horas o incluso días. En nuestro caso, el proceso de alineación con SBERT finalizó en 77 horas.

El resultado del proceso de alineación ha sido un directorio con los 140 archivos de texto que contenían los segmentos origen, los segmentos meta y un índice de fiabilidad (*confidence score*) separados por tabuladores. La Tabla 3 presenta un ejemplo de alineación de segmentos con un índice de fiabilidad alto (cercano al 2):

Segmento original (FR)	Segmento traducido (ES)	Índice de fiabilidad
<i>Les splendides constellations de l'hémisphère austral se dessinent avec une incomparable pureté.</i>	<i>Las espléndidas constelaciones del hemisferio austral se destacan con una pureza incomparable.</i>	1.8712
<i>Ils ont mérité de vous accompagner là où le vent vous poussera.</i>	<i>Ellos merecen ir junto a usted donde el viento los lleve.</i>	1.6910
<i>Si vous avez faim, allez chercher de quoi manger !</i>	<i>¡Si tiene hambre, vaya y busque algo que comer!</i>	1.6618

Tabla 3: Ejemplos de segmentos bien alineados con un índice de fiabilidad alto

En una exploración visual de varios de los archivos, hemos apreciado que prácticamente todos los segmentos con un índice de fiabilidad cercano al 2 estaban bien alineados. No obstante, los segmentos con un índice de fiabilidad bajo (cercano al 1) contenían a menudo errores de alineación. A continuación, se muestran algunos de estos segmentos detectados en la exploración visual:

Segmento original (FR)	Segmento traducido (ES)	Índice de fiabilidad
» <i>a-t-il murmuré.</i>	<i>Cerró la puerta.</i>	1.0109
<i>Nous sommes des bénévoles, passionnés de littérature.</i>	<i>En mi país no adoramos ídolos.</i>	1.0040
<i>C'est là qu'ils se logeront, leur poste étant absolument inhabitable.</i>	<i>SITUACIÓN.</i>	1.0014

Tabla 4: Ejemplos de segmentos mal alineados con un índice de fiabilidad bajo

Cabe mencionar que 1 no es el valor mínimo del índice de fiabilidad, pero MTUOC-aligier-SBERT está configurado para que por defecto no muestre resultados con un índice de fiabilidad menor a 1.

El siguiente paso consistió en la **selección** de los segmentos que compondrían el corpus con la herramienta selectAlignments. Este programa selecciona de forma automática los segmentos alineados con un índice de fiabilidad igual o mayor a uno determinado y los devuelve en un único archivo de texto. Tras la exploración visual, parecía lógico deducir que debían evitarse los segmentos con un índice de fiabilidad cercano al 1, pero desechar estos segmentos implicaría reducir sobremanera el tamaño del corpus y renunciar a una gran cantidad de segmentos que, a pesar de tener un índice de fiabilidad bajo, sí estaban alineados correctamente. De hecho, Moreno (2022) también señaló que «subir el índice de fiabilidad supone perder muchos segmentos» y que «no siempre los segmentos descartados son malos». De igual manera, concluyó que aplicar índices de fiabilidad más altos en la selección «no proporcionaba mejores resultados» (p. 39).

Tomando en consideración estas observaciones, hemos decidido realizar tres selecciones diferentes: una con un índice de fiabilidad de 1 (es decir, sin descartar ningún segmento), otra con un índice de fiabilidad de 1,1 y la última con un índice de fiabilidad de 1,2.

Posteriormente, hemos extraído una muestra de 100 segmentos aleatorios de cada una de estas selecciones para comprobar manualmente el número de segmentos alineados correctamente. Para extraer la muestra de 100 segmentos aleatorios, hemos ejecutado la orden «shuf -n 100 corpus-fra-spa.txt

> random_sample.txt» en la terminal del equipo con Ubuntu. La Tabla 5 recoge, por un lado, el número de segmentos que contenía cada uno de los tres corpus y, por otro, el número de segmentos alineados correctamente de la muestra de 100 segmentos.

	Índice de fiabilidad mínimo	Número total de segmentos	Segmentos de la muestra alineados correctamente
Corpus 1	1	530 082	96/100
Corpus 2	1,1	473 977	98/100
Corpus 3	1,2	394 474	99/100

Tabla 5: Tamaño de los corpus y resultados de la evaluación manual aleatoria

Tras estas operaciones, hemos podido constatar que, al aumentar el índice de fiabilidad exigido, los corpus resultantes tenían un número de segmentos claramente menor, tal y como destacó Moreno (2022). También hemos observado que el número de segmentos alineados correctamente es similar en las tres muestras. Por lo tanto, después de realizar esta comprobación manual, concluimos que la precisión del corpus 1 es lo suficientemente alta como para entrenar el motor y, de esta forma, aprovechar la totalidad del corpus especializado que habíamos creado. De esta forma, descartamos los corpus 2 y 3, y continuamos el proceso con el corpus de 530 082 segmentos.

El siguiente paso ha sido la **limpieza** del corpus eliminando los segmentos repetidos. Para ello, hemos ejecutado la siguiente orden en la terminal de Ubuntu: «cat corpus-fra-spa.txt | sort | uniq | shuf > corpus-uniq-fra-spa.txt». Tras eliminar los segmentos repetidos, el corpus estaba compuesto por 526 988 segmentos.

3.1.2. Ampliación del corpus

Puesto que, según las instrucciones del Dr. Oliver, se requieren entre 5 y 10 millones de segmentos para entrenar un motor de TAN, nos hemos visto en la necesidad de ampliar el corpus especializado combinándolo con un corpus general de mayor tamaño, tal y como lo hicimos con el motor 1 en el transcurso de las prácticas curriculares. Para esta combinación, hemos escogido el mismo

corpus que utilizamos en el entrenamiento del motor 1, CCMatrix¹², un corpus de gran volumen que contiene 285 936 865 segmentos en la combinación francés-español y que se encuentra disponible para su descarga libre en el repositorio de corpus paralelos OPUS¹³. Emplear el mismo corpus general nos permite, además, comprobar que la posible mejora en la calidad de las traducciones proviene del corpus especializado y no de la elección de otro corpus general.

En primer lugar, hemos procedido a la descarga del corpus CCMatrix en formato Moses, que consistía en un archivo comprimido de una carpeta de 53,1 GB. Esta carpeta contenía, entre otros archivos, dos archivos de texto sin formato para cada lengua, pero alineados línea a línea. Hemos unido estos dos archivos en uno solo mediante la orden: «paste CCMatrix.es-fr.fr CCMatrix.es-fr.es > CCMatrix-fra-spa.txt».

Para combinar los dos corpus de los que disponíamos (el corpus especializado creado por nosotros mismos y CCMatrix), hemos utilizado la herramienta MTUOC-corpus-combination¹⁴, un componente del proyecto MTUOC para completar un corpus paralelo de especialidad de un tamaño insuficiente para entrenar un sistema de TA con segmentos extraídos de un corpus mayor. El programa calcula un modelo de lengua a partir de los segmentos en la lengua de partida y selecciona del corpus de gran tamaño los segmentos que sean lo más parecidos posible a los del corpus de especialidad (Oliver, 2022a). Oliver señala que este programa utiliza una base de datos SQLite para almacenar todos los datos, por lo que es capaz de trabajar con corpus extraordinariamente grandes como CCMatrix (2022).

Tal y como hicimos en el proceso de entrenamiento del motor 1 durante las prácticas, hemos indicado una selección de 10 millones de segmentos para el proceso de combinación de corpus. Tras la combinación, el corpus resultante constaba de tres partes:

- Un corpus de entrenamiento (*train-fra-spa.txt*) de 10 216 988 segmentos.
- Un corpus de evaluación (*eval.fr* y *eval.es*) de 5000 segmentos.
- Un corpus de validación (*val-fra-spa.txt*) de 5000 segmentos.

¹² <https://opus.nlpl.eu/CCMatrix.php> Última consulta: 28/04/2023

¹³ <https://opus.nlpl.eu/> Última consulta: 28/04/2023

¹⁴ <https://github.com/aoliverg/MTUOC-corpus-combination> Última consulta: 28/04/2023

En la Tabla 6, se compara el tamaño de los corpus tras la combinación empleados para entrenar los motores 1 y 2. Podemos observar que, mientras que en el caso del motor 1 el corpus especializado supone el 0,38 % del corpus de entrenamiento, en el motor 2, el corpus especializado, compuesto por 132 obras más (489 447 segmentos más), supone el 5,16 % del corpus de entrenamiento. En ambos casos, el corpus de combinación, creado con segmentos extraídos de CCMatrix, es el que ocupa mayor porcentaje.

	Corpus motor 1	Corpus motor 2
Tamaño del corpus especializado	37 541 segmentos (0,38 %)	526 988 segmentos (5,16 %)
Tamaño del corpus de combinación	9 962 459 segmentos (99,62 %)	9 690 000 segmentos (94,84 %)
Tamaño del corpus de entrenamiento	10 000 000 segmentos (100 %)	10 216 988 segmentos (100 %)

Tabla 6: Comparación de los corpus del motor 1 y del motor 2

3.2. Preprocesamiento del corpus

La siguiente fase ha consistido en el preprocesamiento del corpus mediante la herramienta MTUOC-corpus-preprocessing¹⁵ del proyecto MTUOC. El preprocesamiento de corpus está formado por varios pasos como la tokenización, el tratamiento de expresiones numéricas o el alineamiento guiado (*guided alignment*). La herramienta empleada en este paso es en realidad un *script* que realiza todos los pasos del preprocesamiento de corpus uno detrás de otro de manera automatizada según las instrucciones introducidas en un archivo de configuración. A continuación, explicaremos en qué consiste cada paso del preprocesamiento:

La **tokenización** consiste en dividir un texto en tókenes, separando las palabras de los signos de puntuación y las cifras (Martínez, 2023). Por ejemplo, la oración «They're the teacher's best students.» se tokenizaría de la siguiente manera: «They 're the teacher 's best students .» (Oliver, 2022b). El proyecto MTUOC dispone de tokenizadores en varios idiomas, entre ellos el francés y el

¹⁵ <https://github.com/aoliverg/MTUOC-corpus-preprocessing> Última consulta: 02/05/2023

español. También incluye un tokenizador genérico que puede utilizarse en las lenguas que no cuentan con un tokenizador específico. Además, en el caso de los idiomas que distinguen entre mayúsculas y minúsculas, debe asignarse a cada token las mayúsculas y minúsculas como le corresponden independientemente de la posición en la oración o del estilo de capitalización de la oración (Oliver, 2022b). Esta asignación se conoce como *truecasing*. Para ello, se debe disponer de un modelo de *truecasing* entrenado a partir de diccionarios de formas de palabras y de los propios corpus.

El siguiente paso del procesamiento es el **tratamiento de expresiones numéricas, direcciones de correo electrónico y URL**. En lo que respecta a las expresiones numéricas, en los sistemas neuronales las cifras se separan con espacios en blanco en el corpus de entrenamiento y en las oraciones por traducir, y, al recibir una traducción, se eliminan los espacios en blanco (Oliver, 2022b). Las direcciones de correo electrónico y las URL, por su parte, también necesitan un tratamiento especial, pues no son elementos traducibles. Normalmente, las direcciones de correo electrónico se reemplazan en el corpus de entrenamiento y en las oraciones por traducir por el código @EMAIL@ y las URL por el código @URL@. Una vez traducida la oración, los códigos se sustituyen por los correspondientes originales (Oliver, 2022b).

El cálculo y la aplicación de **subpalabras** (*subwords*) también es un paso importante del preprocesamiento. Trabajar con subpalabras se ha convertido en la estrategia más empleada para solventar el problema de la limitación de vocabulario que presentaban los primeros sistemas neuronales (Oliver, 2022b). En estos sistemas, el número de palabras flexionadas capaz de traducir era finito, dado que no eran capaces de traducir palabras desconocidas que no estuvieran en los corpus con los que se habían entrenado. Al trabajar con subpalabras, es decir, fragmentos de palabras, las palabras menos frecuentes se dividen en fragmentos que sí son frecuentes. De esta forma, se pueden representar todas las palabras, ya sean enteras o por fragmentos. Al final del proceso, se unen todos los fragmentos de las subpalabras traducidas (Oliver, 2022b). Para realizar este paso, hemos optado por utilizar el algoritmo *sentencepiece*, un tokenizador y detokenizador no supervisado que puede utilizarse como algoritmo para calcular y aplicar subpalabras, sin requerir ni tokenización ni *truecasing* previo (Oliver, 2022b).

El **alineamiento guiado** es el último de los pasos del preprocesamiento. En este paso se le proporciona al sistema la información de alineación a nivel de palabra o de subpalabra del corpus de entrenamiento para que sea capaz de tratar correctamente las etiquetas XML/HTML, es decir, que en la traducción aparezcan las mismas etiquetas en las posiciones correctas (Oliver, 2022b). En el archivo de configuración de MTUOC-corpus-preprocessing, le hemos indicado al algoritmo que calcule la alineación del corpus de entrenamiento mediante el programa eflomal, que hemos debido instalar en nuestro equipo.

Tras configurar correctamente este archivo de configuración que hemos mencionado, hemos preprocesado el corpus y hemos obtenido los siguientes archivos que necesitaremos en la siguiente etapa del proceso de entrenamiento:

- Los archivos *train.sp.fr* y *train.sp.es* para el entrenamiento.
- Los archivos *val.sp.fr* y *val.sp.es* para la validación.
- Los archivos *train.sp.fr.es.align* y *val.sp.fr.es.align* resultantes del alineamiento guiado.

3.3. Entrenamiento del motor de TA neuronal con Marian

Una vez concluido el preprocesamiento, hemos procedido al entrenamiento del motor utilizando Marian ¹⁶, un motor de TAN escrito en el lenguaje de programación C++ y desarrollado por el equipo Microsoft Translator. Debemos señalar que para entrenar un motor de TA neuronal es imprescindible disponer de una GPU potente, ya que con una CPU convencional el entrenamiento duraría años (Oliver, 2022c). Por este motivo, en este paso hemos realizado una simulación del entrenamiento para comprobar que no se producía ningún error y hemos detenido el entrenamiento al cabo de unos minutos. El verdadero entrenamiento del motor lo ha llevado a cabo el Dr. Oliver en una GPU de las instalaciones de la universidad.

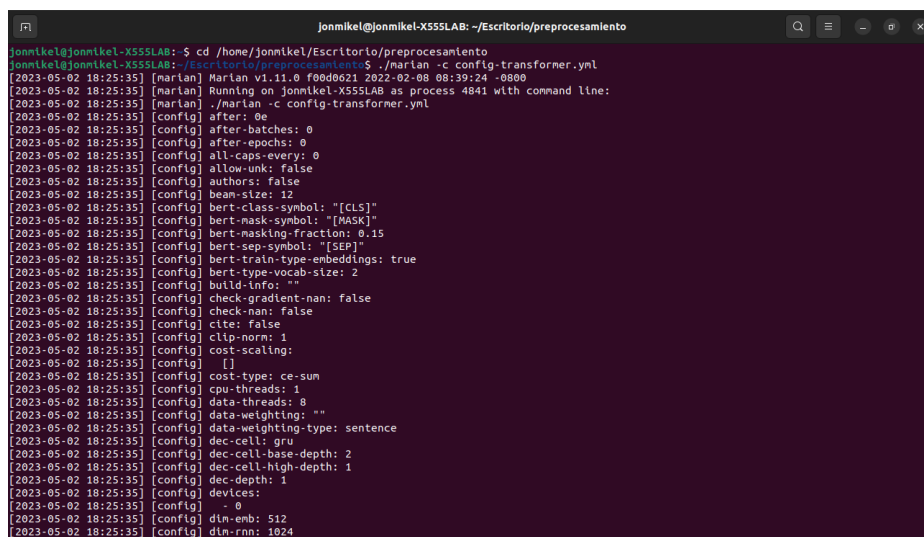
En primer lugar, hemos debido descargar la versión compilada para CPU de Marian, en nuestro caso, la versión Marian CPU V1.11.0, que funciona en los equipos con Ubuntu. Después, hemos colocado en el mismo directorio de la versión compilada de Marian los archivos generados durante el preprocesamiento, mencionados en la sección anterior. Oliver (2022c) describe dos métodos para generar los archivos de configuración con Marian: el

¹⁶ <https://marian-nmt.github.io/> Última consulta: 02/05/2023

entrenamiento básico y el avanzado. El entrenamiento avanzado permite un control más detallado sobre los parámetros de entrenamiento y escoger el tipo de motor neuronal (s2s o Transformer). Para entrenar nuestro motor, hemos optado por un entrenamiento de tipo Transformer. Hemos utilizado la orden «./marian --train-sets train.sp.fr train.sp.es --valid-sets val.sp.fr val.sp.es --vocabs vocab-fr.yml vocab-es.yml --model model.npz --task transformer-base --dump-config > config-transformer.yml» para obtener el archivo de configuración YML en el que hemos realizado las siguientes modificaciones:

- **Métricas de validación:** Hemos seleccionado las métricas *cross-entropy* y *bleu-detok*. Esta última calcula el valor de BLEU sobre los conjuntos de validación posprocesados (Oliver, 2022c).
- **Early-stopping con paciencia de 5, overwrite y keep-best.** El *early-stopping* es la técnica habitual para decidir cuándo detener un entrenamiento. Al utilizar dos métricas de validación, hemos escogido una paciencia (*patience*) de 5. Esto significa que el entrenamiento se detendrá cuando se produzcan cinco procesos de validación iterativos que no consigan mejoras. En relación con este punto, hemos activado las opciones *overwrite* y *keep-best* para que Marian elimine los modelos de pasos de validación anteriores y conserve únicamente el mejor modelo.

Tras terminar de modificar el archivo de configuración YML, hemos procedido a iniciar el entrenamiento con la orden «./marian -c config-transformer.yml». Una vez verificado que no se producía ningún error, hemos detenido el entrenamiento al cabo de unos minutos. La Figura 1 y la Figura 2, que se incluyen a continuación, muestran el inicio del entrenamiento en la terminal de Ubuntu.



```
jonmikel@jonmikel-X555LAB: ~/Escritorio/preprocesamiento
jonmikel@jonmikel-X555LAB:~$ cd /home/jonmikel/Escritorio/preprocesamiento
jonmikel@jonmikel-X555LAB:~/Escritorio/preprocesamiento$ ./marian -c config-transformer.yml
[2023-05-02 18:25:35] [marian] Marian v1.11.0 f0d0621 2022-02-08 08:39:24 -0800
[2023-05-02 18:25:35] [marian] Running on jonmikel-X555LAB as process 4841 with command line:
[2023-05-02 18:25:35] [marian] ./marian -c config-transformer.yml
[2023-05-02 18:25:35] [config] after: 0e
[2023-05-02 18:25:35] [config] after-batches: 0
[2023-05-02 18:25:35] [config] after-epochs: 0
[2023-05-02 18:25:35] [config] all-caps-every: 0
[2023-05-02 18:25:35] [config] allow-unk: false
[2023-05-02 18:25:35] [config] authors: false
[2023-05-02 18:25:35] [config] beam-size: 12
[2023-05-02 18:25:35] [config] bert-class-symbol: "[CLS]"
[2023-05-02 18:25:35] [config] bert-mask-symbol: "[MASK]"
[2023-05-02 18:25:35] [config] bert-masking-fraction: 0.15
[2023-05-02 18:25:35] [config] bert-sep-symbol: "[SEP]"
[2023-05-02 18:25:35] [config] bert-train-type-embeddings: true
[2023-05-02 18:25:35] [config] bert-type-vocab-size: 2
[2023-05-02 18:25:35] [config] build-info: ""
[2023-05-02 18:25:35] [config] check-gradient-nan: false
[2023-05-02 18:25:35] [config] check-nan: false
[2023-05-02 18:25:35] [config] cite: false
[2023-05-02 18:25:35] [config] clip-norm: 1
[2023-05-02 18:25:35] [config] cost-scaling:
[2023-05-02 18:25:35] [config] []
[2023-05-02 18:25:35] [config] cost-type: ce-sum
[2023-05-02 18:25:35] [config] cpu-threads: 1
[2023-05-02 18:25:35] [config] data-threads: 8
[2023-05-02 18:25:35] [config] data-weighting: ""
[2023-05-02 18:25:35] [config] data-weighting-type: sentence
[2023-05-02 18:25:35] [config] dec-cell: gru
[2023-05-02 18:25:35] [config] dec-cell-base-depth: 2
[2023-05-02 18:25:35] [config] dec-cell-high-depth: 1
[2023-05-02 18:25:35] [config] dec-depth: 1
[2023-05-02 18:25:35] [config] devices:
[2023-05-02 18:25:35] [config] - 0
[2023-05-02 18:25:35] [config] dtm-emb: 512
[2023-05-02 18:25:35] [config] dtm-rnn: 1024
```

```
jonmikel@jonmikel-X555LAB: ~/Escritorio/preprocesamiento
[2023-05-02 18:25:35] [config] valid-script-args:
[2023-05-02 18:25:35] [config] []
[2023-05-02 18:25:35] [config] valid-script-path: ""
[2023-05-02 18:25:35] [config] valid-sets:
[2023-05-02 18:25:35] [config] - val.sp.fr
[2023-05-02 18:25:35] [config] - val.sp.es
[2023-05-02 18:25:35] [config] valid-translation-output: ""
[2023-05-02 18:25:35] [config] vocababs:
[2023-05-02 18:25:35] [config] - vocab-fr.yml
[2023-05-02 18:25:35] [config] - vocab-es.yml
[2023-05-02 18:25:35] [config] word-penalty: 0
[2023-05-02 18:25:35] [config] word-scores: false
[2023-05-02 18:25:35] [config] workspace: 2048
[2023-05-02 18:25:35] [config] Model is being created with Marian v1.11.0 f08d0621 2022-02-08 08:39:24 -0800
[2023-05-02 18:25:35] [training] Using single-device training
[2023-05-02 18:25:35] [comm] Compiled without MPI support. Running as a single process on jonmikel-X555LAB
[2023-05-02 18:25:35] [syncd] seed 1683044735
[2023-05-02 18:25:35] [data] Loading vocabulary from JSON/Yaml file vocab-fr.yml
[2023-05-02 18:25:35] [data] Setting vocabulary size for input 0 to 29,514
[2023-05-02 18:25:35] [data] Loading vocabulary from JSON/Yaml file vocab-es.yml
[2023-05-02 18:25:35] [data] Setting vocabulary size for input 1 to 38,238
[2023-05-02 18:25:35] [data] Using word alignments from file train.sp.fr.es.align
[2023-05-02 18:25:35] [memory] Extending reserved space to 2048 MB (device cpu0)
[2023-05-02 18:25:35] [malloc] large alloc 2147483648 bytes == 0x5f92be000 @
[2023-05-02 18:25:35] [training] Using 1 CPUs
[2023-05-02 18:25:35] [training] Training started
[2023-05-02 18:25:35] [data] Shuffling data
[2023-05-02 18:26:05] [data] Done reading 10,516,743 sentences
[2023-05-02 18:27:32] [data] Done shuffling 10,516,743 sentences to temp files
[2023-05-02 18:27:33] [training] Batches are processed as 1 process(es) x 1 devices/process
[2023-05-02 18:27:34] [logits] Applying loss function for 1 factor(s)
[2023-05-02 18:27:34] [memory] Reserving 336 MB, device cpu0
[2023-05-02 18:27:37] [memory] Reserving 336 MB, device cpu0
[2023-05-02 18:27:42] [memory] Reserving 336 MB, device cpu0
[2023-05-02 18:27:42] [parameter] Parameter type float32, optimization type float32, casting types false
[2023-05-02 18:27:42] [memory] Allocating memory for Adan-specific shards
[2023-05-02 18:27:42] [memory] Reserving 672 MB, device cpu0
```

Figuras 1 y 2: Inicio del entrenamiento del motor TAN con Marian

Una vez finalizada la simulación, el entrenamiento, que ha durado 18 horas, se ha llevado a cabo en las instalaciones de la universidad como hemos comentado. En este primer entrenamiento, se ha aplicado la **ponderación de oraciones** mencionada en la Introducción. Tal y como sugieren Rieß *et al.* (2021), «in order to apply sentence weighting to the translation process, one first needs to come up with a method for scoring sentences with respect to how similar they are to in-domain data». En nuestro caso, hemos decidido otorgar a las oraciones que provienen del corpus especializado un peso (*weight*) de 1 y a las que provienen del corpus general, un peso de 0,5. Esto implica que el motor dará prioridad a los segmentos que pertenezcan al corpus especializado frente a los del corpus general. La Figura 3 nos muestra el aspecto del archivo de texto en formato TXT en el que se indica el peso asignado cada segmento del corpus.

```
train-weights.txt
~/Escritorio/preprocesamiento
Guardar
1 1
2 1
3 1
4 1
5 1
6 1
7 1
8 1
9 1
10 1
11 1
12 1
13 1
14 1
15 1
16 1
17 1
18 1
19 1
20 1
21 1
22 1
23 1
24 1
25 1
Texto plano Anchura del tabulador: 8 Ln 1, Col 2 INS
```

Figura 3: Archivo train-weights.txt

Ulteriormente, hemos vuelto a entrenar el motor, pero sin aplicar la técnica de ponderación de oraciones. En el capítulo 4, comprobaremos si esta técnica consigue mejorar la calidad de las traducciones.

3.4. Puesta en marcha del servidor de MTUOC y traducción

Una vez entrenadas ambas versiones del motor, el siguiente paso ha consistido en la puesta en marcha del servidor de MTUOC ¹⁷. Este servidor es una aplicación que se comunica con servidores de traducción Marian, OpenNMT, Moses y ModernMT para ofrecer servicios de traducción a programas cliente, que puede ser MTUOC-Translator¹⁸ o una herramienta TAO (Oliver, 2022d). Este programa cliente envía una petición al servidor MTUOC. Cuando el servidor recibe el segmento que debe traducir, lo preprocesa y lo envía al servidor de traducción (ya sea, Marian, Moses, OpenNMT o ModernMT). El servidor de traducción devuelve el segmento traducido preprocesado y, por último, el servidor MTUOC lo desprocesa y envía la traducción correcta al programa cliente (Oliver, 2022d). El siguiente diagrama (Figura 4) explica el funcionamiento del servidor MTUOC:

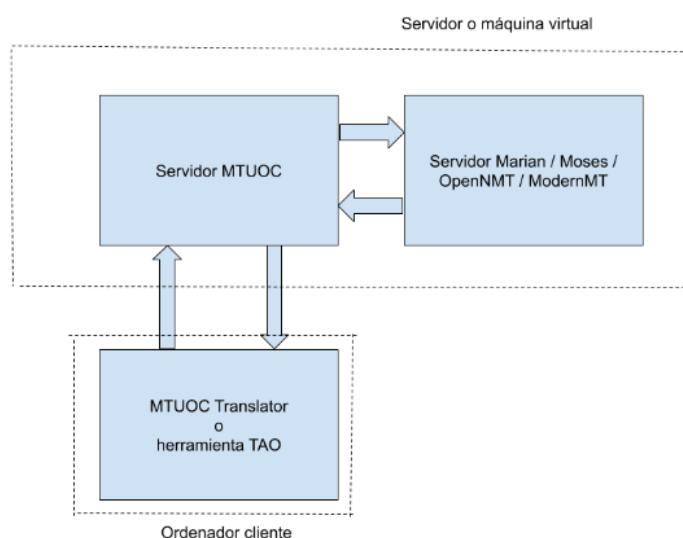


Figura 4: Funcionamiento del servidor MTUOC (Oliver, 2022d)

Después de configurar el archivo *config-server.yaml*, hemos puesto en marcha el servidor ejecutando el *script* MTUOC-server.py. El programa cliente escogido para traducir con el motor entrenado ha sido MTUOC-Translator. Al igual que otras herramientas del proyecto MTUOC, esta también se presenta tanto como

¹⁷ <https://github.com/aoliverg/MTUOC-server> Última consulta: 06/05/2023

¹⁸ <https://github.com/aoliverg/MTUOC-translator> Última consulta: 06/05/2023

script de Python como un ejecutable con interfaz gráfica para Windows. También incluye una versión simplificada que funciona con línea de comandos llamada MTUOC-Translator-lite (Oliver, 2022d). En nuestro caso, hemos optado por ejecutar el *script* de Python en nuestro equipo con Ubuntu.

MTUOC-Translator permite, por una parte, hacer pruebas de traducción con segmentos simples y, por otra, traducir archivos en formato TXT y XLIFF segmentados previamente. La Figura 5 muestra una prueba de traducción en la pestaña «test» del programa.

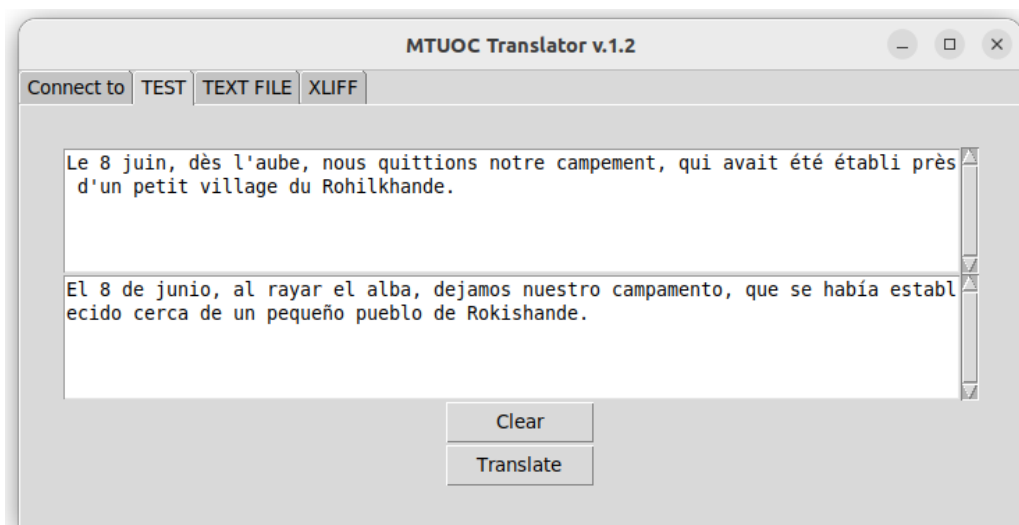


Figura 5: Prueba de traducción en MTUOC-Translator

Una vez comprobado que el motor funciona correctamente, hemos procedido a traducir los 1000 primeros segmentos del archivo *eval.fr* mencionado al final del apartado 3.1. Para obtener esta versión abreviada del archivo (que contenía 5000 segmentos), hemos utilizado la orden «`head -n 1000 eval.fr > eval.fr-curt.txt`». Hemos realizado la misma operación con el archivo *eval.es* para obtener los 1000 segmentos que nos sirven como traducción de referencia en la evaluación del motor.

La traducción de los 1000 primeros segmentos del archivo *eval.fr* en MTUOC-Translator se ha realizado un total de tres veces: la primera, con el motor 1; la segunda, con el motor 2 entrenado con ponderación de oraciones; y la tercera, con el motor 2 entrenado sin ponderación de oraciones.

3.5. Evaluación automática de los motores

La evaluación automática se ha llevado a cabo con la herramienta MTUOC-eval¹⁹. Esta herramienta también se distribuye tanto en versión para línea de comandos como en versión con interfaz de usuario. Hemos utilizado esta última en nuestro equipo con Windows para evaluar automáticamente los siguientes sistemas: motor 1, motor 2 con ponderación de oraciones y motor 2 sin ponderación de oraciones. Las métricas que hemos utilizado en la evaluación han sido BLEU, NIST, WER, TER y la distancia de edición (%EdDist).

En cada ocasión, en la pestaña «Files» del programa, hemos debido seleccionar la traducción de referencia (*References*), los segmentos traducidos automáticamente con cada motor (*Hypothesis*), el tokenizador de la lengua de llegada y los segmentos en lengua origen (*Source*). Incluir este último archivo nos permitía obtener, amén de los resultados de cada métrica, un cálculo detallado para cada segmento en formato XLXS.

Los resultados de esta evaluación automática se presentarán en el apartado 4.1.

3.6. Elaboración de las pruebas de evaluación humana

La evaluación humana del motor de traducción automática se ha realizado mediante la administración de dos encuestas a participantes voluntarias. Para ello, hemos retomado las dos preguntas de investigación que nos planteamos en el apartado 1.4. Para responder a la pregunta P. I. 1., en la que nos cuestionamos si el uso de un corpus especializado de mayor tamaño mejora la calidad de las traducciones del motor entrenado, hemos decidido diseñar una encuesta en la que se comparan el motor 1, entrenado durante las prácticas, y el motor 2 sin ponderación de oraciones. Comparar estos dos motores nos permitirá comprobar que la mejora en la calidad de las traducciones proviene del corpus especializado de mayor tamaño y no de la ponderación de oraciones o de una combinación de ambos motivos.

Por otro lado, en la segunda pregunta de investigación (P. I. 2) nos planteamos si el uso de ponderación de oraciones mejora la calidad de las traducciones de nuestro motor. Para dar respuesta a esta pregunta, en cambio, hemos considerado oportuno diseñar una segunda encuesta en la que se

¹⁹ <https://github.com/aoliverg/MTUOC-eval> Última consulta: 06/05/2023

compara el motor 2 entrenado con ponderación de oraciones con el motor 2 entrenado sin ponderación de oraciones.

Hemos elaborado ambas encuestas con la herramienta Formularios de Google basándonos en estudios similares previos y realizando ligeras modificaciones, tal y como se explicará a continuación.

3.6.1. Características de las pruebas del análisis comparativo

El análisis comparativo es un procedimiento habitual en la evaluación humana de la traducción automática. En este tipo de análisis, se ofrece un segmento original y varias propuestas de traducción, y se ha de determinar cuál de las propuestas es la mejor traducción u ordenarlas (*rank-based evaluation*). Investigadores como Toral y Way (2018) y Briva-Iglesias (2020) han ejecutado este tipo de pruebas. Mientras que estos autores han ofrecido a los participantes tres propuestas de traducción por cada segmento original, autores como Vilar *et al.* (2007) sostienen que un sistema binario de evaluación (dicho de otra forma, ofrecer únicamente dos propuestas de traducción) simplifica la toma de decisiones del evaluador humano y favorece una toma de decisión natural. Mientras que Briva-Iglesias (2020) elabora una prueba en la que los participantes evalúan 100 segmentos aislados, Toral y Way (2018) escogen de forma aleatoria pasajes compuestos de 10 oraciones contiguas con la finalidad de ofrecer mayor contexto a los evaluadores.

En nuestro caso, hemos decidido replicar dos pruebas el sistema binario de Vilar *et al.* (2007) para favorecer la naturalidad de la toma de decisiones de las personas evaluadoras. De esta forma, estas personas solamente han debido escoger la mejor traducción entre dos propuestas por cada fragmento original que se les presentó. Si bien comprendemos el motivo por el cual Toral y Way (2018) escogen series de oraciones contiguas, consideramos que incluir en cada pregunta fragmentos de 10 segmentos seguidos podía alargar en gran medida la prueba y abrumar a las personas evaluadoras que participan voluntariamente. Por esta razón, en nuestras pruebas se les han ofrecido únicamente segmentos aislados como en la investigación de Briva-Iglesias (2020). Creemos que esto ha podido contribuir a agilizar la toma de decisiones en la misma línea de lo que sugieren Vilar *et al.* (2007).

De este modo, en la **primera prueba** se les presentaron a las personas evaluadoras 30 segmentos originales en francés escogidos de forma aleatoria del archivo *eval.fr* del corpus de evaluación y, por cada uno de ellos, dos propuestas de traducción: una proveniente del motor 1 y la otra, del motor 2 entrenado sin ponderación de oraciones. En la **segunda prueba**, en cambio, se les presentaron otros 30 segmentos originales en francés y dos propuestas por cada uno de ellos: una procedente del motor 2 entrenado con ponderación de oraciones y la otra, del motor 2 sin ponderación de oraciones. En ningún momento se les ha dado a conocer el origen de las traducciones a las personas evaluadoras, aunque en la descripción de la prueba sí se ha señalado que se estaba investigando en torno a la calidad que ofrecen motores de TAN en la traducción de textos literarios. Los segmentos originales no se repiten entre ambas pruebas.

Aunque estas dos pruebas se hubieran podido presentar en una sola prueba de 60 segmentos, consideramos que presentarla como dos pruebas independientes permitiría a las evaluadoras realizarlas en días diferentes y, por lo tanto, se reduciría el riesgo de que la fatiga de las evaluadoras afectara al resultado de las pruebas.

Además, para diseñar las preguntas, hemos optado por un diseño cruzado (*crossover design*) y presentar las propuestas de traducción como «ES A» y «ES B», en las que las propuestas de «ES A» pueden provenir tanto de un motor como de otro. Lo mismo ocurre con las propuestas presentadas como «ES B». De esta manera, hemos evitado el sesgo por *order effect* or *sequence effect* que menciona MacKenzie (2013) y que se produciría si las personas evaluadoras percibieran que la propuesta presentada como A es siempre peor que B o viceversa. Para mantener la trazabilidad y saber a qué motor corresponde cada propuesta a la hora de analizar los resultados, hemos elaborado una tabla de MS Excel con las correspondencias entre cada propuesta y el motor del que proceden.

Por último, para determinar qué traducción es la mejor, se les facilitó a las participantes las siguientes indicaciones replicadas del estudio conducido por Toral y Way (2018):

- La mejor traducción será la que suene más natural y conserve mejor el significado del original sin tener en cuenta los errores (léxicos, sintácticos, etc.) que pueda contener.
- Si las dos traducciones conservan por igual el significado del original y suenan natural, la mejor traducción será la que menos errores contenga.

3.6.2. Características de las personas evaluadoras

Estas pruebas han sido administradas a tres personas evaluadoras con estudios universitarios en Traducción y que se dedican profesionalmente a la traducción, a la docencia en estudios de Traducción o a la investigación en Traducción. Las tres participantes tienen el francés como una de sus lenguas de trabajo y han accedido a participar en las pruebas de manera voluntaria.

Hemos estimado que la participación de tres personas posibilitaría el acuerdo entre evaluadores (*inter-annotator agreement*) y que, al evitar extraer conclusiones a partir de lo que juzga únicamente una persona, aumentaría el grado de objetividad de las pruebas.

Antes de realizar dichas pruebas, estas tres personas han debido declarar su consentimiento para el tratamiento de datos personales. En el anexo II, se incluye la declaración de consentimiento y, en los anexos III y IV, las dos encuestas administradas a las participantes.

4. Resultados

En este capítulo, presentamos, por una parte, los resultados de la evaluación automática realizada con la herramienta MTUOC-eval y, por otra, los resultados de la evaluación humana, tras la administración de las dos pruebas descritas.

4.1. Resultados de la evaluación automática

En este apartado, se muestran los resultados de la evaluación automática. Recordamos que en esta evaluación hemos comparado con una traducción humana de referencia las traducciones obtenidas por tres motores de TAN: motor 1, motor 2 con ponderación de oraciones y motor 2 sin ponderación de oraciones. Los resultados se mostrarán en este mismo orden. Asimismo, incluimos la Tabla 7 que proporciona un resumen de los valores óptimos de las métricas empleadas extraída de Martínez (2023). Para obtener una explicación más detallada de cada métrica, véase el apartado 2.2.

Métrica	Explicación del valor óptimo
BLEU	Los valores más cercanos a 1 indican mayor similitud con la traducción de referencia.
NIST	A mayor puntuación, mayor el número de coincidencias con la traducción de referencia.
WER	Los valores más cercanos a 0 indican menor número de modificaciones en relación con la traducción de referencia.
%EdDist	Cuanto menor sea el valor, menor es la distancia de edición para obtener la traducción de referencia.
TER	Los valores más cercanos a 0 indican menor número de modificaciones en relación con la traducción de referencia.

Tabla 7: Resumen de los valores óptimos de cada métrica (Martínez, 2023)

A continuación, en la Tabla 8, presentamos los resultados obtenidos mediante MTUOC-eval. La mejor puntuación obtenida con cada métrica se resalta en negrita y en color verde, y la segunda mejor puntuación, en color azul:

	Motor 1	Motor 2 (con ponderación)	Motor 2 (sin ponderación)
BLEU	0,2725	0.3701	0.3668
NIST	6,314	7.595	7.56
WER	0,6157	0.5252	0.5292
%EdDist	50,7	43.52	43.36
TER	0,5739	0.4863	0.4894

Tabla 8: Resultados de la evaluación automática

En la tabla, podemos observar que el motor 2, entrenado en el contexto del presente trabajo, es el que ha obtenido los mejores resultados y que supera en todas las métricas al motor 1, entrenado durante las prácticas curriculares con un corpus especializado de únicamente 37 541 segmentos.

En lo que respecta a las diferencias entre las dos variantes del motor 2 (con y sin ponderación de oraciones), observamos que el motor 2 entrenado con ponderación de oraciones ofrece los mejores valores en cuatro de las cinco métricas empleadas, mientras que el motor 2 sin ponderación de oraciones solo obtiene la mejor puntuación en una de ellas. Sin embargo, las diferencias

obtenidas en esta evaluación automática no son significativas, ya que las dos variantes del motor presentan valores muy similares.

Los anexos V, VI y VII incluyen una muestra de los resultados detallados obtenidos con MTUOC-eval de 20 segmentos traducidos con el motor 1, el motor 2 con ponderación de oraciones y el motor 2 sin ponderación de oraciones respectivamente. En estas muestras, se resaltan en azul las adiciones y en rojo las eliminaciones necesarias para que la traducción ofrecida por el motor en cuestión coincida por completo con la traducción de referencia.

4.2. Resultados de la evaluación humana

Seguidamente, presentaremos los resultados que ha logrado cada motor en las pruebas realizadas en la evaluación humana. Los datos obtenidos en las encuestas se han interpretado de la siguiente manera:

- Si las tres evaluadoras han considerado, por ejemplo, que la propuesta A es mejor que la B, se le ha otorgado un punto al motor del que procedía la propuesta A.
- Si dos evaluadoras han considerado, por ejemplo, que la propuesta A es mejor que la B y una que la B es mejor que la A, se le ha otorgado el punto al motor del que procedía la propuesta A.

Por consiguiente, y puesto que cada encuesta constaba de 30 preguntas, cada motor ha podido conseguir en cada encuesta un máximo de 30 puntos.

En lo que respecta a la **primera prueba**, en la que se comparaban el motor 1 y el motor 2 sin ponderación de oraciones, el motor 2 sin ponderación de oraciones ha logrado un total de 23 puntos, mientras que el motor 1 solo ha obtenido 7. El siguiente gráfico (Figura 6) muestra el porcentaje de las veces en las que cada motor ha sido considerado el mejor:

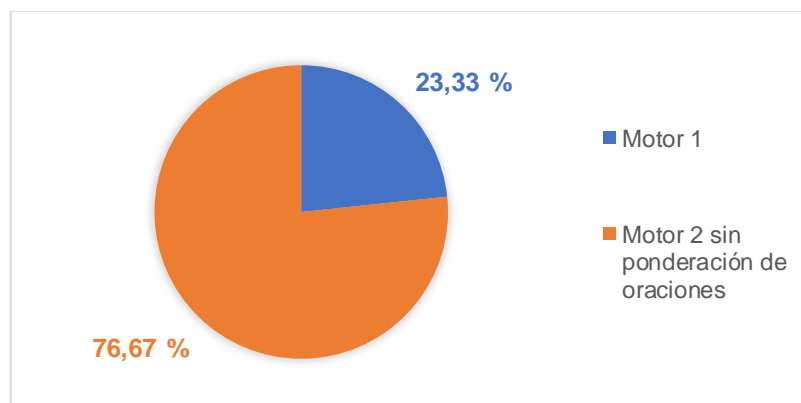


Figura 6: Resultados de la prueba 1 de la evaluación humana

Asimismo, hemos calculado la medida estadística kappa de Fleiss (1971) con el fin de conocer el nivel de acuerdo entre las evaluadoras. Para ello, hemos utilizado la calculadora en línea Online Kappa Calculator²⁰. En esta primera prueba, las evaluadoras han alcanzado un nivel de acuerdo de un 88,22 % (un valor de kappa de margen libre o *free-marginal kappa* de 0,64). Según Landis y Koch (1977), un valor de kappa superior a 0,6 se interpreta como un nivel de acuerdo importante (*substantial*).

La Tabla 9 recoge dos de los ejemplos en los que las evaluadoras han coincidido en que la traducción del motor 2 sin ponderación de oraciones es de una calidad mayor, probablemente, porque mientras que el motor 1 ha devuelto traducciones sin sentido, el motor 2 ha mantenido el significado del original:

FR:	Duval ne vous dit rien, vous le bouleversez tellement qu'il ne trouve pas un mot.
ES (Motor 1):	Duald no te dice nada, te lo vas a detener tanto que no encuentra ni una palabra.
ES (Motor 2 sin PO):	Duval no te dice nada, lo molestas tanto que no encuentra ni una palabra.

FR:	Devinez, lui dis-je, à qui je viens de téléphoner ?
ES (Motor 1):	Adivina, ¿de quién acabo de llamar?
ES (Motor 2 sin PO):	—¿Sabe usted —le dije— a quien acabo de telefonear?

Tabla 9: Ejemplos en los que el motor 2 sin ponderación de oraciones supera al motor 1

Nótese, además, que en el segundo ejemplo el motor 2 ha sabido reproducir correctamente las marcas de diálogo en español. No obstante, también encontramos dos ocasiones en las que las evaluadoras han estimado que el motor 1 ha devuelto una traducción de una calidad superior, como observamos en la Tabla 10:

FR:	Chaque fois que Swann lui avait demandé un service, ne le lui avait-il pas toujours refusé ?
ES (Motor 1):	Cada vez que Swann le había pedido un favor, ¿no se lo había negado siempre?
ES (Motor 2 sin PO):	Cada vez que Swann le había pedido un favor, ¿no lo había negado siempre?

²⁰ <http://justusrandolph.net/kappa/> Última consulta: 29/05/2023

FR:	L'Anglais saisit le paquet, le déficela, enleva les deux feuilles de papier qui l'enveloppaient.
ES (Motor 1):	El inglés cogió el paquete y le quitó las dos hojas de papel que lo envolvía.
ES (Motor 2 sin PO):	El inglés cogió el paquete, el desafío, quitó las dos hojas de papel que lo envolvían.

Tabla 10: Ejemplos en los que el motor 1 supera al motor 2 sin ponderación de oraciones

En ambos casos, los dos motores han ofrecido traducciones muy similares, pero con ciertas diferencias que han hecho que las evaluadoras se inclinen a favor del motor 1. En el primer ejemplo, esta decisión se debe seguramente a que el motor 1 ha incluido el pronombre personal *se* con función de complemento indirecto, mientras que el motor 2 no incluye este pronombre. En el segundo ejemplo, la decisión se debe probablemente a que el motor 2 ha intentado traducir «le déficela» (forma verbal del verbo *déficeler*) como «el desafío», tal vez por su parecido con la palabra *défi*, en vez de como «lo desató». El motor 1, en cambio, ha optado por omitir «le déficela» en su traducción, lo cual hace que el resultado sea mejor para el lector, a pesar del error de concordancia que observamos seguidamente en «envolvía» (en lugar de «envolvían»).

Con relación a los resultados de la **segunda prueba**, en la que se enfrentaban las dos variantes del motor 2 (con y sin ponderación de oraciones), la variante con ponderación de oraciones ha logrado un total de 22 puntos. Por el contrario, la variante sin ponderación de oraciones ha obtenido 8. En la Figura 7, presentamos un gráfico con los porcentajes que corresponden a cada motor en la segunda prueba:

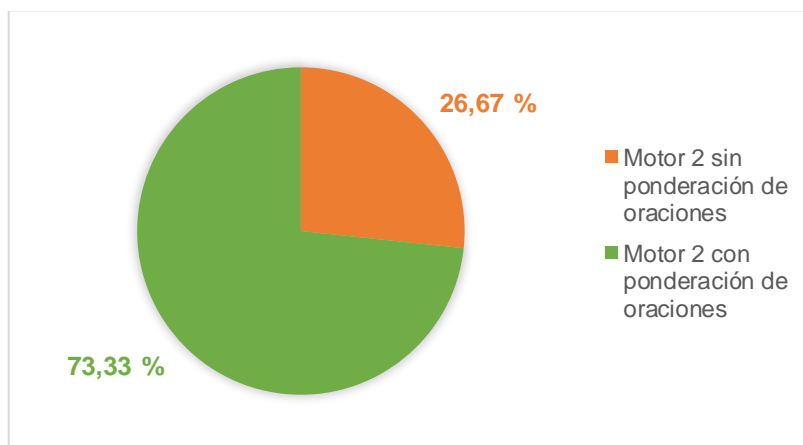


Figura 7: Resultados de la prueba 2 de la evaluación humana

En esta segunda prueba, también hemos calculado el nivel de acuerdo entre evaluadoras mediante la herramienta mencionada. En este caso, el nivel de acuerdo entre evaluadoras ha sido del 66,67 % (un valor de kappa de margen libre de 0,33). Este valor debe interpretarse como un nivel de acuerdo razonable (*fair*) (Landis y Koch,1977). Mientras que, en la primera encuesta, las tres evaluadoras marcaron la misma opción en 21 preguntas, en la segunda, las tres coincidieron en 15 preguntas. Esto explica que el valor de kappa de Fleiss sea menor en la segunda prueba.

La Tabla 11 muestra dos ejemplos en los que las evaluadoras han escogido la traducción propuesta por el motor 2 con ponderación de oraciones como la mejor:

FR:	Dans ces jours de clarté, de chaleur et de sérénité, il y a une certaine heure surtout où il faut admirer le portail de Notre-Dame.
ES (Motor 2 con PO):	En estos días de claridad, de calor y de serenidad, hay una cierta hora sobre todo en que hay que admirar el pórtico de Notre Dame.
ES (Motor 2 sin PO):	En estos días de claridad, de calor y de serenidad, hay una cierta hora sobre todo en que hay que admirar el portal de Nuestra Señora.

FR:	Si votre père est entre les mains des indigènes, ce que fait pressentir d'ailleurs le document, nous le retrouverons.
ES (Motor 2 con PO):	Si vuestro padre está en manos de los indígenas, lo que hace presentir el documento, lo encontraremos.
ES (Motor 2 sin PO):	Si su padre está en manos de los indígenas, lo que presentirán el documento, lo encontraremos.

Tabla 11: Ejemplos en los que el motor 2 con ponderación de oraciones supera al motor 2 sin ponderación de oraciones

Como podemos apreciar, las diferencias entre estas dos variantes del motor son más sutiles. En el primer ejemplo, la única diferencia entre las dos propuestas es que la variante con ponderación de oraciones ha mantenido «Notre Dame» en la traducción (sin guion, tal y como se suele emplear en español). En cambio, la variante sin ponderación de oraciones lo ha traducido como «Nuestra Señora». En el segundo ejemplo, la variante sin ponderación de oraciones no ha sabido traducir «ce que fait pressentir d'ailleurs le document», además de que ha optado por el ustedeo, mientras que la variante con ponderación de oraciones ha

empleado el voseo reverencial propio de los textos literarios que reflejan el lenguaje de otras épocas.

Por último, la Tabla 12 recoge los dos únicos casos en los que las tres evaluadoras han coincidido en que la variante sin ponderación de oraciones presenta una traducción de mayor calidad:

FR:	Dans cette première réunion, on convint de choisir l'avocat Savaron pour candidat, avec un enthousiasme que personne n'aurait pu espérer à Besançon.
ES (Motor 2 con PO):	En esta primera reunión se convino en elegir al abogado Savaron para candidato, con un entusiasmo que nadie hubiera podido esperar en Besanzón.
ES (Motor 2 sin PO):	En esta primera reunión se convino en elegir al abogado Savaron para candidato, con un entusiasmo que nadie hubiera podido esperar en Besançon.

FR:	Mais soudain il parut stupéfait, ses yeux s'agrandirent, et il bégaya :
ES (Motor 2 con PO):	Pero, de pronto, pareció estupefacto, sus ojos se agrandaron, y tartamudeó:
ES (Motor 2 sin PO):	Pero, de pronto, pareció estupefacto, sus ojos se agrandaron y tartamudeó:

Tabla 12: Ejemplos en los que el motor 2 sin ponderación de oraciones supera al motor 2 con ponderación de oraciones

Como observamos, las dos variantes del motor han ofrecido traducciones prácticamente idénticas, a diferencia de ciertos matices. En el primer ejemplo, la variante sin ponderación de oraciones ha mantenido el topónimo en francés (Besançon). La variante con ponderación de oraciones, en cambio, ha utilizado el topónimo en español (Besanzón). En el segundo ejemplo, la variante con ponderación de oraciones ha calcado el uso de la coma del original al incluir una coma antes de la conjunción y, lo cual en este caso se considera un error en español.

Los resultados detallados de cada prueba están recogidos en los anexos VIII y IX.

5. Conclusiones

Tras haber analizado todos los resultados obtenidos tanto en la evaluación automática como en la evaluación humana, en este capítulo trataremos las

conclusiones que extraemos en relación con los objetivos, las preguntas de investigación e hipótesis planteadas en los apartados 1.3. y 1.4., así como las limitaciones del presente trabajo y las posibles vías de investigación que surgen de él.

En cuanto a los **objetivos**, en la etapa inicial del trabajo nos planteamos un objetivo principal y dos secundarios. Por una parte, entrenar un motor de TAN para la traducción de texto literarios del francés al español por medio de un corpus paralelo alineado con el algoritmo SBERT y una técnica de ponderación de oraciones (O. P. 1). Por otra parte, realizar una evaluación automática (O. S. 1) y humana (O. S. 2) de la calidad de las traducciones del motor entrenado, y compararla con la de otro motor entrenado previamente (motor 1) y con la del mismo motor entrenado sin ponderación de oraciones. Estos tres objetivos han sido alcanzados siguiendo los pasos descritos en el capítulo 3. En este sentido, consideramos que las herramientas propias del proyecto MTUOC han resultado de gran utilidad y una elección adecuada, ya que cumplen su función de poner a disposición de todas las personas esta tecnología sin necesidad de poseer conocimientos informáticos avanzados. Aun así, estimamos que el proceso sí exige un dominio mínimo de la línea de comandos y del lenguaje de programación Python.

En lo que respecta a la **primera pregunta de investigación** (P. I. 1), en la Tabla 8 hemos podido observar que las dos variantes del motor 2 obtienen resultados mejores que el motor 1 en todas las métricas de la evaluación automática. Además, los resultados de la primera prueba de la evaluación humana parecen corroborar estos resultados, dado que las evaluadoras consideraron que el motor 2 sin ponderación de oraciones ofrecía mejores traducciones en un 76,67 % de los casos, con un alto nivel de acuerdo entre ellas. Puesto que la diferencia entre ambos motores comparados es el tamaño del corpus especializado (un 5,38 % del corpus total frente a un 0,16 %), los resultados parecen indicar que cuanto mayor es el tamaño del corpus especializado empleado en el entrenamiento de un motor de TAN, mayor es la calidad de las traducciones que ofrece el motor. De esta forma, podríamos concluir que la hipótesis H₁ queda demostrada.

Con relación a la **segunda pregunta de investigación** (P. I. 2), la Tabla 8 del apartado 4.1 no presenta resultados significativos que nos permitan decidir

si la variante con ponderación de oraciones del motor 2 ofrece efectivamente mejores traducciones que la variante sin ponderación de oraciones. A pesar de que la variante con ponderación de oraciones ha obtenido mejores resultados en cuatro de las cinco métricas, los resultados de ambas variantes en la evaluación automática son demasiado similares como para extraer conclusiones a partir de ellos. En este caso, los resultados de la evaluación humana han sido clave para determinar cuál de las dos variantes devuelve traducciones de mayor calidad. Según los resultados obtenidos en la segunda encuesta de la evaluación humana, la variante con ponderación de oraciones ofrece mejores traducciones en el 73,33 % de los casos. Aun así, creemos que es necesario mencionar que el nivel de acuerdo entre las evaluadoras es menor en esta segunda prueba. Esto se debe seguramente a que, como hemos podido apreciar, la diferencia en la calidad de las traducciones entre ambas variantes del motor no es tan evidente como lo era entre los motores comparados en la encuesta 1, ya que, a menudo, se basa en ligeros matices. En cualquier caso, los resultados obtenidos en la evaluación humana parecen señalar que el empleo de una técnica de ponderación de oraciones en la fase de entrenamiento de un motor de TAN mejora la calidad de las traducciones de dicho motor y, por lo tanto, la hipótesis H₂ también quedaría demostrada.

Asimismo, podemos concluir que hemos logrado mejorar los resultados que obtuvimos durante las prácticas curriculares en las que entrenamos el motor 1, entrenando esta vez un motor con un corpus especializado mayor y empleando una técnica de ponderación de oraciones. También hemos contribuido al ámbito de estudio al incorporar un nuevo par de lenguas (francés-español), el algoritmo SBERT en la alineación de un corpus paralelo de contenido literario, una técnica de ponderación de oraciones y una evaluación humana que confirma los resultados de la evaluación automática. Estos elementos no habían sido introducidos en los trabajos de fin de máster presentados en esta universidad hasta la fecha, a excepción del algoritmo SBERT que fue incluido por Martínez (2023), aunque aplicado a un corpus paralelo de textos turísticos. Estas aportaciones revisten especial interés si tenemos en cuenta las vías de investigación propuestas por Moreno (2022), quien, tras entrenar un motor especializado en textos literarios, sugería la necesidad de emplear una técnica de alineación más avanzada y de ejecutar una evaluación humana en futuros

estudios. Puesto que en nuestra investigación no hemos experimentado los problemas derivados de una alineación deficiente a los que se refiere Moreno (2022) en su estudio, podemos concluir que decantarnos por el algoritmo SBERT ha sido una decisión acertada, tal y como parecían sugerir las conclusiones de Martínez (2023).

No obstante, y pese a que los resultados obtenidos en este estudio exploratorio son positivos, somos conscientes de que la calidad de las traducciones que ofrece nuestro motor aún dista mucho de la calidad de las traducciones humanas, máxime en una especialidad en la que la creatividad, el estilo del propio autor y la función estética del lenguaje son primordiales. Tampoco consideramos que las traducciones de nuestro motor puedan suponer un buen punto de partida para la posesición, puesto que, amén de contener errores que exigirían un cierto esfuerzo de posesición, condicionarían al traductor humano que seguramente daría con propuestas de traducción más originales si partiera de cero.

5.1. Limitaciones del trabajo

Este trabajo ha tenido diversas limitaciones que debemos mencionar en este apartado, tanto a la hora de crear el corpus como en otras fases del estudio, como la evaluación.

Por un lado, los derechos de autor de las obras han sido la mayor limitación que hemos tenido a la hora de confeccionar nuestro corpus. Para no infringir derechos de autor, solo hemos utilizado obras que pertenecieran al dominio libre. Estos derechos caducan al pasar 70 años desde el fallecimiento del autor, lo cual nos ha condicionado sobremanera, ya que únicamente nos hemos podido servir de obras antiguas en las que, a menudo, encontrábamos un lenguaje arcaizante y expresiones que han caído en desuso.

Por otro lado, la evaluación humana que hemos llevado a cabo en este estudio no deja de ser un mero acercamiento, puesto que solo hemos realizado un estudio comparativo de un número de segmentos reducido. La evaluación humana es un proceso complejo en el que se pueden ejecutar, además de estudios comparativos más exhaustivos, estudios absolutos en los que se midan aspectos como la precisión y la fluidez de las traducciones. Abarcar todos estos estudios en la evaluación humana nos habría obligado a dejar de lado otras

cuestiones, como la creación del corpus y la evaluación automática, que también considerábamos relevantes, dadas las características propias de un trabajo de fin de máster. Por este motivo, hemos conducido una evaluación humana sencilla que simplemente nos permitiera corroborar los resultados obtenidos en la evaluación automática.

Del mismo modo, hemos decidido no incluir ningún aspecto que guardara relación con la posesición, como el estudio del esfuerzo de posesición o el nivel de productividad. Esta decisión se debe a la complejidad de este proceso, que puede investigarse desde muchos ámbitos y ángulos diferentes. Si hubiésemos decidido incluir alguno de estos aspectos en un trabajo de fin de máster como este, habríamos tenido que centrarnos en el estudio de la posesición descartando otros aspectos de la investigación.

Por último, a pesar de que hemos querido incorporar un nuevo par de lenguas como novedad frente a estudios previos, sentimos que, al emplear técnicas diferentes durante el entrenamiento del motor, nuestros resultados no son comparables con los de motores entrenados en otros estudios como el de Moreno (2022). Para poder realizar comparaciones y estimar, por ejemplo, si la diferencia en la calidad se debe al grado de proximidad entre las lenguas, se deberían entrenar los dos motores en igualdad de condiciones.

5.2. Futuras vías de investigación

Las limitaciones que ha sufrido este trabajo son, a su vez, futuras oportunidades de investigación. Una vez descritas las limitaciones, estas son las posibles vías de investigación que surgen de este estudio:

- Elaborar un corpus con obras más actuales que reflejen un uso del lenguaje más moderno. Para ello, debería estudiarse alguna forma de salvar la cuestión de los derechos de autor, como poseer las obras originales.
- Realizar un análisis comparativo más exhaustivo, evaluando una cantidad de segmentos mayor por más personas evaluadoras. Asimismo, sería interesante observar los motivos concretos a los que se debe el acuerdo o el desacuerdo entre las personas evaluadoras en un número mayor de segmentos.

- Realizar un análisis absoluto de conceptos como la precisión o la fluidez de las traducciones.
- Analizar cuestiones que guarden relación con la posesición, como estudiar el esfuerzo de posesición, el nivel de productividad o las necesidades de los poseedores a la hora de poseer textos literarios traducidos automática.
- Comparar los resultados obtenidos con motores entrenados con corpus paralelos en diferentes pares de lenguas, pero empleando las mismas técnicas en el proceso de entrenamiento. Esto nos permitiría estudiar, por ejemplo, cómo afecta el grado de proximidad entre lenguas a la calidad de las traducciones.

Bibliografia

- Adamopoulou, E. y Moussiades, L. (2020). An Overview of Chatbot Technology. En: *IFIP Advances in Information and Communication Technology* (vol. 584). https://link.springer.com/chapter/10.1007/978-3-030-49186-4_31
- Allen, J. (2003). Post-editing. En H. Somers (ed). *Computers and Translation: A Translator's Guide* (pp. 297-318), Ámsterdam: John Benjamins Publishing
- Bentivogli, L. *et al.* (2016). Neural versus phrase-based machine translation quality : a case study. En: *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing* (pp. 257-267). Austin: Association for Computational Linguistics. <https://aclweb.org/anthology/D16-1025>
- Bernardo, F. L. (2014). Tradução literária homem vs. máquina: um ensaio sobre a tradução online. *Entretextos* (vol. 14, n. 2, pp. 191-204). <https://cutt.ly/lfP43Tf>
- Besacier, L. (2014). *Traduction automatisée d'une oeuvre littéraire: une étude pilote*. <https://cutt.ly/TfP4P54>
- Briva-Iglesias, V. (2020). Traducción automática inglés-catalán: tecnología de vanguardia, calidad y productividad. [Trabajo de fin de máster]. Universitat Autònoma de Barcelona. <https://ddd.uab.cat/record/249932>
- Casacuberta, F. y Peris, A. (2017). *Traducción automática neuronal*. Revista Tradumàtica. *Tecnologies de la Traducció*, n. 15, pp. 66-74. <https://www.raco.cat/index.php/Tradumatica/article/download/n15-casacuberta-peris/422330/>
- Escribe, M. (2019). Human Evaluation of Neural Machine Translation: The Case of Deep Learning. En: Temnikova, I. *et al.* (2019). *Proceedings of the Human-Informed Translation and Interpreting Technology Workshop* (pp. 36-46).
- Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin* (vol. 16, n. 5, pp. 378-382). <https://psycnet.apa.org/record/1972-05083-001>

- Goutte, C. (2006). *Automatic Evaluation of Machine Translation Quality*. https://www.researchgate.net/publication/245575542_Automatic_Evaluation_of_Machine_Translation_Quality
- Guerberof, A. y Toral, A. (2022a). Creativity in translation - Machine translation as a constraint for literary texts. *Translation Spaces* (vol. 11, no. 2, pp. 184-212). <https://www.jbe-platform.com/content/journals/10.1075/ts.21025.gue>
- Guerberof, A. y Toral, A. (2022b). CREAMT: Creativity and narrative engagement of literary texts translated by translators and NMT. En: *Proceedings of the 23rd Annual Conference of the European Association for Machine Translation* (pp. 357–358). Gante: European Association for Machine Translation. <https://aclanthology.org/2022.eamt-1.68.pdf>
- Hearne, M., y Way, A. (2011). Statistical Machine Translation: A Guide for Linguists and Translators: SMT for Linguists and Translators. *Language and Linguistics Compass*, 5 (5), 205-226. <https://doi.org/10.1111/j.1749-818X.2011.00274.x>
- Jones, R. e Irvine, A. (2013). The (Un)faithful Machine Translator. En: *Proceedings of the 7th Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities* (pp. 96-101). Sofía: Association for Computational Linguistics. <https://cutt.ly/vfP487b>
- Koehn, P. (2010). *Statistical Machine Translation*. Cambridge: Cambridge University Press
- Koehn, P. (2011). *What is a Better Translation? Reflections on Six Years of Running Evaluation Campaigns*. Symposium on Translation Careers and Technologies (Tralogy), 3-4 de marzo de 2011, París. <https://aclanthology.org/www.mt-archive.info/Tralogy-2011-Koehn.pdf>
- Landis, J. R. y Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics* (vol. 33, n. 1, pp. 159-174). <https://www.jstor.org/stable/2529310>
- MacKenzie, I. S. (2013). *Human-Computer Interaction: An Empirical Research Perspective*. USA: Elsevier. <http://www.yorku.ca/mack/HCIbook/>

- Martínez Mateo, R. (2014). *A deeper look into metrics for translation quality assessment (TQA): A case study*. Universidad de Castilla-La Mancha <https://dialnet.unirioja.es/servlet/articulo?codigo=5828441>
- Martínez, T. (2023). *Creación de un corpus paralelo y entrenamiento de un motor de traducción neuronal para la traducción de textos turísticos del castellano al alemán en el marco del proyecto MTUOC*. [Trabajo de fin de máster]. Universitat Oberta de Catalunya.
- Moreno, L. (2022). *La traducción automática en la literatura: Creación de un corpus paralelo inglés-español para entrenar y evaluar un sistema neuronal en el marco del proyecto MTUOC*. [Trabajo de fin de máster]. Universitat Oberta de Catalunya. <https://openaccess.uoc.edu/handle/10609/146721>
- O'Brien, S. (2011) Towards Predicting Post-Editing Productivity. *Machine Translation*, 25 (1), pp. 197- 225.
- Oliver González, A. (2021). *MTUOC - Sistemas de traducción automática estadísticos y neuronales—XWiki*. Universitat Oberta de Catalunya. <https://xwiki.recursos.uoc.edu/wiki/mat00001ca/view/3.%20Tutorials/MTUOC%20-%20Sistemas%20de%20traducci%C3%B3n%20autom%C3%A1tica%20estad%C3%ADsticos%20y%20neuronales/>
- Oliver, A. (2023). *MTUOC-aligner* [Python]. <https://github.com/aoliverg/MTUOC-aligner>
- Oliver, A. *et al.* (2022). Bilingual e-books via neural machine translation and their reception. En J. L. Hadley *et al.* (Eds.), *Using Technologies for Creative-Text Translation* (pp. 89-115). Routledge. <https://doi.org/10.4324/9781003094159-5>
- Oliver, A. (2022a). Instrucciones de uso. *MTUOC-corpus-combination*. <https://github.com/aoliverg/MTUOC-corpus-combination/wiki/Instrucciones-de-uso>
- Oliver, A. (2022b). Pasos genéricos del preprocesamiento. *Curso práctico de TA*. <https://github.com/aoliverg/a-practical-course-on-machine-translation/wiki/C1.-Pasos-gen%C3%A9ricos-de-procesamiento>

- Oliver, A. (2022c). Entrenamiento con Marian. *Curso práctico de TA*. https://github.com/aoliverg/a_practical_course_on_machine_translation/wiki/C3.-Entrenamiento-con-Marian
- Oliver, A. (2022d). Puesta en marcha de MTUOC. *Curso práctico de TA*. https://github.com/aoliverg/a_practical_course_on_machine_translation/wiki/F2.-Instalaci%C3%B3n-y-puesta-en-marcha-de-MTUOC
- Quah, C. K. (2006). *Translation and Technology*. Nueva York: Palgrave MacMillan
- Rieß et al. (2021). A Comparison of Sentence-Weighting Techniques for NMT. En: *Proceedings of Machine Translation Summit XVIII: Research Track* (pp. 176-187). Association for Machine Translation in the Americas. <https://aclanthology.org/2021.mtsummit-research.15/>
- Rodriguez, C. (2021). *Creación de motores de traducción automática (estadística y neuronal) inglés-español especializados en el campo de la aviación con la herramienta MTUOC*. [Trabajo de fin de máster]. Universitat Oberta de Catalunya. <https://openaccess.uoc.edu/handle/10609/133768>
- Saldanha, G. y O'Brien, S. (2014). *Research Methodologies in Translation Studies*. Londres: Routledge. <https://doi.org/10.4324/9781315760100>
- Sánchez, M. M. (2017). *Traducción automática y posesición*. Universidad Internacional de Valencia.
- Sánchez, M. M. y Rico, C. (2020). *Traducción automática: conceptos clave, procesos de evaluación y técnicas de posesición*. Granada: Comares
- Santoyo, J. C. (1996). *El delito de traducir*. León: Universidad de León. <https://fddocuments.es/document/el-delito-de-traducir.html?page=1>
- Serrano, R. (2020). Traducción automática y literatura: ¿Enemigas íntimas? *Vasos Comunicantes*. (n. 54). <https://vasoscomunicantes.ace-traductores.org/2020/09/11/traduccion-automatizada-y-literatura-enemigas-intimas/>
- Simard, M. (2020). Building and using parallel texts for translation. En: Alves, F. y Jakobsen, A. (2020). *The Routledge Handbook of Translation and Cognition* (pp. 78-90). Londres: Routledge

- Snover, M., *et al.* (2006). A Study of Translation Edit Rate with Targeted Human Annotation. *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas: Technical Papers* (pp. 223-231). <https://aclanthology.org/2006.amta-papers.25/>
- TAUS (2014). Post-Editing Course.
- Toral, A. y Way, A. (2015a). Translating Literary Text between Related Languages using SMT. In: *Proceedings of the Fourth Workshop on Computational Linguistics for Literature* (pp. 123–132) <https://aclanthology.org/W15-0714/>
- Toral, A. y Way, A. (2015b). Machine-Assisted Translation of Literary Text: A Case Study. *Translation Spaces* (vol. 4:2, pp. 241–268). John Benjamins. <https://cutt.ly/EfP7ord>
- Toral, A. y Way, A. (2018). *What level of quality can neural machine translation attain on literary texts?* <https://arxiv.org/pdf/1801.04962.pdf>
- Vilar, D. *et al.* (2007). Human Evaluation of Machine Translation Through Binary System Comparisons. En: *Proceedings of the Second Workshop on Statistical Machine Translation* (pp. 96-103). Praga: Association for Computational Linguistics
- Youdale, R. (2020). *Using Computers in the Translation of Literary Style. Challenges and Opportunities*. Londres: Routledge

Anexo I: Listado de obras del corpus literario

Abbé Prévost	Manon Lescaut
Alain-Fournier	Le grand Meaulnes
Albert Camus	L'étranger
Albert Camus	La chute
Albert Camus	La peste
Alexandre Dumas	Georges
Alexandre Dumas	La boule de neige
Alexandre Dumas	Le chevalier de Maison-Rouge
Alexandre Dumas	Les compagnons de Jéhu
Alexandre Dumas	Les trois mousquetaires
Alexandre Dumas	Vingt ans après
Alexandre Dumas fils	La dame aux camélias
André Gide	Les faux-monnayeurs
Antoine de Saint-Exupéry	Le petit prince
Antoine de Saint-Exupéry	Lettre à un otage
Antoine de Saint-Exupéry	Terre des hommes
Antoine de Saint-Exupéry	Vol de nuit
Boris Vian	Elles se rendent pas compte
Boris Vian	Et on tuera tous les affreux
Boris Vian	J'irai cracher sur vos tombes
Boris Vian	L'arrache-cœur
Boris Vian	L'automne à Pékin
Boris Vian	L'écume des jours
Boris Vian	L'herbe rouge
Boris Vian	Le loup-garou : suivi de douze autres nouvelles
Boris Vian	Les morts ont tous la même peau
Colette	Claudine à l'école
Colette	Duo
Colette	L'ingénue libertine
Colette	La chatte
Colette	Le blé en herbe
Émile Gaboriau	L'affaire Lerouge

Émile Gaboriau	Le dossier 113
Émile Zola	Germinal
Émile Zola	L'argent
Émile Zola	La curée
Émile Zola	La fortune des Rougon
Émile Zola	Le rêve
Émile Zola	Le ventre de Paris
Émile Zola	Nana
Émile Zola	Thérèse Raquin
Émile Zola	Une page d'amour
Gaston Leroux	La machine à assassiner
Gaston Leroux	La poupée sanglante
Gaston Leroux	Le fantôme de l'Opéra
Gaston Leroux	Le fauteuil hanté
Gaston Leroux	Le mystère de la chambre jaune
Guillaume Apollinaire	L'hérésiarque et cie
Guillaume Apollinaire	Les exploits d'un jeune Don Juan
Guillaume Apollinaire	Les onze mille verges ou les amours d'un hospodar
Gustave Flaubert	Bouvard et Pécuchet
Gustave Flaubert	L'éducation sentimentale
Gustave Flaubert	Madame Bovary
Gustave Flaubert	Salammbô
Gustave Flaubert	Trois contes
Henri Barbusse	L'enfer
Honoré de Balzac	Albert Savarus
Honoré de Balzac	Eugénie Grandet
Honoré de Balzac	Grandeur et décadence de César Birotteau
Honoré de Balzac	La peau du chagrin
Honoré de Balzac	Le Chef-d'œuvre inconnu
Honoré de Balzac	Le colonel Chabert
Honoré de Balzac	Le cousin Pons
Honoré de Balzac	Le curé de Tours
Honoré de Balzac	Le lys dans la vallée

Honoré de Balzac	Le père Goriot
Honoré de Balzac	Pierrette
Honoré de Balzac	Une fille d'Ève
Irène Némirovsky	David Golder
Irène Némirovsky	Jezabel
Irène Némirovsky	Le bal
Irène Némirovsky	Les mouches d'automne
Jean-Jacques Rousseau	Du contrat social
Jules Verne	Autour de la Lune
Jules Verne	Aventures de trois Russes et de trois Anglais
Jules Verne	Bourses de voyage
Jules Verne	Cinq semaines en ballon
Jules Verne	Claudius Bombarnac
Jules Verne	De la Terre à la Lune
Jules Verne	Deux ans de vacances
Jules Verne	Face au drapeau
Jules Verne	Famille sans nom
Jules Verne	Héctor Servadac
Jules Verne	L'archipel en feu
Jules Verne	L'école des Robinsons
Jules Verne	L'île mystérieuse
Jules Verne	La chasse du météore
Jules Verne	La Jangada - Huit cents lieues sur l'Amazone
Jules Verne	La maison à vapeur
Jules Verne	Le Chancellor
Jules Verne	Le château des Carpathes
Jules Verne	Le pays des fourrures
Jules Verne	Le phare au bout du monde
Jules Verne	Le sphinx des glaces
Jules Verne	Le superbe Orénoque
Jules Verne	Le testament d'un excentrique
Jules Verne	Le tour du monde en quatre-vingts jours
Jules Verne	Le village aérien

Jules Verne	Les cinq cents millions de la Béguine
Jules Verne	Les enfants du capitaine Grant
Jules Verne	Les Indes noires
Jules Verne	Les révoltés de la Bounty
Jules Verne	Les tribulations d'un chinois en Chine
Jules Verne	Michel Strogoff
Jules Verne	Nord contre sud
Jules Verne	Robur le Conquérant
Jules Verne	Un billet de loterie
Jules Verne	Un capitaine de quinze ans
Jules Verne	Une ville flottante
Jules Verne	Vingt mille lieues sous les mers
Jules Verne	Voyage au centre de la terre
Marcel Proust	À l'ombre des jeunes filles en fleurs
Marcel Proust	Du côté de chez Swann
Marcel Proust	La prisonnière
Marcel Proust	Le temps retrouvé
Marcel Proust	Les plaisirs et les jours
Marcel Proust	Sodome et Gomorrhe
Marquis de Sade	Justine ou les malheurs de la vertu
Maurice Leblanc	813
Maurice Leblanc	Arsène Lupin gentleman cambrioleur
Maurice Leblanc	L'aiguille creuse
Maurice Leblanc	La comtesse de Cagliostro
Maurice Leblanc	La demoiselle aux yeux verts
Maurice Leblanc	La femme aux deux sourires
Maurice Leblanc	Le bouchon de cristal
Stendhal	L'abbesse de Castro
Stendhal	La chartreuse de Parme
Stendhal	Le rouge et le noir
Stendhal	Les Cenci
Victor Hugo	Les Misérables – Tome I – Fantine
Victor Hugo	Les Misérables – Tome II – Cosette

Victor Hugo	Les Misérables – Tome III – Marius
Victor Hugo	Les Misérables – Tome IV – L'idylle rue Plumet
Victor Hugo	Les Misérables – Tome V – Jean Valjean
Victor Hugo	Notre Dame de Paris
Voltaire	Candide
Voltaire	Le siècle de Louis XIV
Voltaire	Lettres philosophiques
Voltaire	Micromégas

Anexo II: Declaración de consentimiento

Mi nombre es Jon Mikel Peñagarikano Calvo, estudiante de la asignatura Trabajo final de máster (TFM) del máster universitario en Traducción y Tecnologías de la Universitat Oberta de Catalunya. En el marco de la investigación que estoy realizando, pretendo analizar la calidad de traducción que ofrecen diversos motores de traducción automática neuronal al traducir textos literarios del francés al español.

DECLARACIÓN DE CONSENTIMIENTO

Su participación en esta investigación es voluntaria y se llevará a cabo teniendo en cuenta las siguientes condiciones:

- DATOS PERSONALES: se entiende por dato personal toda información relacionada con una persona física que se puede utilizar para identificarla directa o indirectamente (persona física identificada o identificable).
- RESPONSABLE: Jon Mikel Peñagarikano Calvo.
- FINALIDAD: los datos facilitados para llevar a cabo el TFM se utilizan con una finalidad exclusivamente de investigación y sin ánimo de lucro.
- LEGITIMACIÓN: consentimiento de la persona interesada o de quien la represente legalmente.
- DESTINARIOS: no se ceden datos a terceras personas.
- DERECHOS: en cualquier momento se puede ejercer el derecho a retirar el consentimiento dado dirigiéndose al estudiante por medio del correo electrónico jpenagarikano@uoc.edu.

MANIFIESTO:

- Que he recibido la información sobre las características y finalidad perseguidas por el estudiante con el uso de los datos facilitados para la elaboración de su TFM.
- Que el estudiante se compromete a tratar los datos facilitados siguiendo en todo momento los principios y criterios establecidos en la normativa de protección de datos vigente aplicable (Ley orgánica 3/2018, de 5 de diciembre, de Protección de datos personales y garantía de los derechos digitales, y Reglamento de la UE 2016/679, relativo a la protección de las

personas físicas en relación al tratamiento de datos de carácter personal (RGPD)).

- Que el estudiante se compromete a recoger los datos mínimos necesarios para la elaboración de su TFM de acuerdo con el principio de minimización, así como a destruir toda la información de carácter personal facilitada una vez haya terminado la investigación docente que derive en la realización del TFM.
- Que el estudiante se compromete a mantener en secreto la información y los datos personales facilitados ya usarlos exclusivamente con una finalidad docente y de investigación en el marco de la realización del TFM de tal forma que en el trabajo final resultante no sea posible la identificación de la persona titular de estos datos.

POR CONSIGUIENTE:

- He leído las condiciones y consiento participar voluntariamente en la investigación.

Anexo III: Encuesta 1 (motor 1 vs. motor 2 sin ponderación de oraciones)

A continuación, leerá treinta segmentos en francés y, por cada uno de ellos, dos traducciones en español. Por favor, indique cuál de las dos traducciones le parece la mejor.

Por favor, tenga en cuenta las siguientes directrices para decidir qué traducción es mejor:

- Elija la traducción que le suene más natural y que mejor conserve el significado del original sin tener en cuenta los errores (léxicos, sintácticos, etc.) que pueda contener.
- Si las dos traducciones conservan por igual el significado del original y le suenan natural, fijese en el número de errores que contiene cada traducción. La mejor será la que menos errores contenga.

1. **FR:** Marius interrompit la discussion pour s'écrier : – C'est cependant bien agréable d'avoir la croix !

ES A: Marius interrumpió la discusión para exclamar: -¡Sin embargo, es agradable tener la cruz!

ES B: Mariano interrumpió y exclamó: ¡Es muy agradable tener la cruz!

2. **FR:** Cependant, Pécuchet tourmenté par la mésaventure du malaga prit dans l'armoire les boîtes de fer-blanc, fit sauter le couvercle de la première, puis de la seconde, de la troisième.

ES A: Entretanto, Pécuchet atormentado por la desventura del malaga tomó en el armario las cajas de hojalata, hizo saltar la tapa de la primera, luego la segunda, de la tercera.

ES B: Sin embargo, Pedro atormentada por la desgracia del malaga tomó en el armario las cajas de hierro y hizo saltar la tapa de la primera, y la segunda, la segunda.

3. **FR:** Les spéculations les plus sûres sont celles qui reposent sur la vanité, sur l'amour-propre, l'envie de paraître.

ES A: Las especulaciones más seguras son las que se basan en la vanidad, sobre el amor propio, las ganas de aparentar.

ES B: Las especulaciones más seguras son las que reposan sobre la vanidad, el amor propio, las ganas de parecer.

4. **FR:** Duval ne vous dit rien, vous le bouleversez tellement qu'il ne trouve pas un mot.

ES A: Duval no te dice nada, lo molestas tanto que no encuentra ni una palabra.

ES B: Duvald no te dice nada, te lo vas a detener tanto que no encuentra ni una palabra.

5. **FR:** Il est étrange qu'on vienne me reprocher cette gêne pécuniaire qui, hélas, chez un noble romain, n'étonne personne à l'ordinaire...
- ES A:** Es extraño que se venga a reprocharme esta incómoda molestia, que, desgraciadamente, en un noble romano, no sorprende a nadie de lo habitual...
- ES B:** Es extraño que se me pueda reprochar esta molestia pécunica que, por desgracia, en casa de un noble romano, no sorprende a nadie por lo común...
6. **FR:** Félicien ne tenait plus qu'un rien très doux et très tendre, cette robe de mariée, toute de dentelles et de perles, la poignée de plumes légères, tièdes encore, d'un oiseau.
- ES A:** No tenía más que una cosa muy dulce y muy tierna, este vestido de novia, todas las de las hojas y las perlas, la hoja de plumas legadas, incluso de un pájaro.
- ES B:** Félicien no tenía ya más que una nada muy suave y muy tierna, aquel vestido de novia, todos los encajes y perlas, el puñado de plumas ligeras, tibias aún, con un pájaro.
7. **FR:** Tous les deux détournaient la tête, restaient parfois une heure sans se parler, avec l'air de se hair pour des choses enterrées en eux, et sur lesquelles ils ne s'expliquaient point.
- ES A:** Los dos desviaban la cabeza, permanecían a veces una hora sin hablarse, con el aire de odiarse por cosas enterradas en ellos, y sobre las cuales no se explicaban.
- ES B:** Ambos se apartaban de la cabeza, a veces se quedaban sin hablar, con el aire de odiar y con el aire de odiar las cosas enterradas en ellos, y sobre las cuales no se explicaban.
8. **FR:** Hors le service, il reste presque muet, se promène seul et vit mécaniquement.
- ES A:** Fuera del servicio, queda casi mudo, se pasea solo y vive mecánicamente.
- ES B:** Fuera del servicio, se queda casi mudo, camina solo y vive mecánicamente.
9. **FR:** – Qu'est-ce qui a pu faire de vous, ordinairement si froid, un être inspiré, une espèce de prophète de Michel-Ange ?
- ESA:** - ¿Qué es lo que ha hecho de ti, de manera tan fría, un ser inspirado, una especie de profeta de Miguel Ángel?
- ES B:** —¿Qué ha podido hacer de usted, generalmente tan frío, un ser inspirado, una especie de profeta de Miguel Ángel?

- 10. FR:** Devinez, lui dis-je, à qui je viens de téléphoner ?
ES A: —¿Sabe usted —le dije— a quien acabo de telefonear?
ES B: Adivina, ¿de quién acabo de llamar?
- 11. FR:** Les parents du jeune homme écrivirent, pour prendre des informations, une lettre assez dédaigneuse et où l'origine protestante était méprisée.
ES A: Los padres del joven escribieron, para obtener información, una carta bastante desdeñosa y en la que el origen protestante se despreciaba.
ES B: Los padres del joven escribía, para hacer información, una carta bastante deshecha y donde el origen protestante se había despreciado.
- 12. FR:** La garnison de ce fort est peu nombreuse, et une centaine d'hommes déterminés l'obligeraient à se rendre.
ES A: La guarnición de este fuerte es muy poco grande, y un centenar de hombres decididos a obligarla a rendirse.
ES B: La guarnición de este fuerte es poco numerosa y unos cien hombres decididos a obligarle a rendirse.
- 13. FR:** Il faut que je le lui écrive, c'est une de ces choses qu'il doit savoir officiellement avant de se brouiller avec son prince...
ES A: Tengo que escribirle, es una de esas cosas que debe saber oficialmente antes de romper con su príncipe...
ES B: Tengo que escribirlo, es una de esas cosas que debe saber oficialmente antes de rendirse con su príncipe.
- 14. FR:** Le mois de décembre se termina, et avec lui cette année 1867, pendant laquelle les colons de l'île Lincoln venaient d'être si durement éprouvés.
ES A: El mes de diciembre terminó, y con él este año 1867, mientras que los colonos de la isla Lincoln venían de ser tan duros.
ES B: El mes de diciembre terminó, y con él el año 1867, durante la cual acababan de ser tan duramente probados los colonos de la isla Lincoln.
- 15. FR:** Alors un faible soupir s'échappa du cœur du républicain ; il prit ses gants, son chapeau et se rendit à la section.
ES A: Entonces un débil suspiro se escapó del corazón del republicano; cogió sus guantes, su sombrero y se dirigió a la sección.
ES B: Entonces un débil suspiro se escapó del corazón del republicano; tomó los guantes, su sombrero y se fue a la sección.
- 16. FR:** On eût dit un champ de bataille couvert de morts.
ES A: Se hubiera dicho un campo de batalla cubierto de muertos.
ES B: Parecía un campo de batalla cubierto de muertos.

17.FR: Chaque fois que Swann lui avait demandé un service, ne le lui avait-il pas toujours refusé ?

ES A: Cada vez que Swann le había pedido un favor, ¿no se lo había negado siempre?

ES B: Cada vez que Swann le había pedido un favor, ¿no lo había negado siempre?

18.FR: Mais il ne suffisait pas d'aller, il fallait voir en route.

ES A: Pero no había que ir, había que mirar por el camino.

ES B: Pero no bastaba con ir, era preciso ver en el camino.

19.FR: En ce moment, Jasper Hobson saisit le fusil qu'il portait en bandoulière et se tint prêt à tirer.

ES A: En aquel momento, Jasper Hobson cogió el fusil que llevaba en bandolera y se dispuso a disparar.

ES B: En ese momento, Jasper se apoderó de la escopeta que llevaba en bandolera y se puso a disparar.

20.FR: répondit le banquier d'une voix ferme, et, en ce qui vous concerne, voici la dernière opération que nous ferons ensemble...

ES A: —respondió el banquero con voz firme—, y, en lo que a usted concierne, he aquí la última operación que haremos juntos...

ES B: respondió el banquero con voz firme y, en lo que te concierne, esta es la última operación que haremos juntos...

21.FR: Enfin, je vous amuse, ce qui, sans vanité, suppose chez vous une certaine ouverture d'esprit.

ES A: Por último, te estoy divirtiendo, lo que, sin vanidad, piensa en ti una cierta apertura de mente.

ES B: En fin, le divierto, lo cual, sin vanidad, supone en usted una cierta apertura de espíritu.

22.FR: – Vous répondrez que nous avons rejoint James, que nous l'accompagnons dans ses recherches de l'autre côté du fleuve !...

ES A: —Responderá usted que nos hemos reunido con James, que le acompañamos en sus investigaciones al otro lado del río.

ES B: —Responde usted a James, que lo acompañamos en la búsqueda del otro lado del río...

23.FR: Il y eut un moment où les prétextes manquèrent pour justifier sa perpétuelle absence.

ES A: Hubo un momento en que los pretextos faltaron para justificar su eterna ausencia.

- ES B:** Hubo un momento en que los pretextos faltaron para justificar su ausencia.
- 24. FR:** L'événement a prouvé que j'avais raison.
ES A: El evento demostró que tenía razón.
ES B: El acontecimiento ha demostrado que yo tenía razón.
- 25. FR:** L'Anglais saisit le paquet, le défilcela, enleva les deux feuilles de papier qui l'enveloppaient.
ES A: El inglés cogió el paquete y le quitó las dos hojas de papel que lo envolvía.
ES B: El inglés cogió el paquete, el desafío, quitó las dos hojas de papel que lo envolvían.
- 26. FR:** Puis ils gémissaient ; quelques-uns pleuraient tout bas, comme de petits enfants.
ES A: Luego gemía; algunos lloraban en voz baja, como niños pequeños.
ES B: Luego se pusieron a llorar; algunos estaban llorando, como niños pequeños.
- 27. FR:** Kesselbach s'arrêta net au seuil du salon, prit le bras de son secrétaire, et murmura d'une voix inquiète :
ES A: El señor Kesselbach se detuvo en seco en el umbral del salón, cogió del brazo a su secretario y murmuró con voz inquieta:
ES B: El señor Kesselbach se detuvo en el umbral del salón, tomó el brazo de su secretario y murmuró con voz preocupada:
- 28. FR:** En quelques secondes, rendus aphones par un bâillon, aveugles par un bandeau, maîtrisés, ligotés, ils furent emportés rapidement à travers la clairière.
ESA: Al cabo de unos segundos, regresó a través de una mordaza, ciega, bajo control, bajo control.
ES B: En pocos segundos, devueltos por una mordaza, ciegos por una venda, dominadas y amarradas, se llevaron rápidamente a través del claro.
- 29. FR:** Alors lui-même, exténué, il s'écroula sur le pied du lit en essuyant les gouttes de sueur qui mouillaient son front, et il bredouilla :
ES A: Entonces él mismo, extenuado, se desplomó sobre el pie de la cama, secándose las gotas de sudor de su frente, y murmuró:
ES B: Entonces él mismo, extenuado, se desplomó sobre el pie de la cama, enjugándose las gotas de sudor que humedecían su frente y farfulló:

30.FR: Il était peu visible, d'ailleurs ; mais, la nuit venue, il se fût montré comme un point lumineux dans l'espace.

ES A: Por otra parte, era poco visible; pero, por la noche, se había mostrado como un punto luminoso en el espacio.

ES B: Era poco visible; pero, por la noche, se había mostrado como un punto luminoso en el espacio.

Anexo IV: Encuesta 2 (motor 2 con ponderación de oraciones vs. motor 2 sin ponderación de oraciones)

A continuación, leerá treinta segmentos en francés y, por cada uno de ellos, dos traducciones en español. Por favor, indique cuál de las dos traducciones le parece la mejor.

Por favor, tenga en cuenta las siguientes directrices para decidir qué traducción es mejor:

- Elija la traducción que le suene más natural y que mejor conserve el significado del original sin tener en cuenta los errores (léxicos, sintácticos, etc.) que pueda contener.
- Si las dos traducciones conservan por igual el significado del original y le suenan natural, fíjese en el número de errores que contiene cada traducción. La mejor será la que menos errores contenga.

1. **FR:** La fille du banquier Elizundo n'est pas digne de vous !
ES A: ¡La hija del banquero Elizundo no es digna de usted!
ES B: La hija del banquero Elizundo no es digna de vosotros.
2. **FR:** Demain, au lever du jour, il y aura régal pour ceux qui aiment les œufs de tortue et les petites tortues fraîches !
ES A: Mañana, al amanecer, habrá fiesta para aquellos que aman los huevos de tortuga y las pequeñas tortugas frescas.
ES B: Mañana, a la salida del día, habrá festín para aquellos que aman los huevos de tortuga y las pequeñas tortugas frescas.
3. **FR:** Cette particularité ne pouvait émerveiller des gens qui venaient de passer entre les sources chaudes du Waikato.
ES A: Esta particularidad no podía sorprender a las gentes que acababan de pasar entre las aguas cálidas del Waikato.
ES B: Esta particularidad no podía sorprender a la gente que acababa de pasar entre las aguas calientes del Waikato.
4. **FR:** J'aime la fraîcheur des bois et la tranquillité champêtre, comme tu sais ; tu m'as souvent accusé d'être romanesque.
ES A: Me encanta la frescura de los bosques y la tranquilidad campo, como sabes; a menudo me has acusado de ser novelesca.
ES B: Me gusta la frescura de los bosques y la tranquilidad campestre, como tú sabes; me has acusado a menudo de ser novelesco.
5. **FR:** Mais, dans cette obscurité, ils ne pouvaient reconnaître leurs sauveurs.

- ES A:** Pero en aquella oscuridad no podían reconocer a sus salvadores.
ES B: Pero en la oscuridad no podían reconocer a sus salvadores.
6. **FR:** Dans ces jours de clarté, de chaleur et de sérénité, il y a une certaine heure surtout où il faut admirer le portail de Notre-Dame.
ES A: En estos días de claridad, de calor y de serenidad, hay una cierta hora sobre todo en que hay que admirar el portal de Nuestra Señora.
ES B: En estos días de claridad, de calor y de serenidad, hay una cierta hora sobre todo en que hay que admirar el pórtico de Notre Dame.
7. **FR:** Il se haussa sur la pointe des pieds, tendit le bras, fouilla parmi l'agglomération de terreau humide que le temps avait formée au fond du vase, et murmura en frémissant :
ESA: Se levantó de puntillas, tendió el brazo, buscó entre la aglomeración de tierra húmeda que el tiempo había formado en el fondo del vaso, y murmuró estremeciéndose:
ES B: Se alzó sobre la punta de los pies, extendió el brazo, rebuscó entre la agallas húmeda que el tiempo había formado en el fondo del vaso, y murmuró, estremeciéndose:
8. **FR:** Il se souvint alors que, depuis environ quinze jours, il était sevré de tous ces petits soins qui, pendant dix-huit mois, lui avaient rendu la vie si douce à porter.
ES A: Recordó entonces que desde hacía unos quince días había adquirido todos estos cuidados que, durante dieciocho meses, le habían hecho la vida tan dulce de sobrellevar.
ES B: Recordó entonces que desde hacía unos quince días se había levantado de todos esos cuidados que, durante dieciocho meses, le habían hecho la vida tan suave de llevar.
9. **FR:** D'Artagnan s'arrêta devant eux, et ils s'arrêtèrent devant lui.
ES A: D'Artagnan se detuvo ante ellos, y se detuvo delante de él.
ES B: D'Artagnan se detuvo ante ellos y se detuvieron ante él.
10. **FR:** Si votre père est entre les mains des indigènes, ce que fait pressentir d'ailleurs le document, nous le retrouverons.
ES A: Si su padre está en manos de los indígenas, lo que presentirán el documento, lo encontraremos.
ES B: Si vuestro padre está en manos de los indígenas, lo que hace presentir el documento, lo encontraremos.
11. **FR:** Dans cette première réunion, on convint de choisir l'avocat Savaron pour candidat, avec un enthousiasme que personne n'aurait pu espérer à Besançon.

- ES A:** En esta primera reunión se convino en elegir al abogado Savaron para candidato, con un entusiasmo que nadie hubiera podido esperar en Besanzón.
- ES B:** En esta primera reunión se convino en elegir al abogado Savaron para candidato, con un entusiasmo que nadie hubiera podido esperar en Besançon.
- 12.FR:** Il apprit à son oncle de quelle manière les choses s'étaient passées entre nous, et sa générosité le porta sur-le-champ à publier les effets de la mienne.
- ES A:** Aprendió a su tío de qué modo habían pasado las cosas entre nosotros, y su generosidad le llevó al instante a publicar los efectos de la mía.
- ES B:** Aprendió a su tío de qué manera las cosas habían pasado entre nosotros, y su generosidad lo llevó inmediatamente a publicar los efectos de la mía.
- 13.FR:** Il le trouvait trop hébraïque pour ne pas faire mauvais effet.
- ES A:** Era demasiado hebrea para que no hiciera mal efecto.
- ES B:** Le encontraba demasiado hebrea para no hacer mal efecto.
- 14.FR:** Mais soudain il parut stupéfait, ses yeux s'agrandirent, et il bégaya :
- ES A:** Pero, de pronto, pareció estupefacto, sus ojos se agrandaron y tartamudeó:
- ES B:** Pero, de pronto, pareció estupefacto, sus ojos se agrandaron, y tartamudeó:
- 15.FR:** Décidément, c'est un héros de dévouement et de courage, ce Hunt !...
- ES A:** Decididamente, es un héroe de dedicación y valor, este Hunt...
- ES B:** Decididamente, es un héroe de abnegación y valor ese Hunt...
- 16.FR:** Le vieux médecin sourit mélancoliquement de cette véhémence, puis son sourire s'effaça, devant tant de claire et indéniable honte.
- ES A:** El viejo médico sonrió melancólicamente en esta vehemencia, y después su sonrisa desapareció, ante tanto claro e indiscutible.
- ES B:** El viejo médico sonrió melancólicamente de esta vehemencia, luego su sonrisa se borró, ante tanta claridad y innegable vergüenza.
- 17.FR:** – Ni moi ni personne n'y croit, capitaine Guy, et vous êtes le premier que j'aurai entendu soutenir qu'il ne s'agit pas d'un simple roman...
- ES A:** —Ni yo ni nadie cree en ella, capitán Guy, y vos sois el primero que habré oído sostener que no se trata de una simple novela...
- ES B:** —Ni yo ni nadie lo cree, capitán Guy, y sois el primero que habré oído sostener que no se trata de una simple novela...

- 18. FR:** – Écoutez, s'écria Quasimodo croyant vaincre d'un mot son hésitation, venez, monseigneur !
ES A: —Oigan —exclamó Quasimodo creyendo vencer con una palabra su vacilación—, venid, monseñor.
ES B: —Escuche —exclamó Quasimodo creyendo vencer una palabra su vacilación—, venid, monseñor.
- 19. FR:** Je vous ferai souvent réciter des leçons, faites-moi réciter la mienne.
ES A: A menudo os haré recitar lecciones, haced que yo recite la mía.
ES B: A menudo os haré recitar lecciones, venid a recitar la mía.
- 20. FR:** Vers six heures du matin, d'Artagnan se réveilla avec ce malaise qui accompagne ordinairement le point du jour après une mauvaise nuit.
ES A: Hacia las seis de la mañana, D'Artagnan despertó con aquel malestar que acompaña ordinariamente el punto del día después de una mala noche.
ES B: Hacia las seis de la mañana, D'Artagnan se despertó con ese malestar que acompaña ordinariamente el punto del día después de una mala noche.
- 21. FR:** Or, on n'ouvre pas, de l'extérieur, une porte verrouillée.
ES A: No se abre, desde el exterior, una puerta cerrada.
ES B: Sin embargo, no se abre desde el exterior una puerta cerrada.
- 22. FR:** Il les avait trouvées et les avait brûlées avec les autres papiers, dans le petit poêle.
ES A: Los había encontrado y los había quemado con los otros papeles, en la estufa.
ES B: Los había encontrado y los había ardido con los otros papeles, en la estufa.
- 23. FR:** Et s'étendant sur un divan, Nicholl ne tarda pas à ronfler comme un boulet de quarante-huit.
ES A: Y estirándose sobre un diván, Nicholl no tardó en roncar como una bala de cuarenta y ocho.
ES B: Y levantándose sobre un diván, Nicholl no tardó en roncar como una bala de cuarenta y ocho.
- 24. FR:** Bien emmitouflés de fourrures, ces deux êtres charmants avaient l'air de gracieux Esquimaux, lorsqu'ils patinaient ensemble devant le littoral de la Terre-Chaude.

- ES A:** Abordados de pieles, los dos seres encantadores tenían el aspecto de agradable esquimales cuando patinaban juntos delante del litoral de Tierra Caliente.
- ES B:** Vestido bien de pieles, aquellos dos seres encantadores tenían pinta de gracioso a esquimales, cuando patinaban juntos delante del litoral de Tierra Caliente.
- 25. FR:** Son petit frère la rejoignit au moment où, brisée, elle allait se laisser happer à son tour par l'escalier insidieux et réussit à la retenir en empoignant le bas de sa robe.
- ES A:** Su hermano menor la alcanzó en el momento en que, destrozada, iba a dejarse picar a su vez por la escalera insidiosa y consiguió retenerla empuñando el bajo de su vestido.
- ES B:** Su hermano pequeño se reunió con ella en el momento en que, destrozada, se disponía a dejarse vencer por la escalera insidiosa y logró retenerla agarrando la parte inferior de su vestido.
- 26. FR:** La duchesse avait eu le temps de réfléchir ; d'un ton ferme et lent, et comme si elle eût dicté son ultimatum, elle répondit :
- ES A:** La duquesa había tenido tiempo de pensar; en tono firme y lento, y como si hubiera dictado su ultimátum, respondió:
- ES B:** La duquesa había tenido tiempo de reflexionar; con un tono firme y lento, y como si hubiera dictado su ultimátum, respondió:
- 27. FR:** Cela vous va très bien, vos cheveux ébouriffés.
- ES A:** Eso le va muy bien, su pelo revuelto.
- ES B:** Te va muy bien, con el pelo desgreñado.
- 28. FR:** Elle avait beaucoup d'amitié pour Racine ; mais cette amitié ne fut pas assez courageuse pour le protéger contre un léger ressentiment du roi.
- ES A:** Tenía una gran amistad para Racine; pero esa amistad no fue lo suficientemente valiente para protegerle de un ligero resentimiento del rey.
- ES B:** Tenía una gran amistad con Racine, pero esa amistad no fue lo suficientemente valiente como para protegerla contra un leve resentimiento del rey.
- 29. FR:** Elle les quitta un instant, remonta le perron, pour crier à travers pièces toutes ouvertes :
- ES A:** Los dejó un instante, volvió a subir la escalinata, gritando a través de todas las habitaciones abiertas:
- ES B:** Los abandonó un momento, subió la escalinata para gritar a través de todas las habitaciones abiertas:

30.FR: Il traversa ainsi des marais, des rivières, des montagnes, toute cette portion inhabitée du continent que de rares voyageurs ont sillonnée de leurs hardis itinéraires.

ES A: Atravesó así los pantanos, los ríos, las montañas, toda esta porción deshabitada del continente que pocos viajeros han surcado de sus atrevida itinerarios.

ES B: Atravesó así las marismas, de los ríos, de las montañas, toda aquella parte del continente que no hay en el continente más que pocos viajeros que han surcado con sus audaces itinerarios.

Anexo v: Muestra de la evaluación automática del motor 1

Segmento original	Traducción de referencia	Hipótesis	Diferencias
Vers quatre heures du matin, la rapidité de l'appareil s'accrut.	Hacia las cuatro de la mañana aumentó la velocidad.	Hacia las cuatro de la mañana se instaló la velocidad del dispositivo.	Hacia las cuatro de la mañana se instaló umentó la velocidad del dispositivo .
« à chaque pas, les bras et les jambes s'y embarrassaient, et ces lianes, de véritables serpents, enroulaient le corps de leurs tortueux replis.	A cada paso se enredaban los brazos y las piernas en aquellas cuerdas, que como serpientes rodeaban el cuerpo con sus tortuosas espirales.	A cada paso, los brazos y las piernas molestaron, y estos lianes, de verdaderas serpientes, resfrían el cuerpo de sus tortuotes.	A cada paso, se enredaban los brazos y las piernas molestaron , y estos lianes , de verdaderas serpientes , de verdaderas , que como serpientes, rodeaban resfrían el cuerpo de con sus tortuosas espirales .
Que la vie est simple...	Qué sencillo es vivir...	La vida es simple...	La Qué v sencillo lo es v vivir... imple ...
— Briant, demanda Baxter, quelle hauteur penses-tu qu'il conviendrait d'atteindre ?...	—¿A qué altura crees que será preciso subir?	—Murmuró —preguntó Baxter, ¿qué altura crees que sería capaz de alcanzar...?	— Murmuró —preguntó Baxter , ¿A qué altura crees que sería capaz de alcanzar eciso subir?
Comment, après ce que vous venez de lire...	¿Cómo puede opinar eso después de lo que acaba de leer?	Después de lo que acabáis de leer...	D ¿Cómo puede opinar eso después de lo que acabáis de leer?...
Bouvard ayant salué de droite et de gauche, fléchi le menton, et levé les bras, commença.	Bouvard, después de saludar a derecha e izquierda, bajó el mentón, levantó los brazos y comenzó:	Bouvard que saludó de derecha y de izquierda, doblando el mentón, y levantó los brazos, comenzó.	Bouvard, después de saludar de derecha y de izquierda, doblando el mentón, y levantó los brazos, y comenzó:
dit-elle faiblement, si troublée qu'elle chancela et s'assit.	—exclamó ella débilmente, tan turbada que titubeó y se sentó.	Dijo débilmente, tan turbada que se sentó y se sentó.	Dijo —exclamó ella débilmente, tan turbada que se sentó titubeó y se sentó.
Elle le tint serré entre ses bras, en murmurant :	Lo mantuvo apretado entre sus brazos, murmurando:	La colocó entre sus brazos, murmurando:	La col mantuvo o apretado entre sus brazos, murmurando:
Elle lui prit la bouche entre ses dents comme un fruit et mordit violemment ses lèvres.	Joy le atrapó el labio con los dientes y se lo mordió con fuerza, como si fuera una fruta.	Ella le tomó la boca entre los dientes como una fruta y mordió sus labios.	Joy Ella le atrapó el la-bio ca e cont re los dientes y se lo mordió con fuerza, como si fuera una fruta y mordió sus labios .
Gladys tressaillit ; un éclair de rage et presque de folie passa sur ses traits :	Un ataque de rabia, casi de locura, desencajó sus facciones.	Lupin se estremeció; un rayo de rabia y casi locura pasó por sus rostros:	U Lupin se estremeció ; un rayo ataque de rabia, y casi de locura, de pas en ca jó por sus f fac ciones tres :
— Il est trop tard pour remédier à cela.	—Es demasiado tarde para remediar eso.	Es demasiado tarde para remediar esto.	—Es demasiado tarde para remediar esto .
— Étiez-vous autorisé, demanda-t-il sévèrement, à puiser à la caisse selon vos besoins et à faire des avances ?	—¿Estaba usted autorizado -preguntó con severidad- a tomar de la caja lo que necesitase y a conceder adelantos?	—¿Está usted autorizada, preguntó con severidad, a sacar adelante de la caja según sus necesidades y a hacer avances?	—¿Está aba usted autorizada, preguntó con severidad, a tom sacar adelante de la caja lo según que se necesit da ese y a ha con ceder a del ante ces ?
Ils disaient notre chambre, notre tapis, nos fauteuils, même	Decían nuestra habitación, nuestra alfombra, nuestras	Dijeron nuestra habitación, nuestra alfombra, nuestras	Dije decían nuestra habitación, nuestra alfombra, nuestras

elle disait mes pantoufles, un cadeau de Léon, une fantaisie qu'elle avait eue.	butacas, incluso ella decía mis pantuflas, un regalo de León, un capricho que ella había tenido.	sillas, incluso ella decía mis pantuflas, un regalo de León, una fantasía que había tenido.	si butacas, incluso ella decía mis pantuflas, un regalo de León, una c aprichontasía que ella había tenido.
« Ce n'est rien !... »	-¡No es nada!...	"¡No es nada...!	-¡No es nada...!...
Couchez-vous, ou, du moins, mettez-vous dans votre lit tout habillée ; quand j'aurai fini, je frapperai aux carreaux.	Acostaos, o al menos meteos en vuestra cama completamente vestida; cuando haya terminado, golpearé en los cristales.	Vete, o por lo menos ponéis en vuestra cama todo vestido; cuando termine, voy a golpear a las rubias.	Ve Acosteaos, o a por lo menos pmete onéis en vuestra cama cto completamente vestida e ; cuando haya termine e , voy ado, golpearé a en los cr ubistales.
D'une part (et ceci est le petit côté de la chose), on trouve que l'aristocratie semble proportionnellement, dans ce livre, plus accusée de dégénérescence que les autres classes sociales.	Por una parte (y este es el aspecto menos importante del asunto), resulta que este libro parece presentar a la aristocracia más degenerada, proporcionalmente, que las demás clases sociales.	Por un lado (y este es el pequeño lado de la cosa), parece que la aristocracia parece proporcional a este libro, más acusado de degeneración que las demás clases sociales.	Por un lado parte (y este es el aspect queño men lados importante de la cos asunto), resulta que este libro parece pre quesentar a la aristocracia par más de generada, proporcional a este libro, más acusado de de gente, ración que las demás clases sociales.
Le lendemain, 20 mars, le personnel de l'expédition était réuni sur le pier de Camdless-Bay.	Al día siguiente, 20 de marzo, el personal de la expedición estaba reunido al pie del puerto de Camdless-Bay.	Al día siguiente, 20 de marzo, el personal de la expedición se reunió con el servicio de Camdless.	Al día siguiente, 20 de marzo, el personal de la expedición se estaba reunid o con al pie del spu ervicio de Camdless-Bay.
s'écria le vicaire avec un effroi indescriptible.	—exclamó el vicario con un terror indescriptible.	—exclamó el vicarios con un horror indescriptible.	—exclamó el vicarios con un te horror indescriptible.
— Désirez-vous monter dans la charrette ou aller à pied, monsieur ?	—¿Deseáis subir a la carreta o ir a pie, caballero?	¿Te atreves a subir al carro o ir a pie?	—¿ D Te atreves a eáis subir a la carre eta o ir a pie, cab allero?
Le pape, sans être expressément un des confédérés, les animait tous par ses intrigues.	El papa, sin ser de los confederados, los animaba a todos con sus intrigas.	El Papa, sin ser expresadosmente uno de los confedidos, los animaba a todos por sus intrigas.	El p Papa, sin ser expresadosmente uno de los confedierados, los animaba a todos c on f sus intrigas.

Anexo VI: Muestra de la evaluación automática del motor 2 con ponderación de oraciones

Segmento original	Traducción de referencia	Hipótesis	Diferencias
Vers quatre heures du matin, la rapidité de l'appareil s'accrut.	Hacia las cuatro de la mañana aumentó la velocidad.	A las cuatro de la mañana aumentó la velocidad del aparato.	A Hacia las cuatro de la mañana aumentó la velocidad del aparato .
« à chaque pas, les bras et les jambes s'y embarrassaient, et ces lianes, de véritables serpents, enroulaient le corps de leurs tortueux replis.	A cada paso se enredaban los brazos y las piernas en aquellas cuerdas, que como serpientes rodeaban el cuerpo con sus tortuosas espirales.	A cada paso, los brazos y las piernas se embarazaban, y aquellas lianas, de verdaderas serpientes, enrollaban el cuerpo de sus falsuras repletas.	A cada paso, se enredaban los brazos y las piernas se embarazaban , y aquellas cu lianas, de verda deras , que como serpientes, enrode llaban el cuerpo de con sus to falsu rtuosas respira letas.
Que la vie est simple...	Qué sencillo es vivir...	Que la vida es fácil...	Qué e sencil vidallo es v fácil vir ...!...
— Briant, demanda Baxter, quelle hauteur penses-tu qu'il conviendrait d'atteindre ?...	—¿A qué altura crees que será preciso subir?	—Briant —preguntó Baxter—, ¿qué altura cree usted que convendría alcanzar?	— Briant preguntó Baxter —, ¿A qué altura cree usted que se con ven d r í a a l pre can z a i s u b i r ?
Comment, après ce que vous venez de lire...	¿Cómo puede opinar eso después de lo que acaba de leer?	Después de lo que acabas de leer...	D ¿Cómo puede opinar eso después de lo que acabas s de leer?...
Bouvard ayant salué de droite et de gauche, fléchi le menton, et levé les bras, commença.	Bouvard, después de saludar a derecha e izquierda, bajó el mentón, levantó los brazos y comenzó:	Bouvard saludó a derecha e izquierda, dobló la barbilla y levantó los brazos.	Bouvard, después de saludá r a derecha e izquierda, dobló la bajó barbilla y el mentón , levantó los brazos y comenzó :
dit-elle faiblement, si troublée qu'elle chancela et s'assit.	—exclamó ella débilmente, tan turbada que titubeó y se sentó.	—dijo débilmente, tan turbada que se tambaleó y se sentó.	— dijo exclamó ella débilmente, tan turbada que se tam itub aleó y se sentó.
Elle le tint serré entre ses bras, en murmurant :	Lo mantuvo apretado entre sus brazos, murmurando:	Ella lo estrechó entre sus brazos, murmurando:	L E lla lo es man t u v o ap re ch á t a d o entre sus brazos, murmurando:
Elle lui prit la bouche entre ses dents comme un fruit et mordit violemment ses lèvres.	Joy le atrapó el labio con los dientes y se lo mordió con fuerza, como si fuera una fruta.	Ella le tomó la boca entre los dientes como una fruta y se mordió los labios con violencia.	Joy E lla le at em r a p ó el la-bio ca e co n t r e los dientes y se lo mordió con fuerza , como si fuera una fruta y se mordió los labios con violencia .
Gladys tressaillit ; un éclair de rage et presque de folie passa sur ses traits :	Un ataque de rabia, casi de locura, desencajó sus facciones.	Gladys se estremeció; un destello de rabia y casi locura pasó por sus facciones.	U G ladys se estremeció ; un a d e s t a q u e l l o de rabia, y casi de locura, de p a s e n c a j ó por sus facciones.
— Il est trop tard pour remédier à cela.	—Es demasiado tarde para remediar eso.	—Ya es demasiado tarde para ponerle remedio.	— E Y a es demasiado tarde para ponerle remediar eso .
— Étiez-vous autorisé, demanda-t-il sévèrement, à puiser à la caisse selon vos besoins et à faire des avances ?	—¿Estaba usted autorizado -preguntó con severidad- a tomar de la caja lo que necesitase y a conceder adelantos?	—¿Estaba usted autorizado a sacar a la caja de acuerdo a sus necesidades y a hacer avances?	—¿Estaba usted autorizado a preguntó con severidad ed a tomar ade la caja lde acuerdo qa sue s n e c e s i t a d e s e y a

			haconceder avdelantoces?
Ils disaient notre chambre, notre tapis, nos fauteuils, même elle disait mes pantoufles, un cadeau de Léon, une fantaisie qu'elle avait eue.	Decían nuestra habitación, nuestra alfombra, nuestras butacas, incluso ella decía mis pantuflas, un regalo de León, un capricho que ella había tenido.	Decían nuestra habitación, nuestra alfombra, nuestros sillones, incluso ella decía mis zapatillas, un regalo de León, una fantasía que había tenido.	Decían nuestra habitación, nuestra alfombra, nuestraes butacasillones, incluso ella decía mis zapantufillas, un regalo de Leóéon, una cfaprichontasía que ella había tenido.
« Ce n'est rien !... »	-¡No es nada!...	—¡No es nada!...	—¡No es nada!...»
Couchez-vous, ou, du moins, mettez-vous dans votre lit tout habillée ; quand j'aurai fini, je frapperai aux carreaux.	Acostaos, o al menos meteos en vuestra cama completamente vestida; cuando haya terminado, golpearé en los cristales.	Acuéstese, o, por lo menos, poneos en la cama, y cuando haya terminado, golpearé a los cristales.	Acouésteaose, o, apof le menos, metponeos en lvuestra cama, y completamente vestida; cuando haya terminado, golpearé aen los cristales.
D'une part (et ceci est le petit côté de la chose), on trouve que l'aristocratie semble proportionnellement, dans ce livre, plus accusée de dégénérescence que les autres classes sociales.	Por una parte (y este es el aspecto menos importante del asunto), resulta que este libro parece presentar a la aristocracia más degenerada, proporcionalmente, que las demás clases sociales.	Por una parte, y esto es el pequeño lado de la cosa), se encuentra que la aristocracia parece proporcionalmente en este libro más acusada de degeneración que las otras clases sociales.	Por una parte, (y este es el aspecto queño menlados importante de-la cosasunto), se encuentresulta que este libro parece presentar a la aristocracia parmás degenerada, proporcionalmente, -en este libro más acusada de degeneración que las demáetras clases sociales.
Le lendemain, 20 mars, le personnel de l'expédition était réuni sur le pier de Camdless-Bay.	Al día siguiente, 20 de marzo, el personal de la expedición estaba reunido al pie del puerto de Camdless-Bay.	Al día siguiente, 20 de marzo, el personal de la expedición estaba reunido en el muelle de Camdless-Bay.	Al día siguiente, 20 de marzo, el personal de la expedición estaba reunido aalen pie del pmuertoHe de Camdless-Bay.
s'écria le vicair avec un effroi indescriptible.	—exclamó el vicario con un terror indescriptible.	—exclamó el vicario con indescriptible espanto.	—exclamó el vicario con un terror indescriptible-espanto.
– Désirez-vous monter dans la charrette ou aller à pied, monsieur ?	—¿Deseáis subir a la carreta o ir a pie, caballero?	—¿Quieres subir a la carreta o ir a pie?	—¿DQuiesreáis subir a la carreta o ir a pie, caballero?
Le pape, sans être expressément un des confédérés, les animait tous par ses intrigues.	El papa, sin ser de los confederados, los animaba a todos con sus intrigas.	El papa, sin ser expresamente uno de los confederados, los animaba a todos con sus intrigas.	El papa, sin ser expresamente uno de los confederados, los animaba a todos con sus intrigas.

Anexo VII: Muestra de la evaluación automática del motor 2 sin ponderación de oraciones

Segmento original	Traducción de referencia	Hipótesis	Diferencias
Vers quatre heures du matin, la rapidité de l'appareil s'accrut.	Hacia las cuatro de la mañana aumentó la velocidad.	Hacia las cuatro de la mañana aumentó la velocidad del aparato.	Hacia las cuatro de la mañana aumentó la velocidad del aparato .
« à chaque pas, les bras et les jambes s'y embarrassaient, et ces lianes, de véritables serpents, enroulaient le corps de leurs tortueux replis.	A cada paso se enredaban los brazos y las piernas en aquellas cuerdas, que como serpientes rodeaban el cuerpo con sus tortuosas espirales.	En cada paso, los brazos y las piernas se enredaban en él, y aquellas lianas, verdaderas serpientes, derramaban el cuerpo de sus tortuosos recortados.	A En cada paso, se enredaban los brazos y las piernas se enredaban en él, y aquellas cu lianas, verda deras, que como serpientes, ro derramaban el cuerpo de con sus tortuosas es pirales.
Que la vie est simple...	Qué sencillo es vivir...	La vida es sencilla...	Qué La vida es sencilla... o es vivir...
— Briant, demanda Baxter, quelle hauteur penses-tu qu'il conviendrait d'atteindre ?...	—¿A qué altura crees que será preciso subir?	—Briant —preguntó Baxter—, ¿qué altura cree usted que convendría alcanzar?	— Briant —preguntó Baxter —, ¿A qué altura cree usted que se convendría al precañizo subir?
Comment, après ce que vous venez de lire...	¿Cómo puede opinar eso después de lo que acaba de leer?	Después de lo que acabas de leer...	D ¿Cómo puede opinar eso después de lo que acabas s de leer? ...
Bouvard ayant salué de droite et de gauche, fléchi le menton, et levé les bras, commença.	Bouvard, después de saludar a derecha e izquierda, bajó el mentón, levantó los brazos y comenzó:	Bouvard, después de saludar a derecha e izquierda, dobló la barbilla y levantó los brazos.	Bouvard, después de saludar a derecha e izquierda, dobló la bajó barbilla y el mentón, levantó los brazos y comenzó:
dit-elle faiblement, si troublée qu'elle chancela et s'assit.	—exclamó ella débilmente, tan turbada que titubeó y se sentó.	—dijo débilmente, tan turbada que se tambaleó y se sentó.	— dijo exclamó ella débilmente, tan turbada que se tam itubaleó y se sentó.
Elle le tint serré entre ses bras, en murmurant :	Lo mantuvo apretado entre sus brazos, murmurando:	Ella lo sostuvo entre sus brazos, murmurando:	L Ella lo man sostuvo apretado entre sus brazos, murmurando:
Elle lui prit la bouche entre ses dents comme un fruit et mordit violemment ses lèvres.	Joy le atrapó el labio con los dientes y se lo mordió con fuerza, como si fuera una fruta.	Le cogió la boca entre dientes como una fruta y se mordió los labios.	L Joy le cog ió la boca e entre los dientes y se lo mordió con fuerza, como si fuera una fruta y se mordió los labios.
Gladys tressaillit ; un éclair de rage et presque de folie passa sur ses traits :	Un ataque de rabia, casi de locura, desentajó sus facciones.	Gladys se estremeció; un destello de rabia y casi de locura pasó por sus facciones.	U Gladys se estremeció; un a destaque llo de rabia, y casi de locura, de pa sentajó por sus facciones.
— Il est trop tard pour remédier à cela.	—Es demasiado tarde para remediar eso.	Ya es demasiado tarde para remediar esto.	— E Ya es demasiado tarde para remediar esto.
— Étiez-vous autorisé, demanda-t-il sévèrement, à puiser à la caisse selon vos besoins et à faire des avances ?	-¿Estaba usted autorizado -preguntó con severidad- a tomar de la caja lo que necesitase y a conceder adelantos?	—¿Estaba autorizado —preguntó con severidad—, a sacar de la caja según sus necesidades y hacer avances?	—¿Estaba usted autorizado —preguntó con severidad—, a tom sacar de la caja lo según que sus necesidades y ha

			conceder a v delant ces ?
Ils disaient notre chambre, notre tapis, nos fauteuils, même elle disait mes pantoufles, un cadeau de Léon, une fantaisie qu'elle avait eue.	Decían nuestra habitación, nuestra alfombra, nuestras butacas, incluso ella decía mis pantuflas, un regalo de León, un capricho que ella había tenido.	Ellos decían nuestra habitación, nuestra alfombra, nuestras butacas, incluso ella decía mis pantuflas, un regalo de León, un capricho que ella había tenido.	DE Ellos decían nuestra habitación, nuestra alfombra, nuestras butacas, incluso ella decía mis pantuflas, un regalo de Leó éon, un capricho que ella había tenido.
« Ce n'est rien !... »	-¡No es nada!...	—¡No es nada!...	—¡No es nada!... ...
Couchez-vous, ou, du moins, mettez-vous dans votre lit tout habillée ; quand j'aurai fini, je frapperai aux carreaux.	Acostaos, o al menos meteos en vuestra cama completamente vestida; cuando haya terminado, golpearé en los cristales.	Acostarse o, por lo menos, poneos en su cama vestida; cuando yo haya terminado, golpearé a los cristales.	Acostarse se os, por al e menos, met poneos en v uestra cama completamente vestida; cuando yo haya terminado, golpearé a en los cristales.
D'une part (et ceci est le petit côté de la chose), on trouve que l'aristocratie semble proportionnellement, dans ce livre, plus accusée de dégénérescence que les autres classes sociales.	Por una parte (y este es el aspecto menos importante del asunto), resulta que este libro parece presentar a la aristocracia más degenerada, proporcionalmente, que las demás clases sociales.	Por un lado (y esto es el lado pequeño de la cosa), se encuentra que la aristocracia parece proporcionalmente, en este libro, más acusada de degeneración que las otras clases sociales.	Por un lado parte (y este e es el lado pequeño de la cosa), se encuentra que la aristocracia parece proporcionalmente , en este libro, más acusada de degeneración que las demás otras clases sociales.
Le lendemain, 20 mars, le personnel de l'expédition était réuni sur le pier de Camdless-Bay.	Al día siguiente, 20 de marzo, el personal de la expedición estaba reunido al pie del puerto de Camdless-Bay.	Al día siguiente, 20 de marzo, el personal de la expedición estaba reunido en el muelle de Camdless-Bay.	Al día siguiente, 20 de marzo, el personal de la expedición estaba reunido al en pie del puerto le de Camdless-Bay.
s'écria le vicair avec un effroi indescriptible.	—exclamó el vicario con un terror indescriptible.	—exclamó el vicario con un espanto indescriptible—.	—exclamó el vicario con un espan terror indescriptible—.
– Désirez-vous monter dans la charrette ou aller à pied, monsieur ?	—¿Deseáis subir a la carreta o ir a pie, caballero?	—¿Quieres subir a la carreta o ir a pie, señor?	—¿ D Quieres re áis subir a la carreta o ir a pie, scaballe ñero?
Le pape, sans être expressément un des confédérés, les animait tous par ses intrigues.	El papa, sin ser de los confederados, los animaba a todos con sus intrigas.	El papa, sin ser expresamente uno de los confederados, los animaba a todos por sus intrigas.	El papa, sin ser expresamente uno de los confederados, los animaba a todos cp on n sus intrigas.

Anexo VIII: Resultados detallados de la encuesta 1 (motor 1 vs. motor 2 sin ponderación de oraciones)

Pregunta n.º	Evaluadora 1	Evaluadora 2	Evaluadora 3	Mejor propuesta	Motor
1	A	B	A	A	2 (sin PO)
2	A	A	A	A	2 (sin PO)
3	A	B	A	A	1
4	A	A	A	A	2 (sin PO)
5	B	B	B	B	2 (sin PO)
6	B	B	B	B	2 (sin PO)
7	A	A	A	A	2 (sin PO)
8	A	B	B	B	1
9	B	B	B	B	2 (sin PO)
10	A	A	A	A	2 (sin PO)
11	A	A	A	A	2 (sin PO)
12	B	A	B	B	2 (sin PO)
13	A	A	A	A	2 (sin PO)
14	B	B	B	B	2 (sin PO)
15	A	A	A	A	2 (sin PO)
16	B	B	B	B	2 (sin PO)
17	A	A	A	A	1
18	B	B	B	B	2 (sin PO)
19	A	A	A	A	2 (sin PO)
20	A	A	A	A	2 (sin PO)
21	B	B	B	B	2 (sin PO)
22	A	A	A	A	2 (sin PO)
23	A	A	A	A	2 (sin PO)
24	A	B	B	B	2 (sin PO)
25	A	A	A	A	1
26	B	B	A	B	1
27	A	A	A	A	2 (sin PO)
28	B	B	B	B	2 (sin PO)
29	B	A	A	A	1
30	B	B	A	B	1

Anexo IX: Resultados detallados de la encuesta 2 (motor 2 con ponderación de oraciones vs. motor 2 sin ponderación de oraciones)

Pregunta n.º	Evaluadora 1	Evaluadora 2	Evaluadora 3	Mejor propuesta	Motor
1	A	A	A	A	2 (con PO)
2	B	A	A	A	2 (con PO)
3	B	B	A	B	2 (sin PO)
4	B	A	B	B	2 (con PO)
5	A	A	A	A	2 (con PO)
6	B	B	B	B	2 (con PO)
7	A	B	A	A	2 (sin PO)
8	A	A	A	A	2 (con PO)
9	B	B	B	B	2 (sin PO)
10	B	B	B	B	2 (con PO)
11	B	B	B	B	2 (sin PO)
12	B	B	B	B	2 (con PO)
13	B	A	A	A	2 (con PO)
14	A	A	A	A	2 (sin PO)
15	B	A	B	B	2 (sin PO)
16	B	B	A	B	2 (con PO)
17	B	A	B	B	2 (con PO)
18	A	B	A	A	2 (con PO)
19	A	B	A	A	2 (con PO)
20	B	B	B	B	2 (con PO)
21	B	A	A	A	2 (sin PO)
22	A	A	A	A	2 (con PO)
23	A	A	A	A	2 (con PO)
24	B	B	A	B	2 (con PO)
25	A	B	B	B	2 (con PO)
26	B	A	A	A	2 (con PO)
27	A	A	A	A	2 (con PO)
28	A	B	A	A	2 (sin PO)
29	B	B	B	B	2 (con PO)
30	A	A	A	A	2 (con PO)