



Universitat Oberta
de Catalunya

Àrea de Artes y Humanidades

Máster en Traducción Especializada

Trabajo de Fin de Máster

**Evaluación de traducciones realizadas con un modelo neuronal y uno estadístico:
valoración de resultados para los pares francés-español e inglés-español**

Autora:

Cristina Crosby Casali

Tutor:

Antoni Oliver González

Universitat Oberta de Catalunya, junio de 2019

Resumen

Este Trabajo de Fin de Máster compara los sistemas de traducción automática estadística y neuronal, y evalúa las traducciones obtenidas a partir de modelos de ambos sistemas creados en plataformas de código abierto. Por un lado, se presenta un modelo creado con Moses, una estructura que permite crear modelos de traducción automática estadística (TAE). Por otro lado, se presenta un modelo creado con Marian, un conjunto de herramientas que da soporte para la creación de modelos de traducción automática neuronal (TAN). Se utilizan los corpus bilingües en línea MultiUN, que permiten responder a la necesidad de cantidad y calidad de datos para entrenar los modelos de TAN y TAE en los pares francés-español e inglés-español. Las traducciones producidas con estos dos modelos se evalúan con métricas manuales que incluyen un proceso de posesición simple y anotación de errores. Se contrastan los resultados con traducciones generadas con un motor de traducción automática comercial y se califican con métricas automáticas. La evaluación ayuda a definir las fortalezas y debilidades de cada sistema e identificar los errores de traducción más comunes. Según el desempeño de los modelos, es posible dar una valoración que podría permitir considerarlos parte integrante de un proceso más amplio de traducción aumentada, que concibe a la traducción automática como una herramienta y, al traductor, como parte central del flujo de trabajo.

Palabras clave: traducción neuronal, Marian, TAN, traducción estadística, TAE, Moses, traducción aumentada

Abstract

This project compares statistical and neural machine translation systems, and assesses translations generated by models of both systems created on open source frameworks. The first model is created using Moses, a structure that allows the creation of statistical machine translation (SMT) models. While the second model is created using Marian, a toolkit that supports the creation of neural machine translation (NMT) models. MultiUN bilingual corpora are used, which allow to build both SMT and NMT models with quantity and quality of data for the French-Spanish and English-Spanish translations. The resulting translations are assessed using manual metrics, including a light post-editing process and error annotation. The results are benchmarked with a third translation obtained from an open-access machine translation engine. Automatic metrics are calculated with the three translations. The evaluation helps to define the strengths and weaknesses of each system, and to identify the most common translation errors. The performance of the models allows to ascertain if they could be considered an integral part of a broader process of augmented translation, which regards machine translation as a tool, while the translator is at the core of the translation workflow.

Key Words: Neural Machine Translation, Marian, NMT, Statistical Machine Translation, SMT, Moses, augmented translation

Índice

1.	Introducción	4
1.1	Objetivos	4
2.	Sistemas de traducción automática: el sistema basado en estadísticas comparado con el sistema neuronal	5
2.1	Traducción automática estadística: uso de Moses	5
2.2	Traducción automática neuronal: uso de Marian	5
2.3	Literatura previa	6
3.	Análisis de las traducciones	7
3.1	Características de los modelos y motor de traducción	7
3.2	Datos utilizados para el entrenamiento: Corpus MultiUN	8
3.3	Datos de evaluación	9
3.4	Características del análisis	9
3.5	Metodología de evaluación	10
4.	Evaluación de las traducciones realizadas con los modelos propios y comparación con una traducción hecha con un sistema de traducción automática comercial	10
4.1	Comparación de las traducciones de los modelos para el par francés-español	10
4.1.1	Valoración manual de la traducción neuronal realizada con Marian	10
4.1.2	Valoración manual de la traducción estadística realizada con Moses	12
4.1.3	Contraste de las traducciones de los dos modelos con una traducción generada por Google Translate	13
4.2	Comparación de las traducciones de los modelos para el par inglés-español	15
4.2.1	Valoración manual de la traducción neuronal realizada con Marian	15
4.2.2	Valoración manual de la traducción estadística realizada con Moses	16
4.2.3	Contraste de las traducciones de los dos modelos con una traducción generada por Google Translate	18
4.3	Valoración de los modelos con métricas automáticas	20
4.4	Observaciones sobre el desempeño de los dos modelos en ambos pares de idiomas	21
5.	La importancia de la posesión como parte del proceso de traducción automática: traducción aumentada	22
6.	Conclusiones	23
7.	Bibliografía	24

1. Introducción

Los motores de traducción automática comercial como Google Translate (<https://translate.google.com/>) han hecho que la traducción automática sea accesible para todos. Estos motores necesitan enormes cantidades de datos multilingües que se procesan con distintas lógicas (basados en estadísticas y con mecanismos neuronales). Gracias a la disponibilidad de corpus bilingües de acceso libre y de las plataformas, estructuras y conjuntos de herramientas de código libre es posible crear modelos propios de traducción automática.

Los sistemas de traducción automática se han ido desarrollado a partir del estudio de diccionarios y gramáticas, de bases de datos, corpus monolingües y bilingües, glosarios: en general, conocimiento lingüístico. Mientras que los primeros sistemas de traducción basados en frases, reglas o estadísticas requerían entrenamiento e intervención humana, la traducción automática neuronal (TAN o más conocida por sus siglas en inglés, NMT) tiene la gran ventaja de que los datos se entrenan sin necesidad de intervención humana. A pesar de este y otros beneficios, al contrastar este sistema con el paradigma inmediatamente precedente, la traducción automática basada en estadísticas (TAE o más conocida por sus siglas en inglés, SMT), se observa que la primera produce traducciones más fluidas, pero inadecuadas; mientras que la segunda muestra las características opuestas (Wang et al., 2017). A pesar de esto, en la actualidad, la TAN está en su apogeo, y es el enfoque escogido por las empresas más grandes del sector de la traducción, la inteligencia artificial y el aprendizaje profundo. De la mano de sistemas de uso comercial y gracias a aplicaciones de código abierto que permiten su difusión, una tecnología que en sus inicios, hace tan solo 6 años, era considerada costosa y de aplicación muy exigente en términos de recursos, ya se ha convertido en la norma. En este contexto, el objetivo de este trabajo es comparar traducciones generadas por un sistema neuronal y uno estadístico entrenados con corpus bilingües de Naciones Unidas disponibles en línea en la plataforma Opus (<http://opus.nlpl.eu/MultiUN.php>), y contrastarlas con una traducción automática generada por un motor comercial. Esta comparación evalúa el desempeño de los modelos y la calidad de los resultados.

En las siguientes secciones, se presentan los dos sistemas de traducción automática, estadística y neuronal, y una breve descripción de modelos propios preparados en entornos de código abierto, con los cuales se generaron traducciones en los pares francés-español e inglés-español. Se evalúan estas traducciones junto con una tercera realizada con Google Translate. Las evaluaciones incluyen la medición del esfuerzo temporal y técnico, esto incluye una posesición simple realizada por dos lingüistas, junto con una anotación de los errores detectados en las traducciones, y el cálculo de tres métricas automáticas. Finalmente, se trata el tema de la posesición junto con la traducción automática y su relevancia en la actualidad, y del papel del traductor en un proceso denominado *traducción aumentada*, en el cual, a diferencia de lo que es común leer sobre la traducción automática —su rápida evolución y su potencial reemplazo al traductor—, se lo ubica como parte central en un proceso integrado de alto valor agregado.

1.1 Objetivos

- Explorar las características de los sistemas de traducción automática neuronales y estadísticos.
- Comparar las diferencias entre las traducciones realizadas usando un modelo neuronal y otro estadístico con una traducción realizada por un motor comercial en los pares francés-español e inglés-español, mediante métricas automáticas y manuales.
- Diferenciar los tipos de errores encontrados a nivel de posesición en cada uno de los pares de idiomas.
- Determinar la adecuación de cada sistema de traducción automática para el dominio de texto analizado.
- Exponer la posesición y la traducción automática como una parte del proceso de traducción aumentada.

2. Sistemas de traducción automática: el sistema basado en estadísticas comparado con el sistema neuronal

2.1 Traducción automática estadística: uso de Moses

Hasta que se impusieran los sistemas neuronales, la traducción automática estadística era el sistema utilizado en la industria (Forcada, 2017). La investigación en traducción automática estadística empezó en la década de 1980 de la mano de IBM. Como sistema, fue presentada por Microsoft en 1993 en un entorno bayesiano. En este marco, y sobre la base de modelos basados en datos, como se conoce también a este sistema, en 2005, nació Moses, un *toolkit* (conjunto de herramientas) de código abierto y con licencia LGPL que ayuda a elaborar modelos propios y producir secuencias con datos, y que integra recursos no solo propios sino de muchas fuentes. En los modelos estadísticos, cuya principal aplicación son los modelos basados en frases, se usan dos sistemas de datos, corpus paralelos y corpus monolingües; de uno, conformado por segmentos alineados en el idioma original y en el idioma de llegada, el sistema aprende cómo traducir y se obtiene la fluidez del texto resultante; del otro, se logra afinar el estilo del idioma de llegada y la precisión de los términos utilizados. El uso de estos datos monolingües es un mecanismo que asegura que todas las palabras sean traducidas sobre la base de palabras observadas al menos una vez en los datos de referencia.

El primer paso para la preparación del modelo es procesar los datos. Esto incluye la limpieza de la información, aplicación de correcciones, así como alineación de segmentos, tokenización, true-casing, eliminación de segmentos no válidos y eliminación de oraciones demasiado largas. Este primer paso es importante porque, como lo manifestó uno de los creadores de este sistema, Philippe Koehn, la clave para crear un buen sistema — cualquier sistema de traducción automática — es contar con muchos datos y de buena calidad: mientras más se asemeje la calidad de estos datos a la calidad que se pretende obtener en la traducción, mejores serán los resultados (Koehn, 2007). Al ser un modelo de elaboración propia, también es posible ajustar características a la medida de lo que se necesita, como, por ejemplo, preparar un motor como el realizado para este trabajo, que se va a dedicar a la traducción de textos específicos de Derecho Internacional.

Los componentes principales del sistema son la función de entrenamiento de datos y el decodificador. La primera es un conjunto de herramientas programadas en Perl y en C++ que convierten todos los datos en un modelo de traducción automática. En la función de entrenamiento, las oraciones pasan por un proceso de alineación de palabras, realizado con GIZA ++, a partir de lo cual se extraerán las traducciones de frases o se crearán las reglas de jerarquía. En este punto, el sistema toma los datos paralelos e infiere correspondencias de traducción entre los dos idiomas y se estiman las probabilidades. El decodificador está hecho en C++ y toma los datos del modelo y el texto de referencia, y se encarga de generar la traducción. Aquí interviene el modelo lingüístico, que lo constituye el corpus monolingüe, que es utilizado por el decodificador para producir una traducción fluida (aquí se usan herramientas externas que soportan la creación del modelo lingüístico). Finalmente, se afinan los modelos: este proceso consiste en la que los modelos se ponderan entre sí mediante algoritmos. El decodificador se encarga de encontrar las oraciones con la mejor calificación.

El sistema no solo reproduce secuencias de correspondencias en el segmento de llegada (de aquí que se conozca a estos modelos como basados en frases), sino que agrega otras estructuras que va aprendiendo de los datos (el modelo genera traducciones que se escogen según su jerarquía: la calificación interna del modelo). Esta característica se llama traducción factorizada y es la que puede crear la diferencia en los resultados generados.

2.2 Traducción automática neuronal: uso de Marian

En sus inicios, la traducción automática se basaba en reglas, pasó por sistemas basados en frases, hasta que llegó a las estadísticas. Las características de este último sistema —cuyo emblema es su enfoque basado en datos, más precisamente, n-gramas— es el que dio paso a lo que hoy conocemos como traducción neuronal.

El primer sistema completamente neuronal fue introducido por Nal Kalchbrenner y Phil Blunsom en 2013, con un sistema de estructura integral codificador-decodificador (*end-to-end encoder-decoder*) (Peri, Cebrián y Casacuberta, 2017) que marcó el inicio de la traducción automática neuronal como un sistema auxiliado por

aprendizaje que no ha parado de desarrollarse. La evolución de los sistemas neuronales pudo ser corroborada en la Segunda Conferencia de Traducción Automática de 2017 realizada en Dinamarca. En 2016, tan solo algunos de los grupos de trabajo usaban redes neuronales, para el año siguiente, los ganadores y la mayoría de los sistemas de mejor desempeño incluyeron redes neuronales en la construcción de sus modelos (Bojar et al., 2017).

En su informe sobre el estado de la traducción automática neuronal en 2018, el sitio web slator.com, de referencia para conocer sobre el panorama actual en el ámbito de la Traducción, menciona una frase usada comúnmente para describir la traducción neuronal: es una tecnología innovadora de inteligencia artificial que tiene un rendimiento muy superior a la traducción basada en estadísticas (Slator, 2018). La realidad es que los modelos de traducción neuronal evolucionaron a partir de la traducción estadística y, de esta forma nació Marian, un entorno diferenciado de traducción automática neuronal desarrollado a partir de la función estadística del sistema Moses (Junczys-Dowmunt, Grundkiewicz et al., 2018).

Marian es un entorno de programación (*framework*) de traducción automática desarrollado completamente en C++ y con dependencias mínimas. Es un sistema de código abierto de traducción neuronal con licencia MIT. La red neuronal integral (*end-to-end*) mapea textos bilingües para maximizar las probabilidades de traducción. Su estructura consiste en un codificador-decodificador, que además, reduce el tiempo de implementación, por lo que los modelos creados en este sistema pueden seguir la estructura de entrenamiento, evaluación y traducción (Junczys-Dowmunt, Grundkiewicz et al., 2018). En términos más simples, se trata de un modelo que procesa información de corpus bilingües que se relacionan mediante algoritmos de aprendizaje automático, redes neuronales profundas, y un sistema de atención agregada (que analiza el contexto de cada palabra para producirla en el texto de llegada). En los vectores está la principal diferencia de este modelo con el modelo estadístico: el elemento central de los modelos estadísticos son las palabras (n-gramas), en este son los vectores. En este modelo, se predice la palabra siguiente sobre la base de representaciones vectoriales, mientras que en el sistema estadístico, se hace por recomendaciones del modelo según ponderaciones. Una de las fortalezas de los modelos neuronales es la facilidad de aprendizaje del texto de entrenamiento, esto se debe a la arquitectura de las redes neuronales recurrentes: una que sirve para absorber el texto de entrada y otra que genera el texto de llegada. En la actualidad, la mayoría de las aplicaciones neuronales utilizan la estructura codificador-decodificador con el mecanismo de atención (Forcada, 2017).

2.3 Literatura previa

Artículos, entradas de blogs y muchas fuentes académicas consultadas coinciden en que la traducción neuronal es un sistema innovador y de desempeño superior (Bentivogli et al., 2016; Wu et al., 2016; Bojar et al., 2017; Forcada, 2017; Koehn, 2017). Esto se contrasta con otros trabajos que consideran que estas afirmaciones deben ser más prudentes (Kenny, 2018; Castilho et al., 2017). Para la primera, las redes neuronales son una consecuencia obvia de los sistemas de traducción estadística, así que solo se trata de un continuo, no de una evolución. La traducción neuronal demanda una igual cantidad de datos limpios y entrenamiento, de la misma forma que los sistemas estadísticos, y para ambos se coincide en la necesidad de capacitar a cada vez más traductores/poseedores para que sean competentes al momento de integrar la traducción automática en su flujo de trabajo y aprovechar sus beneficios. Para la segunda, se debe limitar las expectativas puestas en los modelos neuronales por la necesidad de la industria de ahorrar costos con la traducción automática, pues los resultados varían muchísimo entre tipos de textos, disciplinas y pares de idiomas. Dos trabajos consultados abordan la idea de incluir elementos de la TAE en los modelos neuronales (Wang et al., 2017; Koehn y Knowles, 2017), pues han detectado los principales errores de estos sistemas: qué ocurre cuando se pretende traducir textos de un campo diferente al de los datos de entrenamiento (limitación del sistema), la cantidad de datos necesarios para lograr que el desempeño sea óptimo (una desventaja para el trabajo con idiomas con bajos recursos), las palabras inventadas por el sistema cuando no se encuentran las palabras en el texto original (aunque se pueda valer de incrustaciones (*embeddings*), representaciones de las palabras que permiten generar otras palabras por su contexto), las oraciones largas (esto impediría reflejar la realidad estructural de los textos en varios idiomas), la alineación de las palabras (se confía en el recurso de codificación por pares de bytes, pero

los nombres propios tienen una estructura no predecible), entre otros, que son recurrentes en las evaluaciones de desempeño de estos modelos. Koehn y Knowles mencionan un reto más de la TAN: la dificultad de interpretar los sistemas y la necesidad de mejorar los análisis más allá de saber cómo interactúan el codificador con el decodificador sobre la base de un mecanismo de atención (Koehn y Knowles, 2017).

Algunos trabajos se dedican a abordar la arquitectura de los sistemas y la creación de modelos (Koehn, 2007; Peris, Domingo y Casacuberta, 2016; Koehn, 2017; Junczys-Dowmunt, Grundkiewicz et al., 2018), cuyo contenido solo es accesible si se cuenta con conocimientos especializados. A manera de contraste, otras investigaciones se han centrado en hacer análisis sobre conjuntos de datos específicos y lograr identificar en qué puede mejorar cada sistema en el aspecto lingüístico (Bentivogli, Bisazza et al., 2016; Burchardt et al., 2017; Castilho et al., 2017; Koponen, Salmi y Nikulin, 2019). En estos trabajos se consideran distintas muestras de datos, idiomas con pocos recursos, pares de idiomas que suelen ser problemáticos por su morfología y muestras de oraciones representativas de la estructura de un idioma particular que permiten definir, cada uno en un tema específico, los fenómenos lingüísticos que se puede esperar de cada sistema de traducción.

Se encontraron pocas investigaciones académicas que lograran explicar la lógica de la traducción neuronal en términos no técnicos (Casacuberta, Peris, 2017; Forcada, 2017), este segundo, sobre todo, expone de forma clara la arquitectura neuronal. Sin embargo, en una búsqueda en la web, lo más comprensible se encuentra en artículos de blogs y análisis de *think tanks* (Zhang, 2017; DePalma, 2017). El nivel de complejidad de la mayor parte de los artículos hace que los sistemas de traducción automática —no en su utilidad, pero sí en su integración— permanezcan alejados de los profesionales de la lengua sin un perfil especializado. La predominancia de la traducción automática en la industria y en los requerimientos laborales, y su integración con los programas de traducción en la actualidad plantean el desafío de una profundización de los conocimientos técnicos de los traductores.

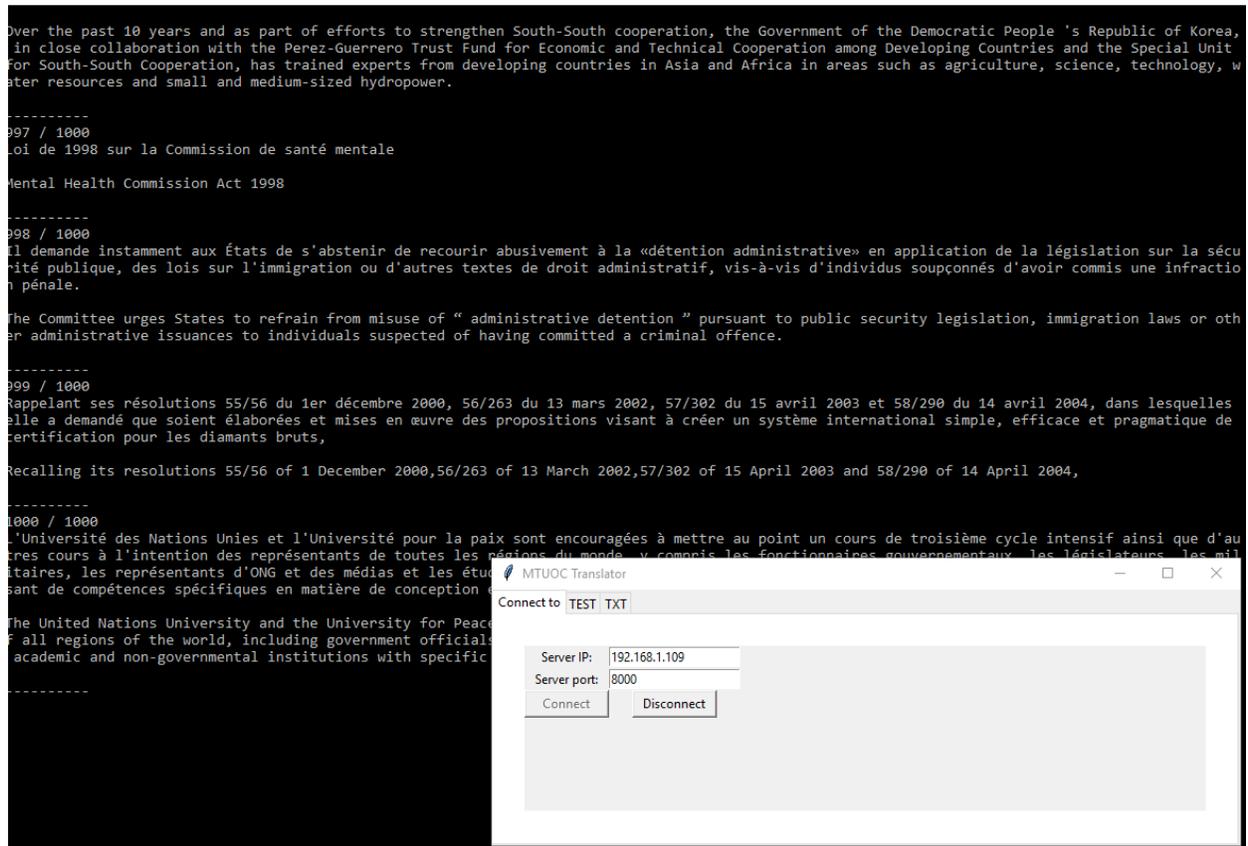
3. Análisis de las traducciones

Este trabajo es realizado por una traductora con conocimientos básicos de programación y aprendizaje automático, por lo tanto, no ahonda en los aspectos técnicos que caracterizan el funcionamiento de estos sistemas, sino en las traducciones generadas. Se escogen dos pares de idiomas: por una parte, francés-español, que son idiomas que forman parte de la misma familia y que, a simple vista, en estructura y terminología, no deberían generar errores en estos aspectos. Por otro lado, se estima otro par, inglés-español que, por ser de distinta familia idiomática, podrían presentar otro tipo de fenómenos y desafíos interesantes a los modelos.

3.1 Características de los modelos y motor de traducción

Tanto el modelo neuronal como el modelo estadístico fueron preparados como parte de un proyecto del LPG (Language Processing Group) de la Universitat Oberta de Catalunya. Están alojados en la página <http://lpg.uoc.edu/MTUOC/>.

Los modelos estadísticos fueron realizados con Moses v 4.0 (actualizada en 2017) (<https://github.com/moses-smt/mosesdecoder>) y los modelos neuronales fueron construidos con Marian v 1.7.6 (<https://marian-nmt.github.io/>). Se utilizó una unidad de procesamiento gráfico (GPU) sin la cual el procesamiento habría sido casi imposible. Estos modelos se realizan en Linux y la redacción del trabajo se realizó en Windows, por lo tanto, para observar su funcionamiento, se trabajó usando una máquina virtual. Se debió instalar un motor de traducción, referido como cliente, que permitió evaluar ambos modelos.



Vista del motor de traducción (cliente) utilizado para ambos modelos

3.2 Datos utilizados para el entrenamiento: Corpus MultiUN

Ambos sistemas fueron entrenados con corpus bilingües, preparados con documentación de las Naciones Unidas recopilada entre 2000 y 2009, y que se pueden descargar en línea en la plataforma de corpus paralelos de acceso libre OPUS (<http://opus.nlpl.eu/MultiUN.php>). Estos grandes corpus bilingües, MultiUN, son archivos que constan de segmentos alineados en cada par de idiomas en donde se eliminaron todos los segmentos duplicados y desalineados. Asimismo, para la preparación de ambos modelos, en el corpus de entrenamiento para ambos pares de idiomas, se eliminaron las oraciones demasiado largas. Como referencia, una oración corta consta de hasta 30 tokens; las oraciones largas, de hasta 60 tokens; y las demasiado largas, de más de 60 tokens. Para el modelo lingüístico monolingüe necesario para el sistema estadístico, se utilizaron los mismos segmentos usados para la alineación en español del corpus bilingüe.

Información de los corpus

	Modelo francés-español	Modelo inglés-español
Segmentos alineados	11.5 millones de pares	11.3 millones de pares
Tokens	francés: 474.3 millones, inglés: 443.5 millones, español: 380.4 millones	

Los segmentos incluyen lenguaje oral y escrito (cada uno se caracteriza por un distinto tono, lo cual beneficia a los modelos, ya que tiene distintas referencias para un mismo tema: informal y formal) y se pueden encontrar documentos oficiales que incluyen textos de declaraciones, leyes, actas, programas del día, informes de proyectos, artículos periodísticos, noticias y anuncios. Abundan los nombres propios (personas, lugares, organizaciones, tratados; los primeros son los que, tal vez, pueden presentar la mayor complicación), fechas, detalles de instrumentos judiciales y cifras.

3.3 Datos de evaluación

Para cada uno de los modelos, los corpus bilingües se han dividido en tres partes: la mayor parte se destinó al entrenamiento y se separaron dos muestras de 1000 segmentos cada una: la primera para validación/optimización y una segunda para evaluación del desempeño de los modelos y posterior revisión. De esta segunda muestra, escogida de forma arbitraria, se seleccionaron 250 segmentos. La evaluación incluye dos conjuntos de traducción de 250 segmentos cada uno y un tercer conjunto de segmentos traducidos con Google Translate.

3.4 Características del análisis

En primer lugar, se analizan las traducciones midiendo el esfuerzo temporal y el esfuerzo técnico, que incluye una posesición simple. En este tipo de posesición, en contraste con la posesición completa, se espera un texto comprensible, más no de carácter publicable.

Para el análisis, se pidió a los poseedores que usaran la herramienta PET (<http://www.clg.wlv.ac.uk/projects/PET/>), de código abierto y escrita en Java, que permite recopilar estadísticas sobre el proceso de posesición, como el número de revisiones y el tipo de cambios. La interfaz es muy intuitiva, y se explicó a los poseedores las principales funcionalidades antes de que iniciaran la tarea.

PET Marian FR-ES by Cristina Crosby	
editing ...	partial: 9s
revisions: 0	total: 0s
<p>Estas consideraciones desempeñan un papel más importante en las decisiones relativas a los recortes presupuestarios, un análisis riguroso de las economías posibles o la preocupación por dotar a la Organización de medios acordados con sus mandatos.</p>	<p>1891...</p> <p>100 saved 09:08:38</p>
<p>La loi constitutionnelle autrichienne contient plusieurs dispositions ayant trait à la liberté de pensée, de conscience et de religion.</p>	<p>La legislación constitucional austriaca contiene varias disposiciones relativas a la libertad de pensamiento, de conciencia y de religión.</p>
<p>Le Conseil rappelle qu'il a insisté dans sa déclaration du 23 mars 2000 sur le processus de désarmement, de démobilisation et de réinsertion qui peut jouer un rôle clef pour stabiliser la situation après un conflit, réduire les risques de nouvelles violences et faciliter la transition vers une situation normale et le développement.</p>	<p>El Consejo recuerda su insistencia en su declaración de 23 de marzo de 2000 sobre el proceso de desarme, desmovilización y reintegración, que puede desempeñar un papel fundamental en la estabilización de la situación después de los conflictos, reducir los riesgos de nuevos actos de violencia y facilitar la transición hacia la normalidad y el desarrollo.</p>
<p>Malheureusement, un certain nombre de mouvements qui n'approuvent pas cette tolérance ont tenté de déstabiliser la situation en troublant l'ordre public lors d'événements tels que les matchs de football.</p>	<p>Lamentablemente, varios movimientos que no están de acuerdo con esa tolerancia han tratado de desestabilizar la situación en medio de acontecimientos como los partidos de fútbol.</p>
<p>VC 193/4 - Lettre datée du 7 octobre 2008, adressée au Secrétaire général par le Représentant permanent de la Lituanie auprès de l'Organisation des Nations Unies</p>	<p>AFRC 193/4- Carta de fecha 7 de octubre de 2008 dirigida al Secretario General por el Representante Permanente de Lituania ante las Naciones Unidas</p>
<p>Ces considérations jouent un plus grand rôle dans les décisions relatives aux coupes budgétaires qu'une analyse rigoureuse des économies possibles ou que le souci de doter l'Organisation de moyens proportionnés à ses mandats.</p>	<p>Estas consideraciones desempeñan un papel más importante en las decisiones relativas a los recortes presupuestarios, un análisis riguroso de las economías posibles o la preocupación por dotar a la Organización de medios acordados con sus mandatos.</p>
<p>Le Comité a également pris note avec intérêt des informations fournies par le secrétariat au groupe de travail de présession, lequel est principalement chargé d'adopter une liste écrite des points à traiter en ce qui concerne les rapports des États parties reçus par le Comité et en attente d'examen.</p>	<p>El Comité también tomó nota con interés de la información proporcionada por la secretaría al grupo de trabajo anterior al período de sesiones, que se encarga principalmente de aprobar una lista por escrito de cuestiones relativas a los informes de los Estados Partes recibidos por el Comité y en espera de su examen.</p>
<p>Le secrétariat devrait informer les États membres de ses réflexions sur les thèmes de la prochaine session de la Conférence ainsi que de la préparation et de la date de celle-ci.</p>	<p>La secretaría debería informar a los Estados miembros de sus reflexiones sobre los temas del próximo período de sesiones de la Conferencia, así como de la preparación y la fecha de ésta.</p>
<p>Avila (Espagne), juillet.</p>	<p>Avila (España), julio.</p>
<p>Iles sont parallèlement fortement contestées par certains chercheurs, activistes politiques et associations de la société civile dans les pays en développement et émergents.</p>	<p>Al mismo tiempo, los investigadores, los activistas políticos y las asociaciones de la sociedad civil en los países en desarrollo y emergentes se oponen enérgicamente a ellas.</p>

Vista del programa PET

```

32 </annotations>
33 </unit>
34 <unit id="2" status="FINISHED" type="pe">
35 <S producer="Cristina Crosby">In that meeting, he underlined that culture, alongside education, should constitute a pillar for the reconstruction
36 <MT producer="Cristina Crosby">En esa reunión subrayó que la cultura, junto con la educación, debía constituir un pilar para la reconstrucción de
37 <annotations revisions="1">
38 <annotation r="1">
39 <PE producer="Natalia Meneses">En esa reunión subrayó que la cultura, junto con la educación, debía constituir un pilar para la reconstrucción
40 <indicator id="unchanged" type="flag">false</indicator>
41 <indicator elapsed="0" id="assignment" length="359" offset="0" t0="0" type="change">En esa reunión subrayó que la cultura, junto con la edu
42 <indicator elapsed="0" id="insertion" length="24" offset="318" t0="0" type="change">algo tan necesario como </indicator>
43 <indicator id="substitution" type="wrap">
44 <action elapsed="0" id="deletion" length="22" offset="342" t0="0" type="change">necesidades básicas de</action>
45 <action elapsed="0" id="insertion" length="2" offset="342" t0="0" type="change">la</action>
46 </indicator>
47 <indicator elapsed="0" id="insertion" length="3" offset="352" t0="0" type="change"> la</indicator>
48 <indicator id="editing" type="time">1m5s,334</indicator>
49 <indicator id="letter-keys" type="count">24</indicator>
50 <indicator id="digit-keys" type="count">0</indicator>
51 <indicator id="white-keys" type="count">5</indicator>
52 <indicator id="symbol-keys" type="count">0</indicator>
53 <indicator id="navigation-keys" type="count">87</indicator>
54 <indicator id="erase-keys" type="count">0</indicator>
55 <indicator id="copy-keys" type="count">0</indicator>

```

Find result - 76 hits

Search "false" (76 hits in 1 file)

C:\Users\cris\Desktop\PET-v2.0\example\demo\Cristina Crosby\PET Marian EN-ES.per (76 hits)

Line 40: <indicator id="unchanged" type="flag">false</indicator>
Line 402: <indicator id="unchanged" type="flag">false</indicator>
Line 590: <indicator id="unchanged" type="flag">false</indicator>
Line 728: <indicator id="unchanged" type="flag">false</indicator>

Vista del informe final generado por PET

Tras la revisión, y luego de preparar un informe con sus anotaciones, los poseedores compararon las tres traducciones (una de cada modelo y la tercera, generada con Google Translate) para ilustrar algunas de las características más evidentes de cada sistema de traducción automática. En una sección posterior, se calcularon las métricas automáticas para las tres traducciones.

3.5 Metodología de evaluación

La evaluación es realizada por dos poseedores (designados como P1 y P2) con experiencia de 14 años en traducción hacia el español y que realizan trabajos de posesición de forma regular. Los dos recibieron la indicación de hacer una revisión de dos muestras de 250 segmentos cada una, primero, comparadas con el texto de referencia y, luego, contrastadas de forma paralela con una tercera traducción realizada por un motor comercial. La revisión realizada es una posesición simple, y esta designación delimita la calidad del texto final: no se espera un texto de carácter publicable. Esta posesición toma en cuenta el esfuerzo temporal, el número de modificaciones realizadas y el tipo de cambios realizados. En cada sección de evaluación, se muestra un detalle del número de cambios que se realizaron en la muestra, mas no el número de segmentos que fueron modificados, se eligió hacer esto porque los segmentos difieren mucho en longitud; asimismo, en un segmento puede haber más de un cambio. Se destaca que la revisión no incluye modificaciones de estilo ni preferencias de redacción.

4. Evaluación de las traducciones realizadas con los modelos propios y comparación con una traducción hecha con un sistema de traducción automática comercial

Se seleccionaron los segmentos de forma aleatoria, por lo que no tienen contexto. Se consideraron dos macrocategorías para la revisión: fluidez (incluye aspectos de gramática, puntuación y terminología) y precisión (cantidad de información del texto original que esté reflejada en la traducción: adición, omisión, error, orden de las palabras). No obstante, los poseedores también tomaron nota de otros fenómenos observados.

4.1 Comparación de las traducciones de los modelos para el par francés-español

4.1.1 Valoración manual de la traducción neuronal realizada con Marian

Durante el proceso de traducción, es frecuente que el motor se trabe cuando está traduciendo algún segmento con números y signos o cuando no encuentra una palabra. Este punto permite destacar una de las fortalezas de los modelos neuronales: hay un desempeño elevado en cuanto a gramática, pero siempre persiste la potencialidad de que se degrade la transferencia léxica.

El tiempo que le tomó traducir 1000 segmentos con un total de 30777 palabras fue de aproximadamente una hora y 30 minutos. El procesamiento se realiza segmento a segmento y con una unidad de procesamiento convencional (CPU).

En la muestra de 250 segmentos, se revisaron un total de 7013 palabras. El primer poseedor (P1) tardó 130 minutos, mientras que el segundo poseedor (P2) tardó 132 minutos.

Detalle del esfuerzo temporal

	P1	P2
Tiempo dedicado a la revisión	2 horas y 10 minutos	2 horas y 12 minutos
Tiempo dedicado a la redacción de observaciones, lectura del informe, interpretación del informe de PET y detalle de los errores	35 minutos	38 minutos

Los poseedores detectaron las mejores traducciones en las oraciones cortas, con mucho contraste con los segmentos largos donde, en su mayoría, se encontraban las omisiones de texto, además de en las enumeraciones. Esta marcada degradación de los modelos de traducción neuronal respecto de las traducciones provistas por modelos estadísticos cuando se trata de segmentos largos ya había sido observada en trabajos anteriores (Bentivogli, Bisazza et al., 2016; Koehn, 2017). A continuación se presentan las anotaciones de errores encontrados durante el proceso de revisión:

Detalle de los errores

	Revisión	Incidencias	
		P1	P2
	Número total de cambios	75	85
	Tipo de errores		
Fluidez	Terminología	15	19
	Puntuación	0	2
Precisión	Omisión	18	28
	Adición	3	4
	Orden de las palabras	3	1

Como se ve en el detalle de los errores, el tipo más frecuente es de omisión y el segundo más frecuente es de terminología. El resto de errores, los que completan el conteo, se atribuyen a espaciado, sobre todo, en el uso de comillas, y en la separación entre estas y los elementos inmediatos, y en numeración usada en documentación.

Los poseedores agregaron estas observaciones:

- No se observaron faltas ortográficas, salvo algunas palabras con ortografía anacrónica (como el adverbio *sólo* o los demostrativos *éste* o *ésta*) que responde a la datación de los textos.
- Las siglas se traducen y se despliegan correctamente en la mayoría de los casos.
- Los nombres propios se tradujeron correctamente.
- Las mayúsculas se usaron de forma adecuada.

- Las traducciones son fluidas, en la mayoría de casos, y hay un reordenamiento de palabras propio.
- Las fechas son traducidas correctamente y en el formato correspondiente.
- Los textos en inglés siguen apareciendo en inglés en la traducción.
- La traducción era inconstante en el uso del símbolo de moneda: a veces se usaba, a veces, no.
- La traducción presentó una gran profusión de sinónimos. No hubo muchas repeticiones. Por ejemplo, para el término *discuter*, se usaron los sinónimos *debatir*, *discutir*, *examinar*, *analizar*.

4.1.2 Valoración manual de la traducción estadística realizada con Moses

La traducción obtenida con este modelo se completó en poco tiempo, en comparación con la generada con Marian: menos de una hora. Se debe hacer mención nuevamente de que se utilizó una CPU para esta parte del proceso y no una GPU, como en la parte de entrenamiento, pero si la traducción se hubiera hecho en bloque y con una GPU, esta se habría completado en tan solo minutos.

Se revisaron un total de 7195 palabras. Aunque la revisión demoró menos tiempo que la traducción generada por Marian, se detectaron muchos errores y más variados. A continuación, se da un detalle del esfuerzo temporal y, luego, la anotación de los errores.

Detalle del esfuerzo temporal

	P1	P2
Tiempo dedicado a la revisión	1 hora y 56 minutos	1 hora y 55 minutos
Tiempo dedicado a la redacción de observaciones, lectura del informe, interpretación del informe de PET y detalle de los errores	30 minutos	30 minutos

Detalle de los errores

	Revisión	Incidencias	
		P1	P2
	Número total de cambios	127	124
	Tipo de errores		
Fluidez	Terminología	2	2
	Gramática	18	20
	Cifras	12	13
Precisión	Omisión	22	22
	Adición	6	6
	Términos no traducidos	39	37
	Error de sentido	18	17

Se observa que la mayoría de los errores se dan en la categoría de términos no traducidos. El sistema no detecta las contracciones con apóstrofes, y estas corresponden a la mayoría de errores en esta categoría. Otros términos que no se traducen son los que están unidos por barras (/) o guiones (-), pero también se incluyen términos de distintas categorías funcionales, por lo tanto, no es posible definir una tendencia.

La siguiente categoría corresponde a la omisión. En este caso, no se reflejaron en el texto de llegada todo tipo de términos (sin una ubicación fija en la oración). Luego, están los errores gramaticales: en esta categoría se incluyen los errores de conjugación verbal (ya sea porque aparecen en infinitivo o no hay concordancia con el sujeto) y de no concordancia de otros elementos de la oración. Con similar número de ocurrencias aparecen los errores de sentido, y en esta instancia se reconocieron muchas faltas en la traducción de la preposición *pour* (*para, por, a, ante*).

Los poseedores hicieron observaciones respecto de otros errores detectados y otras complicaciones del modelo:

- La mayoría de los errores en las cifras ocurren cuando, en una oración, hay más de dos grupos de cifras (las cifras están separadas por espacios, por lo que, probablemente, el sistema las detecta como dos cifras distintas).
- Se detectaron problemas de duplicación de comas y de puntos. Hay un problema persistente en este sentido, porque, por otro lado, en muchos lugares faltan comas.
- Al igual que con el modelo de Marian, se dejan espacios entre las comillas y la palabra inmediatamente posterior o anterior.
- Aunque no se da en todos los segmentos, hay muchas ocurrencias de traducciones literales. Las formas de los gerundios en francés se tradujeron por gerundios en español. Por el tipo de revisión realizada, esto no se marcó como error, pues no inhibió la comprensión.
- Hubo muy pocos errores de terminología. Los instrumentos y nombres de organizaciones fueron traducidos correctamente.

4.1.3 Contraste de las traducciones de los dos modelos con una traducción generada por Google Translate

A continuación, se ponen en oposición algunos ejemplos de las traducciones obtenidas con los modelos de Marian y Moses, y una traducción generada con Google Translate.

Segmento original: <i>C'est sur la base des idées que s'articule le socle préférentiel constitué par la politique de paix et de réconciliation nationale prônée par le Président de la République endossée par le référendum populaire et qui reflète la dimension politique de la lutte contre le terrorisme.</i>	
Traducción producida con Marian	Es sobre la base de las ideas que el Presidente de la República apoya en el referéndum popular y que refleja la dimensión política de la lucha contra el terrorismo.
Traducción producida por Moses	C'est sobre la base de las ideas que se articula la base preferencial establecido por la política de paz y de reconciliación nacional preconizada por el Presidente de la República que fue refrendada por el referéndum popular y que refleja la dimensión política de la lucha contra el terrorismo.
Traducción de Google Translate	Sobre la base de las ideas, se articula la base preferencial formada por la política de paz y reconciliación nacional defendida por el Presidente de la República respaldada por el referéndum popular y que refleja la dimensión política de la lucha contra el terrorismo.

En este caso, se observa un caso de omisión de información, el error más frecuente de Marian, pero en este caso la omisión aparece en el medio del enunciado, por lo tanto, por su posición, no se puede definir una tendencia en estos errores. En la traducción de Moses se observa un caso de no traducción de un término con apóstrofo (pero todos los demás términos fueron traducidos), no concordancia entre términos y falta de comas. La traducción de Google Translate, por otra parte, es correcta y completa.

Segmento original: <i>ONU-Habitat devrait disposer de l'appui financier et technique voulu pour répondre aux demandes croissantes d'assistance, aux niveaux national et régional.</i>	
Traducción producida con Marian	ONU-Hábitat debería contar con el apoyo financiero y técnico necesario para atender las crecientes demandas de asistencia en los planos nacional y regional.
Traducción producida por Moses	El ONU-Hábitat debería contar con el apoyo financiero y técnico necesario para satisfacer las crecientes demandas de asistencia, en los planos nacional y regional.
Traducción de Google Translate	ONU-Hábitat debe contar con el apoyo financiero y técnico necesario para responder a las crecientes demandas de asistencia a nivel nacional y regional.

Este es un caso de traducción correcta generada por el modelo de Marian, aunque utiliza un término un poco frecuente para este contexto. En todas las incidencias de *niveaux*, el término se traduce como *planos*, cuando en el texto de referencia se usa *niveles*. La traducción de Moses incurre en el mismo error, pero es correcta, brinda toda la información del texto original. Google Translate genera una traducción correcta y completa.

Segmento original: <i>Il a discuté de la non-prolifération et du désarmement avec le Président russe Medvedev et le Président des États-Unis Obama et s'est réjoui qu'ils aient le mois précédent conjointement affirmé leur détermination à remplir leurs obligations en vertu de l'article VI du Traité sur la non-prolifération des armes nucléaires (TNP).</i>	
Traducción producida con Marian	El Consejo examinó la no proliferación y el desarme con el Presidente Medvedev y el Presidente de los Estados Unidos de América, y el Presidente de los Estados Unidos Obama y el Sr.
Traducción producida por Moses	Se examinó la cuestión de la no proliferación y el desarme con el Presidente Medvedev de Rusia y el Presidente de los Estados Unidos Barack Obama y qu'ils celebró el mes anterior conjuntamente han afirmado su compromiso de cumplir con sus obligaciones en virtud del artículo VI del Tratado sobre la no proliferación de las armas nucleares (TNP).
Traducción de Google Translate	Discutió la no proliferación y el desarme con el presidente ruso Medvedev y el presidente de los Estados Unidos Obama y se mostró satisfecho de que el mes anterior hubiera afirmado conjuntamente su determinación de cumplir con sus obligaciones en virtud del artículo VI del Tratado sobre la no proliferación de las armas nucleares (TNP).

En este segmento largo, se observa otro caso de omisión en la traducción con Marian: falta más de la mitad del contenido y, además, se duplica todo un sintagma. En la traducción de Moses, aparece un error de no traducción y, en este caso, el contenido que le sigue genera un texto incomprensible y sin conexión con la información previa. La traducción de Google suministra toda la información aunque omite los signos de puntuación necesarios.

Segmento original: <i>La loi sur l'enregistrement des naissances et des décès permet à tout déclarant d'ajouter un patronyme au nom de l'enfant et permet à un enfant dont la naissance a été déclarée de manière anonyme de faire enregistrer son nom dans les sept années qui suivent sa naissance.</i>	
Traducción producida con Marian	La Ley de inscripción de nacimientos y defunciones permite a todo declarante añadir un patronímico en nombre del hijo y permite que un niño cuyo nacimiento haya sido declarado de manera anónima registre su nombre en los siete años posteriores a su nacimiento.
Traducción producida por Moses	La Ley de 'enregistrement de nacimientos y defunciones permite a un declarante añadir un apellido en nombre de 'enfant y permite que un niño cuyo nacimiento fue declarada de manera anónima de inscribir su nombre en los siete años siguientes a su nacimiento.
Traducción de Google Translate	La Ley de registro de nacimientos y defunciones permite a cualquier solicitante de registro agregar un apellido al nombre del niño y permite que un niño cuyo nacimiento se haya declarado de forma anónima tenga su nombre registrado dentro de los siete años. sigue su nacimiento.

En la versión de Marian, se presenta toda la información, aunque no se añade la puntuación necesaria. La versión de Moses muestra dos ocurrencias de términos sin traducir, además de términos con faltas de concordancia. La traducción de Google es incorrecta.

Segmento original: <i>Le Gouvernement continue d'organiser des rencontres bilatérales et de participer aux conférences multilatérales régionales en coordination avec la MINUK.</i>	
Traducción producida con Marian	El Gobierno sigue celebrando reuniones bilaterales y participando en conferencias multilaterales regionales en coordinación con la UNMIK.
Traducción producida por Moses	El Gobierno sigue organizando encuentros bilaterales y participar en las conferencias multilaterales regionales en coordinación con la UNMIK.
Traducción de Google Translate	El Gobierno sigue manteniendo reuniones bilaterales y multilaterales asistir a conferencias regionales en coordinación con la UNMIK.

Este es un ejemplo de un segmento corto traducido correctamente por el modelo de Marian. En la versión de Moses, se observa un verbo no conjugado, pero también, una adaptación correcta de las siglas. La traducción de Google Translate es incorrecta.

En modelos entrenados con los mismos datos, se observa que el modelo con mejor desempeño es el creado con Marian. Aunque la primera categoría de errores de este modelo sea de omisión y es grave, la segunda categoría es de terminología, pero este tipo no inhibió la comprensión del texto. En cuanto se contrastan los dos modelos con la traducción automática generada con Google Translate, la valoración varía ligeramente (en muchos casos en los que el modelo neuronal genera una traducción incorrecta, el de Google Translate produce una traducción completa y comprensible).

4.2 Comparación de las traducciones de los modelos para el par inglés-español

4.2.1 Valoración manual de la traducción neuronal realizada con Marian

Para este par, los 1000 segmentos traducidos incluyeron un total de 27652 palabras. La traducción se completó en casi una hora. La traducción, igual que en los modelos anteriores, se procesa segmento a segmento con un CPU.

En este par, la muestra tenía un total 6903 palabras. El primer poseedor (P1) tardó 110 minutos en revisarla, el segundo poseedor (P2) tardó 120 minutos. Tal como en el caso del par francés-español, la traducción con este modelo tuvo un nivel óptimo según opinión de los dos poseedores.

Detalle del esfuerzo temporal

	P1	P2
Tiempo dedicado a la revisión	1 hora y 40 minutos	2 horas
Tiempo dedicado a la redacción de observaciones, lectura del informe, interpretación del informe de PET y detalle de los errores	28 minutos	30 minutos

Lo más saliente que se observó en la traducción fue que se respetaron las reglas de puntuación y las mayúsculas, pues no hubo errores de este tipo. Las siglas se interpretaron y tradujeron correctamente, como en el par francés-español, y las monedas fueron presentadas aún mejor que en dicha traducción. No hubo errores en la traducción de números ni en las cifras. A continuación se presenta un detalle de los errores encontrados:

Detalle de los errores

	Revisión	Incidencias	
		P1	P2
	Número total de cambios	76	75
	Tipo de errores		
Fluidez	Terminología	19	18
	Gramática	3	2
Precisión	Omisión	18	18
	Adición	4	3
	Orden de las palabras	6	6

Al igual que en el par francés-español, la mayor cantidad de errores se registra en la terminología y en traducciones incompletas (omisiones). Los demás errores son de espaciado, falta de comas y separación entre las comillas y los términos inmediatos, también, en los puntos seguidos a una abreviatura. En este modelo, a diferencia del par francés-español, se encontraron 2 casos de términos que no fueron traducidos (detectados por ambos poseedores). Los errores gramaticales corresponden a conjugaciones verbales incorrectas.

Los poseedores hicieron las siguientes observaciones:

- Algunos casos de omisión fueron preocupantes, ya que en la mitad de ellos faltaba traducir más de la mitad del segmento.
- La mayoría de los errores en el uso de terminología no generaron incompreensión del texto, tan solo se diferenciaron de los términos utilizados en la traducción de referencia (por ejemplo, el uso de *programación*, en lugar de *agenda*).
- No se registraron faltas de ortografía.
- Se observa más consistencia en la traducción de términos que en el par francés-español.

4.2.2 Valoración manual de la traducción estadística realizada con Moses

Como en el par francés-español, la traducción generada con este modelo se completó en poco tiempo: cerca de 40 minutos. En contraste con esto, la revisión y la preparación del informe tomaron más tiempo que con el modelo anterior.

Detalle del esfuerzo temporal

	P1	P2
Tiempo dedicado a la revisión	1 hora y 53 minutos	2 horas y 30 minutos
Tiempo dedicado a la redacción de observaciones, lectura del informe, interpretación del informe de PET y detalle de los errores	40 minutos	45 minutos

A continuación, se presenta un detalle de los errores encontrados, no solo son muchos más que los encontrados en el modelo de Marian, sino que el tipo de errores es más diverso.

Detalle de los errores

	Revisión	Incidencias	
		P1	P2
	Número total de cambios	153	170
	Tipo de errores		
Fluidez	Terminología	13	10
	Gramática	57	70
	Cifras	3	7
Precisión	Omisión	18	25
	Adición	13	10
	Términos no traducidos	8	13
	Error de sentido	41	36

La mayor parte de los errores se registran en la categoría de gramática, que incluye no concordancia verbal y adjetival (muy pocas instancias de errores en artículos). En el caso de los verbos, estos estaban muy alejados del sujeto.

La siguiente categoría de errores con más incidencias son los de sentido. Estos están relacionados con el orden de las palabras, falta de puntuación, interpretación incorrecta y traducción incorrecta (uso de términos sin ninguna relación con el contexto).

En el caso de las omisiones, se incluye la falta de preposiciones y, en muchos casos, del relativo *que*. Los términos no traducidos corresponden a los que están unidos por barras (/) o guiones (-).

En el caso de las cifras, hay problemas con los porcentajes, muchas veces se invierten los números o no aparecen completos en el segmento de llegada.

Los demás errores que completan el conteo son de espaciado, falta de comas y separación entre las comillas y los términos inmediatos, o de un punto seguido a una abreviatura y la siguiente palabra.

Otras observaciones de los poseedores:

- A diferencia del modelo de Marian, donde no se detectaron estos errores, hubo varios casos de minúsculas después de puntos. Esto puede deberse a un error al momento de realizar la segmentación.
- Se registraron algunos problemas con los nombres. En muchos casos, estos aparecían incompletos en la traducción.
- Se detectaron, como en el par anterior, varios casos de duplicaciones de comas, pero no se detecta una posición específica donde ocurra en el enunciado.

4.2.3 Contraste de las traducciones de los dos modelos con una traducción generada por Google Translate

Al poner lado a lado la traducción neuronal junto con la traducción estadística, se observa algo que ya se había anticipado: el modelo estadístico reproduce todo el contenido de la información en el segmento original, mientras que el modelo neuronal se destaca en la forma en que reproduce la información. Se han escogido algunos segmentos para comparar las traducciones de cada sistema.

Segmento original: <i>This will help to determine whether any signature methods were used in the creation of the IED and what level of expertise was needed to assemble the device.</i>	
Traducción producida con Marian	Esto ayudará a determinar si se utilizaron métodos de firma para la creación del artefacto explosivo improvisado y qué nivel de conocimientos especializados se necesitaban para montar el artefacto.
Traducción producida por Moses	Ello ayudará a determinar si se utilizaron los métodos de firma en la creación del artefacto explosivo improvisado y cuál es el nivel de conocimientos necesario para montar el artefacto.
Traducción de Google Translate	Esto ayudará a determinar si alguno de los métodos de firma se utilizaron en la creación de la IED y se necesitaba qué nivel de experiencia para montar el dispositivo.

En este ejemplo, Marian logra reproducir íntegramente la información del original, despliega correctamente la sigla utilizada y lo logra por contexto (IED podría tener otros significados). Se elige una traducción adecuada para *expertise (conocimientos especializados)* y se elige el término adecuado para reemplazar *device*, pues hace referencia directa a la sigla. La traducción de Moses es correcta, comprensible y también traduce correctamente la sigla. En cuanto a la de Google Translate, solamente logra reproducir todos los términos utilizados en el segmento original, pero no despliega la sigla y no se ordena correctamente la segunda oración.

Segmento original: <i>Logistical activities, such as air operations, medical evacuation, movement control and transport management, require that fault-tolerant communication systems be available.</i>	
Traducción producida con Marian	Las actividades logísticas, como las operaciones aéreas, la evacuación médica, el control de la circulación y la gestión del transporte, exigen que se disponga de sistemas de comunicación resistentes a las averías.
Traducción producida por Moses	Actividades logísticas, tales como las operaciones aéreas, las evacuaciones médicas, control de movimientos y gestión de transporte a prueba de fallos, exigen que se disponga de sistemas de comunicaciones.
Traducción de Google Translate	actividades logísticas, tales como las operaciones aéreas, evacuación médica, control de movimiento y la gestión del transporte, requieren que los sistemas de comunicación tolerantes a fallos estar disponible.

En este caso, la formulación que se registra en la traducción de Marian es correcta, junto con el correcto uso de artículos antes de todos los sintagmas en la numeración y una reformulación adecuada para el adjetivo *fault-tolerant (resistentes a las averías)*. La traducción realizada por Moses es errónea, coloca el adjetivo en una posición incorrecta y no transmite lo que dice el original. En cuanto a la traducción de Google Translate, es imprecisa en la conjugación del verbo *to be*, lo que hace que no sea correcta ni comprensible.

Segmento original: <i>Coverage of the Bacille Calmette-Guerin (BCG) immunisation for children below one year in the year 2000 was 99.9 percent and have increased to 100.0 percent in 2005.</i>	
Traducción producida con Marian	La cobertura de la inmunización de los niños menores de un año (BCG) para los niños menores de un año fue del 99,9 % y aumentó al 100,0 % en 2005.
Traducción producida por Moses	La cobertura de la inmunización Bacille Calmette-Guerin (BCG) para los niños de menos de un año en el año 2000 fue del 99.9 % y ha aumentado al 100.0 % en 2005.
Traducción de Google Translate	La cobertura de la inmunización bacilo de Calmette-Guerin (BCG) para los niños menores de un año en el año 2000 fue de 99,9 por ciento y han aumentado a 100,0 por ciento en 2005.

La traducción de Moses es correcta, a pesar de la cacofonía (*de un año en el año*), pues transmite toda la información comunicada en el segmento original. En la versión de Marian, se omite la traducción del nombre propio, el primer año (2000) y cambia el orden de las palabras, por lo tanto, es incorrecta. La versión de Google Translate es errónea pues traduce una parte del nombre propio, pero a partir de eso se traduce la demás información correctamente.

Segmento original: <i>Besides being desirable from an efficiency perspective, gender equality is itself a legitimate policy goal, as better opportunities for women foster human development.</i>	
Traducción producida con Marian	Además de ser conveniente desde el punto de vista de la eficiencia, la igualdad entre los géneros es en sí misma un objetivo político legítimo, ya que mejores oportunidades para las mujeres fomentan el desarrollo humano.
Traducción producida por Moses	Además de ser conveniente desde el punto de vista de la eficiencia, la igualdad entre los géneros es un objetivo de política legítimo, como las mejores oportunidades para promover el desarrollo humano.
Traducción de Google Translate	Además de ser deseable desde una perspectiva de eficiencia, la igualdad de género es en sí mismo un objetivo legítimo de política, como mejores oportunidades para las mujeres el desarrollo humano de crianza.

En el ejemplo a continuación, la versión de Marian logra reflejar correctamente el sentido del texto en el segmento original. La versión de Moses comunica un sentido erróneo en la segunda parte del enunciado, no solo de interpretación, sino en omisión de la traducción del término *women*. La versión de Google es incorrecta por la misma razón que en la versión de Moses y no traduce correctamente el adjetivo *desirable*, que en este contexto es *conveniente*.

Segmento original: <i>According to a technology writer for the Associated Press, the United States had plans to fire a cruise missile tipped with a high-power EMP emitter (the e-bomb) which Jane's Information Group claims "fries the electronics without killing the people".</i>	
Traducción producida con Marian	Según un escritor de tecnología para Associated Press, los Estados Unidos tenían planes de disparar un misil de crucero con una potencia de alta potencia (la bomba electrónica) que Jane ' s Information Group reclama " Las reservas de la electrónica sin matar a la población ".
Traducción producida por Moses	Con arreglo a una tecnología escritor de Associated Press, los planes de los Estados Unidos, ha inclinado a disparar un misil de crucero de alta potencia (EMP) que el emisor e-bomb Jane ' s Information Group " Fries las reclamaciones sin matando a la gente de la electrónica ".
Traducción de Google Translate	De acuerdo con un escritor de tecnología para la Associated Press, Estados Unidos tenía planes para disparar un misil de crucero de punta con un emisor EMP alta potencia (la bomba electrónica), que el Grupo de Información de Jane afirma "papas a la electrónica sin matar a la gente".

Este último ejemplo es un caso interesante pues ninguna de las tres traducciones es correcta, ninguna logra reflejar el sentido de la frase que reproduce una declaración del Jane's Information Group. Además de eso, todas las versiones tienen varios errores (interpretación del nombre propio, redundancia, sinsentidos).

Al comparar la cantidad de errores encontrados en los dos modelos, la traducción realizada con Moses sobrepasa en número de faltas —y en diversidad— a la realizada con Marian. Esto está muy ligado al tiempo destinado a la revisión y al trabajo posterior a la revisión, y nos da una impresión sobre la calidad sin necesidad de mostrar todo la muestra traducida. La traducción del modelo neuronal supera en desempeño al modelo estadístico.

4.3 Valoración de los modelos con métricas automáticas

Las métricas automáticas complementan la evaluación de los poseedores y se centran en la traducción obtenida de un sistema de traducción automática comparada con un texto de referencia. En este trabajo se ha escogido utilizar BLEU, NIST y WER. Para evaluar las traducciones de ambos modelos junto con la de Google Translate, se utilizaron scripts preparados con la biblioteca NLTK en Python (<https://www.nltk.org/>).

BLEU (Bilingual Evaluation Understudy) es una métrica automática que fue introducida por empleados de IBM en 2002 (Papineni, Roukos et al., 2002). Es el método más utilizado por los sistemas de traducción automática para analizar traducciones (Mae Pan, 2016). El criterio de su funcionamiento es que, mientras más se parece la traducción automática a la traducción humana usada como referencia, más alto es el puntaje. El cálculo se basa en el texto de referencia, toma en cuenta solamente métricas y se centra en dos aspectos principales, la precisión (*adequacy*) y la fluidez (*fluency*), ambos medidos en términos de n-gramas: cuenta los n-gramas en la traducción de muestra y los compara con los n-gramas en el texto de referencia. Este tipo de método funciona bien para los sistemas de traducción estadística, ya que se enfoca en la palabra (n-gramas) como unidad de estructura. Los sistemas de traducción neuronal generan resultados que se diferencian de las traducciones de referencia en el orden y términos, sobre todo, por lo cual esta métrica automática podría no detectar las traducciones correctas y, por lo tanto, la puntuación será baja.

La métrica NIST toma su nombre de las siglas del Instituto Nacional de Normas y Tecnología de los EE. UU. (National Institute of Standards and Technology) y es una métrica que se deriva de BLEU, en cuanto a que evalúa la precisión de los n-gramas, pero tiene una diferencia. En lugar de secuencias de n-gramas, se considera la ganancia de información de cada una. Esto quiere decir que el sistema evalúa las coincidencias de n-gramas, pero da mayores calificaciones a las coincidencias más difíciles que a las más fáciles. Esta métrica da un "reconocimiento" con puntaje al sistema de traducción que es capaz de dar la traducción correcta y en el orden

correcto (Zhang et al., 2004). Como BLEU, NIST puede medir las similitudes entre un texto traducido y uno de referencia, pero no evalúa la calidad.

WER mide la distancia de edición (su nombre es el acrónimo de Word Error Rate o índice de errores de palabras), es decir, la cantidad de palabras que deben agregarse, eliminarse o sustituirse en un texto para lograr que este se parezca a un texto de referencia. Al igual que las dos primeras métricas, esta toma la traducción de referencia como modelo de traducción correcta, con lo cual el uso de vocabulario y de estilo similar a la referencia es lo único que podrá garantizar una calificación alta (Tomás et al., 2003).

Se presentan, a continuación, las puntuaciones obtenidas para cada par:

Valoración automática de las traducciones francés-español

	BLEU	NIST	WER
Marian	0.4132	8.607	0.492
Moses	0.4544	8.927	0.512
Google Translate	0.3626	7.927	0.5882

En este par, Moses obtuvo una mayor puntuación con BLEU y NIST, pero no con WER, donde fue superado por Google Translate. Esto se debe a que en este par, a diferencia de los errores detectados en la traducción inglés-español, la mayor parte de faltas correspondieron a términos no traducidos.

Valoración automática de las traducciones inglés-español

	BLEU	NIST	WER
Marian	0.4547	9.262	0.4427
Moses	0.4906	9.483	0.4672
Google Translate	0.4183	8.672	0.4672

En el par inglés-español, las mejores puntuaciones —en todas las métricas— las obtuvo la traducción realizada con Moses. Es posible anticipar esto a partir de la propia naturaleza de las métricas y de los sistemas estadísticos: las traducciones generadas pueden no ser fluidas, pero cuentan con un mecanismo que garantiza que se traduzcan todos los términos.

Según las calificaciones para ambos pares, las traducciones generadas con Moses superan a las generadas con Marian.

4.4 Observaciones sobre el desempeño de los dos modelos en ambos pares de idiomas

La sección anterior contrasta con la evaluación manual, según la cual las mejores traducciones, en ambos idiomas, fueron generadas por el modelo creado con Marian. En muchos casos, también supera al desempeño de Google Translate. El modelo creado con Moses permitió realizar traducciones más rápidamente, pero demandó un mayor esfuerzo temporal por parte de los poseedores. La cantidad de errores encontrados en las traducciones de Moses no solo fue más numerosa, sino también variada, y con diferencias entre cada par de idiomas. En la traducción francés-español, la mayoría fueron de precisión, con muchos términos no traducidos. En la traducción inglés-español, la mayoría de errores fueron de fluidez, en lo referente a concordancia verbal y adjetival. En efecto, como se adelantó en la parte de análisis, los errores encontrados fueron muy distintos en cada par traducido con Moses, no así en las traducciones generadas con Marian.

5. La importancia de la posesición como parte del proceso de traducción automática: traducción aumentada

La necesidad de que una traducción generada con un sistema de traducción automática pase por un proceso de posesición se evidenció en la sección anterior, donde se presentaron algunos ejemplos de sistemas TAN y TAE comparados con una traducción comercial: para poder usar cualquiera de ellos, es necesario que los textos sean revisados.

Las traducciones generadas por sistemas de traducción automática se enfocan en dos aplicaciones: asimilación (identificar la idea principal de un texto) y diseminación (usar la traducción automática como un borrador que se edita para obtener una traducción de calidad publicable) (Forcada, Scarton, Specia, et al., 2018). Las dos demandan un nivel distinto de calidad: la primera podría usar la traducción tal como la genera el sistema, mientras que, para la segunda aplicación, es imperante incluir un paso de posesición.

La posesición es un trabajo de intervención humana sobre una traducción automática que incluye revisión y correcciones lingüísticas. Este trabajo está sujeto a dos criterios: la calidad de los resultados brutos de la traducción automática y la calidad final que se espera del contenido (TAUS, 2010) que, finalmente, dependerá de las necesidades y los requerimientos del usuario/cliente. Estos dos criterios se conjugan con la aplicación que se pretenden dar al resultado (asimilación o diseminación) para que el proceso de posesición pueda ser considerado simple o completo.

Según dos estudios realizados a proveedores de servicios lingüísticos de todo el mundo, uno en 2018 y el otro en 2019, el avance en la calidad de los resultados de traducción automática, la profusión de sistemas de traducción automática en la industria, el creciente uso de herramientas y aplicaciones como parte del proceso de traducción, una tendencia a ahorrar costos y, a la vez, cumplir con demandantes necesidades de traducción ha hecho que la solicitud de servicios de posesición crezca (Diño, 2018; Bond, 2019). Con la demanda de servicios de posesición en alza, se debe recordar que la posesición cuesta entre 40 % y 50 % menos que el servicio de traducción (Lommel, 2017), y esto podría desanimar a muchos profesionales que no cuenten con los conocimientos, que no alcancen el nivel de productividad o que no logren aprovechar los recursos tecnológicos disponibles.

A partir de las traducciones revisadas, se podría concluir que el modelo de traducción neuronal, aún con las mejoras que sea necesario aplicar, tiene un desempeño óptimo. En este trabajo, se decidió preparar un modelo usando el entorno Marian, pero existen muchos otros sistemas de aprendizaje automático de código abierto (Nematus, Tensorflow, Torch, Matlab, MXNet, Theano, PYTorch, entre otros) que permiten elaborar modelos de traducción neuronal propios, con lo cual la traducción neuronal puede dejar de ser un dominio de los investigadores, programadores o de empresas, y se puede acercar a los ciudadanos digitales con conocimientos especializados. Ya se trate de traducciones generadas por modelos de elaboración propia o de traducciones automáticas preparadas con motores comerciales, es innegable que la posesición es una tarea relevante en la cartera de los proveedores de servicios lingüísticos en la actualidad y que incluir la traducción automática en el flujo de trabajo es una idea atractiva.

En línea con esta idea, Common Sense Advisory introduce el concepto de *augmented translation* (traducción aumentada), que se basa en la definición de realidad aumentada, en la cual la inteligencia artificial es solamente una herramienta para mejorar una experiencia, y que coloca al traductor como parte central del proceso de traducción (DePalma, 2017). Así, la traducción automática vendría a ser solamente una herramienta que apuntala los esfuerzos del traductor, quien sería el encargado de escoger o no los resultados propuestos por una aplicación. Esta estructura integrada de traducción aumentada también podría conjugarse con memorias de traducción, instancias de manejo de terminología, un proceso automático de gestión de proyectos y mejoramiento automatizado de contenido (DePalma, 2017). Toda la parte automática se dedicaría al procesamiento de datos y a potenciar la actividad de traducción, el traductor sería el encargado de poseer el resultado de la interacción de toda la estructura. No obstante, para que esta integración pueda fluir, el profesional de la lengua tiene que estar capacitado para trabajar con ella.

Existe otro enfoque que puede resultar interesante en esta interacción del traductor y la tecnología llamado predicción de traducción interactiva (*ITP, Interactive Translation Prediction*) (Diño, 2019), que se fundamenta en una función de autorrellenado y sugerencias de traducción a medida que el profesional escribe. Esta idea base ya había sido presentada hace más de 20 años en un trabajo de Foster, Isabelle y Plamondon (Forcada, 2017) y podría ser otra forma de integrar la funcionalidad de traducción automática en el flujo regular del trabajo.

6. Conclusiones

El modelo de traducción neuronal genera traducciones óptimas —salvo segmentos con errores marcados de terminología y omisión—, pero este desempeño se puede ver menguado en segmentos muy largos. Estos buenos resultados siempre dependerán de que el modelo haya sido entrenado con datos suficientes, de calidad y relacionados con la temática que se pretende traducir. Si no es así, la traducción será deficiente en cualquier caso. El modelo de traducción estadística, en la mayoría de casos, produce traducciones que comunican mucha de la información del segmento original, pero no en todos los casos, pues se detectaron distintos tipos de errores de gramática, en el par inglés-español y de términos no traducidos, en el par francés-español, un fenómeno que se contradice con una característica básica de estos sistemas.

Después de comparar las traducciones en ambos pares de idiomas, tomando en cuenta la cantidad de errores, sumado el esfuerzo temporal y técnico, se concluye que los modelos creados con Marian tienen un mejor desempeño que los de Moses e, incluso, que las traducciones provistas por el motor comercial. Esta valoración contrasta con la calificación de las métricas automáticas que, por su naturaleza, tienden a beneficiar a los modelos estadísticos; con todo, las métricas automáticas de evaluación solo puntúan la semejanza de una traducción automática con una traducción de referencia, pero nunca podrán definir la calidad de una traducción. Las métricas manuales ayudan a subsanar esto, por lo tanto el esfuerzo y las anotaciones de los poseedores sobre una traducción automática deben ser los indicadores de calidad de un modelo. A partir de los esfuerzos temporales y técnicos, es posible colegir que una traducción automática óptima demanda un menor esfuerzo de posesión.

En la revisión participaron solamente dos poseedores, quienes trabajaron en un texto relativamente pequeño (250 segmentos, una muestra ínfima respecto del universo de segmentos usados para el entrenamiento), pero sus anotaciones lograron dar un vistazo de cómo trabaja cada sistema, comprobarlo en ejemplos y contrastarlo con la traducción de un motor comercial. Los tipos de errores y el tamaño de la muestra no permiten definir una tendencia para cada par de idiomas, tan solo permiten observar el desempeño de cada modelo para el conjunto de datos utilizados. La participación de muchos más profesionales en un estudio de mayor escala podría proporcionar resultados más contundentes sobre el tipo de errores que deberían resolverse en cada modelo.

Las traducciones producidas por los sistemas de traducción automática pueden alcanzar un nivel de calidad publicable o no, todo depende de la necesidad del usuario final. Los esfuerzos deben centrarse en mejorar la eficiencia de los modelos y esta depende tanto de la cantidad como de la calidad de datos que se utilicen para entrenar a los sistemas. La eficiencia permite que el traductor/poseedor pueda prestarle más atención al texto original y pulir el estilo, en lugar de enfocarse solamente en los segmentos traducidos, que es el cambio de paradigma que se da entre el trabajo de un traductor y un poseedor. La realidad es que la tecnología va a seguir desarrollándose y, para que la traducción automática pueda aprovecharse y colaborar con la productividad, el traductor tiene que estar capacitado y preparado para integrar esta tecnología en su flujo de trabajo y utilizarla al máximo.

7. Bibliografía

Bentivogli, L., Bisazza, A., Cettolo, M., Federico, M. (2016). *Neural versus Phrase-Based Machine Translation Quality: An in-depth analysis on English-German and English-French*. Fondazione Bruno Kessler. Disponible en: <https://aclweb.org/anthology/D16-1025> [Última fecha de acceso: 03/05/2019]

Bojar, O. et al. (2017). *Proceedings of the Conference on Machine Translation (WMT)*, Volume 2: Shared Task Papers, pp. 169–214. Disponible en: <https://www.aclweb.org/anthology/W17-4717> [Última fecha de acceso: 12/05/2019]

Bond, E. (2019). *Survey Examines Machine Translation Post-Editing Among Freelancers and LSPs*. Slator.com. Disponible en: <https://slator.com/academia/survey-examines-machine-translation-post-editing-among-freelancers-and-lsps/> [Última fecha de acceso: 26/05/2019]

Burchardt, A. et al. (2017). *A Linguistic Evaluation of Rule-Based, Phrase-Based, and Neural MT Engines*. *The Prague Bulletin of Mathematical Linguistics*. Disponible en: <https://ufal.mff.cuni.cz/pbml/108/art-burchardt-macketanz-dehdari-heigold-peter-williams.pdf> [Última fecha de acceso: 03/05/2019]

Casacuberta, F., Peris, A. (2017). *Traducción automática neuronal*. *Revista Tradumàtica*. *Tecnologies de la Traducció*, 15, pp. 66-74. Disponible en: https://ddd.uab.cat/pub/tradumatica/tradumatica_a2017n15/tradumatica_a2017n15p66.pdf [Última fecha de acceso: 31/03/2019]

Castilho, S. et al. (2017). *Is Neural Machine Translation the New State of the Art?* *The Prague Bulletin of Mathematical Linguistics*. Disponible en: <https://ufal.mff.cuni.cz/pbml/108/art-castilho-moorkens-gaspari-tinsley-calixto-way.pdf> [Última fecha de acceso: 3 de mayo de 2019]

DePalma, D. (2017). *Augmented Translation Powers Up Language Services*. *Common Sense Advisory*. Disponible en: <https://csa-research.com/Insights/ArticleID/140/Augmented-Translation-Powers-up-Language-Services> [Última fecha de acceso: 02/05/2019]

Diño, G. (2018). *Reader Polls: In-House Translators, Asia, Post-Editing, 2019 Outlook*. Slator.com. Disponible en: <https://slator.com/features/reader-polls-in-house-translators-asia-post-editing-2019-outlook/> [Última fecha de acceso: 26/05/2019]

Diño, G. (2019). *Reader Polls: Post-Editing, Interactive Prediction, Translation Devices, and M&A*. Slator.com. Disponible en: <https://slator.com/features/reader-polls-post-editing-interactive-prediction-translation-devices-and-ma/> [Última fecha de acceso: 07/06/2019]

Eisele, A., Chen, Y. (2010). *MultiUN: A Multilingual Corpus from United Nation Documents*. Disponible en: http://www.dfki.de/lt/publication_show.php?id=4790 [Última fecha de acceso: 12/05/2019]

Forcada, M. (2017). *Making sense of neural machine translation*. *Universitat d'Alacant*. Disponible en: <https://www.dlsi.ua.es/~mlf/docum/forcada17j2.pdf> [Última fecha de acceso: 03/05/2019]

Forcada, M., Scarton, C., Specia, L. et al. (2018). *Exploring Gap Filling as a Cheaper Alternative to Reading Comprehension Questionnaires when Evaluating Machine Translation for Gisting*, *Proceedings of the Third Conference on Machine Translation (WMT)*, *Volumen 1: Research Papers*, 192–203. Disponible en: <https://aclweb.org/anthology/W18-6320> [Última fecha de acceso: 15/05/2019]

Junczys-Dowmunt, M., Grundkiewicz, R., Dwojak, T. et al. (2018). *Marian: Fast Neural Machine Translation in C++*. Disponible en: <https://www.aclweb.org/anthology/P18-4020> [Última fecha de acceso: 12/05/2019]

- Kenny, D. (2018). Sustaining Disruption? The Transition from Statistical to Neural Machine Translation. *Revista Tradumàtica*. Tecnologies de la Traducció, N. 19, pp. 59-70. Disponible en: https://revistes.uab.cat/tradumatica/article/view/n16-kenny/pdf_58 [Última fecha de acceso: 12/05/2019]
- Koehn, P. et al. (2007). *Moses: Open Source Toolkit for Statistical Machine Translation*. Annual Meeting of the Association for Computational Linguistics (ACL), demonstration session, Praga, República Checa, junio de 2007. Disponible en: <http://www.statmt.org/moses/?n=Moses.Overview> [Última fecha de acceso: 21/04/2019]
- Koehn, P., Knowles, R. (2017). *Six Challenges for Neural Machine Translation*. Johns Hopkins University. Disponible en: <https://www.aclweb.org/anthology/W17-3204> [Última fecha de acceso: 20/05/2019]
- Koponen, M., Salmi, L., Nikulin, M. (2019). *A product and process analysis of post-editor corrections on neural, statistical and rule-based machine translation output*. University of Turku. Disponible en: <https://link.springer.com/article/10.1007/s10590-019-09228-7.pdf> [Última fecha de acceso: 03/05/2019]
- Lommel, A. (2017) *Augmented Translation Puts Translators Back in the Center*. Intelligent Information Blog. Disponible en: <https://intelligent-information.blog/en/augmented-translation-puts-translators-back-in-the-center/> [Última fecha de acceso: 21/04/2019]
- Mae Pan, H. (2016). *How BLEU Measures Translation and Why It Matters*. Slator.com. Disponible en: <https://slator.com/technology/how-bleu-measures-translation-and-why-it-matters/> [Última fecha de acceso: 21/04/2019]
- Papineni, K., Roukos, S., Ward, T., & Zhu, W. J. (2002, July). *BLEU: a method for automatic evaluation of machine translation*. In Proceedings of the 40th annual meeting on association for computational linguistics (pp. 311-318). Association for Computational Linguistics.
- Peris, A., Domingo, M., Casacuberta, F. (2016). *Interactive neural machine translation*. Computer Speech and Language. Disponible en: <http://hdl.handle.net/10251/83641> [Última fecha de acceso: 03/05/2019]
- Peris, A., Casacuberta, F.; (2017). Traducción automática neuronal. *Revista Tradumàtica*. Tecnologies de la Traducció, 15, 66-74. Disponible en: <https://doi.org/10.5565/rev/tradumatica.203> [Última fecha de acceso: 21/04/2019]
- Peris, A., Cebrián, L., Casacuberta, F. (2018). *Online Learning for Neural Machine Translation Post-editing*. Disponible en: <https://arxiv.org/pdf/1706.03196.pdf> [Última fecha de acceso: 21/04/2019]
- Slator (2018). *Slator Neural Machine Translation Report 2018*. Disponible en: <https://slator.com/data-research/slator-neural-machine-translation-report-2018/> [Última fecha de acceso: 12/05/2019]
- TAUS (2010). *MT Post-editing Guidelines*. Disponible en: <https://www.taus.net/academy/best-practices/postedit-best-practices/machine-translation-post-editing-guidelines> [Última fecha de acceso: 26/05/2019]
- TAUS (2019). *Harmonized DQF-MQM Error Typology*. Disponible en: <https://www.taus.net/evaluate/qt21-project#dqf-qt21> [Última fecha de acceso: 21/04/2019]
- Tiedemann, Jörg (2012). *Parallel Data, Tools and Interfaces in Opus*. 8th International Conference on Language Resources and Evaluation (LREC 2012). Disponible en: http://www.lrec-conf.org/proceedings/lrec2012/pdf/463_Paper.pdf [Última fecha de acceso: 14/05/2019]

- Tomás, J., Más, J., Casacuberta, F. (2003). *A Quantitative Method for Machine Translation Evaluation*. Association for Computational Linguistics. Disponible en: <https://www.aclweb.org/anthology/W03-2804> [Última fecha de acceso: 16/06/2019]
- Volkart, L., Bouillon, P., Girletti, S. (2018). *Statistical vs. Neural Machine Translation: A Comparison of MTH and DeepL at Swiss Post's Language Service*. Université de Genève. Disponible en: <https://archive-ouverte.unige.ch/unige:111777> [Última fecha de acceso: 16/05/2019]
- Wang, X., Lu Z., Tu, Z. et al. (2017). *Neural Machine Translation Advised by Statistical Machine Translation*. Association for the Advancement of Artificial Intelligence. Disponible en: <https://arxiv.org/pdf/1610.05150.pdf> [Última fecha de acceso: 12/05/2019]
- Wu, Y. et al. (2016). *Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation*. Google. Disponible en: <https://arxiv.org/pdf/1609.08144.pdf> [Última fecha de acceso: 20/05/2019]
- Yanfang, J., Michael, C., Xiangling, W. (2019). *How does the post-editing of neural machine translation compare with from-scratch translation? A product and process study*. The Journal of Specialised Translation. Disponible en: https://www.jostrans.org/issue31/art_jia.pdf [Última fecha de acceso: 20/05/2019]
- Zhang, Y., Vogel, S., Waibel, A. (2004). *Interpreting BLEU/NIST Scores: How Much Improvement Do We Need to Have a Better System?* Language Resources and Evaluation Conference 2004. Disponible en: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.536.5232&rep=rep1&type=pdf> [Última fecha de acceso: 16/06/2019]
- Zhang, M. (2017). *History and Frontier of the Neural Machine Translation*. Medium. Disponible en: <https://medium.com/syncedreview/history-and-frontier-of-the-neural-machine-translation-dc981d25422d> [Última fecha de acceso: 21/04/2019]