



Universitat
Oberta
de Catalunya

CREACIÓN DE MOTORES DE TRADUCCIÓN
AUTOMÁTICA (ESTADÍSTICA Y NEURONAL) INGLÉS-
ESPAÑOL ESPECIALIZADOS EN EL CAMPO DE LA
AVIACIÓN CON LA HERRAMIENTA MTUOC

Carlota Rodríguez del Rosario

Trabajo Final de Máster
Tutora: Silvia Rodríguez Vázquez
Máster de Traducción y Tecnologías
Universitat Oberta de Catalunya

Junio 2021

RESUMEN

En el presente trabajo de fin de máster se muestra el proceso de creación de dos motores de traducción automática inglés-español, uno estadístico y otro neuronal, especializados en el campo de la aviación a través de la herramienta de libre acceso MTUOC. Se explican detalladamente todos los pasos que componen el proceso: elección de textos, compilación del corpus, procesamiento del corpus, entrenamiento de los motores de traducción, utilización de los mismos y una evaluación final. Esta evaluación final compara los resultados del motor de traducción entrenado en este trabajo con los propuestos por los traductores automáticos de uso generalizado *Google Translate* (como representación de la traducción automática neuronal) y *Yandex Translate* (como representación de la traducción automática híbrida estadística-neuronal). De esta forma, se busca comprobar si conviene entrenar motores de traducción automática especializados para mejorar la calidad terminológica de los traductores automáticos de uso general.

PALABRAS CLAVE: *traducción automática, traducción automática estadística, traducción automática neuronal, entrenamiento de motores de traducción, creación de corpus, procesamiento de corpus, aviación.*

ABSTRACT

This master's final project shows the process of creating two machine translation engines, one statistical and the other neural, specialized in the field of aviation, using the freely available tool MTUOC. All the steps involved in the process are explained in detail: choice of texts, compilation of the corpus, corpus processing, training of the translation engines, use of the engines and a final evaluation. This final evaluation compares the results of the translation engine trained in this work with those proposed by the widely used machine translators *Google Translate* (as a representation of neural machine translation) and *Yandex Translate* (as a representation of statistical-neural hybrid machine translation). In this way, the aim is to test whether it is convenient to train specialized machine translation engines to improve the terminological quality of general-purpose machine translators.

KEY WORDS: *machine translation, statistical machine translation, neural machine translation, translation engine training, corpus creation, corpus processing, aviation.*

ÍNDICE GENERAL

1.INTRODUCCIÓN	4
1.1 CONTEXTO Y MOTIVACIÓN	4
1.2 HIPÓTESIS Y OBJETIVOS.....	5
1.3 METODOLOGÍA	5
1.4 ESTRUCTURA DEL TRABAJO.....	5
2. MARCO TEÓRICO	6
2.1 ESTADO DE LA CUESTIÓN	7
2.2 LA TRADUCCIÓN AUTOMÁTICA: UNA BREVE HISTORIA	8
2.4 TRADUCCIÓN AUTOMÁTICA ESTADÍSTICA.....	13
2.5 TRADUCCIÓN AUTOMÁTICA NEURONAL	18
2.6 EVALUACIÓN DE LA TRADUCCIÓN AUTOMÁTICA	21
3. MARCO METODOLÓGICO	22
3.1 OBJETIVOS E HIPÓTESIS.....	23
3.2 DESARROLLO DE LA HERRAMIENTA	23
<i>Etapa I: Creación del corpus</i>	23
<i>Etapa II: Procesamiento de los corpus</i>	30
<i>Etapa III: Entrenamiento de los motores de traducción</i>	36
<i>Etapa IV: Evaluación y comparación con sistemas de traducción de dominio general</i>	41
4. RESULTADOS	44
4.1 RESULTADOS DE LA TRADUCCIÓN AUTOMÁTICA ESTADÍSTICA	44
4.2 RESULTADOS DE LA TRADUCCIÓN AUTOMÁTICA NEURONAL	48
5. CONCLUSIONES	52
5.1 CONTRIBUCIÓN AL ÁMBITO DE ESTUDIO	52
5.2 LIMITACIONES DEL TRABAJO.....	53
5.3 FUTUROS ESTUDIOS / FUTURAS VÍAS DE INVESTIGACIÓN	53
6. BIBLIOGRAFÍA	54
ANEXO I: PROGRAMAS UTILIZADOS PARA LA CREACIÓN DEL CORPUS Y EL ENTRENAMIENTO DE LOS MOTORES DE TRADUCCIÓN AUTOMÁTICOS	59
ANEXO II: TEXTOS UTILIZADOS PARA EL CORPUS	61
ANEXO III: TEXTOS EVALUACIÓN	71
TEXTO CREADO PARA EVALUAR TRADUCCIONES.....	71
TEXTO TRADUCIDO POR TRADUCTOR HUMANO	73
TEXTO TRADUCIDO CON NMT (MTUOC).....	76
TEXTO TRADUCIDO CON NMT (GOOGLE TRANSLATE)	79
TEXTO TRADUCIDO CON SMT (MTUOC).....	82
TEXTO TRADUCIDO CON SMT (YANDEX TRANSLATE).....	85

ÍNDICE DE ILUSTRACIONES, ECUACIONES Y TABLAS

ILUSTRACIÓN 1 ECUACIÓN 1 FÓRMULA SMT	14
ILUSTRACIÓN 2 ECUACIÓN 2 FÓRMULA CANAL RUIDOSO APLICADA A SMT	14
ILUSTRACIÓN 3 ECUACIÓN 3 CÁLCULO DE N-GRAMAS	15
ILUSTRACIÓN 4 EJEMPLO OPCIONES DE UN MODELO DE TRADUCCIÓN (KOEHN, 2010, 160)	18
ILUSTRACIÓN 5 EJEMPLO DE ARQUITECTURA CODIFICADORA-DECODIFICADORA DE NMT (CASACUBERTA NOLLA & PERIS ABRIL, 2017, P. 69)	20
ILUSTRACIÓN 6 ENTORNO DE TRABAJO EN VISUAL STUDIO EN UNA FASE AVANZADA DEL PROYECTO	27
ILUSTRACIÓN 7 EJEMPLO DE ARCHIVO BATCH DEL CORPUS IATA	28
ILUSTRACIÓN 8 EJEMPLO DE FUNCIONAMIENTO DE PROGRAMA MTUOC-CLEAN-PARALLEL-CORPUS.PY CON EL CORPUS IATA	29
ILUSTRACIÓN 9 SEGMENTO DEL CORPUS ESPECIALIZADO	30
ILUSTRACIÓN 10 RESULTADO DE LA UTILIZACIÓN DEL PROGRAMA MTUOC-COMBINE-CORPUS.PY	33
ILUSTRACIÓN 11 ENTRENAMIENTO DEL MOTOR DE SMT	37
ILUSTRACIÓN 12 CONFIGURACIÓN SERVIDOR MTUOC TRANSLATOR	40
ILUSTRACIÓN 13 EJEMPLO TRADUCCIÓN TEST CON MTUOC TRANSLATOR	41
ILUSTRACIÓN 14 EVALUADOR INTERACTIVO BLEU	43
ILUSTRACIÓN 15 RESULTADOS DEL ANÁLISIS BLEU CON TILDE (SMT)	45
ILUSTRACIÓN 16 EJEMPLO TRADUCCIÓN INCORRECTA MTUOC-TRANSLATOR SMT (BLEU)	45
ILUSTRACIÓN 17 EJEMPLO BUENA ELECCIÓN TERMINOLÓGICA MTUOC-TRANSLATOR SMT (BLEU)	46
ILUSTRACIÓN 18 EJEMPLO FALLO CONCORDANCIA GRAMATICAL MTUOC-TRANSLATOR (BLEU)	47
ILUSTRACIÓN 19 EJEMPLO TERMINOLOGÍA ESPECIALIZADA (BLEU)	47
ILUSTRACIÓN 20 EJEMPLO TERMINOLOGÍA SIN TRADUCIR EN MTUOC-TRANSLATOR (BLEU)	47
ILUSTRACIÓN 21 RESULTADOS ANÁLISIS BLEU DE TILDE (NMT)	48
ILUSTRACIÓN 22 EJEMPLO VARIACIÓN TIEMPO VERBAL (BLEU)	49
ILUSTRACIÓN 23 EJEMPLO DE MEJOR TRADUCCIÓN TERMINOLÓGICA MTUOC-TRANSLATOR (TILDE)	50
ILUSTRACIÓN 24 EJEMPLO II DE MEJOR TRADUCCIÓN TERMINOLÓGICA CON MTUOC-TRANSLATOR (TILDE)	50
ILUSTRACIÓN 25 EJEMPLO TRADUCCIÓN LITERAL GOOGLE TRANSLATE (TILDE)	50
ILUSTRACIÓN 26 EJEMPLO MALA ELECCIÓN TERMINOLÓGICA POR PARTE DE AMBAS TA (TILDE)	51
ILUSTRACIÓN 27 EJEMPLO II MALA ELECCIÓN TERMINOLÓGICA POR AMBAS TA (TILDE)	51
TABLA 1 EJEMPLO MODELO DE LENGUA SMT	16
TABLA 2 EJEMPLO CÁLCULO P(T)	16
TABLA 3 PROGRAMAS PYTHON COMPILACIÓN DEL CORPUS	59
TABLA 4 PROGRAMAS PYTHON PROCESAMIENTO DEL CORPUS	60
TABLA 5 PROGRAMAS PYTHON PUESTA EN MARCHA SERVIDOR NMT	60

1.INTRODUCCIÓN

El propósito de este trabajo es entrenar dos motores de traducción automática (TA), inglés-español, uno estadístico y otro neuronal, especializados en el campo de la aviación a través de las herramientas desarrolladas por el proyecto MTUOC¹ y comparar los resultados de la traducción en bruto entre ellos para determinar cuál se ajusta mejor a las necesidades. Una segunda evaluación se hará comparando los resultados de la traducción de los motores creados, con los resultados propuestos por dos motores de traducción generales de uso libre muy eficaces y populares: *Google Translate* y *Yandex Translate*.

1.1 CONTEXTO Y MOTIVACIÓN

El sector de la aviación tiene el inglés como lengua vehicular, ya que se trata de un campo de alcance mundial estrictamente regulado por organizaciones internacionales como *Organización de Aviación Civil Internacional* (OACI), *European Aviation Safety Agency* (EASA), *Federal Aviation Administration* (FAA), entre otras. Es por ello por lo que es absolutamente necesario que exista una lengua en común en la que se realicen las comunicaciones, se dicten las normativas y se difunda la importancia y necesidad de unas políticas de seguridad.

Aunque el conocimiento de la lengua anglosajona es un requisito imprescindible para trabajar en este sector, nos planteamos lo mucho que facilitaría el trabajo contar con un traductor automático inglés-español fiable; es decir, especializado con el vocabulario del sector, para las constantes publicaciones sobre nueva normativa, regulaciones, investigaciones de accidentes/incidentes aéreos, etc., ya que algunos de los principales motores de traducción automática presentan ciertas carencias en la terminología especializada.

La motivación principal que ha originado esta investigación ha sido su carácter práctico. Se ha buscado aplicar los conocimientos adquiridos a lo largo de los dos años de duración del máster en una investigación que tenga una aplicación práctica como es la compilación de corpus lingüísticos y el entrenamiento y utilización de motores de traducción automática especializada para un determinado campo, en este caso, la aviación.

¹ <https://github.com/aoliverg/MTUOC> [última consulta: 09/06/2021]

1.2 HIPÓTESIS Y OBJETIVOS

El presente trabajo persigue los siguientes objetivos:

- O1. Recopilar corpus bilingües especializados en el campo de la aviación.
- O2. Entrenar dos motores de traducción automática, uno estadístico y otro neuronal, a través de la herramienta MTUOC de la *Universitat Oberta de Catalunya*.
- O3. Determinar cuál ofrece mejores resultados.
- O4. Comparar los resultados de los dos motores entrenados con los resultados ofrecidos por traductores automáticos generales de libre acceso: *Google Translate* y *Yandex Translate*.

Partimos de la hipótesis de que conviene entrenar motores de traducción propios con corpus especializados para obtener mejores resultados que los que se consiguen, a día de hoy, a través de la utilización de motores de traducción generales ya existentes. Es decir, que ofrezcan en sus resultados el correcto uso de la terminología especializada.

Para evaluar los resultados se utilizó el sistema de evaluación BLEU, analizando las traducciones realizadas por los motores de traducción entrenados en este trabajo y comparándolas con la evaluación de las traducciones realizadas por los dos motores de traducción descritos en el marco metodológico: *Google Translate* y *Yandex Translate*.

1.3 METODOLOGÍA

La metodología utilizada en este trabajo se explica extensamente en el apartado sobre el marco metodológico. Se ha seguido la metodología *Design and creation* (Oates, 2005) que considera el desarrollo de una herramienta informática para solucionar un problema. En primer lugar, se creó un corpus paralelo inglés-español con textos del campo de la aviación. En segundo lugar, se entrenaron los motores de traducción automática, tanto neuronal como estadístico a través del proyecto MTUOC. Y, por último, se evaluaron los resultados y se compararon con sistemas de traducción automática de dominio general.

1.4 ESTRUCTURA DEL TRABAJO

El trabajo sigue una estructura orgánica que presenta una introducción con el contexto y la motivación del trabajo, así como las hipótesis y los objetivos que se ha buscado cumplir. El marco teórico presenta el estado de la cuestión y desarrolla, brevemente, aspectos teóricos

importantes para la consecución de este proyecto como la historia de la traducción automática, el uso de corpus paralelos, la traducción automática estadística, la traducción automática neuronal y la evaluación de la traducción automática. En la parte central del trabajo se explica cómo se desarrolló el proyecto, incluyendo explicaciones sobre la creación del corpus, el entrenamiento de los dos motores de traducción con MTUOC, la evaluación de los resultados de la traducción y la comparación con los sistemas de traducción de dominio general y los resultados de las mismas. En la última parte del trabajo se presentan las conclusiones del presente trabajo. Se incluyen también tres anexos; en el primero se recogen los programas en Python utilizados a lo largo del proyecto, en el segundo se incluye la relación de textos utilizada para entrenar el corpus paralelo especializado y, por último, en el tercer anexo, se incluyen el texto en inglés utilizado para la evaluación de los diferentes sistemas de traducción automática y las diferentes propuestas dadas en español: traducción humana, con *Google Translate*, con *Yandex Translate* y con los dos sistemas entrenados en el presente trabajo a través del proyecto MTUOC.

2. MARCO TEÓRICO

En este capítulo planteamos el estado de la cuestión de la traducción automática especializada en el campo de la aviación y presentamos trabajos relacionados con la creación de motores de traducción automática especializados. Incluye también un breve resumen sobre la historia de la traducción automática, el funcionamiento de la traducción automática estadística y el funcionamiento de la traducción automática neuronal. El capítulo finaliza con una aproximación a la evaluación de los sistemas de traducción automática.

Es necesario mencionar que existen muchos tipos de herramientas de traducción, también denominadas herramientas de TAO (CAT, por sus siglas en inglés - *Computer Assisted Translation*) que ofrecen asistencia al traductor para que desenvuelva sus tareas de una manera más rápida y cómoda a través del uso de memorias de traducción y bases de datos terminológicas, entre otras funciones. “Un sistema de TAO integra varias funcionalidades representadas por recursos que permiten, como mínimo, la gestión terminológica, el almacenamiento y reutilización de segmentos previamente traducidos (en memorias de traducción), y la alineación de textos” (Fernández-Rodríguez, 2010, p. 205). No debemos confundir estas herramientas con la TA, pues las herramientas de TAO no realizan la traducción

en sí, sino que ofrecen ayuda al traductor; mientras que la TA sí ofrece una traducción directamente. En el marco del presente trabajo nos centramos únicamente en la TA.

2.1 ESTADO DE LA CUESTIÓN

No se ha encontrado ninguna publicación relacionada con investigaciones sobre traducción automática en el sector de la aviación en la combinación de idiomas inglés-español. No obstante, sí se han identificado empresas privadas que ofrecen sus servicios con bases terminológicas especializadas en aviación, traducciones automáticas y posesión por parte de traductores especializados.

La *Assam University Silchar* (India) ha llevado a cabo investigaciones sobre Procesamiento del Lenguaje Natural (PLN) en el campo de la aviación, estudiando el uso de diferentes herramientas y de traductores automáticos neuronales con la combinación bengalí-inglés. Saptarshi y Purkhyastha (2020) han desarrollado una herramienta de PLN para decodificar fraseología de ATC (*Air Traffic Control*) y la creación de un corpus específico con *OOV words* (*out of vocabulary words*) para la traducción automática dentro de este campo. Estas investigaciones han quedado reflejadas en una serie de artículos publicados durante el año 2020 por estos autores en diferentes revistas científicas y conferencias.

Uno de los artículos (Saptarshi & Purkhyastha, 2020d) explica la implementación de un sistema de traducción automática neuronal en la herramienta *OpenNMT* con corpus del repositorio *TDIL*, añadiendo la implementación de herramientas de pre-procesamiento y post-procesamiento para obtener un buen resultado. En otro de ellos (Saptarshi & Purkhyastha, 2020b) se describe el diseño y la implementación del campo de la aviación en la traducción automática inglés-bengalí a través de corpus paralelos, entrenando un sistema de traducción automática y creando una herramienta de posesión que incluye fraseología especializada del campo de la aviación. En su participación en la conferencia *Conference on Smart Electronics and Communications* Saptarshi y Purkhyastha, (2020a) proponen la creación de una herramienta de PLN especializada en fraseología de Torre de Control Aéreo (ATC por sus siglas en inglés) en el par de lenguas inglés-bengalí. Por último, Saptarshi y Purkhyastha (2020c) publicaron un artículo en el que explican la importancia de controlar la terminología especializada del campo de la aviación para la traducción automática del inglés a otros idiomas (centrándose especialmente en el bengalí). Como vemos, estos artículos giran en torno al mismo tema que proponemos en esta investigación, pero con un par de lenguas diferentes. Ello parece apuntar a que se trata de un tema que está empezando a generar interés en el seno de la

comunidad científica, si bien todavía no existen publicaciones al respecto sobre la combinación de lenguas inglés-español.

En lo que se refiere a corpus lingüísticos bilingües especializados en aviación, únicamente se ha encontrado un corpus monolingüe sobre comunicación en Torre de Control Aéreo (ATC) llamado *Air Traffic Control Complete*, con licencia de *Linguistic Data Consortium* y creado por John J. Godfrey².

En cuanto a investigaciones realizadas sobre el entrenamiento de motores de traducción automática especializada, merece especial atención el trabajo realizado por Molina Baños (2019). El objetivo del mismo consistió en entrenar un motor de traducción automática estadístico especializado en el campo farmacéutico. Molina Baños defendió la importancia de la TA en la traducción, insistiendo para ello en el estudio sobre cómo perciben los traductores este tipo de traducción y las ventajas en su uso. La autora plantea como hipótesis si realmente merece la pena invertir tiempo en la creación de corpus especializados para construir motores de traducción específicos o si, por el contrario, los motores de traducción genéricos gratuitos ofrecen ya un buen resultado. Los resultados positivos apuntados por la autora parecen indicar que el entrenamiento de motores con textos especializados sí puede dar lugar a mejoras en la calidad de los textos en sectores muy específicos. De ahí nuestra hipótesis de partida, que buscaremos confirmar siguiendo la metodología descrita en el capítulo sobre el marco metodológico.

2.2 LA TRADUCCIÓN AUTOMÁTICA: UNA BREVE HISTORIA

La Traducción Automática (TA) es un subcampo de la lingüística computacional dedicado a la utilización de software para la traducción de un texto de una lengua a otra. En este apartado repasaremos brevemente la historia de este campo.

La prototraducción automática se remonta al siglo XVII, cuando Descartes y Leibniz “formularon teorías acerca de la elaboración de diccionarios basados en códigos numéricos universales” (Hutchins, 1995). Inspirado en la creación de una lengua universal que no tuviese ambigüedades y basada en principios lógicos y símbolos icónicos, John Wilkins publicó en 1668 *Essay towards a Real Character and Philosophical Language*.

² <https://catalog.ldc.upenn.edu/LDC94S14A> [Última consulta: 15/05/2021]

Durante siglos quedó dormida la idea de la traducción automática, aunque no la del lenguaje universal. En 1933 surgieron dos patentes simultáneamente en Francia y en Rusia. Ribas (2016) describe el instrumento desarrollado por el francoarmenio Artsouni, que consistía en el diseño de “un dispositivo de almacenamiento en banda de papel de una especie de diccionario multilingüe” (Ribas, 2016, p.4) en el que podía encontrarse el equivalente de una palabra en otro idioma. Smirnov-Troyanskii desarrolló en Rusia una teoría que concebía la TA en tres fases: un editor conocedor de la lengua de origen analizaría la frase de entrada simplificándola a las formas más básicas; a continuación, una máquina transformaría las secuencias simplificadas en secuencias equivalentes en lengua meta y, por último, un editor conocedor de la lengua meta, las reformularía de una manera natural en esa lengua. Troyanskii afirmaba que, si bien el solo contemplaba la utilización de la máquina en la segunda fase, todas las fases podrían llegar a automatizarse (Hutchins, 1995). Las ideas de ambos autores no fueron muy conocidas; de hecho, las de Troyanskii no llegaron a atravesar la frontera rusa.

Las máquinas usadas para descifrar mensajes que se utilizaban en ambos bandos durante la Segunda Guerra Mundial, la más famosa de ellas la alemana Enigma, inspiraron la traducción automática porque se pensó en la traducción como un código que había que descifrar. Warren Weaver, el denominado padre de la traducción automática estadística, lo explica así:

One naturally wonders if the problem of translation could conceivably be treated as a problem in cryptography. When I look at an article in Russian, I say: ‘This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode (Locke & Booth, 1955).

Weaver, en 1949, propuso el desarrollo de la TA a través de varios métodos, entre ellos “el uso de técnicas criptográficas del período de guerra, los análisis estadísticos, la teoría de la información de Shannon y la exploración de la lógica subyacente y las características universales del lenguaje” (Hutchins, 1995, p. 25). Los ordenadores de los años 50 no permitían desarrollar una TA real, pero pronto se desarrollarían suficientemente como para llevarla a cabo.

Abaitua (2002) describe cómo en 1968, Peter Toma desarrolló *Systran* un sistema de traducción automática diseñado inicialmente para traducir el par de idiomas ruso-inglés para uso de las Fuerzas Aéreas de los Estados Unidos, que a día de hoy sigue utilizándose como traductor en línea multilingüe y para el público general. Para evaluar la evolución de la TA se constituyó en 1964 el comité ALPAC (*Automatic Language Processing Advisory Committee*). Tras dos años

de investigación publicaron un informe que concluía que la TA era demasiado lenta, poco precisa y el doble de cara que la traducción humana, lo que desalentó las investigaciones y los inversores en el campo. Pero fuera de los EE.UU. se siguió investigando, especialmente en aquellas regiones multilingües como son Canadá y Europa, con su recién estrenada Comunidad Económica Europea (CEE).

Marín et al. (s. f.), en su introducción histórica a la TA, explican que en 1976 se desarrolló *Météo*, el primer sistema comercial de TA desarrollado por la Universidad de Montreal que traducía partes meteorológicas. La CEE (Comunidad Económica Europea) decidió en 1976 implantar el sistema *Systran* para sus traducciones. A finales de la década, pusieron en marcha el proyecto de TA *Eurotra*, que funcionó desde 1978 hasta 1992. En 1980 un grupo de investigadores de IBM retomaron la idea de Weaver en su proyecto *Candide*. Este proyecto utilizaba el corpus *Canadian Parliamentary Hansard* para entrenar el sistema de traducción automática inglés-francés. La aproximación que hacía ese sistema de TA consistía en descodificar una frase en inglés/francés, encontrando las palabras más probables en francés/inglés a través del modelo de traducción y el modelo de lenguaje. A medida que las ciencias de la computación iban avanzando, se mejoraba la capacidad de los ordenadores y con ello la potencia del cálculo, lo que hizo surgir un nuevo interés en la traducción automática.

A partir del inicio del siglo XXI, la TA vivió una revolución colosal; las principales razones fueron el desarrollo de sistemas de software de libre acceso y la mejora de los hardware (mejora de los ordenadores y su potencia de procesamiento). En 2005 se desarrolla *Apertium*³, una plataforma de traducción automática libre y de código abierto que ofrece un motor de TA independiente de la lengua, herramientas para manejar los datos necesarios y construir un motor de traducción para el par de lenguas que desee el usuario y datos lingüísticos para conjuntos de pares de lenguas.

En 2005 Hieu Hoang, alumno de P. Koehn desarrolló *Moses*⁴; un sistema de traducción automática estadística sucesor de *Pharaoh*, que permite entrenar automáticamente modelos de traducción para cualquier par de lenguas (Koehn, 2009, p. 11), y que, además, se distribuye con licencia libre. *Moses* es tan popular que llegó a ser utilizado en traductores comerciales como *Google Translate* (Oliver, 2018).

³ <https://apertium.org/> [última consulta: 28/05/2021]

⁴ <http://www.statmt.org/moses/index.php?n=Main.HomePage> [última consulta: 04/06/2021]

A partir de 2010, la mejora de los sistemas computacionales permite que pueda desarrollarse la traducción automática neuronal, sobre la que ya se habían desarrollado prototipos teóricos que no habían podido realizarse porque los ordenadores no presentaban suficiente capacidad de procesamiento. Este cambio se produce cuando se comienza a utilizar el GPU⁵ de los ordenadores como procesadores de operaciones computacionales y no como procesadores de gráficos (Oliver, 2018). Algunos de los primeros artículos publicados sobre este tipo de TA, fueron los de Kalchbrenner & Blunsom (2013), Sutskever et al. (2014) y Cho et al. (2014).

Actualmente, se tienen grandes expectativas en que la traducción automática neuronal mejore la calidad de la traducción automática, especialmente por el abaratamiento de costes que esto supondría en la industria de la traducción, como se desprende del artículo de Moorkens (2017). Por otro lado, si bien los sistemas de NMT ofrecen resultados bastante prometedores; pues obtienen mejores resultados en las métricas automáticas de evaluación de traducciones; parece que en las evaluaciones hechas por humanos los resultados no son tan claros (Castilho et al., 2017).

Para concluir este apartado sobre historia de la TA, explicaremos brevemente dos de los principales traductores automáticos gratuitos que existen actualmente en el mercado y que se utilizarán en este trabajo para comparar la calidad de las traducciones: *Google Translate* y *Yandex Translate*.

En 2006 se inició el traductor automático *Google Translate*⁶. En sus comienzos funcionaba con un sistema de traducción automática estadística (véase apartado 2.4 para una definición más amplia de este tipo de arquitectura) entrenada con corpus de las Naciones Unidas y del Parlamento Europeo, y para ofrecer un servicio multilingüe siempre utilizaba el inglés como lengua intermediaria; es decir, para traducir un texto de italiano a alemán, el programa realizaba la traducción italiano-inglés inglés-alemán, aunque el usuario lo percibía como una traducción directa. Lo novedoso de esta herramienta es que no solo traducía oraciones, sino que además traducía documentos y páginas web. De hecho, en los últimos años, se ha incluido una función que permite la traducción inmediata de un texto a través de la cámara del teléfono móvil haciendo aparecer las letras de la lengua meta encima de las letras en lengua origen como texto superpuesto. Esta función también sirve para traducir un párrafo de un libro o unas instrucciones

⁵ El GPU (*Graphic Processing Unit*) es un coprocesador dedicado al procesamiento de gráficos u otras operaciones computacionales que liberan la carga del trabajo del procesador central del ordenador (CPU).

⁶ <https://translate.google.es/?hl=es> [última consulta: 09/06/2021]

haciendo una simple fotografía y seleccionando el texto que queremos traducir. A partir de 2016 *Google* comenzó a usar un nuevo software basado en un motor de traducción automática neuronal (véase apartado 2.5 para más detalles acerca de este nuevo paradigma) denominado *Google Neural Machine Translation*, que, además, ya no utiliza el inglés como lengua intermedia, lo que contribuye a mejorar la calidad de las traducciones con el denominado *Zero-shot* (Schuster et al., s. f.). Este traductor ofrece, además de la traducción de textos, la opción de escribir con el teclado, escribir de manera táctil, reconocimiento de habla, reconocimiento de imágenes (tanto imágenes con texto como reconocimiento de caracteres instantáneo a través de la cámara) y la opción sin conexión en numerosos idiomas; en su página web⁷ se puede ver un claro gráfico en el que especifican cada una de las funciones y si están disponibles o no en una lista de más de 100 idiomas.

*Yandex Translate*⁸ pertenece a Yandex, una compañía tecnológica de origen ruso fundada en el año 2000, aunque no fue hasta 2011 cuando lanzaron un sistema de traducción automática propio⁹. Aunque inicialmente solo traducía inglés/ucraniano-ruso, actualmente traduce en 95 idiomas diferentes. Ofrece, además, la traducción de páginas web, documentos e imágenes. En 2017 introdujeron un modelo de NMT complementario al estadístico, por lo que actualmente tienen un sistema híbrido de TA.

La TA actualmente está totalmente extendida y se está invirtiendo tiempo, esfuerzo y dinero en mejorarla cada día. Algunos traductores llegan a percibirla como una auténtica amenaza, ya que cada vez funciona mejor y hay más empresas dedicadas a su desarrollo y uso. Castilho (2017) explica que debemos tener cuidado con el revuelo y las expectativas puestas en la NMT, pues puede repetirse la situación de 2006, cuando Moses lanzó su Toolkit de libre acceso y se vivió como una verdadera amenaza para los traductores, pues fue publicitado como un sistema que ofrecía una calidad casi equivalente a la de un traductor profesional; lo que generó una ola de desconfianza por parte de los traductores, que comenzaron a ver la TA como una amenaza a sus puestos de trabajo. La TA debería ser considerada como una herramienta más para el traductor y no como una amenaza a su trabajo.

Con todo, un estudio realizado por Torres-Hostenech, Cid-Leal y Presas como coordinadoras en 2016 para ProjeCTA (Torres Hostench et al., 2016) demuestra que el uso de la TA y la

⁷ <https://translate.google.com/intl/es/about/languages/> [última consulta 25/05/2021]

⁸ <https://translate.yandex.com/> [última consulta 10/06/2021]

⁹ <https://yandex.com/company/history/2011> [última consulta 25/05/2021]

posedición no está tan extendido como podría parecer en las agencias de traducción del mercado español. Existe una gran desconfianza hacia la TA, ya sea por la mala imagen que se supone da su uso hacia los clientes, o por el miedo a que el cliente exija un abaratamiento de las tarifas si se usa un sistema de TA. Otra de las razones que parece existir es la brecha tecnológica y económica que hace que las pequeñas empresas no puedan permitirse entrenar con materiales propios sus sistemas de TA. El estudio concluye afirmando que en España no está tan implementado el uso de la TA por la inversión que requiere, tanto económica como tecnológica y humana. Concluye también que se necesita un cambio de mentalidad que demuestre que el uso de estos sistemas puede resultar beneficioso. Debemos tener en cuenta, no obstante, que este estudio fue realizado hace cinco años, así que el panorama de la industria de la traducción en España puede haber evolucionado.

De lo que no cabe duda es de que la traducción automática está viviendo un crecimiento exponencial en los últimos años: la mejora de la traducción automática estadística, la aparición de la traducción automática neuronal y la aplicación del *deep learning* en este campo, sin dejar de mencionar la propuesta más reciente: la traducción automática neuronal no supervisada basada en cuerpos monolingües (Sennrich et al., 2016a).

2.4 TRADUCCIÓN AUTOMÁTICA ESTADÍSTICA

En este apartado vamos a explicar cómo funciona la traducción automática estadística, también conocida como SMT (*Statistical Machine Translation*) por sus siglas en inglés. Se podría definir este tipo de traducción como un conjunto de frases en la lengua de destino, cada una de las cuales está asociada a una lista de traducciones en la lengua de origen para esas frases y a las probabilidades correspondientes (Hearne & Way, 2011).

La traducción automática estadística contempla dos etapas diferenciadas dentro del proceso de traducción: entrenamiento y decodificación. La parte del entrenamiento se basa en extraer un modelo de traducción estadística de un corpus paralelo y un modelo de lengua estadístico de lengua de destino de un corpus monolingüe (Hearne & Way, 2011, p. 1). El modelo de traducción podría interpretarse como un diccionario bilingüe en el que cada posible traducción de una palabra o frase tiene una probabilidad asociada. El modelo de lengua se compone de una base de datos de secuencias de palabras en la lengua de destino, asociadas asimismo a una probabilidad. Se podría decir que la SMT está orientada a encontrar miles de hipotéticas

traducciones y determinar cuál de ellas es la más probable (Hearne & Way, 2011, p. 2). Esto se traduce en que el objetivo de la decodificación es encontrar la traducción con la mejor puntuación (Koehn, 2010, p. 9).

Como indican Jurafsky y Martin (Jurafsky & Martin, 2000), el objetivo de la traducción es la producción de un resultado que maximice alguna función de valor que represente la importancia de la fluidez y la fidelidad. Esto se puede explicar con la fórmula de la ilustración 1:

$$\text{Best - translation } \hat{T} = \operatorname{argmax}_T \text{fluency}(T) \text{faithfulness}(T, S)$$

Ilustración 1 Ecuación 1 fórmula SMT

Tenemos que entender la traducción como la corrección de una frase corrupta en la lengua de destino; es decir, si queremos traducir la frase inglesa *Mary studies Literature* al español, tenemos que pensar en la frase en inglés como una versión corrupta del español, y el objetivo de nuestra fórmula es descubrir la frase correcta en nuestro idioma.

Una de las fórmulas más utilizadas para encontrar la traducción con mayor probabilidad es la basada en el segundo teorema de Shannon o teorema de la codificación de canal y se denomina canal ruidoso o inferencia bayesiana. Esta fórmula (ver ilustración 2) se interpreta como la asunción de tener T traducciones candidatas entre las que se optará por la traducción de salida que tenga la T con la máxima puntuación (argmax_T). Los otros dos componentes son:

- P(T) que se refiere a la probabilidad de que T sea una oración válida en la lengua de destino o, en otras palabras, la cuantificación de la fluidez. Se denomina comúnmente *Modelo de Lengua*.
- P(S|T) que se refiere a la probabilidad de que una oración en la lengua de origen (S) y la traducción candidata (T) sean equivalentes en su significado; es decir, que el significado de S esté también en T. Se denomina comúnmente como *Modelo de Traducción*.

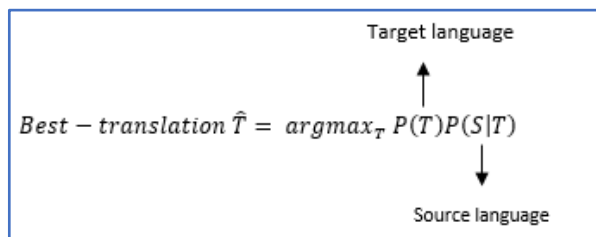


Ilustración 2 Ecuación 2 Fórmula canal ruidoso aplicada a SMT

Los tres elementos que tenemos que tener en cuenta a la hora de implementar esta fórmula son los siguientes (Jurafsky & Martin, 2000):

- Cuantificar la fluidez $P(T)$.
- Cuantificar la fidelidad $P(S | T)$.
- Crear un algoritmo para encontrar la frase que maximiza el producto de estas dos cosas.

El modelo de lengua o $P(T)$ proporciona un modelo del corpus de entrenamiento monolingüe y un método para calcular la probabilidad de una posible cadena que no se haya visto previamente. El modelo de lengua se extrae del corpus de entrenamiento y contempla las frecuencias relativas de cadenas y subcadenas que ocurren en ese corpus, así como la probabilidad asignada a una nueva cadena, basada en las ocurrencias de subcadenas en esa oración (Hearne & Way, 2011, 4).

Un modelo de lengua sencillo con el que explicar esto sería aquel que comprende solo las subcadenas correspondientes a las palabras que tenga un corpus, en el que a cada palabra se le asigne una probabilidad calculada dividiendo el número de apariciones de esa palabra por el número total de palabras que tenga ese corpus (Hearne & Way, 2011, p. 5). Para ello podemos utilizar el modelo de unigramas, pues las subcadenas están formadas por palabras individuales; es decir, se trata de un modelo de predicción de palabras que utiliza $N-1$ palabras para predecir la siguiente. Se basa en asignar a un elemento (fonema, sílaba, palabra...) un número y clasificarlo. Por ejemplo, si tenemos un corpus formado por la frase: “Raskolnikov salió en completa confusión”, y queremos aplicar un modelo de N -gramas basado en 2 palabras, es decir $N=2$, resultará esto:

- Raskolnikov salió
- salió en
- en completa
- completa confusión

En este ejemplo tenemos 4 N -gramas. Podemos saber el número de N -gramas en una frase con la fórmula de la ilustración 3 (X = número de palabras en una frase):

$$Ngrams_K = X - (N - 1)$$

Ilustración 3 Ecuación 3 Cálculo de N-gramas

Entonces si tuviésemos la frase “Raskolnikov salió” el modelo de lengua la aceptaría como correcta, puesto que la subcadena del ejemplo forma parte del corpus. Para no entrar en profundidad en las matemáticas detrás del sistema, podemos resumir que, después de hacer este paso en los corpus, sabremos qué cosas son legibles en la lengua de destino, aunque realmente no estén aún en esa lengua. Todo esto hay que entenderlo desde un corpus que tendría un gran conjunto de oraciones, ya que con una sola frase no tiene ningún sentido realizar este tipo de ejercicios.

Vamos a presentar ahora un ejemplo de un modelo lingüístico muy sencillo. Tomando como corpus la frase *He loves reading adventures books*, haremos un *Modelo de Lengua* basado en unigrama (el modelo de lenguaje más sencillo). Esto significa que lo analizaremos con un N-grama en el que cada tipo de palabra comprende palabras individuales, y averiguaremos la probabilidad de que se den los n-gramas dividiendo el número de apariciones de esa palabra por el número total de tokens de palabras en el corpus.

Corpus: *He loves reading adventures books on Sundays*

Modelo de unigrama:

$P(\text{He})=1/7$	$P(\text{loves})=1/7$	$P(\text{reading})=1/7$	$P(\text{adventure})=1/7$
$P(\text{books})=1/7$	$P(\text{on})=1/7$	$P(\text{Sundays})=1/7$	

Tabla 1 Ejemplo Modelo de Lengua SMT

Ahora podemos multiplicar la probabilidad del modelo por cada palabra con el fin de conocer la probabilidad de cualquier frase. Así, si queremos saber la probabilidad de que la frase *He loves Reading on Sundays* forme parte de nuestro modelo de lenguaje, solo tenemos que multiplicar la probabilidad de cada palabra según el corpus:

$$P(\text{He loves reading on Sundays}) = P(\text{He}) P(\text{loves}) P(\text{reading}) P(\text{on}) P(\text{Sundays})$$

$$P(\text{He loves reading on Sundays}) = 1/7 \times 1/7 \times 1/7 \times 1/7 \times 1/7$$

$$P(\text{He loves reading on Sundays}) = 1 / 16.807$$

$$P(\text{He loves reading on Sundays}) = 71'42\%$$

Tabla 2 Ejemplo cálculo $P(T)$

Este sistema no es muy robusto, ya que, como puede verse, a una longitud más corta de frase corresponde una mayor probabilidad. Además, aunque se escriba una frase agramatical, puede devolver una probabilidad muy alta. De ahí que se utilice el segundo elemento: el *Modelo de Traducción* o $P(S|T)$.

El *Modelo de Traducción* o $P(S|T)$ devuelve la probabilidad de que la traducción candidata (T) y la frase de origen (S) sean equivalentes en su significado. Es decir, la frase en T contiene todo el significado expresado en S. Este modelo ofrece un modelo del corpus de entrenamiento con las oraciones alineadas lengua origen-lengua destino y un método para calcular que S y T sean equivalentes.

Obtener estos datos no es una tarea fácil. Podemos aproximar la probabilidad de que una frase sea una buena traducción como el producto de la probabilidad de que cada palabra de la lengua de destino sea una traducción adecuada de alguna palabra de la lengua de origen (Jurafsky & Martin, 2000, p. 821). Para saber esto, necesitamos conocer la probabilidad de que cada palabra de la lengua de destino se corresponda con cada palabra de la lengua de origen. Es en este punto donde los corpus de textos bilingües se convierten en los protagonistas del proceso.

Los corpus textuales bilingües no especifican qué palabra equivale a cada una, por lo que la primera tarea que se debe realizar es tomar ambos corpus y hacer la ya mencionada alineación. En primer lugar, la alineación de frases; es decir, definir qué frase de la lengua de origen coincide con la frase de la lengua de destino. En segundo lugar, la alineación de palabra: es tal como la alineación de frases, pero palabra por palabra. Este segundo paso es mucho más complejo, y normalmente se realiza mediante el método EM (*Expectation-Maximization*). Como describen (Do & Batzoglou, 2008, p. 898):

The expectation maximization algorithm enables parameter estimation in probabilistic models with incomplete data. [...] The EM algorithm alternates between the steps of guessing a probability distribution over completions of missing data given the current model (known as the E-step) and then reestimating the model parameters using these completions (known as the M-step).

El *Modelo de Traducción* adjudica una probabilidad a cada par (lengua)origen-hipótesis multiplicando las probabilidades de que ese par de frases ocurra en el par acorde a ese modelo. Koehn (2010, 159) presenta un ejemplo (ver ilustración 4) de todas las opciones que se plantean (hipótesis) para la traducción de la oración alemana *Er geht ja nicht nach Hause* al inglés:

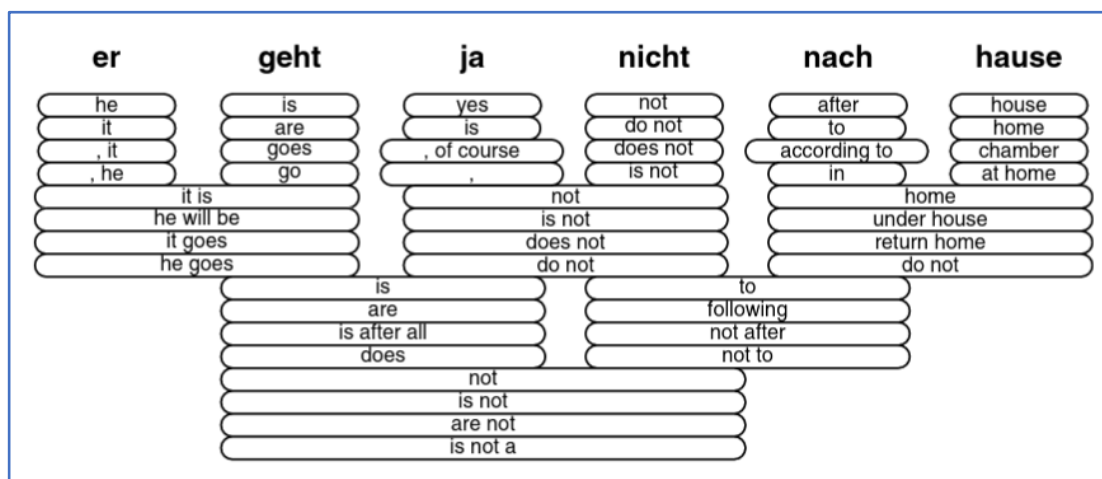


Ilustración 4 Ejemplo opciones de un Modelo de Traducción (Koehn, 2010, 160)

Todas las opciones que aparecen como traducción serían las hipótesis que se plantean para la lengua origen en lengua destino al buscar las subcadenas dentro de la oración; es decir, todas las opciones que plantea el *Modelo de Traducción*. En este punto se utilizaría el algoritmo que determine la mejor traducción, utilizando la probabilidad del *Modelo de Traducción* con el *Modelo de Lengua*. Y de ahí se obtiene la ecuación de la SMT presentada al inicio de este capítulo.

Actualmente existen herramientas para entrenar sistemas de traducción automática estadística personalizados a través de corpus ya creados o de creación propia. Algunos de estos softwares son TILDE¹⁰ o KantanMT¹¹, entre otros. Aunque la aparición de la traducción automática neuronal haya disminuido la presencia de traductores automáticos estadísticos, sigue investigándose en este campo y, además, son numerosos los estudios comparativos para determinar cuál de los sistemas ofrece traducciones de mejor calidad o requiere de un menor tiempo de posesión, como los de (Mutal et al., 2019), (Volkart et al., 2018) o (Wang et al., 2017).

2.5 TRADUCCIÓN AUTOMÁTICA NEURONAL

La Traducción Automática Neuronal, también conocida como NMT (*Neural Machine Translation* por sus siglas en inglés) surge con los avances tecnológicos de los últimos años en inteligencia artificial (IA) y redes neuronales. Las redes neuronales son un modelo

¹⁰ <https://tilde.com/products-and-services/machine-translation> [última consulta: 20/05/2021]

¹¹ <https://www.kantanmt.com/> [última consulta: 20/05/2021]

computacional inspirado en el modelo biológico de las neuronas humanas, que puede definirse con tres características: son “unidades de procesamiento que intercambian datos o información. Se utilizan para reconocer patrones [...] Tienen capacidad de aprender y mejorar su funcionamiento” (Matich, 2001, p. 5), lo que se conoce como aprendizaje automático. El modelo de red neuronal que se utiliza para la NMT es el “modelo dirigido a aplicación [...], su arquitectura está fuertemente ligada a las necesidades de las aplicaciones para la que es diseñada” (Matich, 2001, p. 5).

La NMT intenta construir y entrenar una única y gran red neuronal que lea una frase y produzca una traducción correcta. La mayoría de sistemas de NMT siguen un modelo de codificador-decodificador para cada una de las lenguas que traduce. (Bahdanau et al., 2016, p. 1) explican este proceso de manera simplificada como una red neuronal codificadora que lee y codifica una frase fuente en un vector de longitud fija. A continuación, un decodificador emite una traducción a partir del vector codificado. El sistema completo de codificador-decodificador, que consiste en el codificador y el decodificador para un par de idiomas, se entrena conjuntamente para maximizar la probabilidad de una traducción correcta dada una frase de origen.

Aunque actualmente existen varios paradigmas sobre la aproximación a la NMT, todos coinciden en que las palabras y las frases son representadas de forma numérica mediante vectores. Entender cómo funcionan estas redes neuronales requiere un considerable conocimiento computacional y matemático, así que, para poder entender el funcionamiento de este tipo de traducción utilizaremos la explicación que propone Casacuberta Nolla & Peris Abril (2017, p. 69) y el esquema (ver ilustración 5) con el que ejemplifican esta arquitectura para traducir “a black horse” a “un caballo negro”:

El codificador es una red neuronal (recurrente) que analiza de izquierda a derecha y de derecha a izquierda la frase origen para producir una representación vectorial de la misma. El decodificador genera la frase destino condicionado por la frase de origen. Esta última parte es otra red neuronal recurrente que, en un instante dado, genera una palabra destino tomando como entrada la palabra previamente generada, el estado de la red neuronal en el instante anterior y una representación de la frase de origen obtenida por un alienador (denominado modelo de atención) a partir de la información suministrada por el codificador. El modelo de

atención es a su vez otra red neuronal, en este caso un perceptrón multicapa¹², que permite alinear las palabras destino con las palabras de la frase de origen. Tanto el codificador como el decodificador son redes recurrentes sofisticadas, del tipo LSTM (*long short term memory*) o GRU (*gated recurrent unit*).

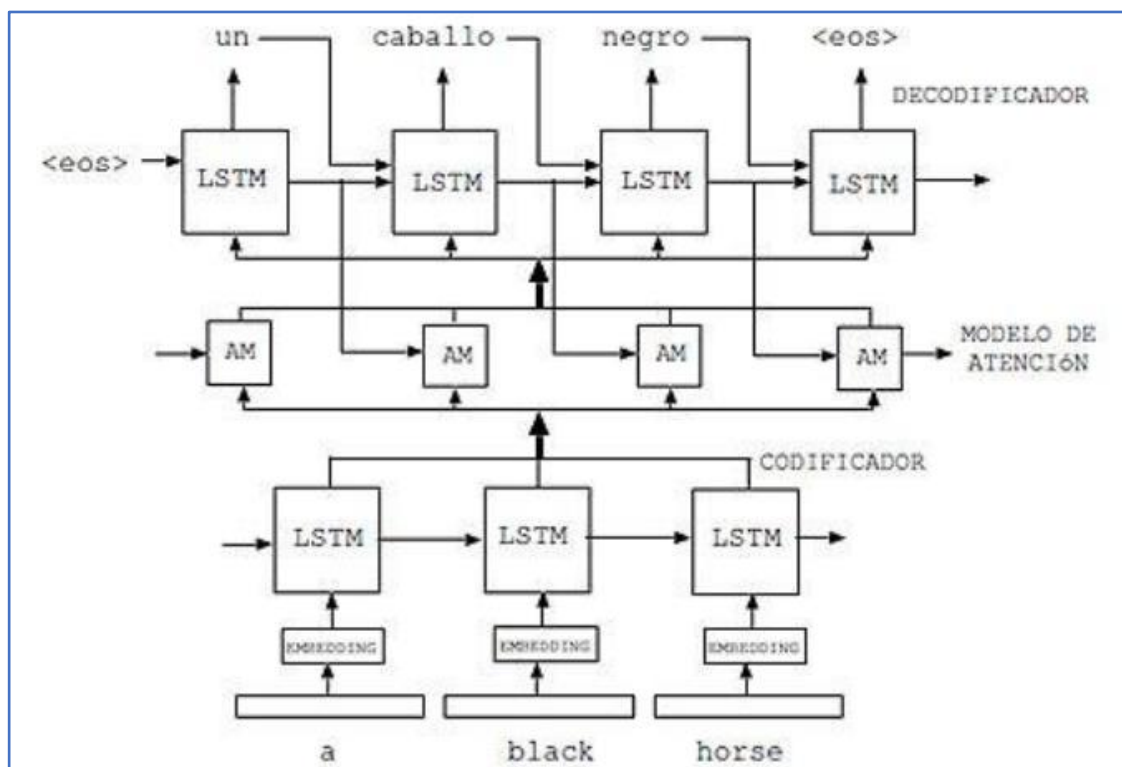


Ilustración 5 Ejemplo de arquitectura codificadora-decodificadora de NMT (Casacuberta Nolla & Peris Abril, 2017, p. 69)

Actualmente hay muchas investigaciones en curso que pretenden mejorar la NMT; algunas de ellas buscan mejorar los resultados de este tipo de traducción con corpus monolingües como los estudios realizados por Sennrich et al. (2016a), con información lingüística (Sennrich & Haddow, 2016) o mediante el aprendizaje conjunto de alineación y traducción (Bahdanau et al., 2016).

Existen softwares de entrenamiento para sistemas de traducción automática neuronal como *OpenNMT*¹³ (entorno de trabajo de código abierto para implantar sistemas de NMT), *MarianNMT*¹⁴ (entorno de trabajo de NMT gratuito escrito en C++), *Google Cloud AutoML*

¹² Capas de neuronas de forma que las salidas de una capa son las entradas de la capa siguiente (Casacuberta Nolla & Peris Abril, 2017, p. 69)

¹³ <https://opennmt.net/> [última consulta: 05/06/2021]

¹⁴ <https://marian-nmt.github.io/> [última consulta: 05/06/2021]

*Translation*¹⁵ (permite crear modelos de traducción personalizados para uso propio) o *TensorFlow*¹⁶ (biblioteca de código abierto en la que se puede aprender a desarrollar y entrenar modelos de aprendizaje automático).

2.6 EVALUACIÓN DE LA TRADUCCIÓN AUTOMÁTICA

Antes de hablar de la evaluación de la traducción automática, es preciso mencionar un proceso que, a día de hoy, resulta indispensable si se utiliza una herramienta de TA desde la perspectiva de un traductor profesional y no un mero usuario: la posesición. Esta fase es indispensable ya que resulta fundamental comprender los “límites relativos a la calidad de las traducciones automáticas” (Pérez Macías, 2017)

La posesición es el proceso mediante el cual se mejora una traducción automática editando, modificando o corrigiendo errores que aparezcan de manera que parezca realizada por un humano y no por una máquina. La posesición no es una mera revisión de la traducción, como explica Somers:

The first thing to note is that post-editing is not an activity restricted to MT output: human translations are generally revised. However, it should be clear that revising a human translation is quite different from post-editing MT output (and so it is useful to keep the two terms 'revising' and 'post-editing' to distinguish these activities). The sociological aspect of this difference has been noted by many commentators - that is, it is easier to wield the metaphorical red pen on the output of some faceless machine than on one's colleague's hard work, especially if one is slightly predisposed to disparage the machine's capabilities - as has the methodological aspect: the machine makes different sorts of mistakes from the human. (Somers, 1997, p. 199)

Para determinar si una etapa de posesición es necesaria, es posible evaluar la calidad de la traducción en bruto de la TA. La evaluación de calidad de una traducción resulta, como cualquier análisis cualitativo, algo compleja. De ahí que surjan sistemas para cuantificar el resultado de las traducciones y valorarlo, lo que lo convierte en un análisis cuantitativo. Como afirma Arnold et al. (Arnold et al., 1994): “La evaluación de los sistemas de traducción automática es una tarea compleja, no solo por la variedad de factores implicados sino porque medir el resultado de la traducción en sí mismo es una tarea difícil”.

¹⁵ <https://cloud.google.com/translate/automl/docs> [última consulta: 05/06/2021]

¹⁶ <https://www.tensorflow.org/> [última consulta: 05/06/2021]

Los traductores profesionales podrán evaluar cualitativamente una traducción a través de metodologías diferentes, como por ejemplo la tradautomaticidad. La tradautomaticidad es un concepto que se aplica a la evaluación de sistemas de traducción automática y se define como el texto de salida que devuelve el traductor automático y que, por su forma, el receptor reconocería que no lo ha hecho un traductor humano (Moré López & Climent Roca, 2006). Entre más ejemplos de tradautomaticidad se encuentren en una traducción, de peor calidad será la misma; pues se notará su ficcionalidad y su carácter automático; es decir, se percibirá que no ha sido realizada por un traductor, sino por una máquina. Sin embargo, en este trabajo nos hemos centrado en la evaluación automática de la traducción.

De acuerdo con Mayor (2009, p. 198), la evaluación de la traducción automática “puede tener dos objetivos: la evaluación absoluta, que da una medida total del comportamiento del sistema; y la evaluación relativa, que permite comparar diferentes sistemas de TA”. Por otro lado, Hutchins (1995) defiende como pruebas de la calidad de una traducción:

- (a) Su fidelidad o precisión, es decir, en qué medida el texto traducido contiene la “misma” información que en el original; (b) su inteligibilidad o facilidad con la que un lector puede entender la traducción; y (c) su estilo o en qué medida la traducción emplea el lenguaje apropiado a su contenido e intención. (Hutchins, 1995)

La evaluación realizada manualmente por revisores o traductores profesionales conlleva un gran gasto económico y, además, no puede reaprovecharse en más de un texto. De ahí que surja la evaluación automática para traducciones automáticas con métodos como BLEU, que ha sido el método elegido para este trabajo. En el marco metodológico se desarrolla el uso de esta herramienta.

3. MARCO METODOLÓGICO

En este capítulo se describe la metodología que se ha utilizado a lo largo de la investigación, incluyendo el proceso de creación del corpus especializado y el entrenamiento de los motores de traducción automática estadístico y neuronal con MTUOC. El marco metodológico finaliza con la evaluación de los resultados de la traducción con la puntuación BLEU, comparando los resultados de la traducción de los motores de traducción creados en este trabajo, con los ofrecidos por dos de los motores de traducción automática de acceso libre más conocidos del momento: *Google Translate* y *Yandex Translate*.

3.1 OBJETIVOS E HIPÓTESIS

La metodología que se sigue en este trabajo recibe el nombre de *Design and Creation*, ya que implica el análisis, diseño y desarrollo de un producto informático [...] que debe contribuir de alguna manera al conocimiento (Oates, 2005). Las investigaciones que siguen esta metodología deben cumplir con los cinco pasos siguientes: concienciación del problema sobre el que se va a investigar, recomendación inicial para la resolución del problema, desarrollo del diseño de la herramienta, evaluación y conclusión (Oates, 2005).

Partimos de la hipótesis de que conviene entrenar motores de traducción propios con corpus especializados para obtener mejores resultados que los que se consiguen, a día de hoy, a través de la utilización de motores de traducción generales ya existentes. Es decir, que ofrezcan en sus resultados el correcto uso de la terminología especializada.

3.2 DESARROLLO DE LA HERRAMIENTA

Etapa I: Creación del corpus

Un corpus lingüístico es una compilación de textos organizados y procesados para su utilización; como define Schmitt, “un corpus es una gran y estructurada colección de textos almacenada en un formato electrónico” (Schmitt, 2010, p. 89). Uno de los elementos principales de la traducción automática son los corpus paralelos (o multilingües en caso de estar trabajando con más de dos idiomas). Estos corpus son textos con sus equivalentes en el otro idioma, segmentados y alineados de forma que se relacionen las oraciones de ambos textos y se traten como equivalentes. Koehn (2005) explica en su investigación sobre EUROPARL que la creación de un corpus paralelo para TA implica, normalmente, cinco pasos:

1. Obtener los datos en bruto (por ejemplo, con la descarga de textos de webs).
2. Extraer y mapear los fragmentos de textos paralelos (alineación de los documentos).
3. Dividir el texto en frases o segmentarlo.
4. Preparar el corpus para los sistemas de TA (normalización y tokenización).
5. Asignar las frases entre ambas lenguas (alineación de frases).

Los corpus paralelos son el contenido que alimentan los motores de traducción automáticos. Existen numerosas páginas web que incluyen como recursos el acceso a corpus paralelos como Opus Corpus¹⁷, TREX¹⁸, TERMACOR¹⁹ o TausData²⁰.

Cuando se crea un corpus se debe tener en cuenta el propósito de la investigación y posterior uso del mismo. Para poder entrenar un motor de traducción automática, hace falta tener un corpus bilingüe de unos cinco millones de segmentos. Como resulta una ardua tarea la recolección de un corpus tan grande, lo que se ha hecho es crear uno especializado en el campo de la aviación con unos 50.000 segmentos y combinarlo con otros dos corpus de libre acceso: *Europarl*²¹ con 1.962.064 segmentos obtenido en *Opus Corpus*²² y *Taus Corona Crisis Corpora*²³, con 879.926 segmentos, descargado de TAUS²⁴.

Europarl es un corpus de libre acceso cuyos textos son las actas del Parlamento Europeo desde 1996 (Koehn, 2005). Es un corpus multilingüe que incluye versiones en 21 de las lenguas europeas: francés, italiano, español, portugués, inglés, alemán, búlgaro, lituano, griego... Este corpus se creó con la intención de extraer y procesar los textos para generar textos alineados para SMT. En este trabajo utilizamos la versión v7 lanzada en 2012. *Europarl* fue descargado en *Opus Corpus*, una herramienta en continuo crecimiento que ofrece corpus paralelos anotados lingüísticamente y herramientas para el procesamiento de datos (tanto monolingües como paralelos) y de búsqueda. El par de lenguas con un mayor número de datos paralelos es español-inglés. En la fecha de publicación del artículo de Tiedemann, contaban con 36 millones de frases paralelas que contienen aproximadamente unos 500 millones de tokens (Tiedemann, 2012).

Se decidió incluir un corpus específico sobre COVID-19 porque, como bien es sabido, este virus ha afectado a nivel mundial creando una situación sin precedentes que ha afectado especialmente al mundo del transporte aéreo. En el último año la mayoría de textos de este sector que salen publicados versan sobre este tema o lo mencionan: nuevas normativas

¹⁷ <https://opus.nlpl.eu/> [última consulta 25/05/2021]

¹⁸ <https://tr-ex.me/traduccion> [última consulta 25/05/2021]

¹⁹ <http://evrokorporus.anyterm.info/k2/index.php?jezik=angl> [última consulta 25/05/2021]

²⁰ <https://datamarketplace.taus.net/> [última consulta 25/05/2021]

²¹ <https://opus.nlpl.eu/Europarl-v8.php> [última fecha de consulta 31/05/2021]

²² <https://opus.nlpl.eu/> [última fecha de consulta 29/05/2021]

²³ <https://md.taus.net/corona> [última fecha de consulta 20/05/2021]

²⁴ <https://www.taus.net/> [última fecha de consulta 31/05/2021]

internacionales, regulación de movimiento, procedimientos específicos que realizar en las aeronaves, etc.

Ahora nos centraremos en explicar cómo ha sido el proceso de creación de la parte especializada del corpus compilado para esta investigación: un corpus bilingüe español-inglés con textos procedentes del campo de la aviación. El criterio principal para la selección de textos ha sido que estuviesen relacionados con la aviación y que hubiesen sido traducidos de una manera fiable. Para ello, las dos fuentes principales de las que se ha obtenido el material textual han sido las bibliotecas de *European Aviation Safety Agency* (EASA), de la *Organización de Aviación Civil Internacional* (OACI) y de *International Air Transport Association* (IATA). Las dos fuentes cumplen con ambas premisas, al estar relacionados con la aviación y ser fiables por almacenar contenido de organismos oficiales. Se creó un documento Excel de control con todos los textos clasificados según el corpus al que perteneciesen, indicando también la web de la que fueron extraídos, el idioma y el título de los documentos. Este documento se puede consultar en el Anexo II del presente trabajo.

Para elegir los textos se visitaron las páginas de EASA²⁵, OACI²⁶ e IATA²⁷, prestando especial atención a aquellos documentos disponibles en español y en inglés. La mayoría de ellos se descargaron en formato .pdf y se guardaron con una codificación que incluyese el nombre de la entidad (EASA, OACI o IATA), el número de texto para que coincidiese con su homólogo en el otro idioma y un sufijo -en/-es que indicase la lengua. Se creó también un cuarto corpus paralelo con los contenidos del manual de operaciones de una aerolínea.

El manual de operaciones de una compañía es un compendio de reglamentaciones que rigen una aerolínea y las instrucciones para que todo el personal de una compañía aérea cumpla con las regulaciones internacionales y sepan actuar tanto en operación normal como en emergencia. El manual de operaciones está regulado por la normativa publicada por los organismos que le la apliquen; es decir, una aerolínea que vuele en España está regida por las normas de Agencia Estatal de Seguridad Aérea (AESA) a nivel nacional y por EASA, OACI e IATA a nivel internacional. Podemos decir que son textos que se refieren a normativa pero que, además, incluyen textos descriptivos e instructivos, lo que enriquece el corpus. Un manual de operaciones tiene siempre la misma estructura a nivel europeo; está dividido en cuatro partes

²⁵ <https://www.easa.europa.eu/regulations> [última consulta: 09/06/2021]

²⁶ <https://www.easa.europa.eu/regulations> [última consulta: 09/06/2021]

²⁷ <https://www.iata.org/en/publications/annual-review/> [última consulta: 09/06/2021]

que se explican a continuación para que, el lector no conocedor de este campo, pueda hacerse una idea del tipo de textos que se han incluido al añadirlo al corpus:

- Parte A: Información general y descriptiva sobre la aerolínea, su organigrama y la estructura de los diferentes departamentos. Incluye cómo se componen las tripulaciones de vuelo, el sistema de seguridad operacional, precauciones de salud, limitaciones de tiempos de vuelo, descripción de los vuelos, sus fases y sus procedimientos, mercancías peligrosas, seguridad, reporte y notificación de incidentes y accidentes, las reglas del aire y un listado con abreviaturas usadas.
- Parte B: Contenido relacionado con aspectos operativos de la aeronave (incluyendo mantenimiento, procedimientos normales, anormales y de emergencia).
- Parte C: Información de las rutas operadas por la aerolínea y categorización de aeródromos.
- Parte D: Todos los aspectos relacionados con el entrenamiento de las tripulaciones y el resto de personal de la aerolínea (cursos que realizar, periodicidad, instructores acreditados, etc.).

El manual de operaciones de una aerolínea es material altamente confidencial, pues incluye elementos de seguridad que solo puede conocer el personal de la compañía. Al tratarse de un material tan delicado, no se mencionará la aerolínea a la que pertenecen los textos ni se dará acceso libre al corpus.

El siguiente paso, una vez se recopilaron todos los textos, fue crear un entorno de trabajo con Visual Studio²⁸ (ver ilustración 6), en el que se incluyeron todos los archivos y programas con los que se trabajaría con el objetivo de que estuviesen localizados todos en el mismo directorio y así evitar problemas durante la programación. En el Anexo I pueden verse todos los programas Python utilizados para la realización del presente trabajo, así como un enlace en el que descargarlos.

Para poder procesar los textos fue necesario transformarlos en .txt, para ello se utilizó el programa pdf2txtTika.py. Para tratar los textos de una forma más sencilla y ordenada, se realizaron tres corpus diferentes, uno por cada una de las entidades. Una vez transformados los documentos en .txt, se dividieron en dos carpetas según el idioma: *Spanish* y *English*.

²⁸ https://code.visualstudio.com/?wt.mc_id=DX_841432 [última consulta: 25/05/2021]

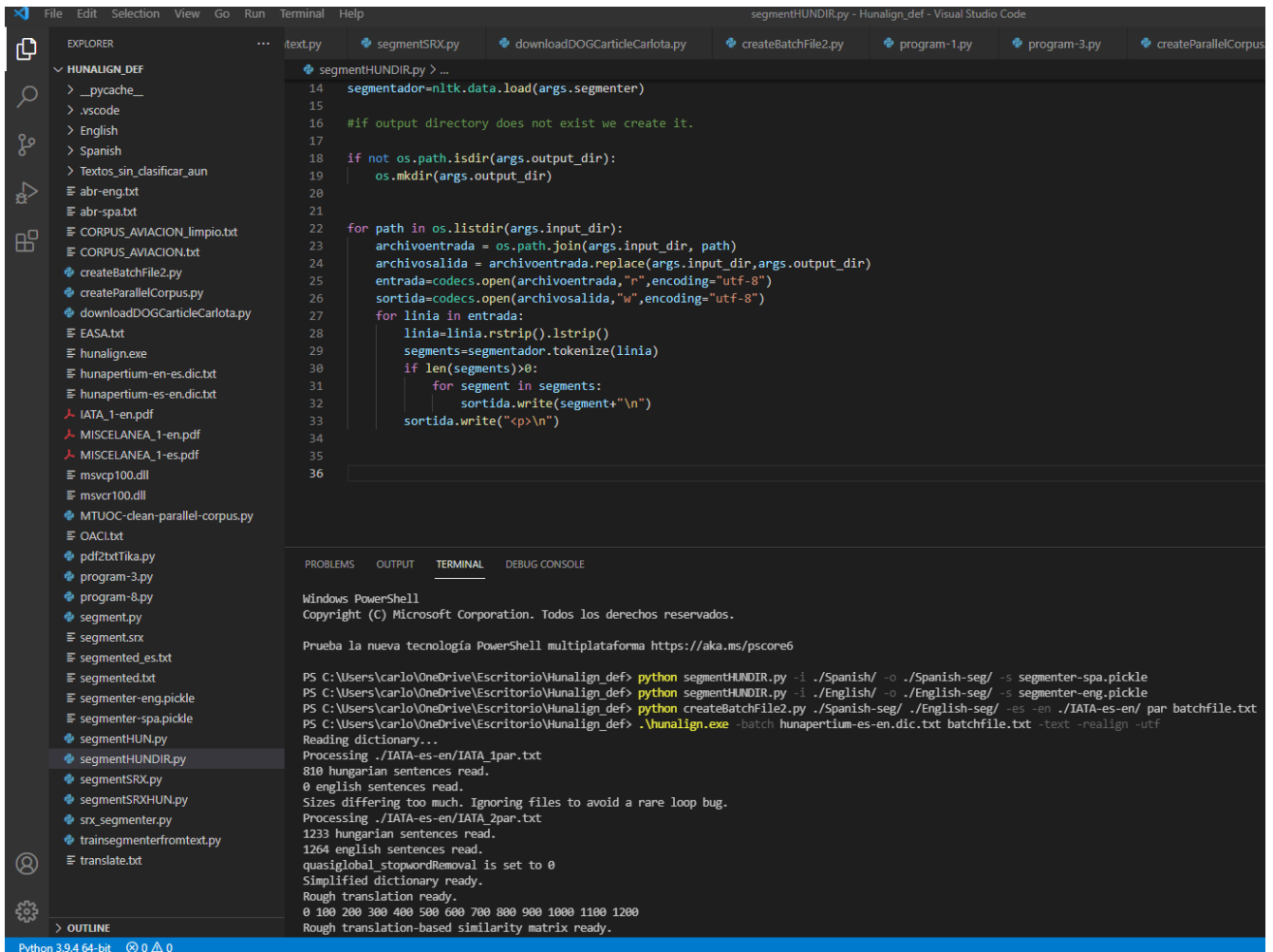


Ilustración 6 Entorno de trabajo en Visual Studio en una fase avanzada del proyecto

El siguiente paso fue segmentar y alinear los conjuntos de textos por carpetas utilizando el programa `segmentHUNDIR.py` (que necesita la descarga previa de los complementos `segmenter-spa.pickle` y `segmenter-eng.pickle`). Los comandos utilizados para ejecutar la segmentación fueron:

```
python segmentHUNDIR.py -i ./Spanish/ -o ./Spanish-seg/ -s segmenter-spa.pickle
python segmentHUNDIR.py -i ./English/ -o ./English-seg/ -s segmenter-eng.pickle
```

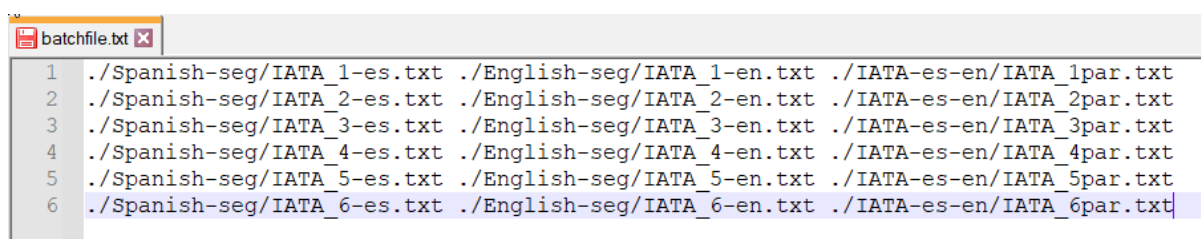
Con esto se obtuvieron dos carpetas nuevas llamadas *Spanish-seg* y *English-seg*, que incluyen los mismos textos que las originales, pero segmentados y con una marca `<p>` al finalizar cada párrafo para facilitar el trabajo que realiza el siguiente programa: *Hunalign*. Cuando ambas carpetas fueron creadas, llegó el turno de la alineación. Para ello utilizamos el programa

createBatchFile.py que lee ambos directorios y para cada archivo en español comprueba si existe en inglés (en nuestro caso). Para ejecutar el programa hay que usar el comando²⁹:

```
python createBatchFile2.py ./Spanish-seg/ ./English-seg/ -es -en ./EASA-es-en/ par
batchfile.txt
```

Este comando nos devolvió un nuevo documento que relaciona los textos con el mismo nombre en ambas carpetas generando una tercera relación con el sufijo que hemos añadido y las direcciones. Es decir, crea un archivo *batch* (ver ilustración 7), también denominado de procesamiento por lotes, así:

```
./Spanish-seg/EASA1-es.txt ./English-seg/EASA1-en.txt ./EASA-es-en/EASA1par.txt
```



```
batchfile.txt x
1 ./Spanish-seg/IATA_1-es.txt ./English-seg/IATA_1-en.txt ./IATA-es-en/IATA_1par.txt
2 ./Spanish-seg/IATA_2-es.txt ./English-seg/IATA_2-en.txt ./IATA-es-en/IATA_2par.txt
3 ./Spanish-seg/IATA_3-es.txt ./English-seg/IATA_3-en.txt ./IATA-es-en/IATA_3par.txt
4 ./Spanish-seg/IATA_4-es.txt ./English-seg/IATA_4-en.txt ./IATA-es-en/IATA_4par.txt
5 ./Spanish-seg/IATA_5-es.txt ./English-seg/IATA_5-en.txt ./IATA-es-en/IATA_5par.txt
6 ./Spanish-seg/IATA_6-es.txt ./English-seg/IATA_6-en.txt ./IATA-es-en/IATA_6par.txt
```

Ilustración 7 Ejemplo de archivo batch del corpus IATA

A continuación, ejecutamos el programa *Hunalign* para que, con el diccionario hunapertium-es-en.dic previamente descargado, devuelva los textos alineados en ambos idiomas. Para que lo realice en todos los textos se necesita activar el modo -batch y añadirle el batchfile.txt generado con el anterior comando. En este caso, el comando utilizado fue el siguiente:

```
.\hunalign.exe -batch hunapertium-es-en.dic.txt batchfile.txt -text -realign -utf
```

El siguiente paso fue utilizar el programa createCorpusParallel.py, introduciendo el directorio de entrada, la puntuación mínima del segmento que se quiere que se incluya y el nombre del

²⁹ En estos comandos se usará el nombre EASA porque son el ejemplo de la creación de ese corpus, pero bastará sustituir el nombre EASA en los comandos por aquel que queramos utilizar.

corpus de salida. Este programa selecciona los segmentos del texto que cumplen con la puntuación dada en el comando, que tengan información textual y donde más de un 75% del contenido del segmento no sean números. Este programa elimina, además, las etiquetas que pudiesen tener los textos, por ejemplo, la <p> que añadimos para poder hacer la alineación. El comando ejecutado fue el siguiente:

```
python createParallelCorpus.py ./EASA-es-en/ 0 EASA.txt
```

Por último, para hacer la limpieza final del corpus utilizamos un programa del proyecto MTUOC llamado MTUOC-clean-parallel-corpus.py. Con este programa se normalizan los apóstrofes y se elimina lo siguiente: etiquetas, las marcas del lenguaje que puedan existir del marcado html, segmentos iguales, los que tengan más de un 60% de números y aquellos que tengan menos de cinco caracteres. En la ilustración 8 vemos un ejemplo del funcionamiento del programa de limpieza de corpus. Para utilizarlo ejecutamos el comando:

```
python MTUOC-clean-parallel-corpus.py -i EASA.txt -o EASA_clean.txt -a --vSL es --vTL en --vNOTL es --vSetLanguages es,en --verbose
```

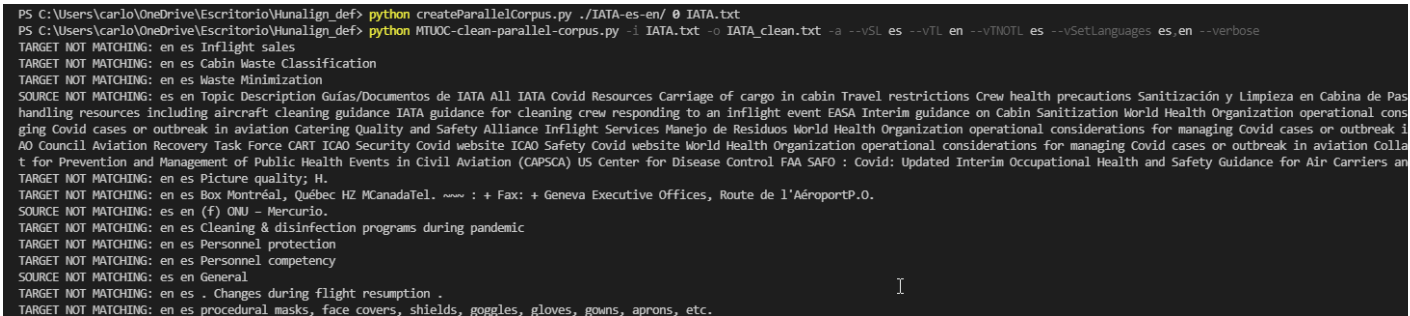


Ilustración 8 Ejemplo de funcionamiento de programa MTUOC-clean-parallel-corpus.py con el corpus IATA

Siguiendo este procedimiento se crearon cuatro corpus: IATA, EASA, OACI y Manual de Operaciones de compañía X. Para unificarlos en un único corpus se copiaron en un único archivo .txt con el nombre CORPUS_AVIACION que tiene, aproximadamente, unos 50.000 segmentos. Como los corpus con los que se combinó eran tabulados de forma que el inglés apareciese primero, se cambió el orden del texto con Visual Studio Code y las expresiones regulares:

```
Ctrl+f: (.+)\t(.+)\n
Reemplazar por: $2\t$1\n
```

Posteriormente este corpus (ver ilustración 9), pasó a denominarse corpusSPE-eng-spa.txt para combinarlo con el corpus general, este paso está explicado en la siguiente etapa del proceso de desarrollo de la herramienta.

Meal(s) and/or snack(s) quality;	Calidad de la(s) comida(s) y/o aperitivo(s);
Selection of movies;	Selección de películas;
Cabin temperature;	Temperatura de la cabina;
Punctuality of departure;	Puntualidad en la salida;
Cleanliness of toilets;	Limpieza de los aseos;
Selection of newspapers/magazines;	Selección de periódicos/revistas;
Sound quality;	Calidad del sonido;
Reliability of entertainment system;	Fiabilidad de los sistemas de entretenimiento;

Ilustración 9 Segmento del corpus especializado

Etapa II: Procesamiento de los corpus

MTUOC es un proyecto del departamento de Artes y Humanidades de la *Universitat Oberta de Catalunya* que ofrece una integración sencilla de los sistemas de traducción automática estadísticos y neuronales. La mayoría de softwares existentes para este propósito son también de acceso libre, pero requieren un elevado conocimiento especializado para su uso (habilidades tecnológicas e integración del sistema desarrollado en los espacios de trabajo) (Oliver, 2020b). MTUOC se diferencia por ofrecer unos *scripts* simplificados de Python y Bash que no requieren de un gran conocimiento tecnológico para su uso, por lo que el primero de los problemas quedaría solucionado. Para la integración de los sistemas, ofrece un servidor que puede imitar el comportamiento de diferentes tipos de servidores (Marian, Moses...), por lo que es adaptable a diferentes herramientas CAT (*Computer Assisted Translation*).

Todos los componentes de MTUOC son de libre acceso y se pueden conseguir en el perfil de GitHub [aoliverg/MTUOC](https://github.com/aoliverg/MTUOC)³⁰; Oliver (2020b) los describe de la siguiente manera:

- Módulos Python: Para la tokenización y otras fases del procesamiento del texto.
- *Scripts* escritos en Python y Bash con sus archivos de configuración:
 - *Scripts* para el pre-procesamiento de corpus
 - *Scripts* para el entrenamiento y la configuración de archivos
 - Evaluación de los *scripts*.

³⁰ <https://github.com/aoliverg/MTUOC> [última fecha de consulta 20/05/2021]

- Servidor MTUOC: es el componente que recibe un segmento del cliente o la herramienta CAT, lo procesa y lo envía al servidor de traducción. Para devolver al cliente la traducción.
- Cliente MTUOC: este componente puede utilizar diferentes formatos de traducción, envía los segmentos al servidor, recibe la traducción y genera el archivo traducido.
- Máquina Virtual MTUOC: Máquina virtual con Linux que incluye todos los componentes necesarios para hacer funcionar MTUOC en un ordenador Windows.
- Motores de traducción pre-entrenados que pueden ser usados libremente con MTUOC.

Una vez obtenidos y procesados los textos que forman parte del corpus, empezamos a trabajar con MTUOC. Para poder trabajar con este proyecto de investigación es necesario tener un entorno de trabajo Linux. Por ello, se descargó el software *Windows Subsystem for Linux*³¹ que da acceso a una consola Ubuntu dentro del sistema operativo de Windows 10. Tras aprender las principales instrucciones para utilizar en el terminal de Linux³² se siguieron los pasos propuestos por (Oliver, 2021b):³³

1. Clonación del repositorio de Github donde se encuentra alojado el proyecto MTUOC³⁴ dentro del terminal Ubuntu. Para tener un acceso correcto se clonó dentro de la carpeta C: de Windows, para ello se utilizó el comando `cd /mnt/c` y una vez dentro del directorio se copió el siguiente comando:

```
git clone https://github.com/aoliverg/MTUOC.git
```

2. Instalación de todos los componentes necesarios para que funcionen correctamente los programas en Python descritos en requirements.txt (jieba, requests, sentencepiece,

³¹ <https://docs.microsoft.com/en-us/windows/wsl/install-win10> [última consulta 18/05/2021]

³²

<https://xwiki.recursos.uoc.edu/wiki/mat21570ca/view/Anexo%201.%20Uso%20b%C3%A1sico%20de%20terminal%20de%20Linux/> [última consulta 18/05/2021]

³³

<http://xwiki.recursos.uoc.edu/wiki/matm4957ca/view/Corpus%20preparation%20and%20preprocessing%20with%20MTUOC/> [última consulta 18/05/2021]

³⁴ <https://github.com/aoliverg/MTUOC.git> [última consulta 18/05/2021]

subword_nmt, nltk, lxml, Flask, PyYAML, SimpleWebSocketServer y websocket_client) mediante el comando:

```
sudo pip install [nombre_del_componente]
```

3. Copia del directorio de trabajo MTUOC en la raíz del sistema mediante el comando:

```
sudo cp -r MTUOC4.0/* /
```

4. Combinación del corpus. Como se ha mencionado anteriormente, para poder entrenar un motor de traducción automática (tanto neuronal como estadístico) es necesario contar con corpus paralelos de gran tamaño. Por tanto, hemos tenido que combinar el corpus creado: CORPUS_AVIACION (ahora denominado corpusSPE-eng-spa.txt) con textos específicos del campo de la aviación, con dos corpus de gran tamaño, tal y como se explicó anteriormente. En primer lugar, unimos los dos corpus generales; para ello descargamos el corpus Europarl-v7 en formato moses. Como lo necesitábamos en un solo documento tabulado, utilizamos en Ubuntu el comando:

```
paste europarl-v7_en.en europarl-v7_es.es > textoL1-L2.txt
```

Unimos el documento resultante con el corpus sobre COVID-19, con lo que se generó un nuevo archivo denominado corpusGEN-eng-spa.txt que fue el utilizado para combinar con el corpus especializado.

Para la combinación del corpus se necesitó descargar previamente el módulo *kenlm* dentro de la carpeta en la que se usó el programa MTUOC-combine-corpus.py con los comandos:

```
git clone https://github.com/kpu/kenlm.git  
sudo pip3 install https://github.com/kpu/kenlm/archive/master.zip
```

El resultado de utilizar el programa MTUOC-combine-corpus.py (ver ilustración 10) fue la obtención de los siguientes documentos, que (Oliver, 2021b) explica de esta manera:

- eval-eng
- -spa: Conjunto de evaluación del corpus especializado.
- train-eng-spa: Un conjunto de entrenamiento compuesto por un subconjunto del corpus especializado y un número de segmentos más similares del corpus general. La similitud se calcula con un modelo lingüístico creado a partir del corpus específico.
- val-eng-spa: Conjunto válido del corpus especializado.

```

carlotardn@DESKTOP-1P7VTF2: /mnt/c/MTUOC/MTUOC-combine-corpus$ python3 MTUOC-combine-corpus.py
STEP 1. Tokenizing SL SPE corpus
STEP 2. Language Model Calculation
*****
/mnt/c/MTUOC4.0/lmplz -o 5 --skip_symbols --discount_fallback < /mnt/c/MTUOC/MTUOC-combine-corpus/slcorpustok.temp > /mnt/c/MTUOC/MTUOC-combine-corpus/lm.arpa.
*****
=== 1/5 Counting and sorting n-grams ===
Reading /mnt/c/MTUOC/MTUOC-combine-corpus/slcorpustok.temp
---5--10--15--20--25--30--35--40--45--50--55--60--65--70--75--80--85--90--95--100
*****
Unigram tokens 1250453 types 31244
=== 2/5 Calculating and sorting adjusted counts ===
Chain sizes: 1:374928 2:379220544 3:711038592 4:1137661696 5:1659090048
Statistics:
1 31244 D1=0.650476 D2=0.951131 D3+=1.32396
2 287454 D1=0.73655 D2=1.13003 D3+=1.44622
3 629138 D1=0.836695 D2=1.25906 D3+=1.44982
4 819369 D1=0.903464 D2=1.41475 D3+=1.52746
5 882077 D1=0.771829 D2=1.66472 D3+=1.17684
Memory estimate for binary LM:
type MB
probing 55 assuming -p 1.5
probing 65 assuming -r models -p 1.5
trie 25 without quantization
trie 13 assuming -q 8 -b 8 quantization
trie 23 assuming -a 22 array pointer compression
trie 11 assuming -a 22 -q 8 -b 8 array pointer compression and quantization
=== 3/5 Calculating and sorting initial probabilities ===
Chain sizes: 1:374928 2:4599264 3:12582760 4:19664856 5:24698156
---5--10--15--20--25--30--35--40--45--50--55--60--65--70--75--80--85--90--95--100
*****
=== 4/5 Calculating and writing order-interpolated probabilities ===
Chain sizes: 1:374928 2:4599264 3:12582760 4:19664856 5:24698156
---5--10--15--20--25--30--35--40--45--50--55--60--65--70--75--80--85--90--95--100
*****
=== 5/5 Writing ARPA model ===
Name:lmplz VmPeak:3960296 kB VmRSS:6800 kB RSSMax:719284 kB user:5.84661 sys:1.80218 CPU:7.6488 real:10.0272
STEP 3. Scores calculation
STEP 4. Tokenizing SL GEN corpus
Loading the LM will be faster if you build a binary file.
Reading /mnt/c/MTUOC/MTUOC-combine-corpus/lm.arpa.en
---5--10--15--20--25--30--35--40--45--50--55--60--65--70--75--80--85--90--95--100
*****
STEP 5. Creation of the selection file
STEP 6. Selection from LM
STEP 7. Splitting the corpus
traintemp.temp 47450
val-eng-spa.txt 1000
eval-eng-spa.txt 1000
STEP 8. Deleting temporal files

```

Ilustración 10 Resultado de la utilización del programa MTUOC-combine-corpus.py

5. Preparación del corpus. Este paso fue común para ambos tipos de traducción automática y consiste en reemplazar direcciones de e-mail y URLs en el corpus, entrenar un *truecaser*³⁵ para la lengua de origen y aplicarlo en los segmentos de origen, entrenar un *truecaser* para la lengua destino y aplicarlo en los segmentos de destino y, por último, limpiar el corpus eliminando segmentos que no cumplan con los números

³⁵ Un *truecaser* se utiliza para determinar correctamente el uso de las mayúsculas en una lengua.

mínimos-máximos determinados por el programa. Para ello se introdujeron previamente los archivos generados en el anterior programa (train-eng-spa.txt y val-eng-spa.txt en la carpeta MTUOC-corpus preparation) y se utilizó el programa MTUOC-prepare.corpus.py con el archivo de configuración config-prepare-corpus.yaml en Ubuntu a través del comando:

```
python3 MTUOC-prepare-corpus.py
```

Este comando devolvió el archivo train-pre-eng-spa.txt y los truecasers trainSL.temp Y trainTL.temp. Estos archivos fueron utilizados en los próximos pasos del proceso de entrenamiento que consistieron en procesar los corpus para cada uno de los dos sistemas de TA.

6. Pre-procesamiento del corpus para SMT. Esta fase de preparación es exclusiva para el pre-procesamiento del corpus utilizado en el traductor automático estadístico. Para llevarlo a cabo se necesitaron los dos archivos creados en el paso anterior. Con este paso se obtuvo el corpus listo para entrenar un sistema de SMT con los archivos denominados train.smt.en train.smt.es y val.smt.en val.smt.es. El comando usado en esta fase fue:

```
python3 MTUOC-SMT-preprocess.py
```

7. Pre-procesamiento del corpus para NMT. Para poder utilizar los programas que procesaran el corpus para el entrenamiento de NMT es necesario instalar previamente los módulos sentencepiece³⁶ y eflomal³⁷. Una vez instalados los módulos se copiarán a la carpeta “MTUOC-NMT-SP-preprocess” los archivos generados anteriormente: train-pre-eng-spa.txt y val-pre-eng-spa.txt y se utilizó el comando:

```
python3 MTUOC-NMT-SP-preprocess.py
```

En MTUOC existen dos formas de procesar el corpus para NMT: con BPE y con SentencePiece. BPE (*Byte Pair Encoding* por sus siglas en inglés) se refiere a la

³⁶ <https://github.com/google/sentencepiece> [última consulta: 03/06/2021]

³⁷ <https://github.com/robertostling/eflomal> [última consulta: 03/06/2021]

codificación de pares de bytes; Sennrich et al. (2016) hicieron una adaptación para utilizarlo como algoritmo de comprensión en la tarea de segmentación de palabras, esto permite la representación de un vocabulario abierto mediante un vocabulario de tamaño fijo de secuencias de caracteres con una longitud variable (Sennrich et al., 2016, p. 1716).

SentencePiece, el método elegido para nuestro trabajo, es un tokenizador de subcadenas independiente del idioma diseñado para el procesamiento de texto basados en procesamiento de textos con tecnología neuronal (Kudo & Richardson, 2018, p. 1). El proceso que siguen los scripts de MTUOC basados en *SentencePiece* son:

- Aprender el modelo utilizando el corpus de entrenamiento (train-pre-eng-spa.txt generado en el preprocesamiento del corpus).
- Generar el vocabulario para aplicar el umbral de vocabulario al codificar el corpus con los modelos.
- Codificar el corpus de entrenamiento con el modelo.
- Codificar el corpus val con el modelo (para ello necesita el archivo val-pre-eng-spa.txt generado en el preprocesamiento del corpus).

Para que funcione perfectamente esta parte del proyecto MTUOC es necesario tener los módulos *SentencePiece*³⁸ y *Eflomal*³⁹ instalados, para que estos funcionen correctamente tendrán que tener instalados a su vez los componentes *Cython*⁴⁰ y *Numpy*⁴¹. Una vez instalado todo, ejecutamos el siguiente comando, que nos devolvió los componentes necesarios para entrenar en la siguiente etapa el motor de traducción neuronal:

```
python3 MTUOC-NMT-SP-preprocess.py
```

³⁸ <https://github.com/google/sentencepiece> [última consulta: 03/06/2021]

³⁹ <https://github.com/robertostling/eflomal> [última consulta: 03/06/2021]

⁴⁰ <https://cython.org/> [última consulta: 03/06/2021]

⁴¹ <https://numpy.org/> [última consulta: 03/06/2021]

Etapa III: Entrenamiento de los motores de traducción Entrenamiento del motor de traducción automática estadístico

El motor de SMT se entrenó a través de Moses⁴² con los programas preparados por el proyecto MTUOC. El funcionamiento de este software consiste en que, una vez entrenado el *Modelo de Traducción*, un algoritmo de búsqueda encuentra cuál es la traducción con mayor probabilidad entre todas las opciones. Los dos componentes principales de Moses son el conjunto de herramientas de entrenamiento y el decodificador. Las herramientas de entrenamiento incluyen “preprocesamiento del corpus, alineación a nivel de palabra para extraer la tabla de traducción de frases, creación del modelo de lengua (de la lengua de llegada) y herramientas para el *tuning*” (Oliver, 2020a). El decodificador busca la puntuación más alta en la lengua de llegada para una frase dada en la lengua de origen. Moses incluye unas herramientas adicionales (Philipp Koehn, 2009, p. 13):

- El servidor Moses que facilita una interfaz de decodificación xml-rpc.
- Una colección de scripts que permiten utilizar Moses para traducir páginas web.
- Y herramientas de análisis que permiten visualizar y analizar los resultados de Moses comparándolos con otra referencia.

Para poder entrenar el SMT a través de MTUOC, tuvimos que instalar Moses; para ello clonamos el repositorio creado específicamente para MTUOC por Oliver⁴³ y lo copiamos en la carpeta raíz con los siguientes comandos:

```
git clone https://github.com/aoliverg/Moses.git
cd /
sudo cp -r Moses /
```

Una vez instalado Moses, accedimos de nuevo a nuestra carpeta de MTUOC training-script-moses y copiamos los archivos generados en la etapa del procesamiento de corpus específico para SMT: train.smt.en train.smt.es val.smt.en val.smt.es. En la ilustración 11 se puede ver una de las fases de entrenamiento del motor de SMT. A continuación ejecutamos el comando necesario para que empezase a entrenar el motor:

```
cd mnt/c/MTUOC/training-script-moses
bash Moses-full-training.sh
```

⁴² <http://www.statmt.org/moses/index.php?n=Main.HomePage> [última consulta: 04/06/2021]

⁴³ <https://github.com/aoliverg/Moses> [última consulta: 04/06/2021]

```

carlotardr@DESKTOP-1P7VTF2:/mnt/c/MTUOC/training-scripts-moses$ bash Moses-full-training.sh
ENTRENAMIENTO DEL MODELO DE LENGUA
mkdir: cannot create directory 'lm': File exists
=== 1/5 Counting and sorting n-grams ===
Reading /mnt/c/MTUOC/training-scripts-moses/train.smt.es
---5---10---15---20---25---30---35---40---45---50---55---60---65---70---75---80---85---90---95---100
tcmalloc: large alloc 3180658688 bytes == 0x564a3a662000 @
*****
Unigram tokens 77115870 types 340792
=== 2/5 Calculating and sorting adjusted counts ===
Chain sizes: 1:4089504 2:378858144 3:710359104 4:1136574464 5:1657504512
tcmalloc: large alloc 1657511936 bytes == 0x564a11c62000 @
Statistics:
1 340792 D1=0.672157 D2=1.01965 D3+=1.30538
2 5065258 D1=0.731288 D2=1.06044 D3+=1.36641
3 19514395 D1=0.803236 D2=1.1181 D3+=1.3747
4 37854307 D1=0.868436 D2=1.18513 D3+=1.38013
5 50988705 D1=0.872477 D2=1.27504 D3+=1.40656
Memory estimate for binary LM:
type MB
probing 2312 assuming -p 1.5
probing 2671 assuming -r models -p 1.5
trie 1114 without quantization
trie 625 assuming -q 8 -b 8 quantization
trie 979 assuming -a 22 array pointer compression
trie 489 assuming -a 22 -q 8 -b 8 array pointer compression and quantization
=== 3/5 Calculating and sorting initial probabilities ===
tcmalloc: large alloc 1427685376 bytes == 0x564a11c62000 @
Chain sizes: 1:4089504 2:81044128 3:390287900 4:908503368 5:1427683740
---5---10---15---20---25---30---35---40---45---50---55---60---65---70---75---80---85---90---95---100
#####
=== 4/5 Calculating and writing order-interpolated probabilities ===
Chain sizes: 1:4089504 2:81044128 3:390287900 4:908503368 5:1427683740
---5---10---15---20---25---30---35---40---45---50---55---60---65---70---75---80---85---90---95---100
#####
=== 5/5 Writing ARPA model ===

Name:lmplz      VmPeak:5252744 kB      VmRSS:4373920 kB      RSSMax:4377176 kB      user:300.849      sys:237.817
BINARIZACIÓN DEL MODELO DE LENGUA
Reading lm.arpa.es
---5---10---15---20---25---30---35---40---45---50---55---60---65---70---75---80---85---90---95---100
*****
SUCCESS
ENTRENAMIENTO DEL MODELO DE TRADUCCIÓN
Using SCRIPTS_ROOTDIR: /Moses/scripts
Using multi-thread GIZA
using gzip
(1) preparing corpus @ Fri Jun 4 21:37:48 CEST 2021
Executing: mkdir -p /mnt/c/MTUOC/training-scripts-moses/working/train/corpus
(1.0) selecting factors @ Fri Jun 4 21:37:49 CEST 2021

```

Ilustración 11 Entrenamiento del motor de SMT

Tras terminar el entrenamiento del motor de traducción, llegó la hora de ponerlo en marcha y probar su funcionamiento. Para ello, fue necesario montar el servidor con Moses creando un directorio con el nombre del motor, en nuestro caso fue Traductor_estadistico_MTUOC. Dentro de ese directorio se introdujeron los siguientes archivos:

- Los archivos del directorio training-script-moses (creados en el paso anterior), exactamente los que se encuentran dentro de la carpeta binarised-model (moses.ini, lm.blm.es, reordering-table-bin.minlexr y phrase-table-bin.minphr).

- Los archivos del directorio MTUOC-server-SMT⁴⁴ (MTUOC-server-SMT.py, MTUOC-stop-server.py y config-server.yaml).
- Ejecutable de Moses⁴⁵.
- El archivo tc.en del directorio de procesamiento de corpus.
- Los tokenizadores del directorio MTUOC 4.0⁴⁶ del idioma que se desee, en este caso MTUOC_tokenizer_eng.py y MTUOC_tokenizer_spa.py.

Una vez estuvieron todos los elementos necesarios en el directorio creado Traductor_estadistico_MTUOC, pusimos en marcha el servidor Moses, para ello necesitamos, una vez más, del software WSL utilizado anteriormente. Accedimos desde Ubuntu al directorio e iniciamos el programa MTUOC-server-SMT.py para conectar el servidor, cuando comprobamos que funcionaba correctamente, iniciamos la interfaz gráfica del traductor MTUOC: *MTUOC Translator v.1.0*. Para utilizarla en Windows 10 bastará con descargar los archivos en un .zip desde la web⁴⁷ y acceder al ejecutable (MTUOC-Translator.exe). El siguiente paso consistió en conectar la interfaz gráfica del traductor con el servidor Moses, para ello necesitamos comprobar que en el archivo config-server.yaml está en la sección *type* el servidor que queremos usar, en nuestro caso fue ModernMT y conectar luego, en la interfaz gráfica la IP y el puerto que indica la terminal de WSL (ver ilustración 12). Una vez se produce la conexión simplemente queda utilizar el traductor.

El motor entrenado con el corpus especializado funciona como un servidor de traducción; esto es que esperará a recibir segmentos en una lengua origen y los devolverá en la de destino. Gracias a la puesta en marcha y configuración que se ofrecen en el proyecto MTUOC, se pueden elegir IP, puerto y servidor de destino y así poder utilizar el motor entrenado en una herramienta CAT como OmegaT o SDL Trados Studio. Oliver, (2021a) explica los protocolos que se aceptan en la configuración:

- MTUOC: un protocolo sencillo propio del proyecto MTUOC
- Moses: el mismo protocolo que utiliza el servidor Moses. Es compatible, por ejemplo, con OmegaT⁴⁸ y SDL Trados Studio⁴⁹ 2017 y 2019

⁴⁴ <https://github.com/aoliverg/MTUOC/tree/master/MTUOC-server-SMT> [última consulta: 14/06/2021]

⁴⁵ <https://github.com/aoliverg/Moses/blob/master/bin/moses> [última consulta: 14/06/2021]

⁴⁶ <https://github.com/aoliverg/MTUOC/tree/master/MTUOC4.0> [última consulta: 14/06/2021]

⁴⁷ <https://github.com/aoliverg/MTUOC-translator> [última consulta: 10/06/2021]

⁴⁸ <https://omegat.org/> [última consulta: 07/06/2021]

⁴⁹ <https://www.trados.com/es/> [última consulta: 07/06/2021]

- ModernMT: el mismo protocolo que utiliza el servidor ModernMT. Es compatible, por ejemplo, con las herramientas de Okapi⁵⁰ (Tikal y Rainbow)
- OpenNMT: el mismo protocolo que utiliza el servidor OpenNMT
- NMTWizard: el mismo protocolo que utiliza el servidor NMTWizard⁵¹

Entrenamiento del motor de traducción automática neuronal

Debido a que nuestro ordenador no tenía las capacidades necesarias para entrenar un motor de traducción automática neuronal, lo entrenó Antoni Oliver⁵², experto en traducción automática y director del Máster de Traducción y tecnologías de la UOC, con los archivos previamente preparados en el paso de procesamiento de corpus para traducción automática neuronal: `smodel.model`, `smodel.vocab`, `tc.en`, `vocab_file.en` y `vocab_file.es`. Para ver las instrucciones sobre cómo entrenar el motor de NMT se puede visitar el Wiki⁵³ del GitHub del proyecto MTUOC.

Una vez recibimos los archivos del motor de traducción ya entrenado, los copiamos en la carpeta MTUOC-server-SP. Los módulos y librerías necesarios para que el motor funcione correctamente están descritos en un archivo incluido llamado `requirements.txt`, así que comprobamos que todos los requisitos estuviesen instalados con el comando:

```
sudo pip3 install -r requirements.txt
```

La siguiente tarea consistió en configurar y poner en marcha el servidor MarianNMT en WSL. MarianNMT, como se explicó en el apartado sobre traducción automática neuronal del marco teórico, es “un marco de traducción automática neuronal eficiente y autónomo con un motor de diferenciación automático integrado basado en gráficos de cálculos dinámicos” (Junczys-Dowmunt et al., 2018, p. 1).

Para configurar y poner en marcha el motor de traducción, es necesario conectarlo a un servidor, para ello hay que modificar el parámetro *type* (en cualquier editor de textos) del archivo `config-server.yaml` escribiendo el que se quiera utilizar, en nuestro caso fue ModernMT. El motor de

⁵⁰ <https://okapiframework.org/> [última consulta: 07/06/2021]

⁵¹ <https://github.com/OpenNMT/nmt-wizard> [última consulta: 07/06/2021]

⁵² <https://www.uoc.edu/webs/aoliverg/ES/curriculum/index.html> [última consulta: 14/06/2021]

⁵³ <https://github.com/aoliverg/MTUOC/wiki/4.-Training-MT-systems-with-MTUOC> [última consulta: 10/06/2021]

traducción comenzará a funcionar con el comando de inicio y lo detendremos con las teclas CTRL+C y escribiendo el comando de fin.

Comando inicio: python3 MTUOC-server-SP.py
Comando fin: python3 MTUOC-stop-server.py

Para iniciar la traducción tuvimos que ejecutar el software *MTUOC Translator* (explicado en el paso anterior). Al iniciarlo configuramos la conexión, para ello escribimos el comando *ifconfig* en el programa previamente iniciado en WSL (una vez esté corriendo el *MTUOC-server-SP.py*), lo que nos devuelve la dirección IP y el puerto que debemos configurar en la interfaz gráfica del *MTUOC Translator* como se ve en la siguiente ilustración (12):

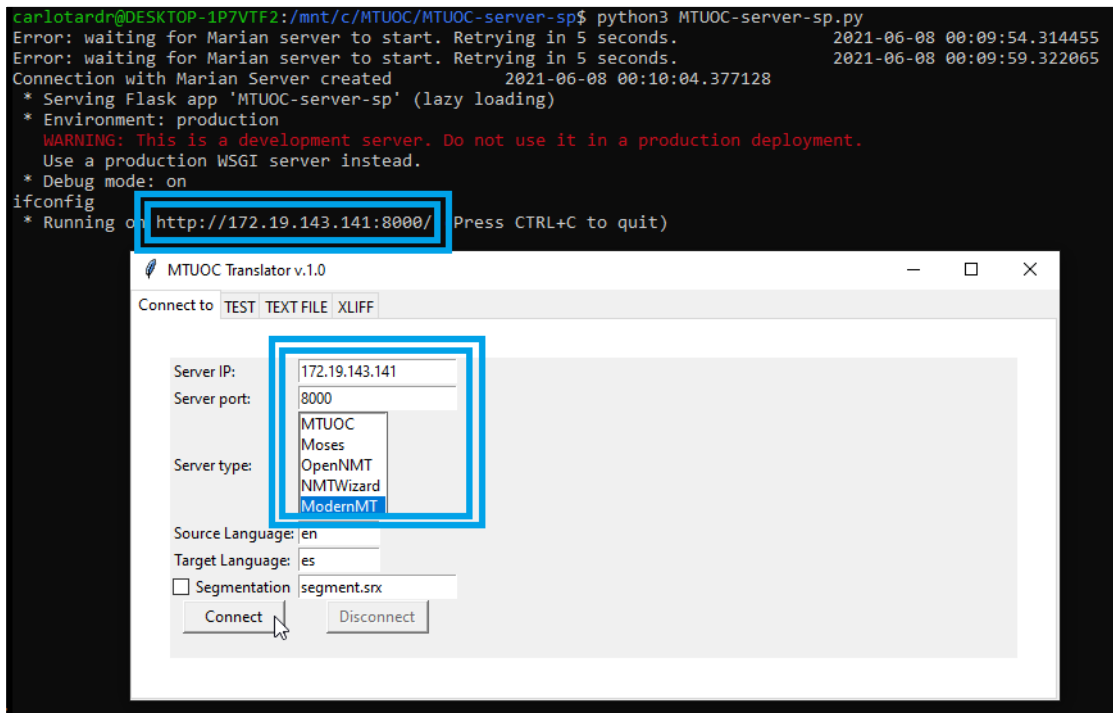


Ilustración 12 Configuración servidor MTUOC Translator

Una vez conectado el traductor al servidor, estará listo para su uso. *MTUOC Translator* admite fragmentos escritos directamente en la pestaña *TEST* (escribiremos el texto en el recuadro superior, cliquearemos en *Translate* y aparecerá la traducción en el recuadro inferior ver ilustración 13), la pestaña *Text File* permite traducir archivos de texto pero que necesitarán segmentarse previamente, para ello en la pestaña *Connect to* se marcará la opción *Segmentation* y se introducirá el archivo *segment.srx* que quiera utilizarse para este fin. Por último, la pestaña *XLIFF* permite traducir archivos de tipo *XLIFF*⁵⁴. Las dos últimas pestañas contemplan la opción de, al traducir, exportar una memoria de traducción en *TMX*⁵⁵.



Ilustración 13 Ejemplo traducción TEST con MTUOC Translator

Etapa IV: Evaluación y comparación con sistemas de traducción de dominio general

Como se mencionó en el marco teórico, la evaluación de la TA es una ardua tarea que puede realizarse de diversas maneras. En este trabajo se ha utilizado el sistema BLEU. Este es un método de evaluación de TA que surge como respuesta al coste de tiempo, esfuerzo y dinero que supone la evaluación humana. Para medir una traducción automática se dice que cuanto más parecida es a una traducción profesional humana, es de mejor calidad (Papineni et al., 2002). Y en eso se basa BLEU, requiriendo para ello de dos ingredientes principales (Papineni et al., 2002):

⁵⁴ XML Localization Interchange File Format (formato basado en XML creado con el propósito de estandarizar la localización).

⁵⁵ Translation Memory eXchange (archivo abierto SML utilizado para intercambiar memorias de traducción en herramientas CAT).

- Una métrica que mida numéricamente la “proximidad a la traducción”
- Un corpus de traducciones humanas de buena calidad como referencia

La tarea principal del sistema de evaluación BLEU es comparar los n-gramas de la frase candidata con los n-gramas de la referencia de la traducción y contar las coincidencias existentes. Mayor et al. explican así el funcionamiento del sistema de evaluación (Mayor, 2009):

BLEU calcula la media geométrica de la precisión de los n-gramas (n=1...4) multiplicada por la penalización de brevedad. La precisión de los n-gramas se calcula dividiendo el número de n-gramas de la traducción del sistema que aparecen en algunas de las traducciones de referencia entre el número de palabras de la traducción del sistema.

El proyecto MTUOC proporciona también una herramienta que permite hacer la evaluación automática de los sistemas entrenados: MTUOC-eval⁵⁶. Esta herramienta está escrita en Python y ofrece una versión en línea de comandos y una interfaz gráfica. Permite calcular métricas automáticas para BLEU, NIST (versión actualizada de BLEU), WER (Word Error Rate), Distancia de edición y TER (Translation Edit Rate). Pero en el presente trabajo, decidimos utilizar la herramienta de evaluación de la plataforma TILDE⁵⁷.

BLEU ha sido el sistema de evaluación elegido en este trabajo para la evaluación de ambos motores de traducción. La elección se ha basado en su popularidad, su aceptación generalizada y por lo sencillo que es utilizarlo, ya que desde plataformas como TILDE se permite la carga de: archivo de origen, traducción humana, resultado de traducción de la TA1 y resultado de traducción de la TA2. Como se ve en la ilustración 14. Una vez se han subido los archivos, se clikea en el botón *score* y devuelve el resultado.

Para poder utilizar BLEU y comparar la traducción automática de un documento, se necesita tener el documento traducido de manera correcta; es decir, por un traductor o al menos por una persona entendida en el tema. Para poder realizar la comparación se cogió el mismo texto y se tradujo a través de cinco métodos diferentes, cada uno de ellos titulado de una manera diferente para que resultase más sencillo su análisis en el apartado de resultados:

- Texto A: Traducción humana/profesional de aviación
- Texto B: Traducción con *Google Translate*

⁵⁶ <https://github.com/aoliverg/MTUOC-eval> [última consulta 10/06/2021]

⁵⁷ <https://www.letsmt.eu/Bleu.aspx> [última consulta 23/05/2021]

- Texto C: Traducción con *Yandex Translate*
- Texto D: Traducción con el motor de traducción neuronal creado en este trabajo
- Texto E: Traducción con el motor de traducción estadístico creado en este trabajo

Ilustración 14 Evaluador interactivo BLEU

En el siguiente capítulo del presente trabajo, se presentan los resultados de la evaluación de las traducciones realizada por la herramienta proporcionada por Tilde. El texto elegido para la evaluación ha sido un compendio de diferentes oraciones, que resultan de interés por su vocabulario especializado, de las *Easy Access Rules for Air Operations*⁵⁸. Se puede ver el texto utilizado íntegro en el Anexo III, en el que también se adjuntan las traducciones propuestas por cada uno de los motores de traducción. Partiendo de una base cuantitativa (resultados de la métrica de BLEU), analizamos algunos segmentos que resultaron de interés desde un punto de vista cualitativo, teniendo en cuenta la elección de los términos, la corrección y concordancia gramatical, etc.

Los dos motores de traducción automática elegidos para esta fase fueron *Google Translate*⁵⁹ y *Yandex Translate*⁶⁰ cabe mencionar que ambas traducciones fueron realizadas en la primera quincena del mes de junio de 2021. Estos traductores presentan la ventaja de ser gratuitos y de tener un acceso directo online. Se prestó también especial atención a cómo tradujo cada uno de

⁵⁸ <https://www.easa.europa.eu/document-library/easy-access-rules/easy-access-rules-air-operations> [última consulta: 10/06/2021]

⁵⁹ <https://translate.google.es/?hl=es> [última consulta: 25/05/2021]

⁶⁰ <https://translate.google.es/?hl=es> [última consulta: 25/05/2021]

los traductores automáticos la terminología especializada del campo, ya que conseguir unas buenas traducciones de esos términos era uno de los propósitos de entrenar motores de traducción especializados. Para ello se creó una tabla con terminología especializada del campo de la aviación y se comparó cómo la tradujo cada uno de los motores de TA.

Como se comentó en el marco teórico, *Google Translate* es un motor de traducción que, aunque en sus inicios fuese un motor de TA estadístico, a partir de 2016 comenzó a usar un software de traducción automática neuronal *Google Neural Machine Translation* (GNMT) basado en Redes Neuronales Recurrentes (RNN).

Yandex Translate ha sido el segundo motor de TA de dominio general elegido porque es uno de los pocos sistemas de traducción que utilicen aún la traducción estadística, dado que la mayoría de sistemas utilizan la NMT.

4. RESULTADOS

En este capítulo analizamos los resultados de las traducciones, la evaluación de los distintos sistemas de traducción aquí analizados (según lo explicado en el anterior capítulo) y mostraremos el ejemplo de parte de la terminología especializada y la resolución de los diferentes traductores. Se mostrarán también algunas ilustraciones en las que se comparan los resultados BLEU por segmentos, mostrando aquellos que resulten de mayor interés. Los resultados de la traducción pueden verse en el Anexo III, así como el texto original en inglés.

4.1 RESULTADOS DE LA TRADUCCIÓN AUTOMÁTICA ESTADÍSTICA

En este apartado comparamos las traducciones realizadas por *MTUOC-Translator* conectado con un motor de traducción automática estadístico entrenado con el corpus especializado de aviación y *Yandex Translate*. Los evaluamos con la herramienta ofrecida por Tilde, ya explicada en el marco metodológico, comparándolos entre sí y con una traducción realizada por una persona conocedora de la terminología aeronáutica. Los resultados que se muestran en color azul en las ilustraciones, corresponden a *MTUOC-Translator*; los resultados en color verde, corresponden a *Yandex Translate*. Los resultados generales son los siguientes (ilustración 15):

BLEU:	48.50	69.76
Precision x brevity:	48.50 x 100.00	69.76 x 100.00
Type	1-gram 2-gram 3-gram 4-gram	1-gram 2-gram 3-gram 4-gram
Individual	75.18 55.16 42.01 31.76	84.53 73.64 65.36 58.20
Cumulative	75.18 64.40 55.85 48.50	84.53 78.90 74.10 69.76

Ilustración 15 Resultados del análisis BLEU con Tilde (SMT)

Del resultado general, podríamos decir que cuantitativamente el resultado de *Yandex Translate* presenta una puntuación más alta en todos los parámetros comparados. Los resultados completos de la traducción de cada uno de los traductores se encuentran en el Anexo III. A continuación, se muestra un análisis de algunos segmentos que resultaron de especial interés por su forma de resolver la traducción:

- En el siguiente segmento (ver ilustración 16) vemos cómo el traductor entrenado en el presente trabajo no tradujo el término *Safety equipment*, esto puede ser debido, entre otras razones a que en los textos de aviación muchas veces se mantienen los términos en lengua inglesa, aunque el texto sea en español; en otro segmento del texto (ver ilustración 17) vemos que se repite la situación cuando no traduce *Normal procedures* y lo mantiene entrecomillado en inglés. Tanto *safety equipment* como *normal procedures* están registrados en el corpus, pues forman parte del vocabulario de el manual de operaciones de cualquier compañía. Vemos que nuestro motor también ha fallado en la forma de traducir *kit*, pues en una traducción al español se debe mantener la palabra de origen neerlandés aceptada por la RAE; mientras que *Yandex Translate* se ha acercado más a los términos usados por el traductor humano.

Sentence 6	BLEU	Length ratio	Text
Source	-	-	'Safety equipment' means equipment installed/carried to be used during day-to-day normal operations for the safe conduct of the flight and protection of occupants (e.g. seat belts, child restraint devices, safety card, safety demonstration kit).
Human	100.00	1.00	" Equipo de seguridad " es el equipo instalado/ transportado para ser utilizado durante las operaciones normales del día a día para la realización segura del vuelo y la protección de los ocupantes (por ejemplo , cinturones de seguridad , dispositivos de retención para niños , tarjeta de seguridad , kit de demostración de seguridad) .
Machine	54.51	1.03	' Safety equipment ' Significa equipos instalados / llevado a usarse durante el día a día las operaciones normales para la realización segura del vuelo y la protección de los ocupantes (por ejemplo , los cinturones de seguridad , los dispositivos de retención para niños , tarjeta de instrucciones de seguridad , botiquín de demostración de seguridad) .
Machine	68.29	0.93	" Equipo de seguridad " : equipo instalado o transportado para ser utilizado durante las operaciones normales cotidianas para la conducción segura del vuelo y la protección de los ocupantes (por ejemplo , cinturones de seguridad , dispositivos de retención infantil , tarjeta de seguridad , kit de demostración de seguridad) .

Ilustración 16 Ejemplo traducción incorrecta MTUOC-Translator SMT (BLEU)

- Otro ejemplo llamativo es el de la ilustración 17, en el que vemos que se repite la no traducción del término entrecomillado, como explicamos en el punto anterior. Pero que, sin embargo, sí traduce *pre-flight check* de la manera entrenada en el corpus; es decir, lo traduce como *comprobaciones pre-vuelo* y no como *comprobaciones previas al vuelo*. Aunque ambos términos son correctos, el más predominante en los textos especializados en aviación es *pre-vuelo*, probablemente por una cuestión de economía del lenguaje.

También en este fragmento podemos observar que *MTUOC-Translator* ha mantenido la voz inglesa *galley*; mientras que *Yandex Translate* lo ha traducido como *cocina*; en el sector de aviación se mantiene la voz inglesa y no se debe traducir este término, al menos en la variante de español de España. Así vemos que, si bien la traducción de *MTUOC-Translator* resulta de peor calidad cuantitativamente y tiene bastantes aspectos que mejorar, ha traducido mejor la terminología especializada del sector.

Sentence 8	BLEU	Length ratio	Text
Source ⤴	-	-	'Normal procedures' means all procedures established by the operator in the operations manual for day-to-day normal operations (e.g. pre-flight briefing of cabin crew, pre-flight checks, passenger briefing, securing of galleys and cabin, cabin surveillance during flight).
Human	100.00	1.00	" Procedimientos normales " son todos los procedimientos establecidos por el operador en el manual de operaciones para las operaciones normales del día a día (por ejemplo , briefing pre - vuelo de la tripulación de cabina , chequeo pre - vuelo , briefing de los pasajeros , aseguramiento de los galleys y la cabina , vigilancia de la cabina durante el vuelo) .
Machine	51.63	0.97	' Normal procedures ' Significa todos los procedimientos establecidos por el operador en el manual de operaciones por día a día las operaciones normales (p . ej . , Briefing pre - vuelo de la tripulación de cabina de pasajeros , comprobaciones pre - vuelo , pasajero Briefing , asegurar de cabina , galleys y vigilancia de la cabina durante el vuelo) .
Machine	60.83	0.95	" Procedimientos normales " : todos los procedimientos establecidos por el operador en el manual de operaciones para las operaciones normales cotidianas (por ejemplo , información previa al vuelo de la tripulación de cabina , comprobaciones previas al vuelo , información a los pasajeros , aseguramiento de las galeras y la cabina , vigilancia de la cabina durante el vuelo) .

Ilustración 17 Ejemplo buena elección terminológica MTUOC-Translator SMT (BLEU)

- Podemos observar que, gramaticalmente, *MTUOC-Translator* comete muchos errores de concordancia gramatical, esto posiblemente se deba a que el modelo de lengua debería ser entrenado con una cantidad mayor de ejemplos para que no sucedan traducciones como la que se puede ver en la ilustración 18:

Sentence 16	BLEU	Length ratio	Text
Source	-	-	These procedures shall be included in the operations manual.
Human	100.00	1.00	Estos procedimientos se incluirán en el manual de operaciones .
Machine	65.80	1.00	Estos procedimientos se incluirá en el manual de operaciones .
Machine	100.00	1.00	Estos procedimientos se incluirán en el manual de operaciones .

Ilustración 18 Ejemplo fallo concordancia gramatical MTUOC-Translator (BLEU)

Un uso de terminología especializada que ha sabido cumplir *MTUOC-Translator*, si bien no en toda la oración, pero sí en el sustantivo, es el ejemplo de la ilustración 19. En el que se observa la correcta traducción del término *slides* como *rampas*; mientras que *Yandex Translate* lo traduce erróneamente como diapositivas. Si bien es cierto que, *arming and disarming* ha sido bien traducido por el traductor ruso, pero no por el creado en el presente trabajo.

Sentence 21	BLEU	Length ratio	Text
Source	-	-	Arming and disarming of slides.
Human	100.00	1.00	Armado y desarmado de rampas .
Machine	19.30	1.00	Armamento y de desarme rampas .
Machine	53.73	1.00	Armado y desarmado de diapositivas .

Ilustración 19 Ejemplo terminología especializada (BLEU)

- Existen varios ejemplos a lo largo de la traducción llevada a cabo por *MTUOC-Translator* en los que directamente no se han traducido términos, como podemos observar en la ilustración 20; probablemente se deba a un fallo en el modelo de traducción que mejoraría con un corpus más amplio, pero esto es solo una hipótesis.

Sentence 26	BLEU	Length ratio	Text
Source	-	-	Any type of dangerous goods accident or incident, or the finding of undeclared or misdeclared dangerous goods should be reported, irrespective of whether the dangerous goods are contained in cargo, mail, passengers' baggage or crew baggage.
Human	100.00	1.00	Se debe notificar cualquier tipo de accidente o incidente con mercancías peligrosas , o el hallazgo de mercancías peligrosas no declaradas o mal declaradas , independientemente de que las mercancías peligrosas estén contenidas en la carga , el correo , el equipaje de los pasajeros o el de la tripulación .
Machine	36.88	0.90	Cualquier tipo de mercancías peligrosas accidente o incidente , o el hallazgo de misdeclared mercancías peligrosas no declaradas o deben ser reportados , independientemente de si las mercancías peligrosas están contenidas en la carga , correo , passengers ' equipaje o equipaje de la tripulación .
Machine	75.34	1.00	Debe notificarse cualquier tipo de accidente o incidente de mercancías peligrosas , o el hallazgo de mercancías peligrosas no declaradas o mal declaradas , independientemente de si las mercancías peligrosas están contenidas en la carga , el correo , el equipaje de los pasajeros o el equipaje de la tripulación .

Ilustración 20 Ejemplo terminología sin traducir en MTUOC-Translator (BLEU)

Como hemos observado en los segmentos elegidos, en términos generales la traducción de *Yandex Translate* resulta de mejor calidad, pues se acerca más a la traducción realizada por una persona; sin embargo, si nos fijamos en la terminología especializada parece que los segmentos en los que ha fallado *Yandex Translate*, son resueltos de una manera más precisa por el motor de traducción creado con un corpus especializado.

4.2 RESULTADOS DE LA TRADUCCIÓN AUTOMÁTICA NEURONAL

En este apartado comparamos las traducciones realizadas por *MTUOC-Translator* conectado con un motor de traducción automática neuronal entrenado con el corpus especializado en aviación y *Google Translate*. Para evaluarlos con la herramienta ofrecida por Tilde, comparamos una traducción del texto realizada por una persona conocedora de la terminología aeronáutica como base con la traducción realizada por *MTUOC-Translator* (resultados en color azul en las ilustraciones de este apartado) y *Google Translate* (resultados en color verde en las ilustraciones de este apartado). Los resultados generales son los siguientes (ver ilustración 21):

BLEU:	57.82	69.39
Precision x brevity:	61.74 x 93.65	70.44 x 98.50
Type	1-gram 2-gram 3-gram 4-gram	1-gram 2-gram 3-gram 4-gram
Individual	81.67 67.02 56.01 47.39	84.70 74.06 66.21 59.29
Cumulative	76.49 69.28 63.15 57.82	83.43 78.01 73.49 69.39
Export data	CSV	

Ilustración 21 Resultados análisis BLEU de Tilde (NMT)

Del resultado general podría inferirse la idea de que el traductor de *Google* ofrece un mejor resultado; no obstante, debemos analizar individualmente los diferentes segmentos para observar que muchos de los errores que empeoran la puntuación de ambos traductores no son errores de traducción como tal. A continuación, analizaremos algunos de los segmentos que presentan elementos de interés:

- Empeoramiento de la traducción debida a la elección de tiempos verbales distintos pero que podrían considerarse válidos de mantenerse constantes en el documento (ver ilustración 22). Vemos que en este caso las tres traducciones serían perfectamente válidas si mantuviesen la forma verbal constante a lo largo de toda la traducción.

Sentence 4	BLEU	Length ratio	Text
Source	-	-	When scheduling cabin crew for a flight, the operator should establish procedures that take account of the experience of each cabin crew member.
Human	100.00	1.00	Al programar la tripulación de cabina para un vuelo , el operador debe establecer procedimientos que tengan en cuenta la experiencia de cada miembro de la tripulación de cabina .
Machine	71.03	0.90	Al programar la tripulación de cabina para un vuelo , el operador deberá establecer procedimientos que tengan en cuenta la experiencia de cada tripulante de cabina .
Machine	90.96	1.00	Al programar la tripulación de cabina para un vuelo , el operador debería establecer procedimientos que tengan en cuenta la experiencia de cada miembro de la tripulación de cabina .

Ilustración 22 Ejemplo variación tiempo verbal (BLEU)

- En las ilustraciones 23 y 24 podemos distinguir cómo MTUOC-Translator devuelve una traducción más precisa que la de *Google Translate*, probablemente por estar entrenado con corpus especializados en el sector de la aviación. En la ilustración 17 vemos cómo MTUOC-Translator traduce correctamente *operating cabin crew* como *tripulación de cabina operativa*; mientras que, *Google Translate* lo malinterpreta como *tripulante de cabina de operación*. En la ilustración 18 tenemos un ejemplo incluso mejor, pues MUOC-Translator traduce *pre-flight checks* como *briefing previo al vuelo* (también válido como traducción), *passenger briefing* como *briefing de pasajeros* y *securing of galleys* como *aseguramiento de galleys*. Mientras que, *Google Translate* traduce estos mismos términos como *controles previos al vuelo*, *información a los pasajeros* y *sujeción de las cocinas*, respectivamente. Si bien los dos primeros ejemplos podrían llegar a entenderse, no representan la terminología más usada en este campo. El tercer ejemplo sin duda es el más claro de los beneficios de entrenar un corpus especializado, pues la traducción ofrecida por el sistema de traducción de uso general es del todo incorrecta.

Sentence 5	BLEU	Length ratio	Text
Source	-	-	The procedures should specify that the required cabin crew includes some cabin crew members who have at least 3 months experience as an operating cabin crew member.
Human	100.00	1.00	Los procedimientos deben especificar que la tripulación de cabina requerida incluya algunos miembros de la tripulación de cabina que tengan , al menos 3 meses de experiencia como miembro de la tripulación de cabina operativa .
Machine	82.08	1.00	Los procedimientos deben especificar que la tripulación de cabina requerida incluye a algunos miembros de la tripulación de cabina que tienen al menos 3 meses de experiencia como miembro de la tripulación de cabina operativa .
Machine	76.84	1.00	Los procedimientos deben especificar que la tripulación de cabina requerida incluye algunos miembros de la tripulación de cabina que tienen al menos 3 meses de experiencia como miembro de la tripulación de cabina de operación .

Ilustración 23 Ejemplo de mejor traducción terminológica MTUOC-Translator (Tilde)

Sentence 8	BLEU	Length ratio	Text
Source	-	-	'Normal procedures' means all procedures established by the operator in the operations manual for day-to-day normal operations (e.g. pre-flight briefing of cabin crew, pre-flight checks, passenger briefing, securing of galleys and cabin, cabin surveillance during flight).
Human	100.00	1.00	" Procedimientos normales " son todos los procedimientos establecidos por el operador en el manual de operaciones para las operaciones normales del día a día (por ejemplo , briefing pre - vuelo de la tripulación de cabina , chequeo pre - vuelo , briefing de los pasajeros , aseguramiento de los galley s y la cabina , vigilancia de la cabina durante el vuelo) .
Machine	51.25	0.88	« Procedimientos normales » : todos los procedimientos establecidos por el operador en el manual de operaciones para operaciones normales diarias (por ejemplo , Briefing previo al vuelo de la tripulación de cabina , comprobaciones previas al vuelo , Briefing de pasajeros , aseguramiento de galley s y cabina , vigilancia de cabina durante el vuelo) .
Machine	61.72	1.06	`` Procedimientos normales `` : todos los procedimientos establecidos por el operador en el manual de operaciones para las operaciones normales del día a día (por ejemplo , información previa al vuelo de la tripulación de cabina , controles previos al vuelo , información a los pasajeros , sujección de las cocinas y la cabina , vigilancia de la cabina durante el vuelo) .

Ilustración 24 Ejemplo II de mejor traducción terminológica con MTUOC-Translator (Tilde)

- Existen también algunos ejemplos en los que BLEU le da una mejor puntuación a la traducción de *Google Translate* pero que, a nuestro criterio, la ofrecida por MTUOC-Translator supera, como puede verse en la ilustración 19; pues en este segmento la utilización de la traducción literal del verbo *to mean* vuelve malsonante a la traducción, este sería un buen ejemplo de la tradautomatización vista en el marco metodológico.

Sentence 9	BLEU	Length ratio	Text
Source	-	-	'Emergency procedures' means all procedures established by the operator in the operations manual for abnormal and emergency situations.
Human	100.00	1.00	" Procedimientos de emergencia " son todos los procedimientos establecidos por el operador en el manual de operaciones para situaciones anormales y de emergencia .
Machine	80.56	0.92	Procedimientos de emergencia : todos los procedimientos establecidos por el operador en el manual de operaciones para situaciones anormales y de emergencia .
Machine	84.67	1.00	" Procedimientos de emergencia " significa todos los procedimientos establecidos por el operador en el manual de operaciones para situaciones anormales y de emergencia .

Ilustración 25 Ejemplo traducción literal Google Translate (Tilde)

- También tenemos ejemplos en los que ambos sistemas de TA han fallado en la traducción de terminología especializada, como es el caso de la traducción incorrecta de *disruptive passengers* que se debe traducir, únicamente, como *pasajeros conflictivos*

o *slides* como *rampas*. Como vemos en las ilustraciones 27 y 28, ambos traductores automáticos fallaron en este ejemplo.

Sentence 21	BLEU	Length ratio	Text
Source	-	-	Arming and disarming of slides.
Human	100.00	1.00	Armado y desarmado de rampas .
Machine	10.40	0.83	Armando y desarmando tobots .
Machine	53.73	1.00	Armado y desarmado de toboganes .

Ilustración 27 Ejemplo II mala elección terminológica por ambas TA (Tilde)

Sentence 19	BLEU	Length ratio	Text
Source	-	-	Situations requiring information to the flight crew may include disruptive passenger.
Human	100.00	1.00	Las situaciones que requieren informar a la tripulación de vuelo pueden incluir a un pasajero conflictivo .
Machine	44.10	0.88	Las situaciones que requieran información a la tripulación de vuelo pueden incluir pasajeros perturbadores .
Machine	51.01	0.88	Las situaciones que requieren información a la tripulación de vuelo pueden incluir pasajeros molestos .

Ilustración 26 Ejemplo mala elección terminológica por parte de ambas TA (Tilde)

Como podemos observar, aunque *Google Translate* haya obtenido una mejor puntuación en un análisis cuantitativo, la traducción llevada a cabo por nuestro motor de traducción entrenado con un corpus especializado resuelve mejor muchas de las apariciones de terminología especializada.

Podemos añadir, en una comparación entre los dos motores de traducción automática elaborados en el presente trabajo, que el sistema de traducción automática neuronal ofrece mejores resultados de traducción, tanto en la traducción general (concordancia gramatical, uso de verbos, etc.) como en la parte especializada (mejor traducción de la terminología especializada). Concluyendo así que, en este caso, el motor de traducción automática neuronal ofrece mejores resultados que uno estadístico entrenado con los mismos datos.

5. CONCLUSIONES

En este último capítulo trataremos las conclusiones a las que se ha llegado tras la realización del trabajo. Está dividido en tres subapartados, la contribución de este trabajo en el ámbito de estudio, las limitaciones que hemos observado y futuras vías de investigación.

Se podría decir que el resultado de la investigación ha resultado, en líneas generales positivo. Afirmando que sí conviene entrenar motores de traducción con corpus especializados para obtener mejores resultados, no obstante, cabe puntualizar que los resultados no han sido óptimos puesto que en la traducción general ofrecen mejores resultados los traductores automáticos generales. Teniendo en cuenta que la traducción automática aún se encuentra en un estadio en el que requiere de la labor de un profesional para la posesición, los resultados de este trabajo son positivos, pues la revisión de terminología especializada supone un mayor problema que la revisión de la concordancia gramatical. No obstante, es necesario mencionar que ninguno de los traductores automáticos presenta unos resultados ideales en traducción especializada en el sector de la aviación. Probablemente, si se repitiese la investigación con un corpus de mayor tamaño, devolvería mejores resultados.

Cabe añadir también que, el motor de NMT ofrece mejores resultados que el motor de SMT en cuanto a traducción general, como traducción terminológica especializada. Esta conclusión resulta de especial interés porque se han comparado estos dos tipos de motores de traducción y sus resultados, tras haber sido entrenados con el mismo conjunto de datos.

5.1 CONTRIBUCIÓN AL ÁMBITO DE ESTUDIO

La novedad que aporta este trabajo es que muestra cómo han sido llevados a cabo todos los pasos del entrenamiento de motores de traducción automática con una herramienta de libre acceso como MTUOC. Estos pasos incluyen la selección de textos para la compilación del corpus, el procesamiento del corpus, el entrenamiento de los motores de traducción, la conexión a los servidores, el uso de la interfaz gráfica del traductor del proyecto (*MTUOC-Translator*) y la evaluación de las traducciones con las herramientas de Tilde.

Esto permite que cualquier investigador que quiera repetir los pasos podrá crear un motor de traducción especializada en otro sector e ir avanzando así en el campo de la traducción automática especializada.

5.2 LIMITACIONES DEL TRABAJO

Las limitaciones de este trabajo, que podrían hacer variar los resultados son las siguientes. En primer lugar, el par de lenguas utilizado, pues no se repetirían los mismos resultados si la combinación fuese otra. En segundo lugar, la selección de herramientas también puede variar el resultado, pues cada traductor automático funciona de manera diferente, así que los resultados variarían no solo si cambiásemos la herramienta de entrenamiento de motores de traducción (MTUOC), sino que también variaría si comparásemos los resultados con otros motores de traducción generales. Por último, habría sido interesante que la evaluación fuese realizada por usuarios finales y no por una métrica automática que, como vimos en el capítulo anterior, devuelve resultados numéricos que no representan la realidad de la traducción.

5.3 FUTUROS ESTUDIOS / FUTURAS VÍAS DE INVESTIGACIÓN

En cuanto a líneas posibles de investigación futuras, podemos plantear las siguientes cuestiones:

- ¿Qué resultados se ofrecería si se repitiese la investigación, pero con un par de lenguas nuevo que no estuviese tan desarrollado en los motores generales de traducción automática como es el par de lenguas inglés-español o con el mismo par de lenguas, pero con una variante del español?
- ¿Qué resultados ofrecería esta misma investigación si los resultados fuesen evaluados por profesionales del sector aeronáutico?
- ¿Qué resultados ofrecería la investigación si se repitiesen las directrices de este trabajo, pero con un corpus mayor que permitiese mejorar los modelos de lengua y los modelos de traducción?
- ¿Qué resultados ofrecería esta investigación si se utilizasen otras herramientas de entrenamiento de motores de traducción, pero el mismo corpus especializado?

6. BIBLIOGRAFÍA

- Abaitua, J. (2002). *Introducción a la traducción automática*.
http://paginaspersonales.deusto.es/abaitua/konzeptu/ta/mt10h_es/
- Arnold, D. J., Balkan, L., Meijer, S., & Sadler, L. (1994). *Machine Translation: An Introductory Guide*. Blackwells-NCC.
- Bahdanau, D., Cho, K., & Bengio, Y. (2016). Neural Machine Translation by Jointly Learning to Align and Translate. <http://arxiv.org/abs/1409.0473>
- Casacuberta Nolla, F., & Peris Abril, Á. (2017). Neural Machine Translation. *Tradumàtica: tecnologies de la traducció*, 15, 66. <https://doi.org/10.5565/rev/tradumatica.203>
- Castilho, S., Moorkens, J., Gaspari, F., Calixto, I., Tinsley, J., & Way, A. (2017). Is Neural Machine Translation the New State of the Art? *The Prague Bulletin of Mathematical Linguistics*, 108, 109-120. <https://doi.org/10.1515/pralin-2017-0013>
- Cho, K., van Merriënboer, B., Bahdanau, D., & Bengio, Y. (2014). On the Properties of Neural Machine Translation: Encoder-Decoder Approaches.
<http://arxiv.org/abs/1409.1259>
- Do, C., & Batzoglou, S. (2008). What is expectation maximization algorithm? *Nature Biotechnology*, 26, 897-899.
- Fernández-Rodríguez, M. (2010). Evolución de la traducción asistida por ordenador y de las herramientas de apoyo a las memorias de traducción. *Sendebarr*, 21, 201-230.
<https://doi.org/10.30827/sendebarr.v21i0.374>
- Hearne, M., & Way, A. (2011). Statistical Machine Translation: A guide for linguists and translators. *Language and Linguistics Compass*, 205-226.
- Hutchins, W. J. (1995). *Introducción a la Traducción Automática*. Visor Lingüística y conocimiento.
- Junczys-Dowmunt, M., Grundkiewicz, R., Dwojak, T., Hoang, H., Heafield, K., Neckermann, T., Seide, F., Hermann, U., Aji, A. F., Bogoychev, N., Martins, A. F. T., & Birch, A. (2018). Marian: Fast Neural Machine Translation in C++. *Proceedings of ACL 2018, System Demonstrations*, 116-121. <https://doi.org/10.18653/v1/P18-4020>
- Jurafsky, D., & Martin, J. (2000). *Speech and language processing*. Prentice Hall.
- Kalchbrenner, N., & Blunsom, P. (2013). Recurrent Continuous Translation Models. *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, 1700-1709. <https://www.aclweb.org/anthology/D13-1176>

- Koehn, P. (2005). Europarl: A Parallel Corpus for Statistical Machine Translation. *MT summit*, 5, 79-86.
- Koehn, P. (2009). *Statistical Machine Translation System User Manual and Code Guide*. 359.
- Koehn, P. (2010). *Statistical Machine Translation*. Cambridge: University Press.
- Kudo, T., & Richardson, J. (2018). SentencePiece: A simple and language independent subword tokenizer and detokenizer for Neural Text Processing.
<http://arxiv.org/abs/1808.06226>
- Locke, W. N., & Booth, A. D. (1955). *Machine Translation of Languages*. Technology Press of the Massachusetts Institute of Technology and Wiley.
- Marín, M., Moreno Sandoval, A., & Sánchez León, F. (s. f.). *El proyecto EUROTRA en el marco de la investigación sobre traducción por ordenador*. Recuperado 26 de mayo de 2021, de
http://www.quadernsdigitals.net/datos_web/articles/telos/telos16/t16proyecto.htm
- Matich, D. J. (2001). *Redes Neuronales: Conceptos Básicos y Aplicaciones*.
- Mayor, A. (2009). Evaluación de un sistema de traducción automática basado en reglas o por qué BLEU sólo sirve para lo que sirve. *Procesamiento del lenguaje natural*, 43, 197-205.
- Molina Baños, A. (2019). *Creación de un motor de traducción automática estadístico (EN>ES) para textos del ámbito farmacéutico. Comparación con otros motores de traducción automática neuronal existentes* [Trabajo de Fin de máster]. Universidad Autónoma de Barcelona.
- Moorkens, J. (2017). Under pressure: Translation in times of austerity. *Perspectives*, 25(3), 464-477. <https://doi.org/10.1080/0907676X.2017.1285331>
- Moré López, J., & Climent Roca, S. (2006). *La tradautomatización: Un concepto aplicado a la evaluación de sistemas de traducción automática*.
<http://rua.ua.es/dspace/handle/10045/3302>
- Mutal, J. D., Volkart, L., Bouillon, P., Girletti, S., & Estrella, P. S. (2019). Differences between SMT and NMT Output—A Translators’ Point of View. En *The Second Workshop on Human-Informed Translation and Interpreting Technology (HiT-IT 2019)*. <https://archive-ouverte.unige.ch/unige:123216>
- Oates, B. (2005). *Researching Information Systems and Computing*. SAGE Publications Ltd.

- Oliver, A. (2018). *Avances recientes en Traducción Automática Neuronal. ¿Cómo se puede beneficiar el traductor humano?*
<https://www.youtube.com/watch?v=oa8xV6FxfnY&t=1578s>
- Oliver, A. (2020a). *Introducción a Moses—XWiki*.
<https://xwiki.recursos.uoc.edu/wiki/mat00001ca/view/1.4.%20Traducci%C3%B3n%20autom%C3%A1tica%20y%20postedici%C3%B3n/2.%20Traducci%C3%B3n%20autom%C3%A1tica%20estad%C3%ADstica%20con%20Moses/2.%20Introducci%C3%B3n%20a%20Moses/>
- Oliver, A. (2020b). MTUOC: easy and free integration of NMT systems in professional translation environments. *European Association for Machine Translation (EAMT). Proceedings 2020.*, 467-468.
- Oliver, A. (2021a). *Marian en funcionamiento—XWiki*.
<https://xwiki.recursos.uoc.edu/wiki/mat21570ca/view/3.%20Traducci%C3%B3n%20autom%C3%A1tica%20neuronal%20con%20Marian%20y%20MTUOC/3.5.%20Marian%20en%20funcionamiento/>
- Oliver, A. (2021b). *Corpus preparation and preprocessing with MTUOC - XWiki*.
<http://xwiki.recursos.uoc.edu/wiki/matm4957ca/view/Corpus%20preparation%20and%20preprocessing%20with%20MTUOC/>
- Papineni, K., Roukos, S., Ward, T., & Zhu, W. J. (2002). Bleu: A method for automatic evaluation of machine translation. *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, 311-318.
- Pérez Macías, L. (2017). *Análisis de las percepciones en torno a la práctica de la posesición en el sector profesional de la traducción en España*.
<https://rio.upo.es/xmlui/handle/10433/6934>
- Ribas, M. M. (2016). *La traducción automática: ¿una amenaza para el traductor del siglo XXI?*
- Saptarshi, P., & Purkhyastha, B. (2020a). An NLP tool for decoding the ATC Phraseology from English to Bengali. *2020 International Conference on Smart Electronics and Communication (ICOSEC)*, 264-270.
- Saptarshi, P., & Purkhyastha, B. (2020b). English to Bengali Neural Machine Translation System for the Aviation Domain. *INFOCOMP: Journal of Computer Science*, 19(2), 78-97.

- Saptarshi, P., & Purkhyastha, B. (2020c). Handling aviation OOV words for machine translation and corpus creation. *International Conference on Smart Electronics and Communication (ICOSEC)*.
- Saptarshi, P., & Purkhyastha, B. (2020d). Detailed analysis of implementation of aviation NMT system and the effects of aviation post-processing tools on TDIL tourism corpus. *Journal of King Saud University-Computer and Information Sciences*.
- Schmitt, N. (2010). *Applied Linguistics*. Routledge.
- Schuster, M., Johnson, M., & Thorat, N. (s. f.). Zero-Shot Translation with Google's Multilingual Neural Machine Translation System. *Google AI Blog*. Recuperado 25 de mayo de 2021. <http://ai.googleblog.com/2016/11/zero-shot-translation-with-googles.html>
- Sennrich, R., & Haddow, B. (2016). Linguistic Input Features Improve Neural Machine Translation. *Proceedings of the First Conference on Machine Translation: Volume 1, Research Papers*, 83-91. <https://doi.org/10.18653/v1/W16-2209>
- Sennrich, R., Haddow, B., & Birch, A. (2016a). Improving Neural Machine Translation Models with Monolingual Data. <http://arxiv.org/abs/1511.06709>
- Sennrich, R., Haddow, B., & Birch, A. (2016b). Neural Machine Translation of Rare Words with Subword Units. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 1715-1725. <https://doi.org/10.18653/v1/P16-1162>
- Somers, H. (1997). A Practical Approach to Using Machine Translation Software. *The Translator*, 3:2, 193-212.
- Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to Sequence Learning with Neural Networks. <http://arxiv.org/abs/1409.3215>
- Tiedemann, J. (2012). Parallel Data, Tools and Interfaces in OPUS. *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC'2012)*.
- Torres Hostench, O., Piqué i Huerta, R., Sánchez-Gijón, P., Aguilar-Amat, A., Martín Mor, A., Rico Pérez, C., Alcina Caudet, A., & Candel Mora, M. Á. (2016). *El uso de traducción automática y posesición en las empresas de servicios lingüísticos españolas: Informe de investigación ProjecTA 2015*. <https://ddd.uab.cat/record/148361>
- Volkart, L., Bouillon, P., & Girletti, S. (2018, noviembre 15). *Statistical vs. Neural Machine Translation: A Comparison of MTH and DeepL at Swiss Post's Language Service*.

Wang, X., Lu, Z., Tu, Z., Li, H., Xiong, D., & Zhang, M. (2017). Neural Machine Translation Advised by Statistical Machine Translation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 31(1), Article 1.

<https://ojs.aaai.org/index.php/AAAI/article/view/10975>

ANEXO I: Programas utilizados para la creación del corpus y el entrenamiento de los motores de traducción automáticos

En este anexo se presentan los programas utilizados para realizar este trabajo. Todos ellos han sido desarrollados en Python (versión 3) y, aunque la mayoría son compatibles con Windows, algunos de los componentes funcionan solo con Ubuntu (como se explicó en el apartado 3 sobre Entrenamiento de los motores de traducción. Se adjunta, además, el enlace donde se encuentran alojados los programas con última fecha de consulta del 08/06/2021.

COMPILACIÓN DEL CORPUS

NOMBRE	USO	LOCALIZACIÓN
pdf2txtTika.py	Transformación pdf en txt	https://xwiki.reursos.uoc.edu/wiki/mat00001ca/view/Research%20on%20Translation%20Technologies/Working%20with%20epub%20files%20using%20Python/
segmentHUNDIR.py	Segmentación de textos (complementos: segmenter-spa.pickle y segmenter-eng.pickle)	https://xwiki.reursos.uoc.edu/wiki/mat00001ca/view/Research%20on%20Translation%20Technologies/Automatic%20text%20alignment%20with%20Hunalign/
Hunalign	Alineación de textos (complementos: hunapertium-es.en.dic)	https://xwiki.reursos.uoc.edu/wiki/mat00001ca/view/Research%20on%20Translation%20Technologies/Automatic%20text%20alignment%20with%20Hunalign/
createBatchFile.py	Creación batch para alinear corpus	https://xwiki.reursos.uoc.edu/wiki/mat00001ca/view/Research%20on%20Translation%20Technologies/Automatic%20text%20alignment%20with%20Hunalign/
createParallelCorpus.py	Creación del corpus	https://xwiki.reursos.uoc.edu/wiki/mat00001ca/view/Research%20on%20Translation%20Technologies/Automatic%20text%20alignment%20with%20Hunalign/
MTUOC-clean-parallel-corpus.py	Limpieza del corpus	https://github.com/aoliverg/MTUOC/tree/master/MTUOC-clean-parallel-corpus

Tabla 3 Programas Python compilación del corpus

PROCESAMIENTO DEL CORPUS

NOMBRE	USO	LOCALIZACIÓN
MTUOC-combine-corpus.py	Combinación de corpus general con especializado	https://github.com/aoliverg/MTUOC/tree/master/MTUOC-combine-corpus
MTUOC-prepare-corpus.py	Limpieza últimos detalles corpus compilado	https://github.com/aoliverg/MTUOC/tree/master/MTUOC-prepare-corpus
MTUOC-SMT-preprocess.py	Procesamiento para entrenamiento de SMT	https://github.com/aoliverg/MTUOC/tree/master/MTUOC-SMT-preprocess
MTUOC-NMT-SP-preprocess.py	Procesamiento para entrenamiento de NMT con SentencePiece	https://github.com/aoliverg/MTUOC/tree/master/MTUOC-NMT-SP-preprocess

Tabla 4 Programas Python procesamiento del corpus

PUESTA EN MARCHA SERVIDOR NMT

NOMBRE	USO	LOCALIZACIÓN
MTUOC-server-SP.py	Inicio de la conexión con el servidor Marian	https://github.com/aoliverg/MTUOC/tree/master/MTUOC-server-SP
MTUOC-stop-server.py	Fin de la conexión con el servidor Marian	https://github.com/aoliverg/MTUOC/tree/master/MTUOC-server-SP

Tabla 5 Programas Python puesta en marcha servidor NMT

ANEXO II:
Textos utilizados para el corpus

CONTROL TEXTOS CORPUS PARALELO INGLÉS-ESPAÑOL DEL CAMPO DE LA AVIACIÓN					
ID_t exto	Nombre	Nombre Corpus	Dirección Web	Idioma	Título texto
1	EASA1-es	EASA	https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX:32018R1139	Español	REGLAMENTO (UE) 2018/1139 DEL PARLAMENTO EUROPEO Y DEL CONSEJO
2	EASA1-en	EASA	https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32018R1139&from=ES	Inglés	REGULATION (EU) 2018/1139 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL
3	EASA2-es	EASA	https://eur-lex.europa.eu/legal-content/ES/TXT/HTML/?uri=CELEX:32014R0376&from=EN	Español	REGLAMENTO (UE) No 376/2014 DEL PARLAMENTO EUROPEO Y DEL CONSEJO
4	EASA2-en	EASA	https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32014R0376&from=EN	Inglés	REGULATION (EU) No 376/2014 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL
5	EASA3-es	EASA	https://eur-lex.europa.eu/legal-content/ES/TXT/HTML/?uri=CELEX:32010R0996&qid=1619170628321&from=EN	Español	REGLAMENTO (UE) No 996/2010 DEL PARLAMENTO EUROPEO Y DEL CONSEJO
6	EASA3-en	EASA	https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32010R0996&qid=1619170628321&from=EN	Inglés	REGULATION (EU) No 996/2010 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL
7	EASA4-es	EASA	https://eur-lex.europa.eu/legal-content/ES/TXT/HTML/?uri=CELEX:32008R0300&qid=1619171763259&from=EN	Español	REGLAMENTO (CE) No 300/2008 DEL PARLAMENTO EUROPEO Y DEL CONSEJO
8	EASA4-en	EASA	https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32008R0300&qid=1619171763259&from=EN	Inglés	REGULATION (EC) No 300/2008 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL
9	EASA5-es	EASA	https://eur-lex.europa.eu/legal-content/ES/TXT/HTML/?uri=CELEX:32019R0494&qid=1619171763259&from=EN	Español	REGLAMENTO (UE) 2019/494 DEL PARLAMENTO EUROPEO Y DEL CONSEJO
10	EASA5-en	EASA	https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32019R0494&qid=1619171763259&from=EN	Inglés	REGULATION (EU) 2019/494 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL
11	EASA6-es	EASA	https://eur-lex.europa.eu/legal-content/ES/TXT/HTML/?uri=CELEX:32005R2111&qid=1619171763259&from=EN	Español	REGLAMENTO (CE) No 2111/2005 DEL PARLAMENTO EUROPEO Y DEL CONSEJO
12	EASA6-en	EASA	https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32005R2111&qid=1619171763259&from=EN	Inglés	REGULATION (EC) No 2111/2005 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL

13	EASA7-es	EASA	https://eur-lex.europa.eu/legal-content/ES/TXT/HTML/?uri=CELEX:32016L0681&qid=1619171763259&from=EN	Español	DIRECTIVA (UE) 2016/681 DEL PARLAMENTO EUROPEO Y DEL CONSEJO
14	EASA7-en	EASA	https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016L0681&qid=1619171763259&from=EN	Inglés	DIRECTIVE (EU) 2016/681 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL
15	EASA8-es	EASA	https://eur-lex.europa.eu/legal-content/ES/TXT/HTML/?uri=CELEX:32019R0712&qid=1619171763259&from=EN	Español	REGLAMENTO (UE) 2019/712 DEL PARLAMENTO EUROPEO Y DEL CONSEJO
16	EASA8-en	EASA	https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32019R0712&qid=1619171763259&from=EN	Inglés	REGULATION (EU) 2019/712 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL
17	EASA9-es	EASA	https://eur-lex.europa.eu/legal-content/ES/TXT/HTML/?uri=CELEX:32004R0785&qid=1619171763259&from=EN	Español	REGLAMENTO (CE) No 785/2004 DEL PARLAMENTO EUROPEO Y DEL CONSEJO
18	EASA9-en	EASA	https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32004R0785&qid=1619171763259&from=EN	Inglés	REGULATION (EC) No 785/2004 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL
19	EASA10-es	EASA	https://eur-lex.europa.eu/legal-content/ES/TXT/HTML/?uri=CELEX:32006R1107&qid=1619171763259&from=EN	Español	REGLAMENTO (CE) N ° 1107/2006 DEL PARLAMENTO EUROPEO Y DEL CONSEJO
20	EASA10-en	EASA	https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32006R1107&qid=1619171763259&from=EN	Inglés	REGULATION (EC) No 1107/2006 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL
21	EASA11-es	EASA	https://eur-lex.europa.eu/legal-content/ES/TXT/HTML/?uri=CELEX:32004R0261&qid=1619171763259&from=EN	Español	Reglamento (CE) n° 261/2004 del Parlamento Europeo y del Consejo
22	EASA11-en	EASA	https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32004R0261&qid=1619171763259&from=EN	Inglés	Regulation (EC) No 261/2004 of the European Parliament and of the Council
23	EASA12-es	EASA	https://eur-lex.europa.eu/legal-content/ES/TXT/HTML/?uri=CELEX:32003R0437&qid=1619171763259&from=EN	Español	Reglamento (CE) n° 437/2003 del Parlamento Europeo y del Consejo
24	EASA12-en	EASA	https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32003R0437&qid=1619171763259&from=EN	Inglés	Regulation (EC) No 437/2003 of the European Parliament and of the Council
25	EASA13-es	EASA	https://eur-lex.europa.eu/legal-content/ES/TXT/HTML/?uri=CELEX:32016R0399&qid=1619171763259&from=EN	Español	REGLAMENTO (UE) 2016/399 DEL PARLAMENTO EUROPEO Y DEL CONSEJO

40	ICAO_6-en	ICAO	https://www.icao.int/safety/aviation-medicine/AvMedSARS/Cleared%20for%20Take%20Off!.pdf	Inglés	CLEARED FOR TAKE OFF! A Pilot's Guide to Returning to Flying
41	ICAO_7-en	ICAO	https://www.icao.int/covid/cart/Documents/067s.pdf	Español	SL 067: ICAO Council Aviation Recovery Task Force (CART) Report and annexed guidance document Take-off: Guidance for Air Travel through the COVID-19 Public Health Crisis
42	ICAO_7-es	ICAO	https://www.icao.int/covid/cart/Documents/067e.pdf	Inglés	SL 067: ICAO Council Aviation Recovery Task Force (CART) Report and annexed guidance document Take-off: Guidance for Air Travel through the COVID-19 Public Health Crisis
43	ICAO_8-es	ICAO	https://www.icao.int/safety/CAPSCA/PublishingImages/Pages/Coronavirus/SL%20050s.pdf	Español	SL 050: Operational measures to ensure safe operations during the COVID-19 pandemic
44	ICAO_8-en	ICAO	https://www.icao.int/safety/CAPSCA/PublishingImages/Pages/Coronavirus/SL%20050e.pdf	Inglés	SL 050: Operational measures to ensure safe operations during the COVID-19 pandemic
45	ICAO_9-es	ICAO	https://www.icao.int/safety/CAPSCA/PublishingImages/Pages/Coronavirus/SL%20047s.pdf	Español	SL 047: Implementation of contingency arrangements to reduce the risks of the spread of COVID-19
46	ICAO_9-en	ICAO	https://www.icao.int/safety/CAPSCA/PublishingImages/Pages/Coronavirus/SL%20047e.pdf	Inglés	SL 047: Implementation of contingency arrangements to reduce the risks of the spread of COVID-19
47	ICAO_10-es	ICAO	https://www.icao.int/safety/CAPSCA/PublishingImages/Pages/Coronavirus/SL%20046s.pdf	Español	SL 046: Adherence to relevant ICAO Annex 9 – Facilitation Standards; and Actions taken by Member States to reduce the spread of the novel coronavirus (COVID-19) by air transport
48	ICAO_10-en	ICAO	https://www.icao.int/safety/CAPSCA/PublishingImages/Pages/Coronavirus/SL%20046e.pdf	Inglés	SL 046: Adherence to relevant ICAO Annex 9 – Facilitation Standards; and Actions taken by Member States to reduce the spread of the novel coronavirus (COVID-19) by air transport
49	ICAO_11-es	ICAO	https://www.icao.int/safety/CAPSCA/PublishingImages/Pages/Coronavirus/SL%202020-015s.pdf	Español	SL 015: State support to prevent the spread of Novel Coronavirus Disease COVID-19
50	ICAO_11-en	ICAO	https://www.icao.int/safety/CAPSCA/PublishingImages/Pages/Coronavirus/SL%202020-015e.pdf	Inglés	SL 015: State support to prevent the spread of Novel Coronavirus Disease COVID-19
51	ICAO_12-es	ICAO	https://www.icao.int/safety/CAPSCA/PublishingImages/Pages/ICAO-References/ES%20TemplateCAPSCAINPSPEdenProd374156v1.pdf	Español	Organización de Aviación Civil Internacional Plan de contingencia de la aviación civil ante emergencias de salud pública
52	ICAO_12-en	ICAO	https://www.icao.int/safety/CAPSCA/PublishingImages/Pages/ICAO-References/EN%20TemplateAviationPHEPNov2010%20(4).pdf	Inglés	International Civil Aviation Organization: Template for a National Aviation Public Health Emergency Preparedness Plan

53	ICAO_13-es	ICAO	https://www.icao.int/safety/GASP/Documents/GASP%20language%20versions%202020/Doc%2010004%202020-2022%20GASP%20SP.pdf	Español	Doc 10004 Plan global para la seguridad operacional de la aviación
54	ICAO_13-en	ICAO	https://www.icao.int/safety/GASP/Documents/Doc.10004%20GASP%202020-2022%20EN.pdf	Inglés	Doc 10004 Global Aviation Safety Plan
55	ICAO_14-es	ICAO	https://www.icao.int/safety/GASP/Documents/Doc%2010131/10131_sp.pdf	Español	Doc 10131 Manual de elaboración de planes nacionales y regionales de seguridad operacional de la aviación
56	ICAO_14-en	ICAO	https://www.icao.int/safety/GASP/Documents/Doc%2010131/10131_en.pdf	Inglés	Doc 10131 Manual on the Development of Regional and National Aviation Safety Plans
57	ICAO_15-es	ICAO	https://www.icao.int/safety/GASP/Documents/Circ358_sp.pdf	Español	Circular 358 Listas de verificación de los planes nacionales y regionales de seguridad operacional de la aviación
58	ICAO_15-en	ICAO	https://www.icao.int/safety/GASP/Documents/358_en.pdf	Inglés	Circular 358 Regional and National Aviation Safety Plan Checklists
59	ICAO_16-es	ICAO	https://www.icao.int/safety/Implementation/Documents/GENERIC%20TERMS%20OF%20REFERENCE%20OF%20REGIONAL%20AVIATION%20SAFETY%20GROUPS%20(RASGs).sp.pdf	Español	ATRIBUCIONES GENERALES DE LOS GRUPOS REGIONALES DE SEGURIDAD OPERACIONAL DE LA AVIACIÓN (RASG)
60	ICAO_16-en	ICAO	https://www.icao.int/safety/Implementation/Documents/GENERIC%20TERMS%20OF%20REFERENCE%20OF%20REGIONAL%20AVIATION%20SAFETY%20GROUPS%20(RASGs).en.pdf	Inglés	GENERIC TERMS OF REFERENCE OF REGIONAL AVIATION SAFETY GROUPS (RASGs)
61	ICAO_17-es	ICAO	https://www.icao.int/safety/Implementation/Documents/GENERIC%20TERMS%20OF%20REFERENCE%20OF%20PLANNING%20AND%20IMPLEMENTATION%20REGIONAL%20GROUPS%20(PIRGs).sp.pdf	Español	ATRIBUCIONES GENERALES DE LOS GRUPOS REGIONALES DE PLANIFICACIÓN Y EJECUCIÓN (PIRG)
62	ICAO_17-en	ICAO	https://www.icao.int/safety/Implementation/Documents/GENERIC%20TERMS%20OF%20REFERENCE%20OF%20PLANNING%20AND%20IMPLEMENTATION%20REGIONAL%20GROUPS%20(PIRGs).en.pdf	Inglés	GENERIC TERMS OF REFERENCE OF PLANNING AND IMPLEMENTATION REGIONAL GROUPS (PIRGs)

63	ICAO_18-es	ICAO	https://www.icao.int/safety/GASP/GASP%20Library/Previous%20GASP%20documentation/2014-2016%20Edition/GASP%202014%202016%20edition%20ES.pdf	Español	Plan global para la seguridad operacional de la aviación 2014–2016
64	ICAO_18-en	ICAO	https://www.icao.int/safety/GASP/GASP%20Library/Previous%20GASP%20documentation/2014-2016%20Edition/2014%202016%20GASP%20EN.pdf	Inglés	2014–2016 Global Aviation Safety Plan
65	ICAO_19-es	ICAO	https://www.icao.int/safety/GASP/GASP%20Library/Regional%20aviation%20safety%20plans/SAMSP_sp_29may2019.pdf	Español	P L A N D E S E G U R I D A D O P E R A C I O N A L D E L A R E G I O N S A M (S A M S P) [2019]
66	ICAO_19-en	ICAO	https://www.icao.int/safety/GASP/GASP%20Library/Regional%20aviation%20safety%20plans/SAMSP_en_29may2019.pdf	Inglés	S A M R E G I O N S A F E T Y P L A N (S A M S P) [2019]
67	ICAO_20-es	ICAO	https://www.icao.int/safety/GASP/GASP%20Library/Previous%20GASP%20documentation/2017-2019%20Edition/10004_es.pdf	Español	Doc 10004 [2017-2019]
68	ICAO_20-en	ICAO	https://www.icao.int/safety/GASP/GASP%20Library/Previous%20GASP%20documentation/2017-2019%20Edition/10004_en.pdf	Inglés	Doc 10004 [2017-2019]
69	ICAO_21-es	ICAO	https://www.icao.int/safety/State%20of%20Global%20Aviation%20Safety/ICAO_SGAS_book_SP_SEPT2013_final_web.pdf	Español	Situación de la seguridad de la aviación mundial [2013]
70	ICAO_21-en	ICAO	https://www.icao.int/safety/State%20of%20Global%20Aviation%20Safety/ICAO_SGAS_book_EN_SEPT2013_final_web.pdf	Inglés	State of Global Aviation Safety [2013]
71	ICAO_22-es	ICAO	https://www.icao.int/safety/Documents/ICAO_State-of-Global-Safety_web_SP.pdf	Español	Situación de la seguridad de la aviación mundial [2011]
72	ICAO_22-en	ICAO	https://www.icao.int/safety/Documents/ICAO_State-of-Global-Safety_web_EN.pdf	Inglés	State of Global Aviation Safety [2011]
73	ICAO_23-es	ICAO	https://www.icao.int/safety/airnavigation/OPS/CabinSafety/Documents/10148_es.pdf	Español	Doc 10148 Manual de instrucción periódica de la tripulación de cabina durante la COVID-19 [2020]

74	ICAO_23-en	ICAO	https://www.icao.int/safety/airnavigation/OPS/CabinSafety/Documents/Doc.10148.alltext.en.pdf	Inglés	Doc 10148 ICAO Handbook for Cabin Crew Recurrent Training during COVID-19 [2020]
75	IATA_1-es	IATA	https://www.iata.org/contentassets/df216feeb8bb4d52a3e16befe9671033/iata-guidance-cabin-operations-during-post-pandemic_spanish.pdf	Español	Material de Orientación General para las Operaciones en Cabina de Pasajeros Durante y Post-Pandemia Edición 4 - 08 Sep 2020
76	IATA_1-en	IATA	https://www.iata.org/contentassets/df216feeb8bb4d52a3e16befe9671033/iata-guidance-cabin-operations-during-post-pandemic.pdf	Inglés	Guidance for Cabin Operations During and Post Pandemic Edition 4 – 08 Sep 2020
77	ICAO_24-es	ICAO	https://www.icao.int/Meetings/avseconf/Documents/HLCA%20-%20Communique%2014%20September%202012%20SP.pdf	Español	COMUNICADO DE LA CONFERENCIA DE ALTO NIVEL SOBRE SEGURIDAD DE LA AVIACIÓN (Montreal, 12 al 14 de septiembre de 2012)
78	ICAO_24-en	ICAO	https://www.icao.int/Meetings/avseconf/Documents/HLCA%20-%20Communique%2014%20September%202012.pdf	Inglés	COMMUNIQUE OF THE ICAO HIGH-LEVEL CONFERENCE ON AVIATION SECURITY Montréal, 12 to 14 September 2012
79	ICAO_25-es	ICAO	https://www.icao.int/Meetings/avseconf/Documents/Principles%20on%20Air%20Cargo%20and%20Mail%20Security%20SP.pdf	Español	PRINCIPIOS DE LA SEGURIDAD DE LA CARGA AÉREA Y EL CORREO PRESENTADOS PARA APROBACIÓN A LA CONFERENCIA DE ALTO NIVEL SOBRE SEGURIDAD DE LA AVIACIÓN (HLCAS) DE LA ORGANIZACIÓN DE AVIACIÓN CIVIL INTERNACIONAL (OACI)
80	ICAO_25-en	ICAO	https://www.icao.int/Meetings/avseconf/Documents/Principles%20on%20Air%20Cargo%20and%20Mail%20Security%20EN.pdf	Inglés	PRINCIPLES ON AIR CARGO AND MAIL SECURITY PRESENTED FOR ENDORSEMENT TO THE INTERNATIONAL CIVIL AVIATION ORGANIZATION (ICAO) HIGH-LEVEL CONFERENCE ON AVIATION SECURITY (HLCAS)
81	ICAO_26-es	ICAO	https://www.icao.int/SAM/Lists/SFActivityProducts/DispForm.aspx?ID=2&ContentTypeId=0x0100820CE6B1CB9B2246B99482F92F1707B4	Español	GUÍA PARA LA REAPERTURA GRADUAL DE ACTIVIDADES COMERCIALES DEL AEROPUERTO
82	ICAO_26-en	ICAO	https://www.icao.int/SAM/Lists/SFActivityProducts/DispForm.aspx?ID=1&ContentTypeId=0x0100820CE6B1CB9B2246B99482F92F1707B4	Inglés	GUIDANCE FOR THE GRADUAL REOPENING OF AIRPORT COMMERCIAL ACTIVITIES
83	ICAO_27-es	ICAO	https://www.icao.int/SAM/eDocumentsrestore/Marco%20Estrategico%20SAM%20ES%20Completo.pdf	Español	MARCO ESTRATEGICO de la Región SAM en respuesta a COVID-19
84	ICAO_27-en	ICAO	https://www.icao.int/SAM/eDocumentsrestore/Strategic%20	Inglés	STRATEGIC FRAMEWORK in response to COVID-19 in SAM

			Framework%20SAM%20EN%20Completo.pdf		
85	ICAO_28-es	ICAO	https://www.icao.int/SAM/eDocumentsstore/19PBN%20Programa%20Implantacion%20RUTA%20Ruta%20spa.pdf	Español	PROYECTO DE IMPLANTACIÓN PBN OPERACIONES EN RUTA A CORTO PLAZO REGIÓN SAM
86	ICAO_28-en	ICAO	https://www.icao.int/SAM/eDocumentsstore/19PBN%20RUTA%20Route%20Implementation%20Programme%20eng.pdf	Inglés	PBN IMPLEMENTATION PROJECT EN-ROUTE OPERATIONS SHORT TERM SAM REGION
87	ICAO_29-es	ICAO	https://www.icao.int/SAM/eDocumentsstore/1ATFM%20CONOPS%20CARSAM%20Spa.pdf	Español	Concepto Operacional para la Gestión de la Afluencia del Tránsito para las Regiones Caribe y Sudamérica (CONOPS ATFM CAR/SAM)
88	ICAO_29-en	ICAO	https://www.icao.int/SAM/eDocumentsstore/1ATFM%20CONOPS%20CARSAM%20ENG.pdf	Inglés	Caribbean/South American Air Traffic Flow Management Concept of Operation (CAR/SAM ATFM CONOPS)
89	ICAO_30-es	ICAO	https://www.icao.int/SAM/eDocumentsstore/1Optimizacion%20ATSRO%20Programa%20Implantacion.pdf	Español	Programa para la Optimización de la Red de Rutas ATS en la Región Sudamericana
90	ICAO_30-en	ICAO	https://www.icao.int/SAM/eDocumentsstore/1Optimizacion%20ATSRO%20Implementation%20Programme.pdf	Inglés	Programme for Optimising the ATS Route Network in the South American Region
91	ICAO_31-es	ICAO	https://www.icao.int/SAM/eDocumentsstore/2003plannacsarspa.pdf	Español	MATERIAL DE ORIENTACIÓN PARA LA ELABORACIÓN DE UN PLAN NACIONAL DE BÚSQUEDA Y SALVAMENTO
92	ICAO_31-en	ICAO	https://www.icao.int/SAM/eDocumentsstore/2003plannacsareng.pdf	Inglés	GUIDANCE MATERIAL FOR THE PREPARATION OF A NATIONAL SEARCH AND RESCUE PLAN
93	ICAO_32-es	ICAO	https://www.icao.int/SAM/eDocumentsstore/2ATFM%20Manual%20Spa%20MAR10.pdf	Español	MANUAL DE GESTIÓN DE AFLUENCIA DEL TRÁNSITO AÉREO PARA LAS REGIONES CARIBE/SUDAMÉRICA (MANUAL ATFM CAR/SAM)
94	ICAO_32-en	ICAO	https://www.icao.int/SAM/eDocumentsstore/2ATFM%20Manual%20Eng%20MAR10.pdf	Inglés	CARIBBEAN/SOUTH AMERICAN AIR TRAFFIC FLOW MANAGEMENT MANUAL (CAR/SAM ATFM MANUAL)
95	ICAO_33-es	ICAO	https://www.icao.int/SAM/eDocumentsstore/3ATFM%20SAM%20Roadmap%20Spa.pdf	Español	HOJA DE RUTA PARA LA GESTIÓN DE AFLUENCIA DEL TRÁNSITO AEREO EN LA REGION SAM
96	ICAO_33-en	ICAO	https://www.icao.int/SAM/eDocumentsstore/3ATFM%20SAM%20Roadmap%20Eng.pdf	Inglés	SAM ROADMAP FOR AIR TRAFFIC FLOW MANAGEMENT

97	ICAO_34-es	ICAO	https://www.icao.int/SAM/eDocumentsrestore/4CDM%20Manual%20Spa.pdf	Español	MANUAL DEL PROCESO DE TOMA DE DECISIONES EN COLABORACIÓN PARA LA REGION SUDAMERICANA (SAM)
98	ICAO_34-en	ICAO	https://www.icao.int/SAM/eDocumentsrestore/4CDM%20Manual%20Eng.pdf	Inglés	MANUAL ON THE COLLABORATIVE DECISION-MAKING PROCESS FOR THE SOUTH AMERICAN (SAM) REGION
99	ICAO_35-es	ICAO	https://www.icao.int/SAM/eDocumentsrestore/AMHS%20Guia.pdf	Español	Guía de Orientación para la Interconexión Operativa de Sistemas AMHS en la Region SAM
100	ICAO_35-en	ICAO	https://www.icao.int/SAM/eDocumentsrestore/AMHS%20Guide.pdf	Inglés	Guide for the Operational Interconnection of AMHS Systems in the SAM Region
101	ICAO_36-es	ICAO	https://www.icao.int/SAM/eDocumentsrestore/Guia%20QMS%20MET-Spa.pdf	Español	GUÍA PARA APOYAR LA IMPLANTACIÓN DEL SISTEMA DE GESTIÓN DE LA CALIDAD EN LOS SERVICIOS DE METEOROLOGÍA AERONÁUTICA (QMS/MET) EN LAS REGIONES CAR/SAM (Borrador 2011)
102	ICAO_36-en	ICAO	https://www.icao.int/SAM/eDocumentsrestore/Guia%20QMS%20MET-Eng.pdf	Inglés	GUIDE TO SUPPORT THE IMPLEMENTATION OF THE QUALITY MANAGEMENT SYSTEM IN AERONAUTICAL METEOROLOGY SERVICES (MET/QMS) IN THE CAR/SAM REGIONS (DRAFT 2011)
103	ICAO_37-es	ICAO	https://www.icao.int/SAM/eDocumentsrestore/GUIDE-TRG-HR-CNSATM-S.pdf	Español	“MANUAL GUIA PARA LA CAPACITACION DE RECURSOS HUMANOS SOBRE LOS SISTEMAS CNS/ATM”
104	ICAO_37-en	ICAO	https://www.icao.int/SAM/eDocumentsrestore/GUIDE-TRG-HR-CNSATM-E.pdf	Inglés	“GUIDANCE MANUAL FOR THE TRAINING OF HUMAN RESOURCES ON THE CNS/ATM SYSTEMS”
105	ICAO_38-es	ICAO	https://www.icao.int/SAM/eDocumentsrestore/PropEnmdFasid-S%20Comp.pdf	Español	Propuesta de enmienda al Plan de Navegación Aérea CAR/SAM (Doc 8733), Volumen II, FASID (Serie No. SAM 07/3-MET)
106	ICAO_38-en	ICAO	https://www.icao.int/SAM/eDocumentsrestore/PropEnmdFasid-E%20Comp.pdf	Inglés	Proposal for amendment to ICAO CAR/SAM Air Navigation Plan (Doc 8733), Volume II, FASID (Serial No. SAM 07/3-MET)
107	ICAO_39-es	ICAO	https://www.icao.int/SAM/eDocumentsrestore/REDDIGII%20Pol%C3%ADticaEnrutamientoSAM.pdf	Español	RED DE TELECOMUNICACIONES AERONAUTICAS REGION SAM (REDDIG II) POLITICA DE ENRUTAMIENTO PARA LA REGIÓN SAM
108	ICAO_39-en	ICAO	https://www.icao.int/SAM/eDocumentsrestore/REDDIGII%20SAMRoutingPolicy.pdf	Inglés	AERONAUTICAL TELECOMMUNICATION NETWORK OF THE SAM REGION (REDDIG II) ROUTING POLICY FOR THE SAM REGION
109	ICAO_40-es	ICAO	https://www.icao.int/SAM/eDocumentsrestore/SAMSP_sp_03Jun2019.pdf	Español	PLAN DE SEGURIDAD OPERACIONAL DE LA REGION SAM (SAMSP) Mayo 2019

110	ICAO_40-en	ICAO	https://www.icao.int/SAM/eDocumentsrestore/SAMSP_en_03Jun2019.pdf	Inglés	S A M R E G I O N SAFET Y PLAN (S A M S P) May 2019
111	ICAO_41-es	ICAO	https://www.icao.int/SAM/eDocumentsrestore/VolcanicAshContingencyPlan_S%20SAM.pdf	Español	PLAN DE CONTINGENCIA SOBRE CENIZA VOLCANICA PARA LA REGION SUDAMERICANA DE LA OACI (VACP/SAM)
112	ICAO_41-en	ICAO	https://www.icao.int/SAM/eDocumentsrestore/VolcanicAshContingencyPlan_E%20SAM.pdf	Inglés	ICAO SOUTH AMERICAN REGION VOLCANIC ASH CONTINGENCY PLAN (VACP/SAM)
113	ICAO_42-es	ICAO	https://www.icao.int/Meetings/ansconf2000/Documents/wp02s.pdf	Español	CONFERENCIA SOBRE LOS ASPECTOS ECONÓMICOS DE LOS AEROPUERTOS Y LOS SERVICIOS DE NAVEGACIÓN AÉREA (Montreal, 19 - 28 de junio de 2000) ORGANIZACIÓN DE LA CONFERENCIA
114	ICAO_42-en	ICAO	https://www.icao.int/Meetings/ansconf2000/Documents/wp02e.pdf	Inglés	CONFERENCE ON THE ECONOMICS OF AIRPORTS AND AIR NAVIGATION SERVICES (Montreal, 19 - 28 June 2000) ORGANIZATION OF THE CONFERENCE
115	ICAO_43-es	ICAO	https://www.icao.int/Meetings/ansconf2000/Documents/wp03s.pdf	Español	Situación económica de los aeropuertos, proveedores de servicios de navegación aérea y su relación financiera con los transportistas aéreos y otros usuarios
116	ICAO_43-en	ICAO	https://www.icao.int/Meetings/ansconf2000/Documents/wp03e.pdf	Inglés	Economic situation of airports, air navigation service providers and their financial relationships with air carriers and other users
117	IATA_2-es	IATA	https://www.iata.org/contentassets/c81222d96c9a4e0bb4ff6ced0126f0bb/annual-review-2013-spanish.pdf	Español	Annual Review 2013
118	IATA_2-en	IATA	https://www.iata.org/contentassets/c81222d96c9a4e0bb4ff6ced0126f0bb/iata-annual-review-2013-en.pdf	Inglés	Annual Review 2013
118	IATA_3-es	IATA	https://www.iata.org/contentassets/90f8038b0eea42069554b2f4530f49ea/dgr-62-es-appendix-h.pdf	Español	APÉNDICE H — DISPOSICIONES DEL ENFOQUE DE CAPACITACIÓN BASADO EN LA ADQUISICIÓN DE COMPETENCIAS
119	IATA_3-en	IATA	https://www.iata.org/contentassets/90f8038b0eea42069554b2f4530f49ea/dgr-62-en-appendix-h.pdf	Inglés	APPENDIX H— DANGEROUS GOODS TRAINING GUIDELINES—COMPETENCY-BASED TRAINING AND ASSESSMENT APPROACH
120	IATA_4-es	IATA	https://www.iata.org/contentassets/b08040a138dc4442a4f066e6fb99fe2a/dgr62-addendum1-es.pdf	Español	Reglamentación sobre Mercancías Peligrosas de la IATA 62a edición (español) En vigencia a partir del 1 de enero de 2021 ENMIENDA Publicación en 4 enero de 2021

121	IATA_4-en	IATA	https://www.iata.org/contentassets/b08040a138dc4442a4f066e6fb99fe2a/dgr62-addendum1-en.pdf	Inglés	IATA Dangerous Goods Regulations 62nd Edition (English) Effective 1 January 2021 ADDENDUM REVISION 1 Posted 18 January 2021
122	IATA_5-es	IATA	https://www.iata.org/contentassets/094560b4bd9844fda520e9058a0fbe2e/spa_aircraft_cleaning_guidance_covid.pdf	Español	Limpieza y Desinfección de la Aeronave durante y después de la Pandemia
123	IATA_5-en	IATA	https://www.iata.org/contentassets/094560b4bd9844fda520e9058a0fbe2e/aircraft-cleaning-guidance-covid.pdf	Inglés	Aircraft cleaning and disinfection during and post pandemic
124	IATA_6-es	IATA	https://www.iata.org/contentassets/094560b4bd9844fda520e9058a0fbe2e/spanish_ground_handling_guideline_covid.pdf	Español	Material de Orientación General para las Operaciones en Rampa durante el COVID19 Ed. 6, 12 de Febrero, 2021
125	IATA_6-en	IATA	https://www.iata.org/contentassets/094560b4bd9844fda520e9058a0fbe2e/ground_handling_guideline_covid.pdf	Inglés	Guidance for ground handling during and post COVID-19 Ed. 6, 12 Feb 2021
126	MISCELANEA_1-es	MISC	https://flightsafety.org/wp-content/uploads/2020/06/Wellbeing-ES-final.pdf	Español	Guía sobre bienestar para el profesional de la aviación
127	MISCELANEA_1-en	MISC	https://flightsafety.org/wp-content/uploads/2020/04/Guide-to-Wellbeing.pdf	Inglés	An aviation professionals guide to wellbeing
128	MO_ParteA-es	MO	Manual de Operaciones de compañía (Parte A)	Español	No disponible por motivos de confidencialidad
129	MO_ParteA-en	MO	Airline Operations Manual (Part A)	Inglés	No disponible por motivos de confidencialidad
130	MO_ParteB-es	MO	Manual de Operaciones de compañía (Parte B)	Español	No disponible por motivos de confidencialidad
131	MO_ParteB-en	MO	Airline Operations Manual (Part B)	Inglés	No disponible por motivos de confidencialidad
132	MO_ParteC-es	MO	Manual de Operaciones de compañía (Parte C)	Español	No disponible por motivos de confidencialidad
133	MO_ParteC-en	MO	Airline Operations Manual (Part C)	Inglés	No disponible por motivos de confidencialidad

ANEXO III: Textos evaluación

TEXTO CREADO PARA EVALUAR TRADUCCIONES

For the operation of aircraft with an MOPSC of more than 19, at least one cabin crew member shall be assigned when carrying one or more passenger(s).

For the purpose of complying with point (a), the minimum number of cabin crew members shall be the greatest number amongst the following:

The number of cabin crew members established during the aircraft certification process in accordance with the applicable certification specifications, for the aircraft cabin configuration used by the operator.

When scheduling cabin crew for a flight, the operator should establish procedures that take account of the experience of each cabin crew member.

The procedures should specify that the required cabin crew includes some cabin crew members who have at least 3 months experience as an operating cabin crew member.

‘Safety equipment’ means equipment installed/carried to be used during day-to-day normal operations for the safe conduct of the flight and protection of occupants (e.g. seat belts, child restraint devices, safety card, safety demonstration kit).

‘Emergency equipment’ means equipment installed/carried to be used in case of abnormal and emergency situations that demand immediate action for the safe conduct of the flight and protection of occupants, including life preservation (e.g. drop-out oxygen, crash axe, fire extinguisher, protective breathing equipment, manual release tool, slide-raft).

‘Normal procedures’ means all procedures established by the operator in the operations manual for day-to-day normal operations (e.g. pre-flight briefing of cabin crew, pre-flight checks, passenger briefing, securing of galleys and cabin, cabin surveillance during flight).

‘Emergency procedures’ means all procedures established by the operator in the operations manual for abnormal and emergency situations.

For this purpose, ‘abnormal’ refers to a situation that is not typical or usual, deviates from normal operation and may result in an emergency.

Safety and emergency equipment:

Each cabin crew member should receive realistic training on and demonstration of the location and use of all safety and emergency equipment carried, including: life jackets, infant life jackets and flotation devices, emergency lights including torches, slide rafts...

After completion of aircraft type specific training and operator conversion training on an aircraft type, each cabin crew member shall complete appropriate supervised familiarisation on the type before being assigned to operate as a member of the minimum number of cabin crew required in accordance with ORO.CC.100.

For the purpose of this Annex, the competent authority exercising oversight over operators subject to a certification or declaration obligation or specialised operation authorisation shall be for operators having their principal place of business in a Member State, the authority designated by that Member State.

The operator shall specify flight planning procedures to provide for the safe conduct of the flight based on considerations of aircraft performance, other operating limitations and relevant expected conditions on the route to be followed and at the aerodromes or operating sites concerned.

These procedures shall be included in the operations manual.

The operator shall establish and maintain dangerous goods training programmes for personnel as required by the technical instructions.

Cabin crew and technical crew use their own discretion to determine whether the situation is related to safety or security matters and whether to call the flight crew.

Situations requiring information to the flight crew may include disruptive passenger.

An operator should establish procedures to be followed by cabin crew covering at least:

Arming and disarming of slides.

Prevention and detection of cabin, galley and toilet fires.

Actions to be taken when turbulence is encountered.

The crew member shall report to the commander any fault, failure, malfunction or defect which the crew member believes may affect the airworthiness or safe operation of the aircraft including emergency systems, if not already reported by another crew member.

The commander shall be responsible for the safety of all crew members, passengers and cargo on board, as soon as the commander arrives on board the aircraft, until the commander leaves the aircraft at the end of the flight.

Any type of dangerous goods accident or incident, or the finding of undeclared or misdeclared dangerous goods should be reported, irrespective of whether the dangerous goods are contained in cargo, mail, passengers' baggage or crew baggage.

TEXTO TRADUCIDO POR TRADUCTOR HUMANO

Para la operación de aeronaves con un MOPSC superior a 19, se asignará al menos un miembro de la tripulación de cabina cuando se transporten uno o más pasajeros.

A efectos del cumplimiento del punto a), el número mínimo de miembros de la tripulación de cabina será el mayor de entre los siguientes:

El número de miembros de la tripulación de cabina establecido durante el proceso de certificación de la aeronave de acuerdo con las especificaciones de certificación aplicables, para la configuración de la cabina de la aeronave utilizada por el operador.

Al programar la tripulación de cabina para un vuelo, el operador debe establecer procedimientos que tengan en cuenta la experiencia de cada miembro de la tripulación de cabina.

Los procedimientos deben especificar que la tripulación de cabina requerida incluya algunos miembros de la tripulación de cabina que tengan, al menos 3 meses de experiencia como miembro de la tripulación de cabina operativa.

"Equipo de seguridad" es el equipo instalado/transportado para ser utilizado durante las operaciones normales del día a día para la realización segura del vuelo y la protección de los ocupantes (por ejemplo, cinturones de seguridad, dispositivos de retención infantil, tarjeta de seguridad, kit de demostración de seguridad).

"Equipo de emergencia" es el equipo instalado/transportado para ser utilizado en caso de situaciones anormales y de emergencia que exijan una acción inmediata para la realización segura del vuelo y la protección de los ocupantes, incluida la preservación de la vida (por ejemplo, oxígeno de salida, hacha de emergencia, extintor de incendios, equipo de protección respiratoria, herramienta de liberación manual (MRT), rampa-balsa).

"Procedimientos normales" son todos los procedimientos establecidos por el operador en el manual de operaciones para las operaciones normales del día a día (por ejemplo, briefing pre-vuelo de la tripulación de cabina, chequeo pre-vuelo, briefing de los pasajeros, aseguramiento de los galleys y la cabina, vigilancia de la cabina durante el vuelo).

"Procedimientos de emergencia" son todos los procedimientos establecidos por el operador en el manual de operaciones para situaciones anormales y de emergencia.

A estos efectos, "anormal" se refiere a una situación que no es típica o habitual, que se desvía del funcionamiento normal y que puede dar lugar a una emergencia.

Equipo de seguridad y emergencia:

Cada miembro de la tripulación de cabina debe recibir una formación realista sobre la ubicación y el uso de todos los equipos de seguridad y emergencia que se lleven, incluyendo: chalecos salvavidas, chalecos salvavidas para bebés y dispositivos de flotación, luces de emergencia, incluyendo linternas, rampas balsa...

Una vez completada la formación específica del tipo de aeronave y la formación de conversión del operador en un tipo de aeronave, cada miembro de la tripulación de cabina completará una familiarización supervisada adecuada en el tipo antes de ser asignado a operar como miembro del número mínimo de tripulantes de cabina requerido de acuerdo con ORO.CC.100.

A efectos del presente anexo, la autoridad competente que ejerza la supervisión de los operadores sujetos a una obligación de certificación o declaración o a una autorización de operación especializada será, para los operadores que tengan su centro de actividad principal en un Estado miembro, la autoridad designada por dicho Estado miembro.

El operador especificará los procedimientos de planificación del vuelo para garantizar la seguridad del mismo, basándose en consideraciones relativas a las prestaciones de la aeronave, a otras limitaciones operativas y a las condiciones pertinentes previstas en la ruta a seguir y en los aeródromos o lugares de operación correspondientes.

Estos procedimientos se incluirán en el manual de operaciones.

El operador establecerá y mantendrá programas de formación en materia de mercancías peligrosas para el personal, tal y como exigen las instrucciones técnicas.

La tripulación de cabina y la tripulación técnica utilizan su propio criterio para determinar si la situación está relacionada con asuntos de seguridad operacional o security y si deben llamar a la tripulación de vuelo.

Las situaciones que requieren informar a la tripulación de vuelo pueden incluir a un pasajero conflictivo.

El operador debe establecer procedimientos que debe seguir la tripulación de cabina que cubran, como mínimo:

Armado y desarmado de rampas.

Prevención y detección de incendios en la cabina, el galley y los aseos.

Acciones a tomar cuando se encuentren turbulencias.

El miembro de la tripulación informará al comandante de cualquier avería, fallo, mal funcionamiento o defecto que el miembro de la tripulación crea que puede afectar a la aeronavegabilidad o al

funcionamiento seguro de la aeronave, incluidos los sistemas de emergencia, si no ha sido ya informado por otro miembro de la tripulación.

El comandante será responsable de la seguridad de todos los miembros de la tripulación, de los pasajeros y de la carga a bordo, desde que llegue a bordo de la aeronave, hasta que abandone la aeronave al final del vuelo.

Se debe notificar cualquier tipo de accidente o incidente con mercancías peligrosas, o el hallazgo de mercancías peligrosas no declaradas o mal declaradas, independientemente de que las mercancías peligrosas estén contenidas en la carga, el correo, el equipaje de los pasajeros o el de la tripulación.

TEXTO TRADUCIDO CON NMT (MTUOC)

Para la operación de aeronaves con un MOPSC de más de 19, se asignará al menos un miembro de la tripulación de cabina cuando transporte uno o más pasajeros(s).

A efectos de cumplir con el punto , letra a), el número mínimo de tripulantes de cabina será el mayor número entre los siguientes:

El número de miembros de la tripulación de cabina establecidos durante el proceso de certificación de la aeronave de acuerdo con las especificaciones de certificación aplicables, para la configuración de cabina de avión utilizada por el operador.

Al programar la tripulación de cabina para un vuelo, el operador deberá establecer procedimientos que tengan en cuenta la experiencia de cada tripulante de cabina.

Los procedimientos deben especificar que la tripulación de cabina requerida incluye a algunos miembros de la tripulación de cabina que tienen al menos 3 meses de experiencia como miembro de la tripulación de cabina operativa.

«Equipo de seguridad»: equipos instalados/portados para ser utilizados durante las operaciones normales diarias para la realización segura del vuelo y la protección de los ocupantes (por ejemplo, cinturones de seguridad, dispositivos de retención infantil, tarjeta de seguridad, kit de demostración de seguridad).

«Equipo de emergencia»: equipo instalado/portado para usarse en caso de situaciones anormales y de emergencia que requieran una acción inmediata para la conducta segura del vuelo y la protección de los ocupantes, incluida la conservación de la vida (por ejemplo, oxígeno de gota, ejes de choque, extintora de incendios, equipo de respiración protectora, herramienta de liberación manual, rampa-balsa).

« Procedimientos normales»: todos los procedimientos establecidos por el operador en el manual de operaciones para operaciones normales diarias (por ejemplo, Briefing previo al vuelo de la tripulación de cabina, comprobaciones previas al vuelo, Briefing de pasajeros, aseguramiento de galleys y cabina, vigilancia de cabina durante el vuelo).

Procedimientos de emergencia: todos los procedimientos establecidos por el operador en el manual de operaciones para situaciones anormales y de emergencia.

Para este propósito, «abual» se refiere a una situación que no es típica o habitual, se desvíe de la operación normal y puede resultar en una emergencia.

Equipos de seguridad y emergencia:

Cada miembro de la tripulación de cabina debe recibir instrucción realista sobre y demostración de la ubicación y uso de todos los equipos de seguridad y emergencia transportados, incluyendo: manzanas de vida, manijones para la vida infantil y dispositivos para flotar, luces de emergencia incluyendo t ...

Tras completar el entrenamiento específico del tipo de aeronave y el entrenamiento de conversión del operador en un tipo de aeronave, cada miembro de la tripulación de cabina deberá completar una familiarización controlada apropiada sobre el tipo antes de ser asignado para operar como miembro del número mínimo de tripulantes de cabina requerido de acuerdo con oro.CC.100.

A efectos del presente anexo, la autoridad competente que ejerza la supervisión sobre los operadores sujetos a una obligación de certificación o declaración o autorización de operación especializada será para los operadores que tengan su principal lugar de actividad en un Estado miembro, la autoridad designada por dicho Estado miembro.

El operador especificará los procedimientos de planificación del vuelo para prever la realización segura del vuelo basándose en consideraciones de rendimiento de la aeronave, otras limitaciones operativas y las condiciones previstas pertinentes en la ruta a seguir y en los aeródromos o sitios operativos afectados.

Estos procedimientos se incluirán en el manual de operaciones.

El operador establecerá y mantendrá programas de entrenamiento de mercancías peligrosas para el personal según lo requerido por las instrucciones técnicas.

La tripulación de cabina y la tripulación técnica utilizan su propia discreción para determinar si la situación está relacionada con cuestiones de seguridad o seguridad y si llamar a la tripulación de vuelo.

Las situaciones que requieran información a la tripulación de vuelo pueden incluir pasajeros perturbadores.

Un operador debe establecer procedimientos a seguir por el personal de cabina que cubra al menos:

Armando y desarmando tobots.

Prevención y detección de incendios en cabina, galley y inodoros.

Acciones que deben tomarse cuando se encuentre turbulencia.

El tripulante informará al Comandante sobre cualquier fallo, fallo, mal funcionamiento o defecto que el tripulante considere que puede afectar a la aeronavegabilidad o a la operación segura de la aeronave, incluidos los sistemas de emergencia, si no es ya reportado por otro miembro de la tripulación.

El Comandante será responsable de la seguridad de todos los miembros de la tripulación, pasajeros y carga a bordo, tan pronto como el Comandante llegue a bordo del avión, hasta que el Comandante abandone el avión al final del vuelo.

Cualquier tipo de accidente o incidente de mercancías peligrosas, o el hallazgo de mercancías peligrosas no declaradas o no debería ser informado, independientemente de si las mercancías peligrosas están contenidas en carga, correo, equipaje de pasajeros o equipaje de tripulación.

TEXTO TRADUCIDO CON NMT (GOOGLE TRANSLATE)

Para la operación de aeronaves con un MOPSC de más de 19, se asignará al menos un miembro de la tripulación de cabina cuando lleve uno o más pasajeros.

A los efectos de cumplir con la letra a), el número mínimo de miembros de la tripulación de cabina será el mayor de los siguientes:

El número de miembros de la tripulación de cabina establecido durante el proceso de certificación de la aeronave de acuerdo con las especificaciones de certificación aplicables, para la configuración de la cabina de la aeronave utilizada por el operador.

Al programar la tripulación de cabina para un vuelo, el operador debería establecer procedimientos que tengan en cuenta la experiencia de cada miembro de la tripulación de cabina.

Los procedimientos deben especificar que la tripulación de cabina requerida incluye algunos miembros de la tripulación de cabina que tienen al menos 3 meses de experiencia como miembro de la tripulación de cabina de operación.

"Equipo de seguridad": equipo instalado / transportado para ser utilizado durante las operaciones normales del día a día para la conducción segura del vuelo y la protección de los ocupantes (por ejemplo, cinturones de seguridad, dispositivos de retención infantil, tarjeta de seguridad, equipo de demostración de seguridad).

`` Equipo de emergencia ``: equipo instalado / transportado para ser utilizado en caso de situaciones anormales y de emergencia que exigen una acción inmediata para la conducción segura del vuelo y la protección de los ocupantes, incluida la preservación de la vida (p. Ej., Suministro de oxígeno, hacha de emergencia, extintor de incendios , equipo de protección respiratoria, herramienta de liberación manual, balsa deslizante).

`` Procedimientos normales ``: todos los procedimientos establecidos por el operador en el manual de operaciones para las operaciones normales del día a día (por ejemplo, información previa al vuelo de la tripulación de cabina, controles previos al vuelo, información a los pasajeros, sujeción de las cocinas y la cabina, vigilancia de la cabina durante el vuelo). vuelo).

"Procedimientos de emergencia" significa todos los procedimientos establecidos por el operador en el manual de operaciones para situaciones anormales y de emergencia.

Para este propósito, "anormal" se refiere a una situación que no es típica o habitual, se desvía del funcionamiento normal y puede resultar en una emergencia.

Equipo de seguridad y emergencia:

Cada miembro de la tripulación de cabina debe recibir capacitación realista y demostración de la ubicación y el uso de todo el equipo de seguridad y emergencia que lleve, incluidos: chalecos salvavidas, chalecos salvavidas para bebés y dispositivos de flotación, luces de emergencia, incluidas linternas, balsas deslizantes ...

Después de completar el entrenamiento específico de tipo de aeronave y el entrenamiento de conversión del operador en un tipo de aeronave, cada miembro de la tripulación de cabina deberá completar la familiarización supervisada apropiada sobre el tipo antes de ser asignado para operar como miembro del número mínimo de tripulantes de cabina requerido de acuerdo con ORO.CC .100.

A efectos del presente anexo, la autoridad competente que ejerza la supervisión de los operadores sujetos a una obligación de certificación o declaración o una autorización de operación especializada será, para los operadores que tengan su domicilio social principal en un Estado miembro, la autoridad designada por ese Estado miembro.

El explotador especificará los procedimientos de planificación de vuelo para garantizar la conducción segura del vuelo basándose en consideraciones de rendimiento de la aeronave, otras limitaciones operativas y las condiciones esperadas pertinentes en la ruta a seguir y en los aeródromos o lugares de operación en cuestión.

Estos procedimientos se incluirán en el manual de operaciones.

El operador establecerá y mantendrá programas de capacitación sobre mercancías peligrosas para el personal según lo requieran las instrucciones técnicas.

La tripulación de cabina y la tripulación técnica usan su propia discreción para determinar si la situación está relacionada con cuestiones de seguridad o protección y si deben llamar a la tripulación de vuelo.

Las situaciones que requieren información a la tripulación de vuelo pueden incluir pasajeros molestos.

El operador debería establecer procedimientos que deberá seguir la tripulación de cabina que cubran al menos:

Armado y desarmado de toboganes.

Prevención y detección de incendios en cabinas, cocinas e inodoros.

Acciones a tomar cuando se encuentran turbulencias.

El miembro de la tripulación informará al comandante sobre cualquier falla, falla, mal funcionamiento o defecto que el miembro de la tripulación crea que puede afectar la aeronavegabilidad o el funcionamiento seguro de la aeronave, incluidos los sistemas de emergencia, si no lo ha informado ya otro miembro de la tripulación.

El comandante será responsable de la seguridad de todos los miembros de la tripulación, pasajeros y carga a bordo, tan pronto como el comandante llegue a bordo de la aeronave, hasta que el comandante abandone la aeronave al final del vuelo.

Se debe informar de cualquier tipo de accidente o incidente con mercancías peligrosas, o el hallazgo de mercancías peligrosas no declaradas o mal declaradas, independientemente de si las mercancías peligrosas están contenidas en la carga, el correo, el equipaje de los pasajeros o el equipaje de la tripulación.

TEXTO TRADUCIDO CON SMT (MTUOC)

Para la operación de aeronaves con una MOPSC de más de 19, al menos un miembro de la tripulación de cabina deberán ser asignados a la hora de llevar una o más pasajeros (s).

A efectos de cumplimiento de la letra a), el número mínimo de miembros de la tripulación de cabina deberá ser el mayor número entre los siguientes:

El número de miembros de la tripulación de cabina durante el proceso de certificación de las aeronaves en conformidad con las especificaciones de certificación, para la configuración de la cabina del avión utilizado por el operador.

En la programación de la tripulación de cabina de vuelo, el explotador debería establecer procedimientos que tener en cuenta la experiencia de cada miembro de la tripulación de cabina de pasajeros.

Los procedimientos deben especificar que la tripulación de cabina de pasajeros incluye algunos diputados que han de la tripulación de cabina de pasajeros al menos 3 meses experiencia como un miembro de la tripulación de cabina de pasajeros.

‘Safety equipment’ Significa equipos instalados / llevado a usarse durante el día a día las operaciones normales para la realización segura del vuelo y la protección de los ocupantes (por ejemplo, los cinturones de seguridad, los dispositivos de retención para niños, tarjeta de instrucciones de seguridad, botiquín de demostración de seguridad).

‘Emergency equipment’ Significa equipos instalados / llevado a emplear en caso de las situaciones anormales y de emergencia, que exigen una acción inmediata para la realización segura del vuelo y protección de ocupantes, incluida la conservación de la vida (por ejemplo, el abandono de oxígeno, accidente hacha, extintor de incendios, equipo respiratorio de protección, herramienta de liberación manual, rampa - balsa).

‘Normal procedures’ Significa todos los procedimientos establecidos por el operador en el manual de operaciones por día a día las operaciones normales (p. ej., Briefing pre - vuelo de la tripulación de cabina de pasajeros, comprobaciones pre - vuelo, pasajero Briefing, asegurar de cabina, galleys y vigilancia de la cabina durante el vuelo).

‘Emergency procedures’ Significa todos los procedimientos establecidos por el operador en el manual de operaciones para las situaciones anormales y de emergencia.

Para este propósito, 'abnormal' se refiere a una situación que no es típica o es habitual, se desvía de operación normal y puede resultar en una emergencia.

Equipos de seguridad y emergencia:

Cada miembro de la tripulación de cabina deben recibir formación sobre realista y demostración de la ubicación y uso de todos los equipos de seguridad y emergencia llevado, incluyendo: chalecos salvavidas, infantil chalecos salvavidas y de flotación de dispositivos, luces de emergencia, incluidas linternas tobogán - balsa...

Después de la terminación de tipo de aeronave y entrenamiento de conversión del operador de formación específica por tipo de aeronave, cada miembro de la tripulación de cabina deberán completar la familiarización supervisado adecuado al tipo antes de ser asignado para operar como miembro del número mínimo de tripulación de cabina requerido, de acuerdo con oro.CC.100.

Para los fines del presente anexo, la autoridad competente el ejercicio de supervisión más de operadores sujetos a una certificación o declaración obligación o la operación especializados autorización deberá ser para los operadores tenga su centro de actividad principal en un Estado miembro, la autoridad designada por dicho Estado miembro.

El operador deberá especificar los procedimientos de planificación de los vuelos para prever la realización segura del vuelo basado en consideraciones de performances de aeronaves, de otras limitaciones operativas y las condiciones previstas pertinentes a la ruta a seguirse y en los aeródromos o lugares de explotación.

Estos procedimientos se incluirá en el manual de operaciones.

El operador deberá establecer y mantener los programas de entrenamiento de mercancías peligrosas para el personal, tal como se estipula en las instrucciones técnicas.

Tripulación de cabina de pasajeros y de la tripulación técnica a utilizar su propia discreción para determinar si la situación está relacionada con la seguridad o las cuestiones de seguridad y si llamar a la tripulación de vuelo.

Situaciones que requieren información a la tripulación de vuelo pueden incluir inadecuado de los pasajeros.

El operador deberá establecer los procedimientos a seguir por la tripulación de cabina de pasajeros que abarquen al menos:

Armamento y de desarme rampas.

La prevención y la detección de la cabina, galley y inodoro incendios.

Las acciones que deben tomarse cuando se encuentre turbulencia debe extremarse.

El miembro de la tripulación deberá informar al Comandante de cualquier avería, fallo, malfuncionamiento o defecto que la opinión de miembros de la tripulación pueden afectar a la aeronavegabilidad o a la operación segura de la aeronave, incluidos los sistemas de emergencia, si no ya reportados por otro miembro de la tripulación.

El Comandante será responsable de la seguridad de todos los miembros de la tripulación, los pasajeros y la carga a bordo, tan pronto como el Comandante llega a bordo de la aeronave, el Comandante de la aeronave hasta que la abandone al final del vuelo.

Cualquier tipo de mercancías peligrosas accidente o incidente, o el hallazgo de misdeclared mercancías peligrosas no declaradas o deben ser reportados, independientemente de si las mercancías peligrosas están contenidas en la carga, correo, passengers' equipaje o equipaje de la tripulación.

TEXTO TRADUCIDO CON SMT (*YANDEX TRANSLATE*)

Para la explotación de aeronaves con un MOPSC superior a 19, se asignará al menos un miembro de la tripulación de cabina de pasajeros cuando transporte uno o más pasajeros.

A efectos del cumplimiento de la letra a), el número mínimo de miembros de la tripulación de cabina de pasajeros será el mayor de los siguientes::

El número de miembros de la tripulación de cabina de pasajeros establecido durante el proceso de certificación de la aeronave de conformidad con las especificaciones de certificación aplicables, para la configuración de cabina de la aeronave utilizada por el operador.

Al programar la tripulación de cabina de pasajeros para un vuelo, el operador debe establecer procedimientos que tengan en cuenta la experiencia de cada miembro de la tripulación de cabina de pasajeros.

Los procedimientos deben especificar que la tripulación de cabina de pasajeros requerida incluye algunos miembros de la tripulación de cabina de pasajeros que tengan al menos 3 meses de experiencia como tripulantes de cabina de pasajeros operativos.

"Equipo de seguridad": equipo instalado o transportado para ser utilizado durante las operaciones normales cotidianas para la conducción segura del vuelo y la protección de los ocupantes (por ejemplo, cinturones de seguridad, dispositivos de retención infantil, tarjeta de seguridad, kit de demostración de seguridad).

"Equipo de emergencia": equipo instalado o transportado para ser utilizado en caso de situaciones anormales y de emergencia que exijan una acción inmediata para la conducción segura del vuelo y la protección de los ocupantes, incluida la preservación de la vida (por ejemplo, oxígeno de descarga, hacha de choque, extintor de incendios, equipo de protección respiratoria, herramienta de liberación manual, balsa deslizante).

"Procedimientos normales": todos los procedimientos establecidos por el operador en el manual de operaciones para las operaciones normales cotidianas (por ejemplo, información previa al vuelo de la tripulación de cabina, comprobaciones previas al vuelo, información a los pasajeros, aseguramiento de las galeras y la cabina, vigilancia de la cabina durante el vuelo).

"Procedimientos de emergencia": todos los procedimientos establecidos por el operador en el manual de operaciones para situaciones anormales y de emergencia.

Para este propósito, "anormal" se refiere a una situación que no es típica o habitual, se desvía de la operación normal y puede resultar en una emergencia.

Equipo de seguridad y emergencia:

Cada miembro de la tripulación de cabina de pasajeros debe recibir una formación realista y una demostración de la ubicación y el uso de todos los equipos de seguridad y de emergencia transportados, incluidos: chalecos salvavidas, chalecos salvavidas para bebés y dispositivos de flotación, luces de emergencia, incluidas antorchas, balsas deslizantes...

Una vez completado el entrenamiento específico de tipo de aeronave y el entrenamiento de conversión del operador en un tipo de aeronave, cada miembro de la tripulación de cabina de pasajeros deberá completar una familiarización supervisada adecuada sobre el tipo antes de ser asignado para operar como miembro del número mínimo de tripulación de cabina de pasajeros requerido de conformidad con ORO.CC.100.

A efectos del presente anexo, la autoridad competente que ejerza la supervisión de los operadores sujetos a una obligación de certificación o declaración o a una autorización de operación especializada será, para los operadores que tengan su centro de actividad principal en un Estado miembro, la autoridad designada por dicho Estado miembro.

El operador especificará los procedimientos de planificación del vuelo para garantizar la seguridad del vuelo basándose en consideraciones de rendimiento de la aeronave, otras limitaciones operativas y las condiciones previstas pertinentes en la ruta que se vaya a seguir y en los aeródromos o lugares de operación de que se trate.

Estos procedimientos se incluirán en el manual de operaciones.

El operador establecerá y mantendrá programas de capacitación sobre mercancías peligrosas para el personal, según lo requieran las instrucciones técnicas.

La tripulación de cabina de pasajeros y la tripulación técnica utilizarán su propia discreción para determinar si la situación está relacionada con cuestiones de seguridad o protección y si deben llamar a la tripulación de vuelo.

Las situaciones que requieran información a la tripulación de vuelo pueden incluir pasajeros perturbadores.

El operador debe establecer procedimientos que debe seguir la tripulación de cabina de pasajeros que abarquen al menos:

Armado y desarmado de diapositivas.

Prevención y detección de incendios en cabinas, cocinas y baños.

Acciones a tomar cuando se encuentre turbulencia.

El miembro de la tripulación informará al comandante de cualquier fallo, fallo, mal funcionamiento o defecto que el miembro de la tripulación considere que puede afectar a la aeronavegabilidad o a la operación segura de la aeronave, incluidos los sistemas de emergencia, si no ha sido ya notificado por otro miembro de la tripulación.

El comandante será responsable de la seguridad de todos los miembros de la tripulación, pasajeros y carga a bordo, tan pronto como el comandante llegue a bordo de la aeronave, hasta que abandone la aeronave al final del vuelo.

Debe notificarse cualquier tipo de accidente o incidente de mercancías peligrosas, o el hallazgo de mercancías peligrosas no declaradas o mal declaradas, independientemente de si las mercancías peligrosas están contenidas en la carga, el correo, el equipaje de los pasajeros o el equipaje de la tripulación.