LA TRADUCCIÓN AUTOMÁTICA DENTRO DEL CONTEXTO DE UNA LENGUA MINORIZADA.

¿QUÉ TIPO DE MOTOR SE ADAPTA MEJOR AL CASO ESPECIAL DEL GALLEGO?

María do Campo Bayón

TFM

Tutora: María Pilar Sánchez Gijón

Facultat de Traducció i Interpretació, UAB, 2019.

UAB
Universitat Autònoma de Barcelona

In the end, these pieces show that the concept of minority is worth exploring because it expires innovation in translation and research.

Lawrence Venuti.

Resumen

El presente trabajo de fin de máster tiene como objetivo evaluar la percepción de adecuación de tres tipos diferentes de motores de traducción automática dentro del contexto de una lengua minorizada. El trabajo parte del análisis teórico de la relación existente entre traducción automática y lenguas minorizadas, centrándose específicamente en el par de idiomas evaluado, español-gallego. Para realizar dicha evaluación, se emplea un diseño mixto con tres métodos distintos (BLEU, encuesta y análisis de errores) para extraer datos cuantitativos y cualitativos sobre un texto de marketing del ámbito eléctrico traducido con un motor basado en reglas, un motor estadístico y un motor neuronal. Una vez realizada cada evaluación por separado, se triangulan los resultados para determinar qué motor proporciona mejores resultados. Finalmente, a partir del análisis de los datos, se extrae una serie de conclusiones que confirman o refutan las hipótesis de partida.

Palabras clave: *traducción automática, lengua minorizada, lengua con recursos reducidos, traducción automática neuronal, traducción automática estadística, traducción automática basada en reglas, evaluación de la traducción automática, errores de traducción automática*

Abstract

The aim of this master's degree Dissertation is to assess the perception of adequacy of three different types of machine translation engines within the context of minoritized languages. The Dissertation is based on the theoretical analysis of the relationship between machine translation and minoritized languages, with special focus on the assessed pair of languages, Spanish-Galician. To perform this evaluation, a mixed design with three different metrics (BLEU, survey and error analysis) is used to extract quantitative and qualitative data about a marketing text from the electric field translated with a rule-based engine, a phrase-based engine and a neuronal engine. Once each evaluation is individually conducted, the results are triangulated to determine which engine provides the best results. Finally, based on the data analysis, a number of conclusions is extracted to confirm or refute the starting hypotheses.

***Keywords**: machine translation, minority language, less-resourced language, neural machine translation, phrase-based machine translation, rule-based machine translation, MT evaluation, MT errors*

Resum

El present treball de fi de màster té com a objectiu avaluar la percepció d'adequació de tres tipus diferents de motors de traducció automàtica dins el context d'una llengua minoritzada. El treball parteix de l'anàlisi teòrica de la relació existent entre traducció automàtica i llengües minoritzades, centrant-se específicament en el parell d'idiomes avaluat, espanyol-gallec. Per realitzar aquesta avaluació, s'empra un disseny mixt amb tres mètodes diferents (BLEU, enquesta i anàlisi d'errors) per extreure dades quantitatives i qualitatives sobre un text de màrqueting de l'àmbit elèctric traduït amb un motor basat en regles, un motor estadístic i un motor neuronal. Un cop realitzada cada avaluació per separat, es triangulen els resultats per determinar quin motor proporciona millors resultats. Finalment, a partir de l'anàlisi de les dades, s'extreu una sèrie de conclusions que confirmen o refuten les hipòtesis de partida.

Paraules clau: *traducció automàtica, llengua minoritzada, llengua amb recursos reduïts, traducció automàtica neuronal, traducció automàtica estadística, traducció automàtica basada en regles, avaluació de la traducció automàtica, errors de traducció automàtica*

LISTA DE ABREVIATURAS

BLEU	*BiLingual Evaluation Understudy*
CAT	*Computer-Aided Translation*
HAMT	*Human Aided Machine Translation*
LP	*Lengua de partida*
LD	*Lengua de destino*
MAHT	*Machine Aided Human Translation*
MQM	*Multidimensional Quality Metrics*
MT	*Machine Translation*
NMT	*Neuronal Machine Translation*
PBMT	*Phrase-based Machine Translation*
RBMT	*Rule-based Machine Translation*
TA	*Traducción Automática*
TAE	*Traducción Automática Estadística*
TAO	*Traducción Asistida por Ordenador*
TAN	*Traducción Automática Neuronal*
TO	*Texto origen*
TT	*Texto traducido*

Índice de contenidos

ÍNDICE DE FIGURAS

ÍNDICE DE TABLAS

INTRODUCCIÓN

En los últimos años, estamos asistiendo a la fuerte irrupción de los motores neuronales en el mercado, pujando especialmente contra los estadísticos por asentarse y marcar una nueva era dentro de la traducción automática. La idea de este trabajo surgió a raíz de observar en la empresa en la que trabajo la necesidad de determinar qué tipo de motor era el más adecuado para ciertas combinaciones lingüísticas y por qué. Debido a que no todas las combinaciones respondían de la misma manera, se impuso la necesidad de desarrollar métodos de evaluación que permitieran discernir qué motor aplicar en cada caso concreto. Consecuentemente, esto me llevó a preguntarme cuál era el escenario actual de la traducción automática en mis dos lenguas maternas y a cuestionarme si era viable apostar por traducción automática neuronal en una combinación lingüística como la de español-gallego que cuenta con bastante bagaje y buenos resultados en los otros motores, que ya tiene establecido un flujo de trabajo y que no dispone de grandes recursos textuales.

Como hablante de la lengua gallega y como traductora español-gallego, me parecía de especial interés dotar a la investigación de la dimensión minorizada de dicha lengua, ya que condiciona por completo su realidad, incluido el mundo de la traducción. Me interesé por los recursos disponibles, por las estrategias de entrenamiento y por las iniciativas realizadas hasta el momento. Por tanto, la idea principal adquirió un nuevo prisma y se encuadró dentro de la relación dependiente entre traducción automática y lenguas minorizadas.

En resumen, debido a que el gallego como lengua minorizada tiene unas particulares especiales, mi motivación es que este trabajo abra una nueva línea de investigación sobre adecuación de motores de traducción automática en lenguas con recursos lingüísticos reducidos y qué posibilidades de mejora existen.

OBJETIVOS E HIPÓTESIS DE TRABAJO

1. Objetivos

El objetivo general de este trabajo es analizar qué tipo de motor de traducción automática se percibe como más adecuado en el contexto de una lengua minorizada como el gallego. Para alcanzar este objetivo principal se establecerán los siguientes objetivos específicos:

- Evaluar qué tipo de motor de traducción automática proporciona mejores resultados según la métrica automática BLEU.
- Evaluar qué tipo de motor de traducción automática proporciona mejores resultados según una evaluación humana (encuesta de percepción de la calidad a poseditores profesionales).
- Evaluar qué tipo de motor de traducción automática proporciona mejores resultados según una clasificación de errores (MQM).

2. Hipótesis

La pregunta de investigación subyacente sobre la que gira este trabajo es: ¿cómo se posiciona la traducción obtenida mediante un motor de TA neuronal frente a la obtenida con motores estadísticos o por reglas? Las siguientes preguntas específicas ayudarán a dar una respuesta a la pregunta principal:

- ¿Qué motor se posiciona mejor según la métrica de evaluación automática BLEU?
- ¿Cuál es el motor que perciben como más adecuado poseditores profesionales?
- ¿Qué motor arroja mejores resultados siguiendo la métrica MQM de clasificación de errores?
- ¿Se producen los mismos errores de traducción en los distintos motores?
- ¿Coincide la evaluación automática, con la humana y con los errores detectados?
- ¿Cuáles son los puntos fuertes y los puntos débiles de cada uno de los motores?
- ¿Se obtienen los mismos resultados en las distintas métricas en segmentos muy cortos y muy largos?

En lo que respecta a esta investigación, la hipótesis inicial es que la traducción obtenida mediante el motor neuronal español-gallego de Google Translator arroja un mejor resultado en su conjunto. Además de esta hipótesis inicial, se derivan dos subhipótesis más. La primera consiste en que la traducción automática neuronal provocará errores diferentes de traducción que hasta ahora no se daban en los otros motores, como una mayor presencia de falsas traducciones. La segunda consiste en que la calidad de los motores será peor en segmentos muy cortos y muy largos.

CAPÍTULO I

MARCO TEÓRICO Y ANTECENTES

En este capítulo se abordan los conceptos que constituyen el cuerpo teórico de este trabajo. En primer lugar, se ofrecerá una imagen introductoria de la traducción automática (de ahora en adelante, TA) y se presentarán los motores de TA que servirán como instrumento para dar respuesta a las preguntas planteadas. Después, se explicará qué se entiende por lenguas minorizadas y por qué se ha decidido utilizar este término deliberadamente. Asimismo, se presentará la relación existente entre ambos conceptos. Por un lado, se explicarán cuáles son las ventajas y problemas que los expertos han destacado hasta el momento con respecto a las lenguas minorizadas, y, por el otro, se citarán casos concretos de éxito aplicando TA en lenguas con recursos lingüísticos reducidos y, específicamente, se expondrá el panorama actual de la TA en lengua gallega.

Por último, se realizará un recorrido sobre la evaluación en traducción automática. Se expondrán las ventajas e inconvenientes de los distintos tipos de evaluación y se justificará la elección de las métricas elegidas para evaluar la calidad de los motores de TA.

1. La traducción automática y las lenguas minorizadas

Cualquier investigación entre traducción (en este caso, TA) y lenguas minorizadas debe asentarse en una buena definición de ambos conceptos. Los estudios en lingüística computacional y tradumática nos han dado definiciones de TA, pero no disponemos de una definición en el ámbito de la traductología para lengua minorizada, por lo que se debe recurrir a otros campos, como el de la sociolingüística. Una vez asentados los conceptos claves que se utilizarán a lo largo de este trabajo, se explicará cómo las tecnologías del lenguaje se han puesto a favor de las lenguas con menos recursos, se citarán casos concretos y se cerrará este apartado con referencias específicas dentro del marco de la lengua gallega.

1.1. La traducción automática y sus motores.

La TA nace de la constante necesidad de superar la histórica barrera del lenguaje. Gracias a los avances tecnológicos se pudieron desarrollar herramientas destinadas a la automatización del proceso de traducción. Como explica Hutchins & Somers (1992: 3), el término aceptado hoy en día en inglés es *Machine Translation* y es el nombre que reciben:

> *computerised systems responsible for the production of translations from one natural language into another, with or without human assistance.*

Inicialmente también se propusieron términos como *mechanical translation* y *automatic translation,* pero ya han sido superados por el término *MT,* si bien, este término se ha mantenido en las traducciones a otros idiomas, como en el caso del castellano. Por tanto, se opta por nombrar *Traducción Automática (TA)* al término inglés *MT*. Berner define la TA como:

> *Machine Translation (MT) is the use of computer software to translate text or speech from one natural language into another. Like translation done by humans, MT does not simply involve substituting words in one language for another, but the application complex linguistic knowledge: morphology, syntax, semantics, and understanding of concepts such as ambiguity. (2003: 7)*

Para el diseño de este trabajo, se han elegido tres motores de TA diferentes que representan los tres enfoques modernos que hoy en día se aplican a la TA. Ordenados cronológicamente, se ha seleccionado un motor de TA basado en reglas (TABR, en adelante), un motor de TA estadística (TAE, en adelante) y un motor de TA neuronal (TAN, adelante). Si bien no es objeto de este trabajo explicar las diferencias entre los distintos tipos de TA, si es de interés explicar cómo funciona el motor específico, especialmente desde una perspectiva orientada al acceso a los recursos.

1.1.1. OpenTrad Apertium

Apertium nace como resultado de aunar proyectos de subvención pública en los que han participado distintas universidades del estado español (Universitat d'Alacant, Universidade de Vigo, Universitat Politècnica de Catalunya, Euskal Herriko Unibertsitatea y Universitat Pompeu Fabra) y empresas (Eleka Ingeniaritza Linguistikoa, Imaxin Software, Elhuyar Fundazioa y Prompsit Language Engineering). La plataforma se asienta sobre los avances cosechados por el grupo de investigación

Transducens de la Universitat d'Alacant, que ya había desarrollado un sistema catalán-español denominado interNOSTRUM y del traductor portugués-español Tradutor Universia. Inicialmente, se ideó para lenguas emparentadas (como catalán–español o portugués–gallego), pero desde el lanzamiento de la versión 2 en 2006 el motor de traducción usa un sistema de transferencia más avanzado que permite manejar rasgos lingüísticos presentes en lenguas no tan emparentadas (como español–inglés). Apertium está escrito principalmente en C++ y puede compilarse y ejecutarse sobre el sistema operativo Linux, aunque también sería posible adaptarlo a otro sistema operativo. Se diseñó como un sistema de transferencia superficial, tal y como lo explican Armentano-Oller & Forcada (2006: 52):

> *The engine is a classical shallow-transfer or transformer system consisting of an 8-module assembly line;*

Armentano-Oller et al. (2007: 9) ofrecen un claro ejemplo de cómo se relacionan los distintos módulos en Apertium:

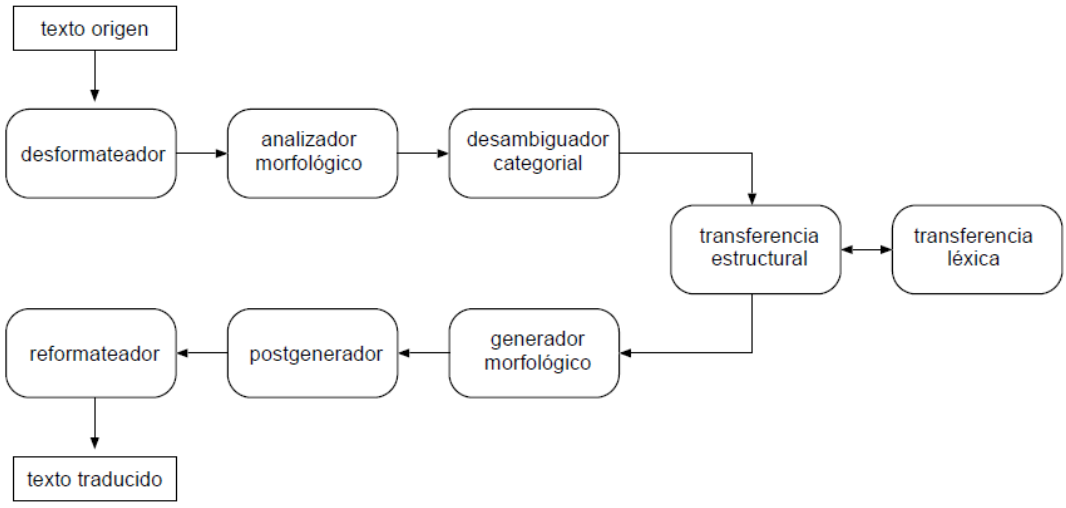


Ilustración 1 - Los ocho módulos del sistema Apertium

El desformateador aísla el texto susceptible de traducción de las etiquetas de formato (XML, HTML, etc.), encapsulándolas entre corchetes. Posteriormente, el resto de los módulos interpretará esos corchetes como si se tratasen de blancos entre palabras.

El analizador morfológico segmenta el texto en formas superficiales (unidades léxicas) y arroja para cada una de ellas una o más formas léxicas (lemas e información de flexión morfológica).

El desambiguador léxico categorial selecciona uno de los análisis que una palabra ambigua tiene definidos. Para poder saber cuál escoger en cada momento se basa en el contexto utilizando un modelo estadístico.

El módulo de transferencia léxica se encarga de gestionar un diccionario bilingüe y lo ejecuta en el módulo de transferencia estructural.

El módulo de transferencia estructural detecta y trata sintagmas que requieren un tratamiento especial debido a cambios de género y número, reordenamientos, cambios preposicionales, etc.

El generador morfológico flexiona adecuadamente una forma superficial en la lengua de destino.

El posgenerador se encarga de operaciones ortográficas en la lengua de destino, como por ejemplo las contracciones y los apóstrofos.

El reformateador devuelve las etiquetas de formato originales al texto traducido.

Una de las características principales que más adecúan Apertium para la investigación de este trabajo es el hecho de que estamos ante una plataforma de código abierto. Tanto los componentes principales del motor como los pares de lenguas y datos lingüísticos necesarios para que el sistema funcione se pueden encontrar en SouceForge bajo licencia GPL o Creative Commons 2.5. Actualmente, existe una versión operativa en el par es-gl, de la que se puede disponer y adaptar a un entorno específico, como será el caso de este trabajo.

1.1.2. ModernMT v. 2.5

ModernMT es un proyecto que recibió financiación del Programa Macro de Investigación e Innovación de la Unión Europea Horizonte 2020 para desarrollar un nuevo sistema de TA aprovechando todos los avances efectuados en este sector. Tal y como afirman sus creadores, se buscaba romper con 4 barreras:

> *MMT was designed and developed to overcome four technology barriers that have so far hindered the wide adoption of machine translation software by end-users and language service providers: (1) long training time before a MT system is ready to use; (2) difficulty to simultaneously handle multiple domains; (3) poor scalability with data and users; (4) complex installation and set-up.*(Bertoldi et al., 2017: 87)

Inicialmente, se integró la tecnología de TAE existente, principalmente basada en Moses, para hoy en día avanzar hasta ofrecer un motor de TA neuronal adaptativo. Si bien, en este trabajo se utilizará la versión 2.5 que es la última versión estadística antes de dar el salto hacia el motor neuronal (a partir de la 3.0).

La estructura de ModernMT se conforma mediante una red de nodos y cada nodo está compuesto por módulos que se relacionan entre sí. Los nodos de trabajo son:

- Gestor de etiquetas: se eliminan las etiquetas xml para poder procesarlas mejor en el motor y luego se vuelven a poner cuando se entrega la traducción.
- Gestor de expresiones numéricas: las expresiones numéricas se transforman en marcadores. Para luego poder revertirlo, cada marcador se asocia a su expresión numérica en un mapa.
- Tokenizador y detokenizador: este nodo convierte cada frase en una secuencia de tokens separados por un espacio y viceversa. Admite 45 lenguas a través de una única entrada.
- Vocabulario central: el funcionamiento interno asocia cada palabra a un ID entero gestionado por un vocabulario conjunto en el que no se fijan equivalentes tanto para la lengua original como de traducción.
- Analizador de contexto: compara el contexto de entrada proporcionado con el corpus de entrenamiento para identificar la mejor coincidencia.
- Alineador de palabras: está basado en *FastAlign*.
- Decodificador: es una versión mejorada del decodificador basado en ejemplos implementado en Moses. Se diferencia en que genera y puntúa posibles traducciones según el contexto de la frase de entrada.
- Modelo de traducción: reimplementación mejorada de la propuesta de ejemplos de Germann (2015) para Moses.

> *Its original implementation creates a phrase table at run-time by sampling sentences from the pool of word-aligned parallel data with a uniform distribution, extracting phrase pairs from them, and computing their scores on the fly. The new version provides two enhancements. First, instead of a suffix array, it relies on a DBbacked prefix index of the data pool, thus allowing for fast updates (i.e., insertions and deletions of word-aligned parallel data). Second, it keeps track of the domains from which phrase pairs are extracted and performs ranked sampling extracted phrases are ranked by their relevance (via the domain they were observed in). Translation scores*

> *are then obtained by going down the ranked list until a sufficient number of samples has been observed. Hence, by associating with all sentence pairs of each domain the corresponding weight, the TM selects and scores phrase pairs giving priority to the best-matching domain. The TM scores are the forward and backward probabilities at lexical and phrase level; the phrase level probabilities are weighted according to the domain weights.* (Bertoldi et al., 2017: 89).

- Modelo de reorganización léxica: A partir de los cómputos extraídos de las oraciones de muestra, sus correspondientes alineaciones de palabras y los cómputos generales almacenados en la BD se generan unos resultados que se emplean para determinar qué tipo de orientación se utiliza.
- Modelo de lengua: combina tanto un modelo de lengua estático con uno que se adapta al contexto.
- Gestor: controla la comunicación de todos los componentes.

Finalmente, y al igual que el modelo anterior, ModernMT también es relevante desde el punto de vista de una lengua minorizada porque también se trata de una herramienta de software libre. Se puede acceder y descargar el código desde https://github.com/ModernMT/MMT. ModernMT incluye una API REST que permite que el usuario controle cada función de MMT a través de una interfaz sencilla y potente.

1.1.3. Google Neural Machine Translation

Google Neural Machine Translation es un sistema de traducción automática neuronal desarrollado por Google. En noviembre de 2016, se lanzó el nuevo motor que consistía en añadir una red neuronal artificial al sistema existente hasta entonces Google Translate para mejorar la fluidez y la exactitud.

El modelo de arquitectura sigue el marco de aprendizaje secuencia a secuencia con atención. Se compone de tres elementos: una red de codificación, una red de decodificación y una red de atención. El codificador transforma una frase original en una lista de vectores a los que les asigna un símbolo. Siguiendo esta lista de vectores, el decodificador produce de uno en uno cada símbolo hasta que llega al símbolo especial de final de oración. Ambos módulos están conectados a través del módulo de atención que es el que le permite al decodificador centrarse en las distintas partes de la oración original mientras se está ejecutando.

En cuanto a la asignación de valores, se usan letras minúsculas en negrita para identificar los vectores, letras mayúsculas en negrita para las matrices, letras mayúsculas en cursiva para representar conjuntos, letras mayúsculas para secuencias y letras minúsculas para presentar los símbolos individuales de cada secuencia. Si, por ejemplo, *X* e *Y* representan una oración de origen y de llegada, $X = \mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_M$ sería la secuencia de símbolos M presentes en la oración original e $Y = \mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3, \ldots, \mathbf{y}_N$, la secuencia de símbolos N de la oración de llegada. Por tanto, el codificador realizaría la siguiente función: $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_M$ = codificadorRNN($\mathbf{x}_1, \ldots, \mathbf{x}_M$). En esta ecuación, $\mathbf{x}_1$, $\mathbf{x}_2$, etc. son la lista de vectores. El número de unidades de la lista se corresponde con el número de símbolos de la oración original. Así, utilizando la regla de la cadena, la posibilidad condicional de una secuencia P(Y|X) se podría formular así:

$$P(Y|X) = P(Y|\mathbf{x_1}, \mathbf{x_2}, \mathbf{x_3}, ..., \mathbf{x_M}) = \prod_{i=1}^{N} P(y_i|y_0, y_1, y_2, ..., y_{i-1}; \mathbf{x_1}, \mathbf{x_2}, \mathbf{x_3}, ..., \mathbf{x_M})$$

Ilustración 2. Fórmula probabilidad condicional de una secuencia (Wu et al., 2016:3)

Donde $\mathbf{y}_0$ es el símbolo especial de inicio de oración que se coloca en cada frase traducida.

Durante la inferencia, se calcula la probabilidad del siguiente símbolo dependiendo de la codificación de la oración original y de la secuencia traducida decodificada obtenidas hasta el momento:

$$P(y_i|y_0, y_1, y_2, y_3, ..., y_{i-1}; \mathbf{x_1}, \mathbf{x_2}, \mathbf{x_3}, ..., \mathbf{x_M})$$

Ilustración 3. Fórmula de probabilidad de GNMT (Wu et al., 2016:3)

El decodificador combina una red RNN y una capa *softmax*. La red de decodificación RNN produce un estado oculto $\mathbf{y}_i$ para el siguiente símbolo que se va a predecir, que posteriormente va a la capa *softmax* para generar una distribución de probabilidad de los símbolos candidatos de salida.

El módulo de atención funciona con una fórmula similar a la reflejada en la Ilustración 2. En concreto, $\mathbf{y}_{i-1}$ es el resultado del decodificador RNN del tiempo de decodificación anterior. El contexto de atención al del paso actual se computa de acuerdo con las siguientes fórmulas (Wu et al., 2016:4):

$$s_t = AttentionFunction(\mathbf{y}_{i-1}, \mathbf{x}_t) \quad \forall t, \quad 1 \leq t \leq M$$

$$p_t = \exp(s_t) / \sum_{t=1}^{M} \exp(s_t) \quad \forall t, \quad 1 \leq t \leq M$$

$$\mathbf{a}_i = \sum_{t=1}^{M} p_t . \mathbf{x}_t$$

Ilustración 4. Fórmula de atención de GNMT

A continuación, se puede ver un esquema de la arquitectura de GNMT, recogida en Wu et al. (2016: 4).

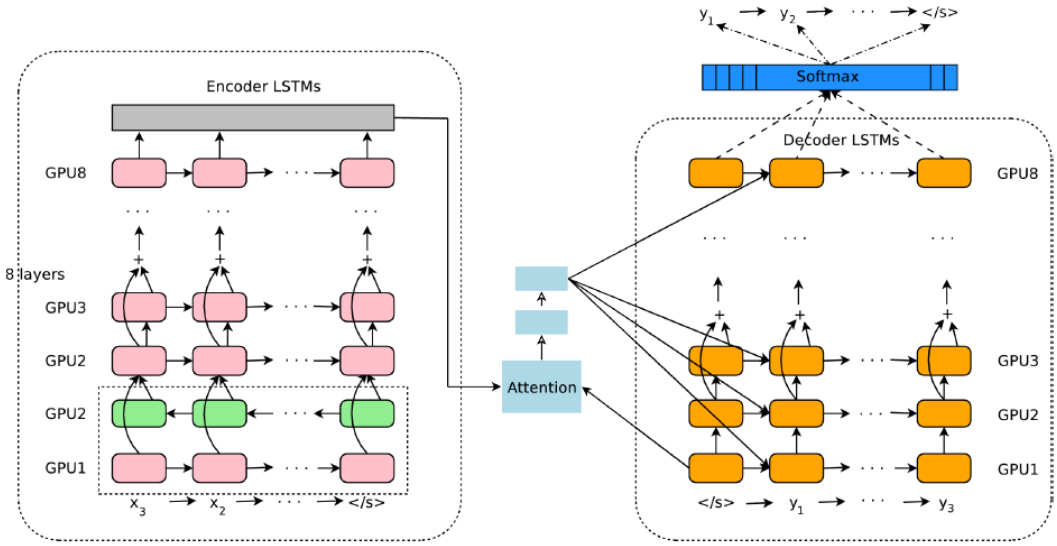


Ilustración 5. Modelo de arquitectura de GNMT

A diferencia de los anteriores, el motor de TAN de Google es propietario. Sin embargo, permite una función que resulta interesante desde el punto de vista de una lengua con recursos reducidos: el motor de TA de Google puede ser multilingüe. Dentro de las ventajas de un tipo de motor así cabe destacar la posibilidad de aplicar los mismos parámetros a varias lenguas a la vez y lo que Johnson et al. (2017: 340) denominan *zero-shot translation*. Esta opción permite crear un modelo que aprenda a traducir entre pares de lenguas, por ejemplo, si se entrena un motor multilingüe neuronal portugués > inglés e inglés > español, se podrían generar traducciones razonables para el par portugués > español. Resulta por tanto interesante en combinaciones con recursos reducidos, como puede ser el caso de la combinación inglés > gallego, para aprovechar las traducciones puente por ejemplo de inglés > español y español > gallego. Así lo explican los autores:

> *The most straight-forward approach of translating between languages where no or little parallel data is available is to use explicit bridging, meaning to translate to an intermediate language first and then to translate to the desired target language. The intermediate language is often English as xx→En and En→yy data is more readily available. The two potential disadvantages of this approach are: a) total translation time doubles, b) the potential loss of quality by translating to/from the intermediate language. (*Johnson et al., 2017: 345)

Si bien, tal y como expresa el autor, habría que tener en cuenta la pérdida potencial de calidad provocada por los posibles errores heredados de la lengua intermedia que se replicarían en las traducciones finales obtenidas.

1.2. Las lenguas minoritarias y minorizadas

Otro de los conceptos clave que rige este trabajo es definir qué se entiende por «lengua minorizada» y por qué no se emplea el término «lengua minoritaria» o «lengua en peligro de extinción». Se puede decir que en el mundo existen más de 7000 lenguas habladas y que, de estas, al menos la mitad se extinguirán en las próximas generaciones, ya que no pasan a la siguiente generación como lenguas maternas. A estas lenguas se les denomina en inglés *endangered languages* (Austin & Sallabank, 2011: 1). A su vez, Cronin entiende el concepto de «minoritario» aplicado a una lengua como una cuestión dinámica y no estática. «*"Minority" is the expression of a relation not an essence*» (Cronin, 1995: 86). Podemos encontrar una definición en la Carta europea de las lenguas regionales o minoritarias, en la que se establece como lengua minoritaria o regional las: (i) practicadas tradicionalmente sobre un territorio de un Estado por ciudadanos de ese Estado que constituyen un grupo numéricamente inferior al resto de la población del Estado; (ii) diferentes de la(s) lengua(s) oficial(es) de ese Estado; y (iii) no se incluyen ni los dialectos de la(s) lengua(s) oficial(es) del Estado ni las lenguas de los emigrantes.

Sin embargo, el término lengua minoritaria no se ajusta con precisión a, por ejemplo, las lenguas cooficiales del estado español, o en el caso particular de este trabajo, al gallego. Por esta razón, creemos conveniente optar conscientemente por el término «lengua minorizada», tal y como lo define Díaz Fouces (2005: 96):

> *1. Their territorial boundaries are often under debate. Secessionist tensions are very frequent, as in the case of Valencian vs. Catalan or Galician vs. Portuguese or, to some extent with Corsican, Faroese, Frisian, Occitan, Walloon, and others.*

> *2. The standards of their linguistic structure are not yet fully consolidated or are open to question.*
> *3. As far as the allocation of use is concerned (i.e. in what situations they can be or are used), the interference of the dominant languages is very strong. The members of the minorised linguistic community have to abide by patterns of bilingual behaviour, whereas the members of the dominant community can use their own language in all or most circumstances within the same territory.*
> *4. Users tend to associate the dominant language with higher expectations of social promotion and the minorised language with less prestigious environments, adjusting their linguistic behaviours accordingly.*
> *5. The exchange of linguistic products with other communities is usually contaminated by the dominant language acting as a filter and as a symbolic and practical border.*

En este sentido, también resulta interesante para el objetivo de este trabajo intentar delimitar cuándo se puede considerar que una lengua tiene recursos reducidos (*less-resourced language*). Para ello, tomaremos los seis niveles que establecen Alegria et al. (2011). En este artículo los autores jerarquizan las lenguas de la siguiente manera:

1. Primer nivel: inglés. Idioma que cuenta con más usuarios en Internet, con más páginas web escritas en inglés y con más recursos de tecnología del lenguaje.

2. Segundo nivel: chino, español, japonés, portugués, alemán, árabe, francés, ruso y coreano. Idiomas que se sitúan entre los 10 más empleados en Internet.

3. Tercer nivel: alrededor de 70 lenguas que cuentan con algún recurso de tecnología del lenguaje registrado.

4. Cuarto nivel: aproximadamente 300 lenguas con algún recurso léxico en línea registrado en yourdictionary.com.

5. Quinto nivel: 2014 lenguas con un sistema de escritura.

6. Sexto nivel: en este nivel se agrupan a todas las lenguas orales.

Por tanto, podemos concluir que las dos lenguas de estudio de este trabajo se sitúan entre el nivel 2 y el nivel 3, lo que supondrá unas particularidades específicas que se tratarán en el apartado 2.3.

2. La TA y su relación con las lenguas minorizadas

La relación existente entre traducción, y, por ende, TA, y una lengua minorizada es, cuando menos, ambigua, ya que al mismo tiempo que desempeña un papel crucial en el proceso de normalización lingüística y cultural, expone la debilidad de la lengua de llegada (Millán-Varela, 2000).

Como se ha podido vislumbrar en el apartado anterior, las lenguas más usadas el mundo cuentan con cada vez más productos de ingeniería del lenguaje. Sin embargo, se ha creado una considerable brecha con respecto a las lenguas menos usadas. Tal y como expresa Cronin (1995: 96), hoy en día una lengua minorizada no se mide en número de hablantes o en la existencia de una infraestructura de publicación, sino en la implantación de la lengua en el desarrollo tecnológico. La siguiente afirmación del autor amplía lo explicado más arriba sobre la ausencia de un carácter estático o permanente en una lengua minorizada.

> *The differentials mentioned above mean that a language's status is always provisional and that changes in technology, for example, can result in it becoming a minority language that is SL*[1] *intensive as it imports more and more material into the language.* (Cronin, 1995: 97-98)

Este apartado, se centra en distinguir cuáles son las técnicas más empleadas para optimizar los recursos de TA en lenguas minorizadas, cuáles son los principales problemas a los que se enfrentan y qué experiencias se han realizado en los últimos años, concretamente en la combinación español-gallego.

2.1. La tecnología del lenguaje al servicio de las lenguas minorizadas: estrategias y desafíos.

Parafraseando a Somers (1997: 8-11), dentro de las aportaciones que la tecnología del lenguaje ha hecho a las lenguas minorizadas, cabe destacar, de más simple a más compleja, el procesamiento de palabras, la separación silábica y las fuentes; la extracción de listas monolingües de palabras de textos existentes; los diccionarios y los tesauros; el uso de corpus bilingües; el desarrollo de descripciones lingüísticas y los

.

motores de traducción automática por reglas. Hoy en día, habría que completar esta lista con los motores de traducción estadística y neuronal, además, de con los avances en procesamiento del lenguaje natural (reconocimiento óptico de caracteres, herramientas de etiquetado, lematizadores, decodificadores, etc.).

Sin embargo, tal y como señalan distintos autores, son muchos los desafíos a los que se enfrentan las lenguas minorizadas en mayor o en menor medida. Keegan (2011) los agrupa en seis grandes áreas:

- Falta de personas: aquí se engloba la ausencia de personas que ayuden a crear tecnologías del lenguaje, la ausencia de hablantes de la lengua, la ausencia de conocimientos sobre la lengua, la ausencia de personas que quieran usar herramientas modernas, la ausencia de personas que quieran crear y responsabilizarse de herramientas, la ausencia de personas con conocimientos sobre cómo crear herramientas y la ausencia de personas que consideren que su deber es almacenar los recursos del lenguaje en un entorno electrónico.
- Falta de contenido: se refiere especialmente a la falta del contenido digital necesario para construir tecnologías del lenguaje.
- Falta de financiación: el diseño de tecnologías del lenguaje está supeditado a la inversión monetaria y de tiempo.
- Falta de unidad: en muchas ocasiones, las comunidades de lenguas minorizadas no se ponen de acuerdo en una única forma del lenguaje o de forma escrita. También se puede dar el caso de que no exista una autoridad específica que tome estas decisiones.
- Falta de apoyo gubernamental.

Parece, por tanto, evidente que buena parte de las estrategias empleadas pasen por maximizar el uso de herramientas que ofrecen las tecnologías del lenguaje para crear motores que puedan aumentar la producción de traducciones, especialmente, hacia las lenguas minorizadas. Existen numerosos ejemplos al respecto, aunque debido a su claridad y cercanía con las lenguas de este trabajo, resulta interesante la estrategia que desarrolló el grupo de investigación IXA de la Universidad del País Vasco para procesar una lengua minorizada como el euskera, que forzosamente debe enfocarse de manera diferente a una lengua mayoritaria como el inglés. Este grupo de investigación diseñó una estrategia de cinco niveles (Aguirre et al., 2001: 4-5).

La primera fase denominada *Laying Foundations* se centra en crear y recompilar un repositorio de corpus de textos en bruto sin etiquetar, marcadores, base de datos léxica (lemas y afijos), diccionarios legibles para máquinas, descripciones morfológicas, corpus del habla y descripción de fonemas. La segunda fase se centra en las herramientas básicas. Consiste en desarrollar herramientas estadísticas para el tratamiento de los corpus, analizadores morfológicos, etiquetadores y lematizadores; conseguir un procesamiento del habla a nivel de palabras; mejorar el corpus existente etiquetando construcciones con partes del discurso y lemas; y mejorar la base de datos léxica para incorporar partes del habla e información morfológica. La tercera fase incluye herramientas de complejidad media: creación de un entorno para la integración de la herramienta, correctores, rastreadores web, sintaxis superficial, versiones estructuradas de diccionarios, diccionario bilingüe integrado con un procesador de texto común para consultar en línea; y mejora de la base de datos léxica al añadirle unidades léxicas de múltiples palabras. La cuarta fase tiene que ver con herramientas avanzadas. En esta fase se añade texto etiquetado sintácticamente al corpus, correctores gramaticales y de estilo, integración de diccionarios en editores de texto, base de conocimiento léxico-semántico (WordNet), desambiguación de significado, procesamiento del habla a nivel de oración y sistemas de aprendizaje del lenguaje. La quinta fase denominada *Multilinguality and General Applications* (Aguirre et al., 2001:4) consiste en etiquetar semánticamente el texto después de la desambiguación, recuperación y extracción de información, ayudas de traducción, sistemas de diálogo y base de conocimiento de relaciones léxico-semánticas multilingües y sus aplicaciones.

Como podemos apreciar, buena parte de las estrategias para solventar los desafíos que suponen las lenguas minorizadas residen en recurrir al amplio abanico de tecnologías de las que disponemos para adaptarlas al contexto de cada lengua en particular. Sin embargo, tal y como afirma, Martin-Mor (2017: 377) una de las herramientas más potentes de las que puede disponer una lengua minorizada es la TA y se puede emplear tanto para propósitos de asimilación como de divulgación. De esta forma, la TA revertiría positivamente en las lenguas minorizadas gracias a su potencial contribución para divulgar la lengua. Si bien es cierto que muchas de las lenguas minorizadas solo pueden optar a un motor de TABR, ya que carecen de los recursos documentales necesarios para crear un motor TAE o TAN. Existen varios ejemplos de proyectos que intentan superar esa barrera documental, por ejemplo, combinar corpus especializados

con corpus generalizados para mejorar los resultados del motor, como en la tesis de Xu (2016). Otra posible estrategia es el proyecto descrito en Doğru et al. (2018) en el que se usaron rastreadores web y herramientas de alineación para poder crear un corpus médico de inglés a turco a partir de una web de una revista.

> *In this study, we report the automatic and semiautomatic methods we use for creating domain-specific (medical) custom translation memories as well as bilingual terminology lists which include web-crawling, document alignment in CAT tools and term extraction.* (Doğru et al., 2018:13)

También resulta interesante la combinación de estrategias empleada por Aydin & Özgür (2014: 180-192) con la que pretenden ampliar los datos de entrenamiento para un motor estadístico inglés-turco mediante corpus paralelos que no pertenecen al dominio del corpus.

> *The method first scores the sentences in the out-of-domain corpus based on their similarities to the in-domain corpus using a language modelling approach. Then, it adapts the vocabulary saturation filter technique, which has recently been proposed in (Lewis and Eetemadi, 2013) for reducing the training data and model sizes, to the domain adaptation problem. The proposed approach is applied to English-Turkish machine translation by using n-gram based as well as dependency parse tree-based language modelling and improvements in terms of BLEU scores are achieved.* (Aydin & Özgür, 2014: 181)

En la misma línea que los enfoques mencionados anteriormente está la idea de usar corpus comparables para intentar reducir la brecha documental entre pares de lenguas desconectadas (especialmente, si no pasan en ningún momento por el inglés) como en los proyectos de Vacalopoulou et al. (2012: 1-6) o Gornostay et al. (2012: 35-38). En el caso de estos últimos, la herramienta desarrollada está pensada para compilar y usar corpus comparables, por ejemplo, recabados de Internet, para evaluar y extraer propuestas de términos del corpus comparable y para alinear.

> *TTC is a three-year project and its main concept is that parallel corpora are scarce resource and comparable corpora can be exploited in the terminology extraction task. (Gornostay et al. 2014: 35)*

Por último, otra de las posibles estrategias para optimizar los recursos existentes en una lengua es reaprovechar los recursos de otra lengua próxima, como puede ser el caso del portugués y del gallego, utilizándola como lengua puente (véase apartado 2.3).

2.2. Trabajos anteriores con TA y lenguas minorizadas

Numerosos investigadores han centrado su objeto de estudio en el ámbito de la traducción automática y las lenguas minorizadas. En este sentido, cabe destacar el trabajo realizado por Bowker (2008) en el que explora la traducción automática como solución parcial para cumplir con las necesidades de traducción de las comunidades lingüísticas minoritarias de Canadá. Bowker concluye que la TA puede ser una ayuda complementaria para los traductores y una ventaja para acercar textos a las comunidades minoritarias. Asimismo, también cabe destacar los esfuerzos realizados por el Special Interest Group on Speech and Language Technology for Minority Languages (SALTMIL, en inglés) de ISCA centrado en promover la investigación y el desarrollo en tecnología del habla y del lenguaje para lenguas menos usadas, especialmente dentro de Europa. Sus objetivos principales son llevar a cabo conferencias y talleres, promover debates electrónicos a través de Internet, mantener una base de investigadores activos y proporcionar un canal de comunicación entre investigadores de lenguas minoritarias y tecnologías del lenguaje. Los talleres de 2006 en Genoa (*Strategies for developing machine translation for minority languages*) y de 2014 en Reikiavik (*Free/open-Source Language Resources for the Machine Translation of Less-Resourced Languages*) son de especial interés para este trabajo ya que se centran en la TA. También dentro del ámbito europeo, recibe una especial mención la European Association for Machine Translation (EAMT) encargada de realizar congresos sobre TA junto con sus dos organizaciones hermanas, la Association for Machine Translation in the Americas (AMTA) y la Asia-Pacific Association for Machine Translation (AAMT), que conforman la International Association for Machine Translation (IAMT). Por otro lado, dentro del mundo editorial, destaca la revista *mTm. minor Translating major - major Translating minor - minor Translating minor* centrada en promover investigaciones sobre las particularidades de la traducción de lenguas mayoritarias a minoritarias y viceversa, así como entre lenguas minoritarias.

Dentro del estado español, se han llevado a cabo varias iniciativas centradas en crear motores de TA para las lenguas cooficiales tanto en el ámbito académico como empresarial. Un buen ejemplo es el grupo de investigación IXA de la Euskal Herriko Unibertsitatea que lleva 25 años trabajando en el ámbito de tecnología del lenguaje. Además de varias herramientas específicas de NLP (véase apartado 2.1), cuenta con un motor por reglas, OpenTrad, un motor estadístico, EUSMT y actualmente, están

realizando experimentos con traducción automática neuronal de español a euskera. De la misma forma que la Universidad del País Vasco, también existen proyectos de TA centrados en el catalán. Contamos así con proyectos de traducción con Traducción Automática Estadística y Posedición (ProjectTA) de la Universidad Autónoma, centrado en investigar el proceso de traducción con TAE y proponer pautas, o el sistema por reglas interNOSTRUM español-catalán desarrollado por el grupo Transducens de la Universitat d'Alacant.

Sin duda, el esfuerzo más significativo en cuanto a TA y lenguas minorizadas es Apertium, plataforma de código abierto que incluye las herramientas y programas necesarios para construir y ejecutar sistemas de traducción automática basados en reglas (véase apartado 1.1.1).

También cabe destacar el proyecto coordinado TUNER (2016-2018), financiado por el Ministerio de Economía, Industria y Competitividad. IXA coordina el proyecto y cuenta con la participación de otros grupos de investigación como el TALN (Tractament Automàtic del Llenguatge Natural) y IULATERM (Léxico, terminología, discurso especialitzado e ingeniería lingüística) de la Universidad Pompeu Fabra; el TALG (Tecnoloxías e Aplicacións da Lingua Galega) de la Universidade de Vigo; el GRIAL, (Grupo de Investigación Interuniversitaria en Aplicaciones Lingüísticas) de la Universitat de Barcelona y de la Universitat Oberta de Catalunya; y el Grupo de Procesamiento del Lenguaje Natural y Recuperación de Información de la Universidad Nacional de Educación a Distancia (UNED). También participan investigadores de la Elhuyar Fundazioa y Vicometch. El proyecto está centrado en la investigación y el desarrollo de tecnologías de adaptación a dominio. El objetivo es usar estas tecnologías para mejorar las herramientas de NLP en distintas lenguas del estado español y diferentes dominios, es especial, salud y turismo. Entre las demostraciones que se ha hecho de estas tecnologías están UMLS Mapper, prototipo para identificar términos médicos en español y mapearlos a los conceptos del compendio de terminología médica UMLS (Pérez, Cuadros & Rigau, 2018); AsisTerm, prototipo para entender mejor términos biomédicos complejos en textos cortos (http://scientmin.taln.upf.edu/scielo); Lingaliza, página web para probar el nuevo abanico de herramientas de NLP proporcionados por IXA en gallego (http://sli.uvigo.gal/lingaliza/); Analhitza, web para acceder fácilmente al conjunto de herramientas proporcionadas por ixaKat para euskera y español (Otegi et al., 2017); un detector de unidades centrales para textos científicos;

una herramienta para procesar los artículos SEPLN (Saggion et al., 2017) y PDFdigest, herramienta para extraer contenido de archivos pdf (Saggion et al., 2017). Por último, fruto de su trabajo, se han creado recursos como el Galnet. WordNet abierto para el gallego (Gómez & Solla, 2018), el corpus SensoGal inglés-gallego (Solla & Gómez, 2017) y el corpus de registros clínicos IULA en español (Marimon, Vivaldi & Bel, 2017). (Agerri et al., 2018:163-166).

2.3. El caso especial del gallego: estado de la cuestión

Como ya hemos venido diciendo a lo largo de este marco teórico, podemos considerar el gallego como una lengua minorizada que plantea una serie de desafíos a la hora de desarrollar motores de TA debido a la falta de recursos. Existen trabajos previos centrados en la creación de recursos para el procesamiento automático de la lengua gallega. Entre ellos, cabe destacar GalNet, WordNet gallego (Gomez Guinovart & Solla Portela, 2017), SemCor gallego (Solla Portela & Gómez Guinovart, 2017), distintos trabajos en terminología (Solla Portela & Gómez Guinovart, 2015) y anotación de corpus grandes (Gomez Guinovart & López Fernández, 2009). En el ámbito del procesamiento del lenguaje natural existen herramientas para el idioma gallego como Freeling (Padro & Stanilovsky, 2012) y Linguakit (Gamallo & García, 2017). Sin embargo, tal y como pone de manifiesto Agerri et al. (2018: 2322):

> *However, it is remarkable the lack of linguistic processors that are available for other less-resourced languages, such as statistical tools for lemmatization or Named Entity Recognition.*

Teniendo esto en cuenta, desde hace unos años se han venido desarrollando distintas iniciativas para intentar dotar a la lengua gallega de estos recursos. Un buen ejemplo de esto es el motor de TA por reglas creado en 1998 por el *Centro Ramón Piñeiro para a Investigación en Humanidades* principalmente para ser usado en las instituciones públicas de Galicia y para la traducción de textos periodísticos (Diz Gamallo, 2001). Asimismo, otro motor basado en reglas, en este caso, en Apertium, es el traductor Gaio desarrollado por la Xunta de Galicia para la traducción de textos generales, jurídicos y administrativos. Otra buena iniciativa es el grupo de investigación TALG (Tecnoloxías e Aplicacións da Lingua Galega) de la Universidade de Vigo. En su página web, podemos acceder al traductor automático OpenTrad apertium español <-> galego, a la plataforma RILG (Recursos Integrados da Lingua Galega), al corrector ortográfico en línea de gallego OrtoGal, Galnet, la red léxico-semántica WordNet de gallego, a la

DBpedia del gallego, al Dicionario CLUVI inglés-galego, al Dicionario de sinónimos do galego, al Corpus Lingüístico da Universidade de Vigo (CLUVI) orientado cara la traducción del gallego, al Corpus Técnico do Galego (CTG) de orientación terminológica especializada, a la Termoteca - Banco de Datos Terminológico de la Universidade de Vigo, a la Neoteca - Banco de Datos de Neologismos de la Universidade de Vigo, al diccionario de toponimia gallega Aquén, y al etiquetador FreeLing para el análisis lingüístico automático del gallego. También es llamativo, el esfuerzo realizado por Agerri et al. (2018: 2322-2325) en el desarrollo de recursos de NLP para el gallego.

> *In this work we aim to contribute to the development of NLP resources for Galician by providing a new manually revised corpus for POS tagging and lemmatization, and a new manually annotated corpus for Named Entity Recognition. This would allow us to train previous unavailable statistical tools for the processing of Galician texts. As a result of our effort, seven new linguistic processors for Galician are presented: a rule-based tokenizer and statistical tools for POS tagging, lemmatization, Named Entity Recognition, Named Entity Disambiguation, Wikification and graph-based Word Sense Disambiguation.* (Agerri et al., 2018: 2322-2325)

Asimismo, el *Centro Ramón Piñeiro para a Investigación en Humanidades* juntamente con la E.T.S. Informática de Ourense (Universidad de Vigo) y la E.T.S. Telecomunicaciones (Universidad de Vigo) desarrollaron un motor de TA por transferencia y un motor de TA estadístico para la combinación español-gallego entrenado con un corpus de más de un millón de palabras del ámbito periodístico (Iglesias et al., 2010).

Un buen resumen de lo expuesto anteriormente es la siguiente conclusión del informe sobre el gallego en la era digital de García Mateo & Arza Rodríguez (2012: 31):

> *As táboas 10–13 mostran que, grazas a programas para o financiamento das tecnoloxías da linguaxe dos gobernos español e galego nas últimas décadas, o galego está ao nivel da maioría das outras linguas europeas. É semellante a outras linguas cun número similar de falantes, como Finlandia ou Noruega, a pesar de que estes son idiomas oficiais de estados da Unión Europea. Non obstante, os recursos e ferramentas para as tecnoloxías da linguaxe no caso do galego aínda non chegan á cobertura e calidade de recursos e ferramentas comparables no caso da lingua española, que se sitúa nunha boa posición en case todas as áreas das tecnoloxías da lingua.*

Para terminar, cabe destacar el sistema Carvalho desarrollado por Pichel Campos et al. (2009). Se trata de un proyecto que realizó la empresa imaxin|software junto con Igalia Free Software y la Universidade de Santiago. Aunque la combinación idiomática no coincide con la evaluada en este trabajo, sí resulta interesante debido a la estrategia usada para entrenar este sistema. Carvalho es un sistema de traducción estadística inglés-gallego construido a partir del corpus paralelo inglés-portugués EuroParl. Como se menciona más arriba, uno de los principales problemas a los que se enfrentan las lenguas minorizadas es la falta de corpus paralelos lo suficientemente relevantes. Ante la escasez de recursos inglés-gallego, los autores, siguiendo las teorías de Eugene Coseriu o Cunha & Cintra en las que se considera que gallego, portugués y brasileño son tres variedades del mismo sistema lingüístico, investigaron si era posible usar el corpus EUROPARL inglés-portugués para conseguir un ingenio de traducción estadística entre el inglés-gallego. Para ello, en primer lugar, convirtieron los corpus inglés-portugués a inglés-gallego usando el motor de TAR, Opentrad portugués-gallego. Las palabras no detectadas por el traductor se enviaron a un conversor ortográfico entre la grafía etimológica e histórica que usa el portugués y la grafía castellanizada del gallego. Posteriormente mediante Moses y Giza++ obtuvieron modelos de lenguaje de su prototipo. Los resultados obtenidos parecen abrir la puerta a la posibilidad de usar recursos lingüístico-computacionales del portugués para construir recursos, herramientas y aplicaciones para el gallego normativo ILG-RAG. De hecho, Google ha incorporado en los últimos años el gallego en su catálogo de recursos lingüísticos. Para ello, su motor fue entrenado con corpus paralelos inglés-portugués convertidos a la ortografía gallega.

3. Métodos de evaluación de la calidad de la traducción automática

Podemos resumir que la evaluación de TA gira generalmente en torno a valorar si es lo suficientemente aceptable como para suponer un ahorro en la productividad de un traductor humano. Existen numerosas referencias sobre qué es calidad en traducción y sobre la evaluación del resultado de la TA (Hutchins & Somers, 1992; Arnold et al., 1994; White, 2003; King, Popescu-Belis & Hovy, 2003; Coughlin, 2003; Babych & Hartley, 2004; Hamon et al. 2007; Melby et al., 2014; Fields et al., 2014; Koby et al., 2014;). La evolución en la evaluación de TA ha ido siempre a la par del desarrollo de

las tecnologías de TA. Durante los inicios de la evaluación de TA, se recurría necesariamente a evaluadores humanos. Sin embargo, el uso de jueces humanos traía consigo ciertos problemas como la subjetividad, el coste y el tiempo. En un intento de superar estas cuestiones, se desarrollaron métricas de evaluación automáticas como las que listaremos a continuación. Parece, por tanto, previsible que en un futuro cercano se evalúe la TA exclusivamente con métodos automáticos. En este apartado se introducen las claves principales y los principios básicos de la evaluación de calidad de la traducción automática. Además de listar las diferentes metodologías existentes, también se detallarán los resultados de autores que han optado por evaluar diferentes motores.

3.1. Evaluación humana

Desde su inicio, se ha contado con jueces humanos para evaluar el resultado de la TA. Aunque este tipo de evaluación se considera a menudo subjetiva o inconsistente, son, sin embargo, el estándar de oro al que aspiran las métricas automáticas. Habitualmente, la evaluación se diseña de tal forma que unos evaluadores humanos puntúan si una traducción es buena o no, según su juicio, habitualmente segmento a segmento. Los criterios de evaluación más usados son fidelidad y comprensibilidad. Fidelidad consiste en saber si el significado del original que se ha transferido al texto traducido es preciso y adecuado, sin pérdidas, adiciones o distorsiones. Por otra parte, la comprensibilidad es una característica exclusiva del texto traducido y se refiere a la facilidad con la que se entiende una traducción. Ambos criterios se puntúan según una escala. El uso de escalas en evaluación de la TA es muy habitual, como la de cinco puntos que se muestra en la Tabla 1.

Tabla 1. Escala de cinco puntos de LDC (2002)

FIDELITY	**INTELLIGIBILITY**
5 All	Flawless
4 Most	Good
3 Much	Non-native
2 Little	Disfluent
1 None	Incomprehensible

Por su parte, la clasificación de traducciones mide la preferencia humana mediante la clasificación de posibles candidatos de traducción en lugar de clasificar según una serie

de atributos de calidad. A los evaluadores se les pide que ordenen las traducciones de la mejor a la peor. También se puede dar la variante de elegir la opción de preferencia entre todos los candidatos o ninguno si no se aprecian diferencias de calidad entre las distintas opciones.

> *Ranking of translation is another method for quality assessment, but this is more common in the field of machine translation quality assessment where a number of candidate translations might be produced by one or more machine translation systems and the evaluators are asked to Rank them according to specific criteria (e.g. fluency, adequacy, compressibility).* (Saldanha & O'Brien, 2014: 102)

El rendimiento general de un motor de TA se obtiene por tanto dependiendo del número de veces que sus candidatos se sitúan mejor que los otros. Este tipo de evaluación ha sido la evaluación humana oficial de los talleres de TAE desde 2008 (Callison-Burch et al. 2008: 70-106).

Finalmente, al contrario que en los casos anteriores, en el análisis de errores se juzga la traducción desde el otro espectro, la mala calidad. Consiste en identificar errores de traducción y en estimar la cantidad de trabajo necesario para corregir una traducción en bruto hasta un estándar aceptado como una traducción (Hutchins & Somers 1992: 164). Saldanha & O'Brien (2014: 101) lo definen así:

> *Typically, the typology has a list of error types (e.g. language, terminology, style), which are sometimes categorized as major or minor, are awarded penalties and are weighted, with some errors considered more serious than others.*

Está considerado como un método más fiable porque identificar errores es más objetivo y consistente que puntuar una traducción. Otro argumento a favor es que los resultados que proporciona este modelo son de mayor utilidad y más comprensibles para las partes interesadas como desarrolladores de sistemas o usuarios (Kit & Wong, 2015: 223). Sin embargo, su dificultad reside precisamente en saber identificar y clasificar los errores, lo que lleva a cierta subjetividad por parte del evaluador y contradice una de las ventajas mencionadas anteriormente. Así lo explican Kit & Wong (2015: 223) cuando recogen las ideas de Arnold et al. (1994) y Flanagan (1994: 65-72): «*[…] what constitutes an error is a subjective matter, involving human factors like evaluators' tolerance of imperfections in a sentence and their preference of expression. Second, classifying errors into pre-defined categories is often problematic, because of the unclear boundaries of error types that are closely interlinked in nature*». También recogen esta

idea Saldanha & O'Brien (2014: 101-102): «*Even though error categories can be well-defined, it is often a matter of subjective judgement as to whether an error falls into one category or another […]. Likewise, the classification of an error as major or minor is subjective, even if definitions are provided*». Esta idea de inconsistencia y subjetividad también se encuentra reflejada en Williams (2009: 6) y en Koby et al. (2014: 415-416) cuando afirman lo siguiente:

> *Our last point of agreement concerns our conviction that both the language industry and translation studies urgently need a method to measure translation quality as objectively as possible. That method should emphasize identifying problems that can be corrected. Any effort to measure translation quality is doomed to confusion without an explicit definition of translation quality.*

3.2. Métodos automáticos

La evaluación automática de resultados de TA implica usar métricas cuantitativas sin intervención humana durante la ejecución del proceso. Está pensada para superar las deficiencias de la evaluación humana, esto es, por un lado, la subjetividad e inconsistencia, y, por el otro, el coste monetario y de tiempo. «*Automatic metrics serve thus as a desirable solution providing a quick and cost-effective means for trustable estimation of the quality of MT output*». (Kit & Wong, 2015: 225)

Dentro de las distintas métricas automáticas existentes, se puede diferenciar entre BLEU, NIST, METEOR, TER y ATEC. A continuación, describiremos las métricas más empleadas en la industria:

BLEU (Papineri et al. 2001) es la métrica más usada y se basa en la premisa de que cuanto más se acerque la traducción automática a la traducción humana, mejor es. Calcula el número de n-gramas (secuencia de palabra(s) consecutiva(s) de distinta longitud) que coinciden de la traducción en bruto con respecto a una o más traducciones de referencia. La TA debe coincidir con la traducción de referencia en la selección de palabras, en el orden de palabras y en la longitud para obtener un buen resultado. Por tanto, no se compara el resultado de la TA con el texto original, si no con una traducción de referencia. El resultado es siempre mayor o igual que cero y menor o igual que uno. Cuanto más se acerque al uno, de mejor calidad se considerará.

NIST (Doddington 2002: 138-145) nace como una revisión de la métrica BLEU. Mientras que BLEU pondera todos los n-gramas de la misma forma, NIST le da un

mayor peso a los n-gramas que son informativos. Cuanto menos se repita un n-grama en la traducción de referencia, más informativo se considerará.

TER (Snover at al. 2006: 223-231) es una métrica de evaluación basada en la cuantificación de la distancia de edición entre dos cadenas de texto. Cuanto menor sea el número de operaciones requeridas para transformar una cadena en la otra, mejor será la evaluación. Se puede usar para medir el esfuerzo de posedición necesario para convertir un candidato en referencia. TER se formula como el número mínimo de inserciones (INT), eliminaciones (DEL), sustituciones (SUB) y cambios (SHIFT) de palabras necesarios para cada candidato:

$$TER = \frac{\text{INT} + \text{DEL} + \text{SUB} + \text{SHIFT}}{\text{N}}$$

Para este trabajo, se ha optado por usar la métrica automática BLEU. Para ello, se ha recurrido al *Interactive BLEU score evaluator* del sistema TILDE, que permite introducir tanto el archivo original como el archivo correspondiente a la traducción humana y el archivo con la traducción automática.

3.3. Estudios anteriores sobre el mismo tema

Tras la irrupción en el mercado de los motores neuronales, varios investigadores han centrado sus estudios en comparar motores por reglas y estadísticos (especialmente, estos últimos) con los motores neuronales. A continuación, se resumirán las conclusiones de sus investigaciones, ya que servirán como paradigma de la investigación exploratoria que se realizará en este trabajo.

Bentivogli et al. (2016) realizaron una comparación entre TAN y TAE analizando 600 frases procedentes de transcripciones del IWSLT de Ted Talks (lenguaje hablado) traducidas del inglés al alemán. La métrica empleada fue una métrica automática sobre las traducciones poseditadas en las que se centraron en errores automáticos, léxicos y de orden de palabras. La conclusión obtenida fue que la principal ventaja de la TAN es un mejor orden de palabras, especialmente de los verbos.

Toral & Sánchez-Cartagena (2017) realizaron un análisis automático multifacético basado en traducciones humanas independientes en nueve pares de idiomas del campo de especialización de las noticias. El análisis consistió en medir el grado de coincidencia con las referencias, la fluidez, el grado de reordenación y también tres grandes clases de

errores: morfológicos, de orden y léxicos. Los principales descubrimientos confirmaron los resultados de la publicación anterior (reducción de errores morfológicos y de orden en la TAN). Ambas publicaciones coinciden en reportar una degradación de la calidad de la TAN en segmentos largos.

Popović (2017), por su parte, es el primero en comparar TAN y TAE desde el punto de vista de problemas relacionados con la lengua en ambas direcciones de la combinación inglés-alemán. La investigación arrojó dos conclusiones principales: (a) la principal ventaja de la TAN es la gestión del orden de los verbos, colocaciones inglesas, compuestos alemanes, artículos y estructura de la frase; (b) aunque la TAN en general arroja menos errores, se detectan errores complementarios en las preposiciones, nombres ambiguos ingleses y formas verbales continuas inglesas antiguas.

Por su parte, Burchardt et al. (2017) realizaron un análisis de los puntos fuertes y débiles de distintos motores de TA mediante una suite de prueba en ambas direcciones de la combinación inglés - alemán. El programa comparaba los motores con respecto a una posedición humana en distintos criterios (ambigüedad, composición, terminología, subordinación, tiempos verbales, etc.). La conclusión general es que los motores TAE obtienen un mejor resultado general. Además, también se observa que las buenas traducciones de algunos motores TAN guardan semejanza con las traducciones del sistema TAE, lo que abre la puerta a combinaciones híbridas.

Castilho et al. (2017b) comparan la calidad de sistemas de TAN y TAE en tres estudios distintos (productos de comercio electrónico, patentes y cursos MOOC) con diferentes combinaciones de idiomas empleando métricas humanas (adecuación según escala de puntos, clasificación de candidatos, análisis de errores) y automáticas (BLEU, METEOR; TER y chrF3). Los datos obtenidos en los tres estudios independientes evidenciaban que, a pesar de los buenos resultados de la TAN en las métricas automáticas, cuando se añadía a la comparación las evaluaciones humanas, los resultados no eran tan claros. En los dos primeros estudios, la TAE superaba a la TAN. En el caso de los cursos MOOC, explicado en detalle en Castilho et al. (2017b), se realizó una evaluación comparativa de un motor de TAE y un motor de TAN en cuatro pares de lenguas distintos usando la interfaz del programa PET y combinando, como se menciona más arriba, métricas automáticas y humanas (clasificación según adecuación y fluidez, análisis de errores y esfuerzo técnico y temporal de posedición). Los

resultados arrojan una preferencia por la TAN en la métrica de clasificación en todas las lenguas, textos y longitud de los segmentos. Además, se produce una mejora en la fluidez percibida y se producen menos errores. Sin embargo, los resultados no están tan claros con respecto a la adecuación, ya que la TAN concurría en un mayor número de errores de omisión, adición y falsas traducciones. Tampoco se demostró que el esfuerzo general de posedición se redujera con respecto a la TAE. La conclusión extraída por los autores es que la TAN promete ser un buen sistema de TA, pero es necesario que se invierta tiempo en mejorarla.

Por su parte, Doherty, O'Brien & Carl (2010) aportan un nuevo enfoque a la evaluación en TA combinando la evaluación humana y automática con la técnica del *eye tracking* para medir la facilidad de lectura de una traducción en bruto. La conclusión extraída es que los datos obtenidos con *eye tracking*, en especial el tiempo de visualización y el cálculo de fijación se corresponden bastante bien con la evaluación humana y automática. Por el contrario, se demostró que la duración de fijación y la dilatación de la pupila son unos indicadores menos fiables con respecto a la dificultad de lectura. Gracias a las conclusiones de esta investigación, parece prometedor el uso de esta técnica como un método automático de evaluación de TA.

Finalmente, es de especial interés para el objeto de investigación de este trabajo los enfoques de López Pereira (2018) y Sánchez-Gijón, Moorkens & Way (2019) sobre percepción de la calidad de la TA. El trabajo de investigación de López Pereira se centra en determinar la percepción, productividad y esfuerzo de posedición (tiempo y número de ediciones) de seis traductores cuando utilizan un sistema TAE y un sistema TAN. El punto de partida de la investigación es la percepción que los traductores tienen de los dos motores para posteriormente saber cuál prefieren, qué tipo de errores y problemas presenta cada sistema y cómo los resolverían. Para llevar a cabo la evaluación de un manual de instrucciones y una página web de márquetin del inglés al español utiliza las herramientas DQF (Dynamic Quality Framework). Los resultados muestran que los traductores prefieren considerablemente el motor neuronal frente al estadístico y que la TAN es más adecuada y fluida. Sánchez-Gijón et al. también analiza el grado de aceptabilidad de la TA entre un grupo de traductores profesionales de la combinación inglés-español, replicando un estudio similar realizado por Moorkens & Way, 2016. A partir de los resultados obtenidos, se concluye que los traductores están más

predispuestos a aceptar el resultado de la TA cuando los segmentos se presentan sin metadatos.

De lo expuesto anteriormente, se puede apreciar que no se puede aplicar un único sistema de evaluación de la TA, ya que todos cuentan con ventajas e inconvenientes. Por esta razón, se ha optado por medir la percepción de la calidad de distintos motores de TA mediante una encuesta realizada a poseditores profesionales (clasificación de candidatos) y completarla con un análisis de errores para dotar de una mayor información a los datos y con una evaluación automática BLEU para intentar compensar la subjetividad inherente a cualquier evaluación humana.

CAPÍTULO II

METODOLOGÍA

El presente capítulo recoge la metodología que se va a aplicar para investigar cual es el motor que mejor se adecúa a la combinación español-gallego. El objeto de estudio por tanto es la calidad percibida mediante distintas métricas de un texto especializado procesado por los motores Google Translate, OpenTrad Apertium y ModernMT v. 2.5 (variables). Para alcanzar este objetivo, se ha diseñado una investigación exploratoria, aunque con paradigma, con la finalizad de explorar las características del objeto estudiado y comprender la situación, ya que la búsqueda de antecedentes pone de manifiesto que el objeto de este TFM ha sido poco estudiado.

La primera parte consistirá en elegir el documento que se va a procesar en los tres motores distintos y que proporcionará la empresa de traducción CPSL Language Solutions de entre uno de sus encargos reales recibidos. Este documento en español tendrá alrededor de 500 palabras y terminología específica del ámbito eléctrico. A continuación, se crearán y entrenarán el sistema TABR y TAE. En el caso del motor OpenTrad, se descargará la versión estable para la combinación español-gallego, se instalará en un entorno Ubuntu y se entrenará con terminología específica del ámbito de especialización del documento elegido. Por su parte, se creará de cero un motor estadístico en la versión 2.5 de MMT y se entrenará con una memoria de traducción del ámbito de especialización y con un corpus de 6 millones de palabras del ámbito jurídico-administrativo. Por último, la empresa de traducción que colabora con este trabajo me dará acceso al motor neuronal Google Translator para la combinación español-gallego a través de la API de Google para SDL Trados Studio. Una vez se tenga acceso a todos los motores, se procesará el texto y se iniciará la siguiente parte del proyecto.

La segunda parte de la investigación se divide en varias fases. En primer lugar, se evaluarán los resultados de los tres motores aplicando la métrica automática BLEU para obtener datos cuantitativos. Seguidamente, se diseñará una encuesta orientada a obtener información cuantitativa sobre la percepción de calidad en la que participarán más de 10 poseditores profesionales de español-gallego. Una vez esté hecho, se analizarán las tres traducciones en bruto aplicando la clasificación de errores MQM para obtener datos

cualitativos. Las tres evaluaciones tienen un enfoque temporal transversal, ya que se referirán al presente y no se recogerán durante una secuencia temporal. Posteriormente, se recogerán todos los resultados, se analizarán por separado y se triangularán para obtener un estudio paralelo y convergente (diseño mixto). Los resultados recogidos con los distintos instrumentos (BLEU, encuesta, clasificación de errores, traducciones) se interpretarán en Excel mediante estadística descriptiva.

La tercera parte consistirá en repetir las tres mismas evaluaciones centrándose específicamente en segmentos muy cortos y muy largos.

1. Elección del documento origen

El documento original que se eligió finalmente para realizar las pruebas forma parte de dos encargos reales de la empresa. Se trata de dos textos de marketing con terminología propia de una empresa proveedora de servicios de electricidad y gas y también terminología del ámbito jurídico. El texto resultante es un texto en español de 594 palabras en el que se ha anonimizado la marca comercial, el nombre del producto ofrecido y la dirección web de la empresa. La elección de estos dos textos se debe en particular al marcado estilo comercial, a la presencia de frases cortas y largas y a la terminología específica del sector. También ha influido en su elección el hecho de disponer de la traducción humana de calidad con la que poder realizar la métrica BLEU. (Véase el texto original en el Anexo I: texto original

y el texto traducido y revisado por traductores profesionales en el Anexo II: texto de la traducción humana

2. Creación y entrenamiento de los motores

2.1. Apertium

Como ya se ha mencionado en el marco teórico, Apertium es un motor de traducción automática por reglas de software libre basado en la filosofía Unix. Aunque existe una versión que se puede ejecutar en Windows, debido a que se iban a realizar cambios sobre la versión lanzada para la combinación español-gallego, se optó por realizar una

instalación clásica de Apertium. Para ello, se utilizó una máquina virtual con un sistema operativo Ubuntu 18.10. Para conectarse desde Windows a esta máquina, se utilizó el programa Putty 0.71 y para el intercambio de archivos entre una máquina y otra, se utilizó el programa Samba 4.10.3.

Desde el terminal, se realizó la instalación de Apertium y del par de lenguas necesario para este trabajo mediante los siguientes comandos:

```
sudo apt-get -f install apertium
sudo apt-get -f install apertium-all-dev
sudo apt-get -f install apertium-spa
sudo apt-get -f install apertium-glg
sudo apt-get -f install apertium-es-gl
```

Figura 1. Comandos de instalación de Apertium

Una vez se instaló el motor y el par de lenguas, el siguiente paso consistió en realizar una primera prueba de traducción para ver cómo respondía el motor. Para ello, se ejecutó el siguiente comando:

```
apertium -d. -f txt es-gl ficheroorigen.txt ficherotraducido.txt
```

Figura 2. Comando de traducción de Apertium

Esta primera prueba puso en evidencia la necesidad de adaptar ciertos términos a la terminología propia del cliente y la necesidad de añadir nuevas entradas a los diccionarios. Cabe señalar que en el caso de este trabajo la estructura de los datos lingüísticos es la siguiente:

- Dos paquetes de datos monolingües: uno para la lengua de partida y otro para la lengua de llegada. En este caso, se identifican con el nombre de directorio apertium-spa y apertium-glg. Cada directorio monolingüe contiene, al menos:

a. Un diccionario monolingüe: apertium-spa.spa.dix y apertium.glg.glg.dix
b. Definición, agrupación y reglas para etiquetas morfológicas: se identifica con la extensión tsx, apertium-spa.spa.tsx y apertium-glg.glg.tsx
c. Un diccionario de reglas de posgeneración: apertium-spa.post-spa.dix o apertium-glg.post-glg.dix

- Un paquete de datos bilingües: en el caso que nos ocupa el directorio se llama apertium-es-gl e incluye:

a. Un diccionario bilingüe: apertium-es-gl.es-gl.dix

b. Reglas de transferencia para el sentido de traducción es->gl: apertium-es-gl.es-gl.t1x

c. Reglas de transferencia para el sentido de traducción gl->es: apertium-es-gl.gl-es.t1x

Por tanto, el primer paso consistió en extraer terminología propia del cliente a partir de su memoria de traducción. Mediante la función de extracción terminológica de la herramienta TAO memoQ se seleccionó una lista de posibles candidatos. Posteriormente, se fueron eliminando aquellos candidatos que producían ruido en la extracción hasta quedarse con una lista de 30 candidatos en formato Excel. A partir de esta lista final, se realizaron búsquedas en los diccionarios de Apertium para determinar si había que introducir una nueva entrada o modificar una entrada existente si ambos diccionarios monolingües contenían la entrada.

Un ejemplo de entrada modificada es el término «finca» que se puede traducir por «leira» cuando se refiere a un terreno de labranza, pero que se tiene que traducir como «predio» cuando se refiere a una propiedad inmueble. En este caso, lo más correcto sería utilizar el segundo término por lo que se modificó la entrada en el diccionario bilingüe español-gallego. Se abrió el diccionario en el editor de textos Notepad++ y se buscó la entrada para finca-leira. A continuación, se sustituyó «leira» por «predio».

```
<e><p><l>finca<s n="n"/></l><r>predio<s n="n"/></r></p></e>
```

Figura 3. Modificación de entrada en el código del diccionario bilingüe español-gallego

Se introdujeron nuevas entradas especialmente en construcciones de varias palabras que se conocen en Apertium como multipalabras: plan ahorro online, margen de comercialización fijo llano, término fijo, término variable, término de potencia, término de energía, término de energía valle, energía de suministro continuo, energía limpia, energía punta, energía activa, etc.

Por ejemplo, para plan de ahorro online, se introdujo la siguiente línea:

```
<e><p><l>plan<s n="n"/><s n="m"/><b/>de<b/>ahorro<s n="n"/><s
n="m"/><b/>online<s n="adj"/></l><r>plan<s n="n"/><s
n="m"/><b/>de<b/>aforro<s n="n"/><s n="m"/><b/>online<s
n="adj"/></r></p></e>
```

Figura 4. Modificación mulipalabras en Apertium

En este caso, era necesario utilizar la etiqueta <b/> para marcar el espacio y también la categoría gramatical y flexión de cada palabra del grupo.

Una vez se introdujeron todas las modificaciones, el siguiente paso consistió en volver a compilar los archivos para que el motor los pudiese interpretar bien. Para ello, se recurrió primero al comando ./autogen.sh para instalar los comandos necesarios y luego a make para compilar el paquete lingüístico y a sudo make install para instalarlo. Se compilaron tanto el directorio monolingüe de cada lengua como el directorio bilingüe.

Una vez ya modificado el motor a nuestras necesidades, se empleó el comando -m para que el motor pudiese utilizar la memoria de traducción del cliente como referencia. Por tanto, el comando final utilizado para traducir el documento fue:

```
apertium -d. -f txt -m memoriacliente.tmx es-gl ficheroorigen.txt
ficherotraducido.txt
```

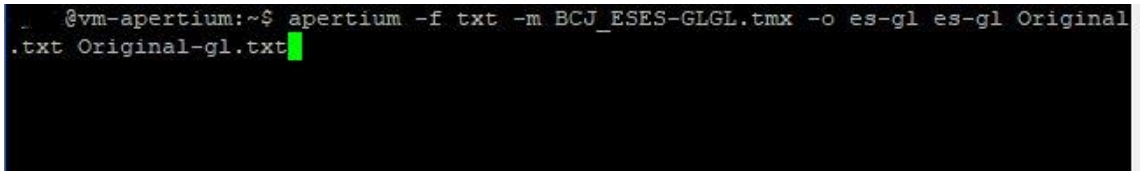


Ilustración 6. Comando de traducción en Apertium

2.2. MMT 2.5

Al igual que con Apertium, para crear el motor de MMT versión 2.5, se utilizó otra máquina virtual con sistema operativo Linux. Para acceder a la máquina virtual también se empleó el programa Putty y para el intercambio de archivos, se recurrió al programa Filezilla.

El primer paso para crear el motor estadístico de MMT para la combinación español-gallego fue preparar la memoria del cliente y los corpus recompilados con los que entrenar el motor. La memoria del cliente estaba en formato propietario de SDL Trados Studio, .sdltm, por lo que primero se utilizó ese mismo programa para exportarla a formato .tmx. Aunque MMT puede procesar tmx, el formato habitual para entrenar el

motor es un corpus bilingüe, es decir, dos corpus monolingües con el mismo número de líneas para que el contenido se alinee automáticamente. Por esta razón, una vez se obtuvo el formato .tmx se utilizó el programa libre Olifant para realizar una serie de pasos con el objetivo de arreglar ciertas cuestiones que pudiesen causar problemas a la hora de entrenar el motor:

- Saltos de línea: una memoria tmx puede contener distintos tipos de saltos de línea en las oraciones: saltos tipo Windows (CRLF o \r\n mediante una expresión regular), Linux (LF o \n) o incluso el carácter Unicode de separador de línea . Como el contenido de las líneas de cada corpus debe coincidir, se deben quitar estos saltos de línea. Por lo tanto, se reemplazó la expresión regular \n por un espacio.
- Eliminación de segmentos vacíos
- Aunque MMT tiene un nodo para procesar etiquetas, lo más seguro es quitar cualquier etiqueta de la memoria para facilitar la interpretación de las cadenas y su alineación.
- Comprobación de posibles problemas como contenido en otras lenguas, segmentos con una diferencia considerable de tamaño entre el original y la traducción, segmentos enteros con texto no traducible, etc.
- Arreglo de espacios al principio ("^ +") y final del segmento (" +$").
- Arreglo de múltiples espacios (" +").
- Eliminación de los atributos de la tmx para reducir el peso de la memoria.

Cuando la memoria estuvo arreglada, se utilizó el programa Rainbow de OkapiTools para exportar la memoria en corpus bilingüe:

- Se seleccionó el par de lenguas deseado en la pestaña *Languages and Encodings*.
- Desde *Utilities* > *Conversion Utilities* > *File Format Conversion*, se seleccionó Parallel Corpus Files.
- La codificación del corpus debe ser UTF8 o UTF8-BOM.
- Los saltos de línea en el corpus deben ser de tipo Linux por lo que se utilizó el programa Notepad++ para cambiarlos, ya que si se hacía desde Rainbow los corrompía.

Una vez lista la memoria, el siguiente paso consistió en buscar corpus que sirvieran como base para entrenar al motor. Una buena página en la que encontrar recursos es http://opus.nlpl.eu/. Sin embargo, si se busca el par español-gallego, el resultado que se obtiene es el siguiente:

corpus	doc's	sent's	es tokens	gl tokens	XCES/XML	raw	TMX	Moses	mono	raw	ud	alg	dic	freq			other files
GNOME v1	1702	0.9M	4.5M	4.5M	xces es gl	es gl	tmx	moses	es gl	es gl		alg smt		es gl		sample	
OpenSubtitles v2018	292	0.2M	1.8M	1.7M	xces es gl	es gl	tmx	moses	es gl	es gl		alg smt	dic	es gl	query	sample	xces/alt
KDE4 v2	1432	0.2M	1.4M	1.4M	xces es gl	es gl	tmx	moses	es gl	es gl		alg smt	dic	es gl	query	sample	
OpenSubtitles v2016	235	0.2M	1.4M	1.3M	xces es gl	es gl	tmx	moses	es gl	es gl		alg		es gl		sample	
Tatoeba v2	1	2.8k	1.7M	30.7k	xces es gl	es gl	tmx	moses	es gl	es gl				es gl	query	sample	
Ubuntu v14.10	412	0.1M	0.6M	0.5M	xces es gl	es gl	tmx	moses	es gl	es gl		alg smt	dic	es gl		sample	
OpenSubtitles v2011	15	5.5k	42.7k	39.2k	xces es gl	es gl										sample	
OpenSubtitles v2012	1	0.1k	0.9k	0.8k	xces es gl	es gl				es gl				es gl		sample	
total	**4090**	**1.6M**	**11.5M**	**9.5M**	**1.6M**		**1.6M**	**1.6M**									

Ilustración 7. Resultados obtenidos en Opus para la combinación es-gl

Los corpus son o bien de software o bien de subtítulos y no están marcados con un color verde oscuro que se corresponde con un buen nivel de fiabilidad. Por esta razón, se recurrió al Corpus Lingüístico da Universidade de Vigo, CLUVI. El CLUVI contiene 23 millones de palabras de seis corpus paralelos principales de cinco áreas de especialidad (ficción, software, divulgación científica, derecho y administración) en cinco combinaciones idiomáticas diferentes (gallego-español, inglés-gallego, francés-gallego, inglés-gallego-español-francés, español-gallego-catalán-euskera). El corpus que más se adaptaba al tipo de texto evaluado en este trabajo es el corpus administrativo de español-gallego de seis millones de palabras y que se puede descargar bajo licencia Creative Commons[2]. El corpus se descarga en formato .xml, por lo que fue necesario realizar los mismos pasos que con la memoria de traducción: arreglo de problemas y exportación a corpus bilingüe en formato Moses.

Una vez que ya se tenían todos los materiales preparados para entrenar el motor, el siguiente paso fue conectarse a la máquina virtual y crear el motor. Para ello, se ejecutaron los siguientes comandos:

```
cd /mmt
```
Para acceder a la carpeta en donde están almacenados los archivos de MMT.

```
./mmt status
```
Como solo se puede tener un único motor en marcha, primero es necesario comprobar el estado de los otros motores, para saber si hay que apagar alguno.

```
./mmt stop -e [NOMBRE DEL MOTOR]
```

[2] Enlace de descarga del corpus: https://repositori.upf.edu/handle/10230/20051

Para parar el motor que estuviese encendido en ese momento.

```
./mmt create es-ES gl-ES -e [NOMBRE DEL MOTOR] [Ruta que siempre tiene que empezar por corpus/é]
```

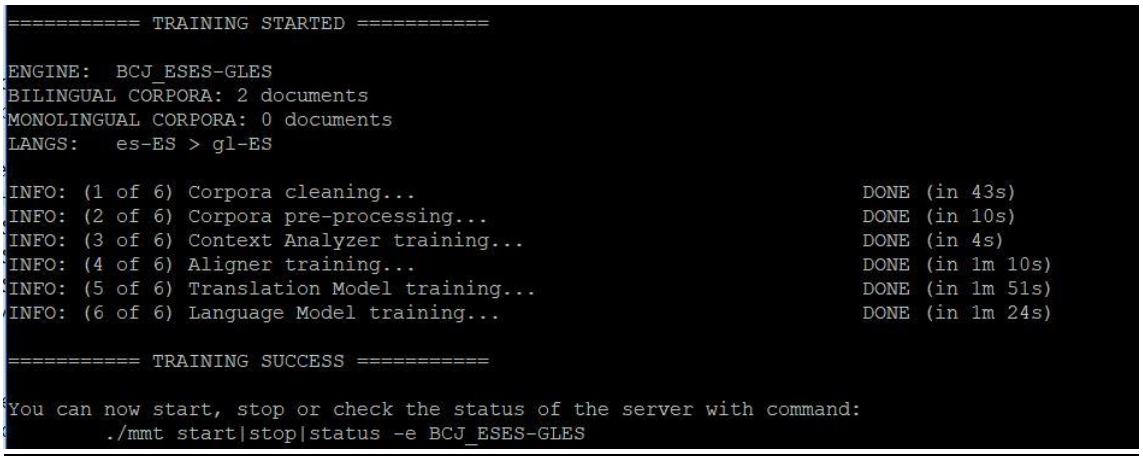


Ilustración 8. Entrenamiento del motor MMT

Una vez que el motor ya está creado, hay que iniciar el motor y ejecutar la orden de traducción:

```
./mmt start -e [NOMBRE DEL MOTOR]
```

Para iniciar un motor.

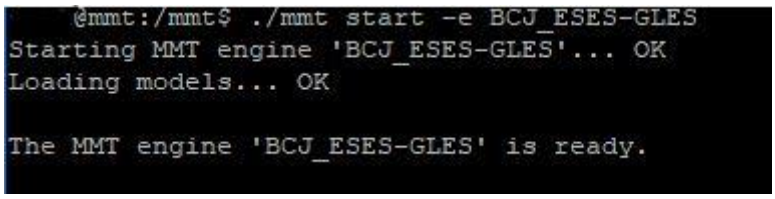


Ilustración 9. Comando para iniciar un motor en MMT

```
./mmt translate -e [NOMBRE DEL MOTOR] --batch --txt < [ruta: files/in/Ficherorigen.txt] > [path: files/out/Ficherotraducido.txt]
```

Para ejecutar el comando de traducción.

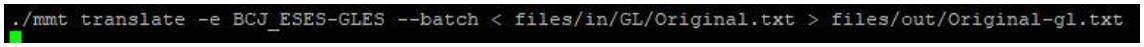


Ilustración 10. Comando de traducción en MMT

2.3. Google Translator

Como ya se ha mencionado en el capítulo anterior, Google Translator es un servicio de pago de la empresa Google. Para poder acceder a su motor neuronal, es necesario contar con una cuenta de usuario y una API Key.

El proceso de entrenamiento del motor neuronal en este caso lo realiza la empresa propietaria a partir de los datos generados a diario, por lo que no se realizó ningún cambio al motor existente. El proceso de traducción en el motor es diferente a los anteriores ya que no se recurrió a ningún comando, sino que se empleó el programa SDL Trados Studio y se lanzó la traducción desde su interfaz. Para ello, primero se creó un proyecto nuevo de traducción y en las opciones de memoria se seleccionó *Use* > *Google Cloud Translation API*. Cuando seleccionas esta opción, te sale una ventana emergente en la que tienes que introducir la API Key y seleccionar si quieres usar un motor estadístico o neuronal. En nuestro caso, se seleccionó el motor neuronal.

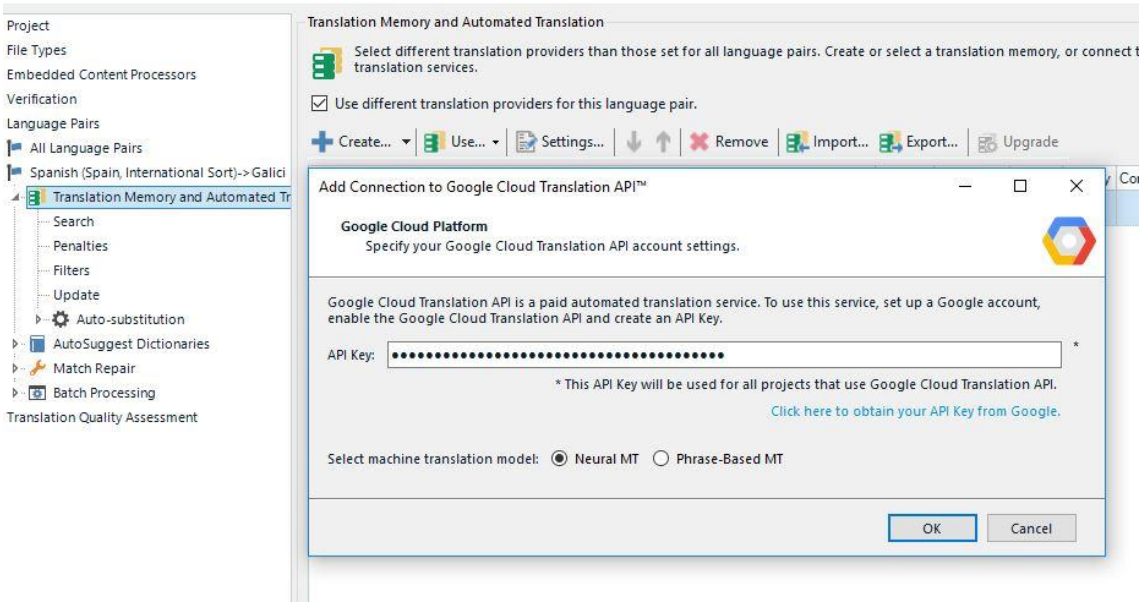


Ilustración 11. Google Cloud Translation API en SDL Trados Studio

Al hacer clic en OK, en la pestaña de memorias se agrega el motor de Google:

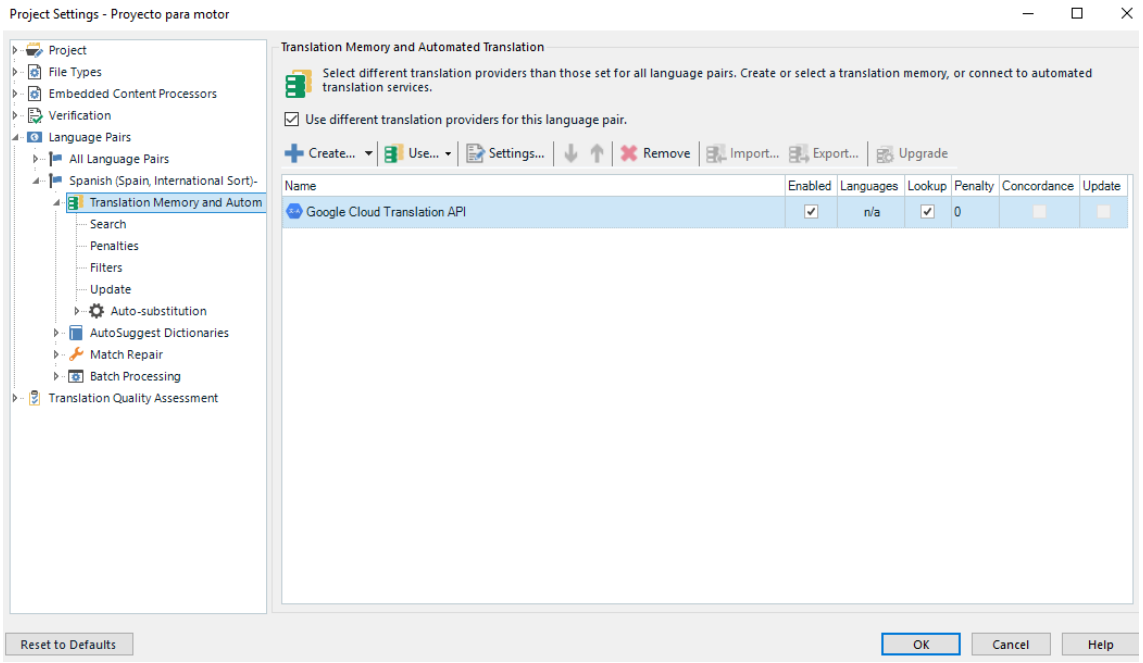


Ilustración 12. Pestaña de memorias de traducción de SDL Trados Studio

Una vez que se ha seleccionado el motor de traducción, el siguiente paso consistió en realizar una tarea por lotes para preparar el archivo y pretraducirlo con el motor neuronal de Google. En este paso, es importante marcar la opción de *Apply automated translation* de los ajustes de Pretraducción.

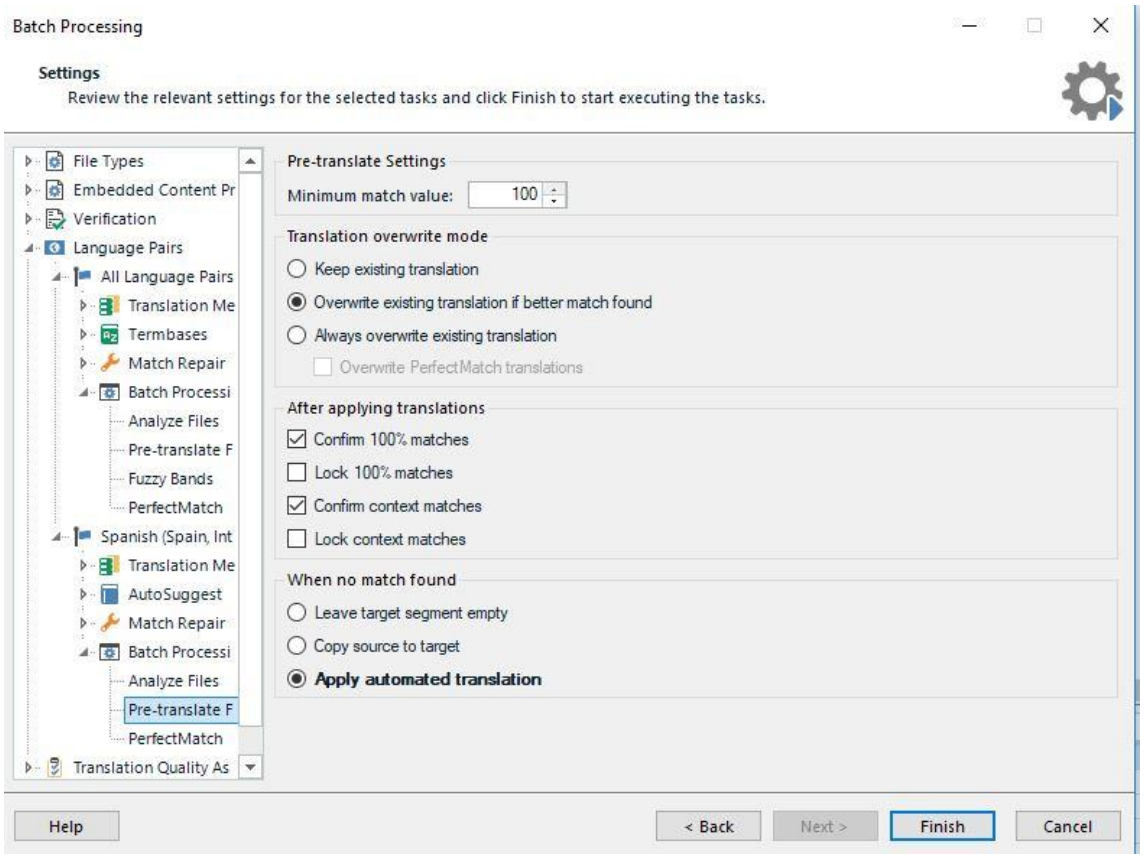


Ilustración 13. Ajustes de pretraducción en SDL Trados Studio

Una vez que el proceso termina, cuando se abre el archivo en el modo Editor, se puede ver que todos los segmentos aparecen marcados con el símbolo AT.

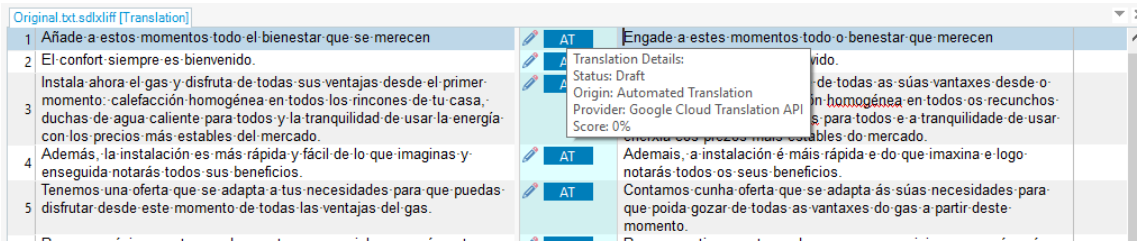


Ilustración 14. Modo vista Editor en SDL Trados Studio

Tras comprobar que el archivo se ha pretraducido entero con el motor de Google, el siguiente paso fue generar las traducciones de destino para obtener el archivo original txt traducido automáticamente en gallego.

3. Preparación de las métricas

3.1. BLEU

Para obtener datos de evaluación siguiendo una métrica BLEU, se usó la herramienta Interactive Bleu de Tilde. Esta herramienta permite realizar una comparación entre el resultado del motor de traducción y la traducción humana de referencia. También permite añadir un segundo motor a la comparación. Sin embargo, como no es posible añadir un tercer motor, se optó por extraer resultados por separado y luego compararlos fuera de la herramienta. Asimismo, la herramienta realiza un análisis frase a frase, especialmente relevante para realizar la evaluación de los segmentos muy largos y muy cortos.

La interfaz es muy sencilla. Basta con adjuntar en formato txt el resultado de la TA y la traducción humana de referencia para obtener los resultados.

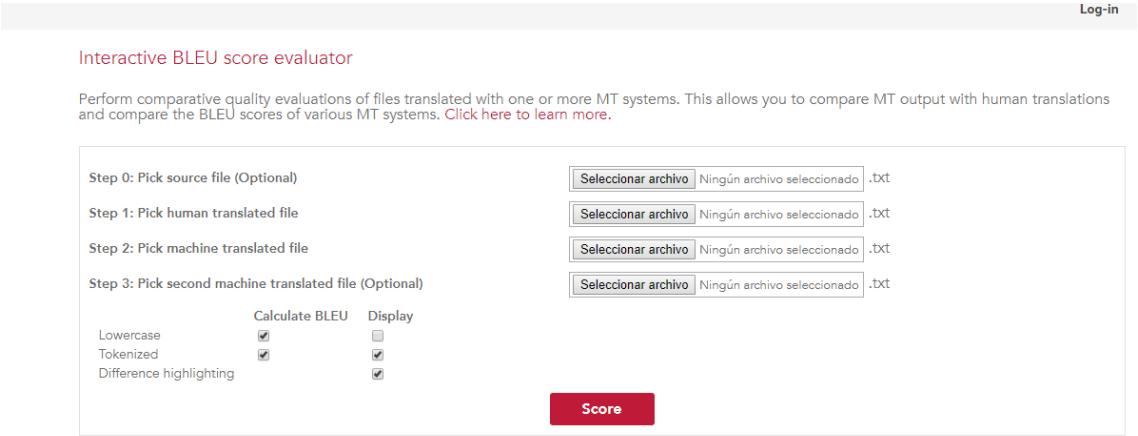


Ilustración 15. Interactive BLEU score evaluator

Por tanto, en primer lugar, se realizó una evaluación separada por cada motor evaluando el texto entero y, en segundo lugar, se extrajo la información individual de los segmentos muy largos y muy cortos para poder realizar la segunda parte de este trabajo.

3.2. Encuesta

La encuesta se preparó con la herramienta Google Forms y se mantuvo abierta a respuestas del 13 de marzo al 13 de abril. El objetivo de la encuesta era evaluar la percepción de calidad que tenían los poseditores profesionales de la combinación español-gallego sobre la traducción en bruto obtenida mediante los tres motores

diferentes. La encuesta se dividió en 4 grandes secciones. En la primera sección se explica el motivo de realizar la encuesta, se establece el tiempo orientativo para realizarla (no más de 30 minutos), se informa de su carácter anónimo y se pide el consentimiento expreso para participar en ella. La siguiente sección consiste en recoger datos geográficos sobre los participantes para poder situar su lugar de procedencia. A continuación, la siguiente sección consiste en la clasificación de los resultados de la TA según el criterio del participante. Para no extender el tiempo de la encuesta demasiado y perder la atención del participante, se seleccionaron 14 segmentos de los 30 de los que se compone el texto original. Se excluyeron de la evaluación aquellos segmentos que eran iguales entre los motores o que el resultado era totalmente inadecuado. Dentro de los 14 segmentos seleccionados, se incluyeron segmentos muy largos y muy cortos para poder analizar específicamente este tipo de segmentos. La pregunta se estructuró de la siguiente forma:

1. Segmento de partida[3].
2. Resultado de TA 1.
3. Pregunta de respuesta cerrada sobre si aprovecharía este resultado de TA para poseditar.
4. Resultado de TA2.
5. Pregunta de respuesta cerrada sobre si aprovecharía este resultado de TA para poseditar.
6. Resultado de TA3.
7. Pregunta de respuesta cerrada sobre si aprovecharía este resultado de TA para poseditar.
8. Casilla de verificación para ordenar según preferencia del 1 al 3 el mejor resultado de traducción.

[3] El segmento original se estructura como forma de pregunta en la encuesta, por lo que deja un espacio para respuesta que se ha marcado como no obligatorio y en el que no se espera respuesta.

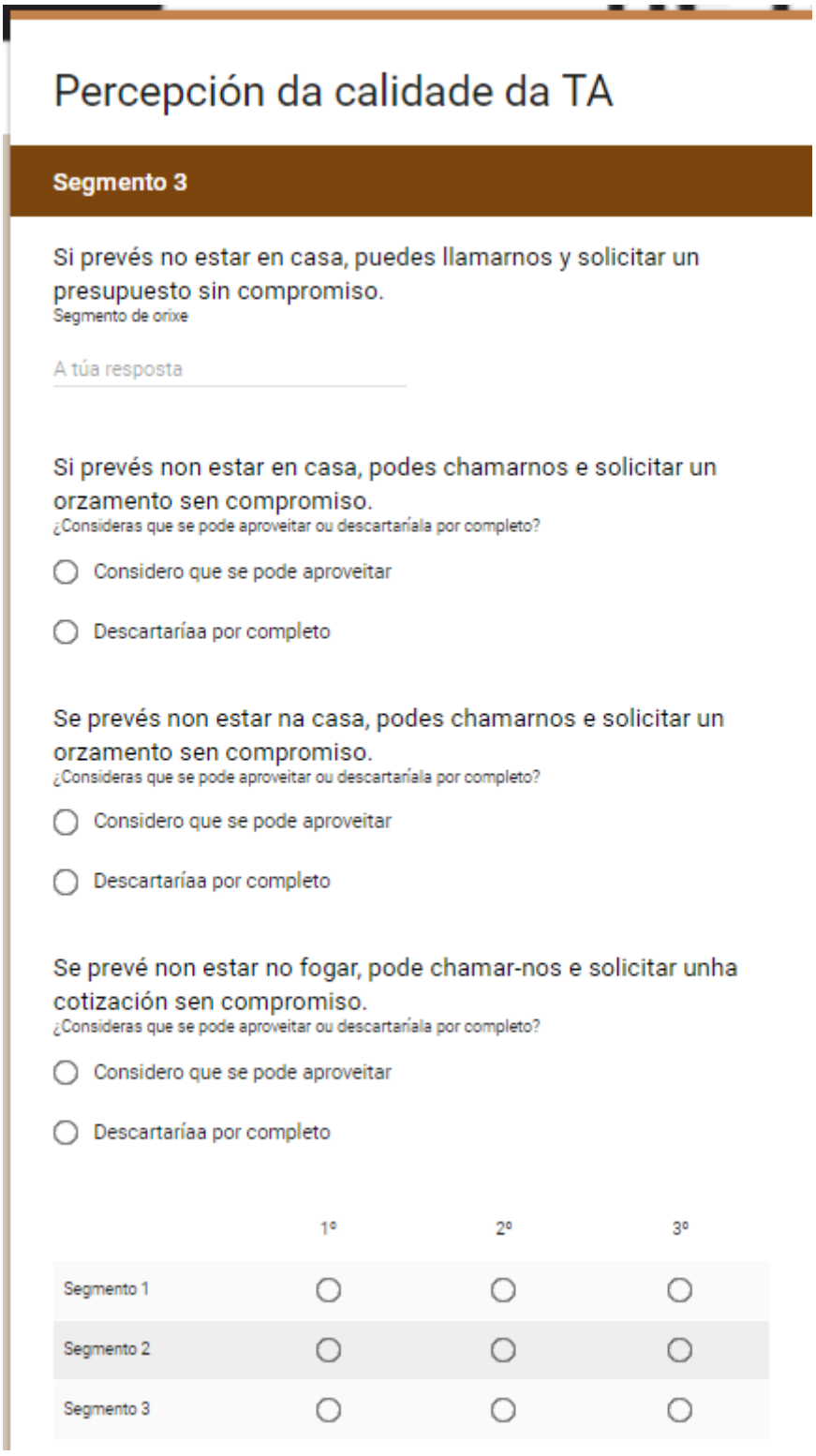


Ilustración 16. Ejemplo de pregunta de la encuesta

Cada pregunta se dividió en una subsección para fomentar la concentración en cada uno de los segmentos y no saturar al participante con los otros segmentos.

La última sección se compone de preguntas de respuesta cerrada destinadas a extraer datos sobre la actividad profesional del participante y su percepción general de la TA para dotar a los resultados de la clasificación anterior de un mayor contexto. Se pidió a los participantes que respondieran sobre si estaban o no familiarizados con el campo de especialización del texto, si traducen de forma habitual el tipo de texto evaluado, si están habituados a trabajar con TA (según una escala Liker 5 siendo el 1 Nunca y el 5 Siempre). En cuanto a la percepción de la TA; las preguntas estaban destinadas a saber si consideraban que la TA agilizaba su trabajo, si consideraban la TA como un riesgo profesional y si consideraban la TA como un recurso de traducción como lo puede ser las memorias de traducción y los glosarios.

3.2.1. Elección de los participantes

Para seleccionar a los participantes de la encuesta, se recurrió a la base de datos de traductores profesionales de la empresa de traducción que colaboró con el proceso de investigación de este trabajo. Dentro de su sistema propio de gestión de proyectos, se realizó una búsqueda filtrando por la combinación español-gallego y si ofrecían servicio de posedición. La lista resultante contenía 28 recursos a los que se les envió la encuesta.

Además de esta lista de recursos, se realizó una búsqueda en el portal Proz.com bajo los mismos criterios anteriores y se añadieron 41 recursos más a la lista existente.

Finalmente, se envió la encuesta a 69 poseditores profesionales de la combinación español-gallego, de los que finalmente participaron 15.

3.3. MQM

Multidimensional Quality Metrics (MQM) es un marco para la evaluación y clasificación de errores de traducción desarrollado para el proyecto financiado por la UE, QTLaunchPad por DFKI y sus socios. MQM no ofrece una métrica de evaluación de traducción fija, sino que se trata de un marco flexible para definir métricas específicas para una tarea de traducción empleando una terminología y tipos de errores comunes. El conjunto de tipos de errores de traducción se ha ido actualizando y, actualmente, se divide en 10 categorías generales (precisión, fluidez, estilo, terminología, compatibilidad, diseño, internacionalización, convenciones locales, veracidad y otros) y 100 subcategorías. Por tanto, no hablamos propiamente de métrica MQM, sino de métrica conforme con MQM (Burchardt & Lommel, 2014: 4).

La guía de MQM recomienda, en primer lugar, diseñar una métrica que vaya acorde con el tipo de texto evaluado y con el propósito de la investigación. También debe ser lo suficientemente granular como para arrojar resultados específicos a la investigación y al mismo tiempo poder tenerlo en la cabeza a la hora de realizar la evaluación. Por lo tanto, en nuestro caso, se descartaron aquellos errores que no fuesen relevantes para el trabajo, como por ejemplo las categorías que tenían que ver con internacionalización (errores relacionados con la preparación del contenido original para su traducción y localización), compatibilidad (categoría que ya no se recomienda usar en el marco y que tenía que ver con plazos de entrega, funcionalidad del software, etc.), convenciones locales (no existen cuestiones relacionas con la adecuación de los números en el texto evaluado), veracidad (problemas relacionados con la adecuación del contenido traducido al público destinatario) y diseño (errores relacionados con la presentación del documento). Además de que algunas de las categorías anteriores no se pueden aplicar al texto seleccionado, el motivo principal para descartarlas fue la necesidad de centrarse en aspectos lingüísticos para poder extraer conclusiones sobre si los motores eran correctos desde un punto de vista léxico y gramatical. Así, las categorías generales seleccionadas fueron precisión, fluidez, estilo y terminología. Dentro de la categoría precisión se eliminó la subcategoría relativa a la coincidencia exacta con la TM no adecuada ya que no se ha empleado ninguna TM para pretraducir el archivo. En cuando a la categoría fluidez, se han eliminado las siguientes subcategorías: contenido (errores relacionados con la presentación de la información), guía de estilo, codificación de caracteres, caracteres no permitidos, problemas relacionados con el patrón, ordenación, referencias cruzadas o enlaces rotos y tabla de contenidos/índice. Asimismo, también se ha eliminado de la categoría de estilo, la subcategoría relativa al estilo del cliente porque no se dispone de ella. Finalmente, tanto en la categoría de estilo como en la categoría de terminología, se han eliminado las subcategorías relacionadas con una tercera parte, ya que solo se dispone de la terminología del cliente.

Esta fue la lista final de tipos y subtipos de errores:

Tabla 2. Lista de tipos y subtipos de errores basada en MQM

Accuracy
Addition
Mistranslation
Omission
Untranslated
Fluency
Ambiguity
Unintelligible
Spelling
Typography
Grammar
Word-form
Part-of-speech
Agreement
Tense-mood-aspect
Word-order
Function-words
Terminology
Inconsistent with termbase
Inconsistent with domain
Style
Register
Awkward
Unidiomatic

Una vez que la métrica está diseñada, se recomienda elaborar un diagrama de decisiones con los tipos de errores que se deben aplicar con el objetivo de facilitar la evaluación y ayudar a distinguir las diferencias entre cada uno de los errores. Se puede consultar el diagrama adaptado de Burchardt & Lommel (2014: 17) y Mercader-Alarcón & Sánchez-Martínez (2016: 184) en el Anexo III: diagrama de decisiones (MQM).

Los autores de la guía emplean para anotar cada error la herramienta TAO libre Translate5. Debido a que el volumen de segmentos evaluados en este trabajo es inferior al empleado por Burchardt & Lommel (2014), la anotación se realizó a mano.

CAPÍTULO III

ANÁLISIS DE LOS DATOS

1. Datos de la evaluación automática

La métrica automática que se ha elegido para evaluar los tres motores de traducción automática es BLEU. La razón de su elección se debe a que ya se contaba con la traducción realizada por un traductor humano y revisada por otro traductor humano. Dicha traducción ha pasado los controles de calidad de la empresa y ha sido entregada al cliente, por lo que se ha tomado como correcta. Este método de evaluación toma como referencia la traducción humana, por tanto, la TA debe coincidir con la traducción de referencia tanto en las palabras seleccionadas como en el orden y en la longitud. En cada análisis, halla la precisión de n-gramas entre la traducción en bruto y la traducción humana segmento a segmento introduciendo ciertas penalizaciones para intentar corregir el resultado. La media geométrica que se emplea es la siguiente:

$$BLEU = PB \cdot exp\left(\sum_{n=1}^{N} w_n \log P_n\right)$$

Ilustración 17. Media geométrica de BLEU

El resultado final siempre es mayor o igual que cero y menor o igual que uno. Cuanto más cerca se sitúe del uno, mejor estará catalogada la TA.

1.1. Datos de la evaluación conjunta

La puntuación final obtenida a partir del análisis de los 30 segmentos de los que se compone el documento evaluado es la siguiente:

Tabla 3. Resultados BLEU globales

	Neural Google Translate	MMT v 2.5	Apertium
BLEU	43,71	75,27	76,39
Precision x brevity	44,76 x 97,64	76,23 x 98,75	77,73 x 98,28

Como se puede apreciar en la tabla comparativa, el motor que mejor resultados arroja es Apertium, aunque le sigue de cerca MMT v2.5. Por el contrario, el motor neuronal de Google obtiene un resultado bastante inferior a los otros dos. Habitualmente, se considera que la métrica es favorable si supera el 50 por ciento, por lo que este último motor no pasaría la prueba. Por otra parte, se establece que, si el resultado se sitúa entre el 60 y el 80, hay garantías para recomendar el uso de la TA. Si bien, es poco habitual que se den resultados de BLEU tan altos. Una posible explicación de este fenómeno podría ser la combinación lingüística, ya que hablamos de lenguas muy próximas que comparten estructuras y que benefician el tipo de análisis realizado con BLEU. Otra posible justificación la da el género textual, ya que nos encontramos ante un género muy estandarizado y homogéneo, con pocas probabilidades de variabilidad. Por otra parte, buena parte de los errores se han neutralizado gracias al entrenamiento previo.

Por todo lo expuesto anteriormente y tomando como referencia los resultados expuestos en la tabla anterior, se podría recomendar los motores de Apertium y de MMT v2.5. Profundizando en el análisis de cada motor, Apertium ha traducido 9 de los 30 segmentos exactamente igual que la traducción de referencia (resultado 100). Asimismo, la mayoría de los segmentos se sitúan entre el 89 y 50. Finalmente, únicamente 4 segmentos no superan el 50.

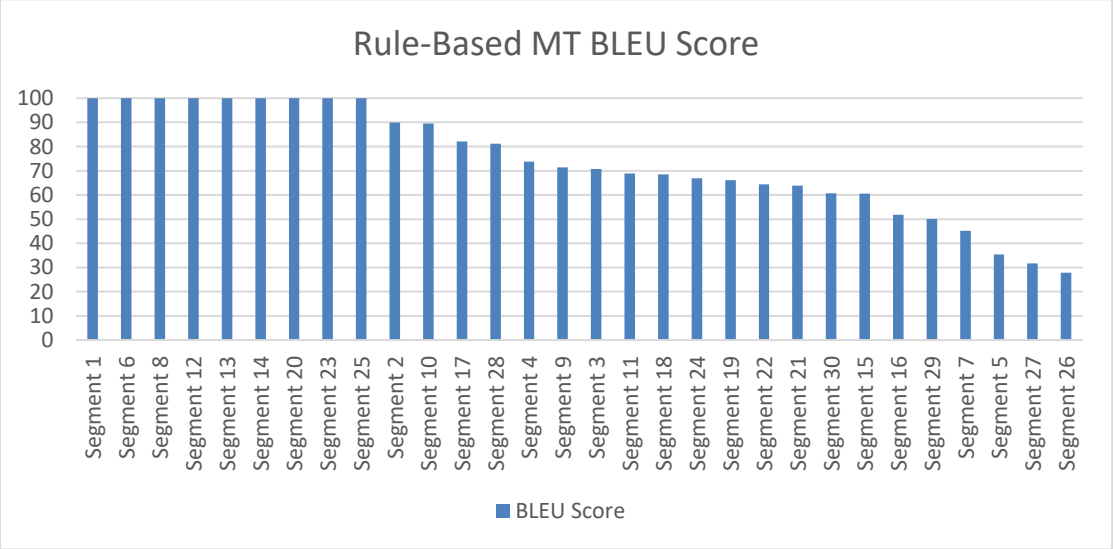


Ilustración 18 - Resultados BLEU para el motor por reglas

En el caso de MMT v2.5, el resultado es bastante similar. Encontramos 5 segmentos traducidos exactamente igual que la traducción de referencia. Por otra parte, 19

segmentos oscilan entre el 89 y el 60 y, finalmente, 6 segmentos se sitúan por debajo de 50.

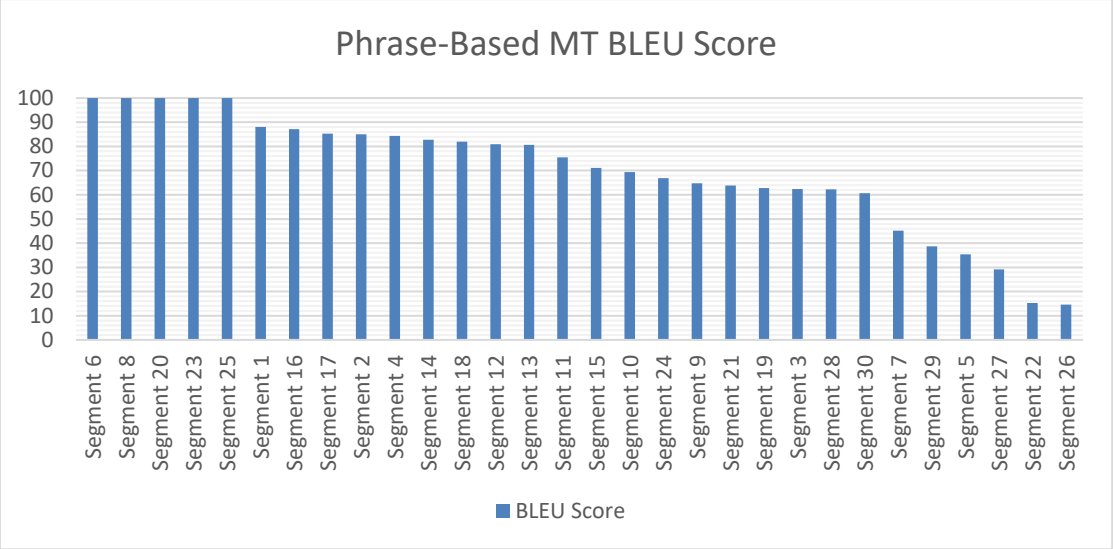


Ilustración 19. Resultados BLEU para el motor estadístico

El escenario cambia totalmente si observamos los resultados de Google Translator, el segmento que mejor resultado obtiene una puntuación de 87. No hay ningún segmento que coincide por completo con la traducción de referencia. Asimismo, encontramos 10 segmentos entre el 79 y el 57. Como podemos apreciar, la mayoría de los segmentos (19) se encuentra entre el 48 y el 7.

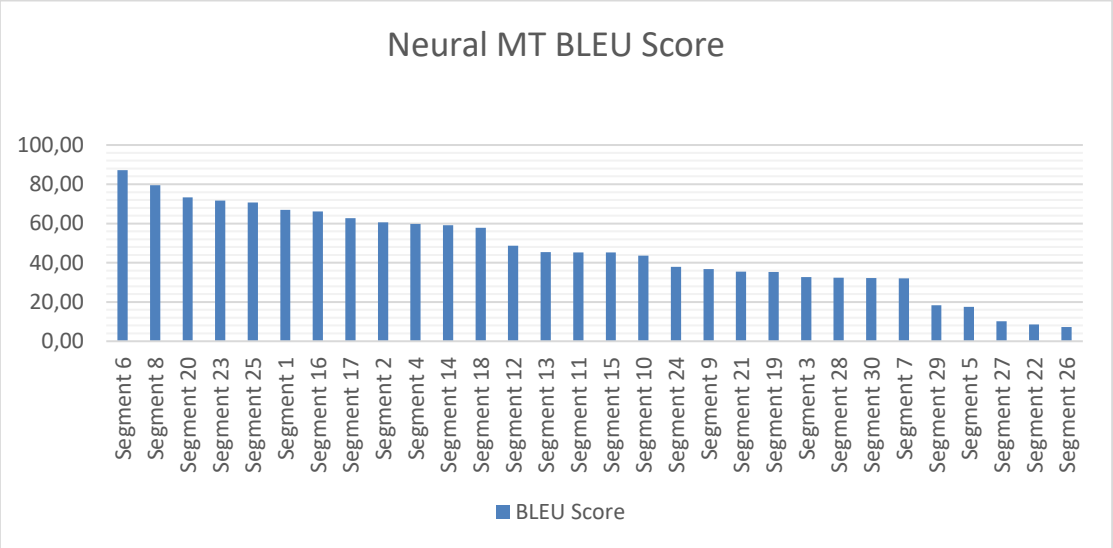


Ilustración 20. Resultados BLEU para el motor neuronal

1.2. Datos de los segmentos muy cortos

De los 30 segmentos evaluados, 10 de ellos tienen menos de cinco palabras por segmento. Estos son los resultados obtenidos en estos segmentos para cada uno de los motores:

Tabla 4. Resultados BLEU para segmentos muy cortos

	Neural Google Translate	MMT v 2,5	Apertium
Segmento 5	21,02	35,36	35,36
Segmento 6	35,36	100	100
Segmento 7	37,99	45,18	45,18
Segmento 20	100	100	100
Segmento 21	63,89	63,89	63,89
Segmento 22	64,32	15,21	64,32
Segmento 23	100	100	100
Segmento 24	66,87	66,87	66,87
Segmento 25	100	100	100
Segmento 30	60,65	60,65	60,65

En general, se mantiene la tónica descrita en el apartado anterior. Apertium obtiene un resultado favorable en 8 de los 10 segmentos, MMT v2.5 en 7 y Google en 6. Debido a la naturaleza de los segmentos cortos, nos encontramos con una relativa homogeneidad en los índices BLEU, por ejemplo, MMT y Apertium coinciden al 100% con la traducción de referencia y salvo en el segmento 22 obtienen el mismo porcentaje. Esto hará que, como veremos más adelante en los apartados 2 y 3, coincidan en el tipo de error cometido y en la valoración sobre si se aprovechan o no.

1.3. Datos de los segmentos muy largos

El texto contiene 5 segmentos de más de 30 palabras por segmento. Al igual que en el caso anterior, se ha prestado especial atención a estos segmentos y se han extraído los resultados en una tabla independiente.

Tabla 5. Resultados BLEU para segmentos muy largos

	Neural Google Translate	MMT v 2.5	Apertium
Segmento 2	59,04	85,01	89,88
Segmento 4	32,44	84,38	73,75

Segmento 17	43,67	85,34	82,16
Segmento 18	45,44	81,93	68,42
Segmento 19	32,64	62,77	66,05

Esta vez el motor que obtiene mejores resultados es MMT v 2.5, aunque no supera en mucho a Apertium. Ambos motores obtienen un resultado favorable en todos los segmentos muy largos (rango 60-80). Sin embargo, Google Translate solo tiene un segmento con una puntuación de 62. Los otros 4 segmentos oscilan del 45 al 32, por lo que es el motor que peores resultados obtiene en estos segmentos. De la misma forma que en los segmentos muy cortos, un menor índice BLEU está relacionado con una mayor presencia de errores y con una peor calificación por parte de los poseditores.

2. Datos de la encuesta

Los datos cuantitativos recogidos en la encuesta se estructuran en tres secciones diferenciadas. A continuación, se presentarán los resultados obtenidos de las valoraciones de los participantes en ella.

2.1. Datos geográficos

Con el objetivo de analizar la distribución geográfica de los participantes, se les pidió que indicaran su ciudad de residencia. Como se puede apreciar en la siguiente gráfica, mayoritariamente se encuentran en las principales ciudades gallegas, en Madrid y en Barcelona, aunque una buena parte de ellos está en ciudades del extranjero, como Bruselas o Tokio.

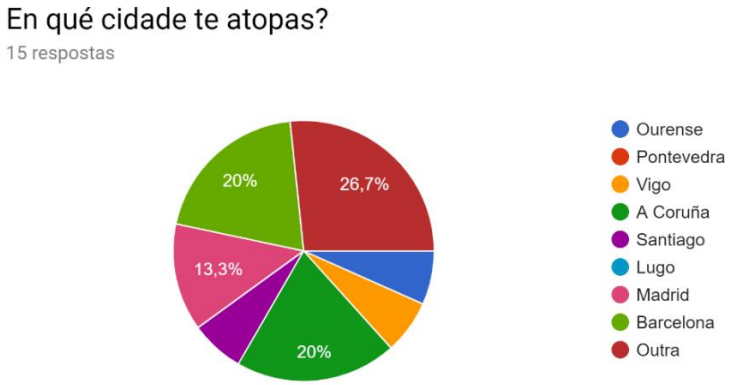


Ilustración 21. Distribución geográfica de los participantes.

2.2. Datos relativos a la profesión y a la TA

Con el objetivo de contextualizar la actividad profesional de los participantes y sus consideraciones sobre la TA, se creó una cuarta sección para recabar este tipo de información.

- *Datos sobre la actividad profesional*

El 93,3 % de los participantes están familiarizados con el tipo de texto evaluado. Si bien, el 60 % indica que no lo traducen de forma habitual. Por último, se empleó una escala de Lykert para indicar el nivel de frecuencia con el que los traductores emplean traducción automática. Los participantes indicaron mayoritariamente que nunca la empleaban (posición 1), aunque si se registraron respuestas entre la posición 2 y 3 de la escala. No se registró ninguna respuesta que indicara que siempre usan TA.

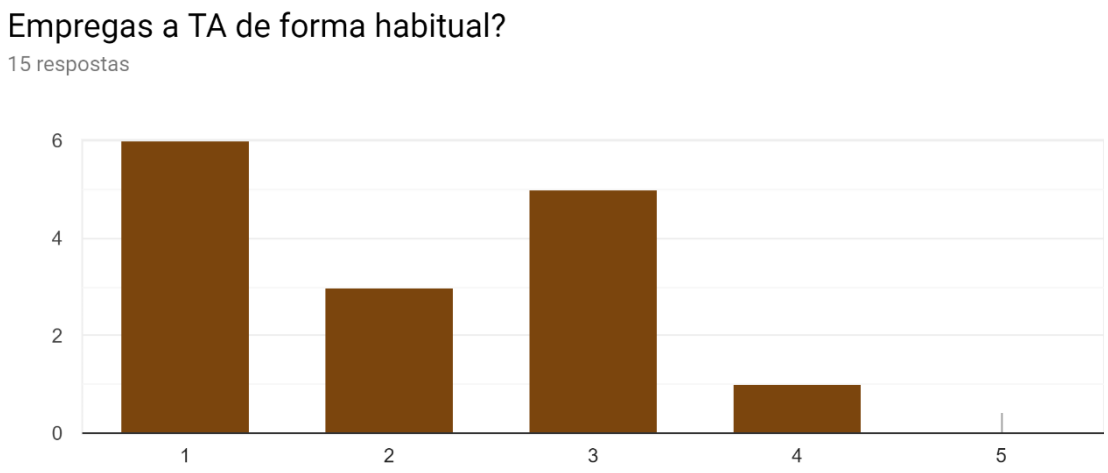


Ilustración 22. Escala de Lykert sobre la frecuencia del empleo de TA

- *Datos sobre la percepción de la TA*

Con respecto a la percepción que los participantes tienen sobre la TA, el 53,3 % de los encuestados consideran que la TA agiliza su trabajo, aunque al mismo tiempo, se registra el mismo porcentaje para señalar la TA como un riesgo para su profesión. A pesar de ello, el 66,7 % de los participantes consideran la traducción como un recurso más de traducción similar a las memorias de traducción, a los glosarios y a las herramientas TAO.

Ilustración 23. Percepciones de la TA

2.3. Clasificación de segmentos

Los participantes ordenaron según validez los 14 segmentos evaluados. En primer lugar, clasificaron según su criterio la traducción en bruto que más adecuada les parecía, y, en segundo lugar, se les pidió que indicaran si esa traducción se podría aprovechar para poseditar o no.

- *Segmento 1*

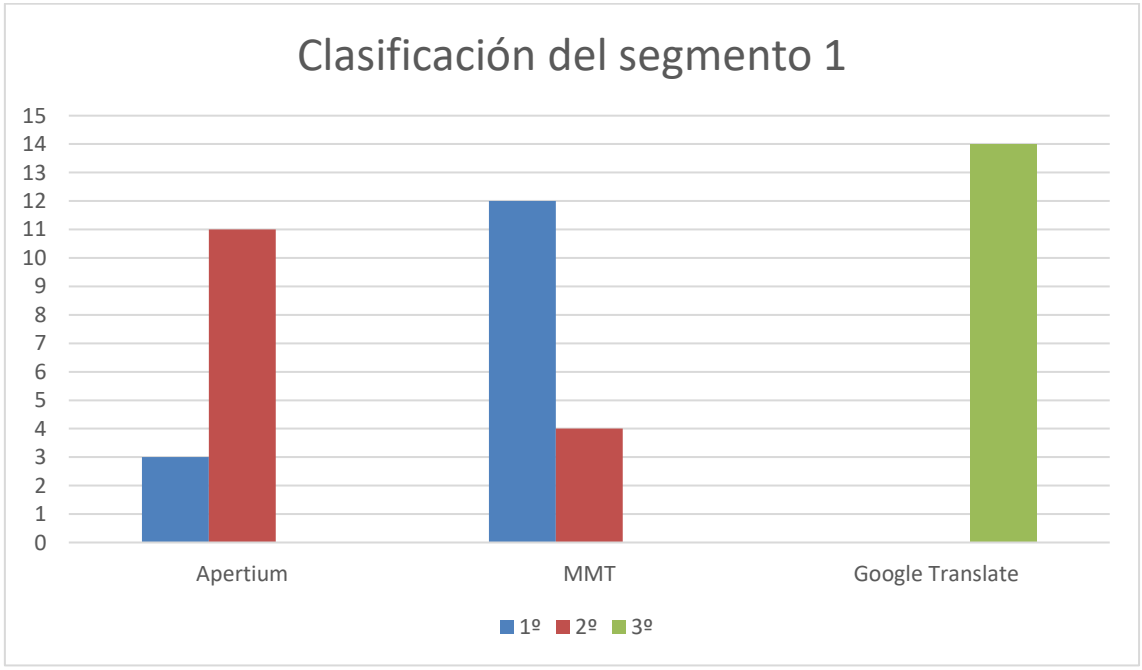


Ilustración 24. Clasificación del segmento 1

La mayoría de los participantes sitúan la versión de MMT como primera opción, seguida por Apertium. Todos los participantes coinciden en que la peor opción es la de Google Translate.

Si observamos los datos obtenidos en la pregunta sobre si aprovecharían el resultado del motor o no para poseditar, se puede apreciar la correlación existente entre el primer y segundo puesto en la clasificación con la categoría de apto para poseditar. Por el contrario, solo un 20% de los participantes contestaron que aprovecharían la opción ofrecida por Google Translate. Sin embargo, la traducción de Google resulta interesante porque muestra cuáles son los tipos de error que los poseditores no aceptan como aptos para poseditar el segmento. En este caso, nos encontramos con palabras no traducidas, cambios de registro, una falsa traducción, una omisión y ausencia del artículo determinado. Por tanto, se registran seis errores en un segmento de 44 palabras, mientras que MMT y Apertium solo cometen un error.

Tabla 6. Aprovechamiento para posedición del segmento 1

	Apertium	**MMT**	**Google Translate**
Porcentaje de respuesta afirmativas	80%	93,3%	20%
Aprovechamiento	Apto	Apto	No apto

- *Segmento 2*

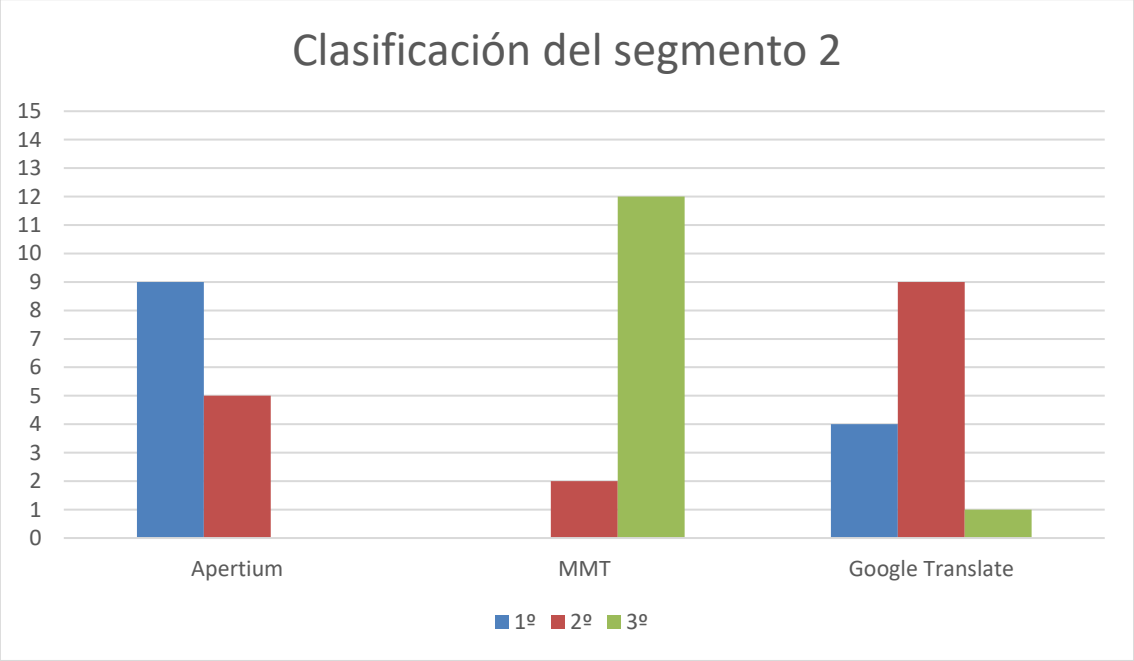


Ilustración 25. Clasificación del segmento 2

En este segmento, la primera opción es Apertium y la segunda, Google Translate. Por su parte, MMT es la peor opción y todos los participantes coinciden en que no se podría aprovechar para poseditar. Si comparamos este segmento con el análisis de errores,

vemos que efectivamente MMT no resuelve bien cuestiones gramaticales (formas verbales incorrectas, orden de palabras incorrecto y no realiza la contracción de la preposición *a* con el articulo determinado en su flexión del femenino plural), lo que produce un estilo muy poco natural y difícil de comprender. Además, llama especialmente la atención que, a pesar de la diferencia en la clasificación entre Apertium y Google, ya que este último llega a registrar hasta 1 vez como tercera posición, el resultado de aprovechamiento sea idéntico. Se observa por tanto cierta inconsistencia en el segmento de Google. Una posible explicación de este fenómeno podría ser que, a diferencia de las otras dos traducciones más literales, Google reordena la oración y cambia el registro, provocando un estilo nuevo. Al no haber unas instrucciones claras del tipo de traducción deseada, cada poseditor entiende el encargo según su experiencia personal y pondera consecuentemente de forma diferente la relación entre precisión, fluidez y estilo y esfuerzo de posedición.

Tabla 7. Aprovechamiento para posedición del segmento 2

	Apertium	**MMT**	**Google Translate**
Porcentaje de respuesta afirmativas	86,7%	0%	86,7%
Aprovechamiento	Apto	No apto	Apto

- *Segmento 3*

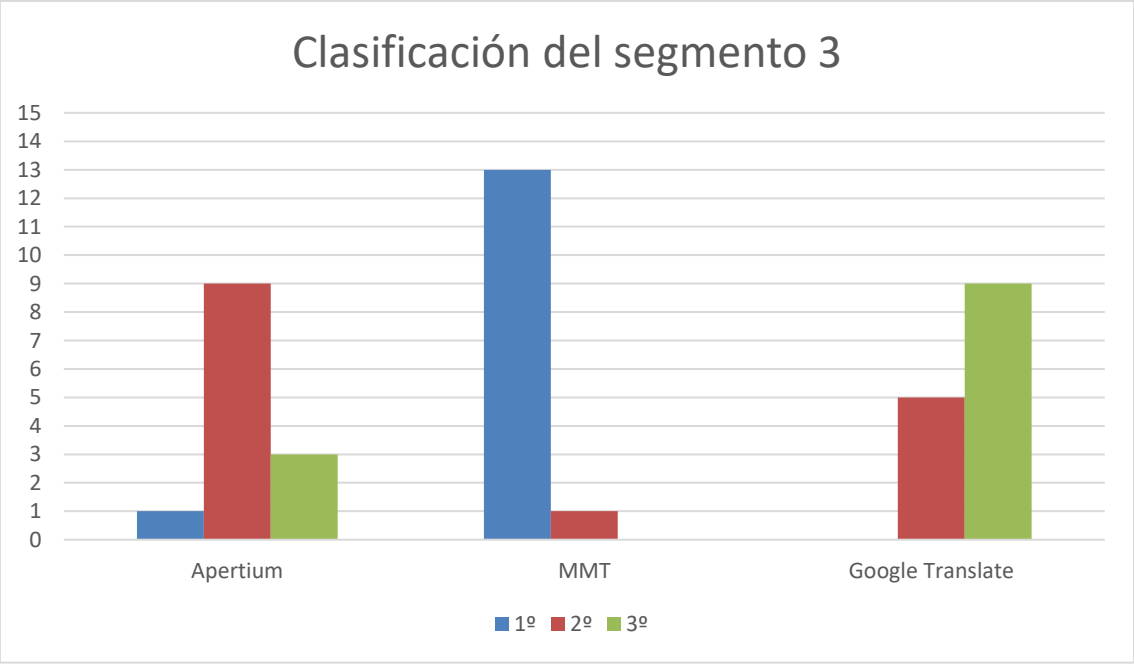


Ilustración 26. Clasificación del segmento 3

Los resultados del segmento 3 muestran que la primera opción es MMT y que el 100 % de los participantes la utilizarían. La siguiente opción es Apertium y, finalmente, la tercera, Google.

Si observamos la tabla de aprovechamiento, vemos que cuantas más personas seleccionan la primera opción, mayor probabilidad hay de que consideren ese segmento apto para poseditar. Por otra parte, llama la atención que a pesar de que 3 personas hayan seleccionado Apertium como tercera opción, el índice de aprovechamiento sea tan alto. De hecho, sorprende que un mismo segmento pueda ser seleccionado como 1ª opción y 3ª opción. Si nos fijamos atentamente en el segmento, vemos que Apertium comete una falsa traducción al interpretar la conjunción *si* (se, en gallego) como el adverbio sí. También registra un error gramatical al no incluir el artículo determinado que rige la preposición *en* para una formulación más natural en gallego de la construcción «en casa». Por su parte, MMT no registra errores y Google cambia el registro de tú a usted y comete un error de ortografía al introducir un guion entre la forma verbal y el pronombre: *chama-nos* en lugar de *chámanos*. Estamos, por tanto, de nuevo ante un claro ejemplo de variabilidad de criterio a la hora de enfrentarse a una posedición, ya que algunos poseditores penalizan más los errores de precisión de Apertium frente a los de registro/ortografía de Google.

Tabla 8. Aprovechamiento para posedición del segmento 3

	Apertium	**MMT**	**Google Translate**
Porcentaje de respuesta afirmativas	80%	100%	40%
Aprovechamiento	Apto	Apto	No apto

- *Segmento 4*

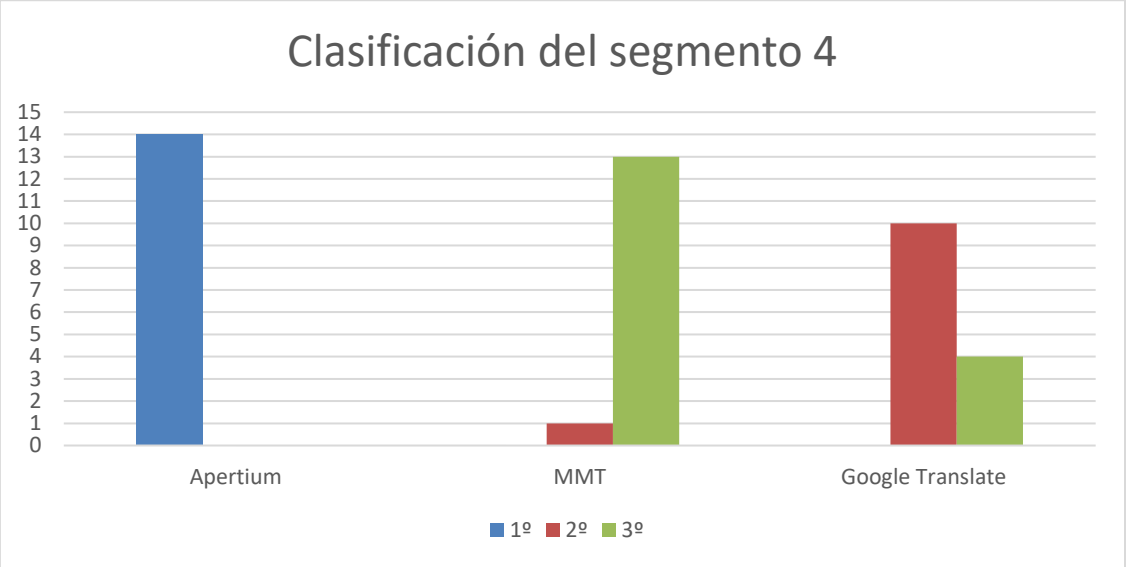


Ilustración 27. Clasificación del segmento 4

Se puede apreciar una clara unanimidad en la clasificación de los segmentos. Sin embargo, es la primera vez que solo se aprovecharía una de las opciones. Este resultado indica que los errores cometidos por los otros dos sistemas son lo suficientemente graves como para desaprovechar la opción ofrecida (omisiones, malas construcciones verbales, ausencia de pronombres, etc.) aunque Google resuelva mejor que la de MMT. Este segmento resulta interesante porque evidencia uno de los puntos fuertes del motor basado en reglas y uno de los ámbitos en los que más hay que hacer hincapié a la hora de entrenar los motores estadísticos y neuronales: el buen análisis morfológico y el correcto trasvase de las estructuras gramaticales.

Tabla 9. Aprovechamiento para posedición del segmento 4

	Apertium	**MMT**	**Google Translate**
Porcentaje de respuesta afirmativas	100%	0%	20%
Aprovechamiento	Apto	No apto	No apto

- *Segmento 5*

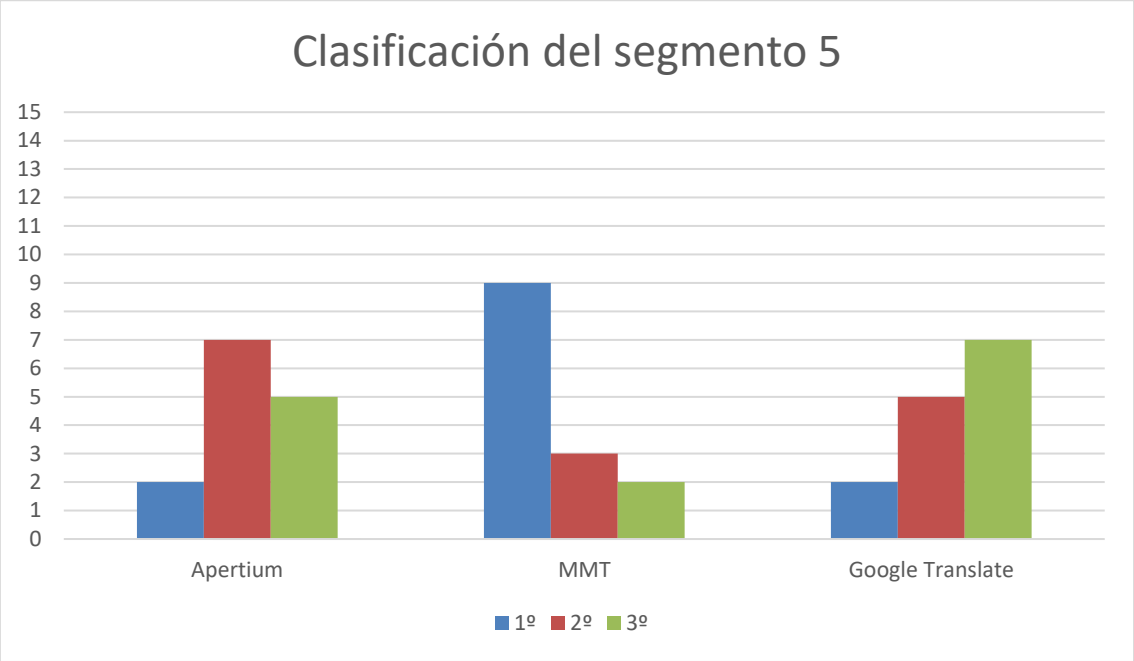


Ilustración 28. Clasificación del segmento 5

En este segmento, las opiniones están más divididas, aunque podríamos decir que predomina como primera opción la traducción ofrecida por MMT, seguida por Apertium y luego, Google. En todos los casos, aprovecharían las opciones ofrecidas por los distintos motores, especialmente la de MMT en la que el 100 % de los participantes coinciden. Estamos ante otro claro ejemplo de los distintos criterios que influyen a cada poseditor a la hora de realizar su tarea. Por ejemplo, resulta incoherente que a pesar de que Google se ha seleccionado como 3ª opción más veces que Apertium, obtenga, sin embargo, una mayor probabilidad de aprovechamiento. Aunque queda fuera de los objetivos de este trabajo, este tipo de segmentos resultan muy interesantes para evaluar los perfiles de los poseditores y su forma de trabajar.

Tabla 10. Aprovechamiento para posedición del segmento 5

	Apertium	**MMT**	**Google Translate**
Porcentaje de respuesta afirmativas	53, 3%	100%	66,7%
Aprovechamiento	Apto	Apto	Apto

- *Segmento 6*

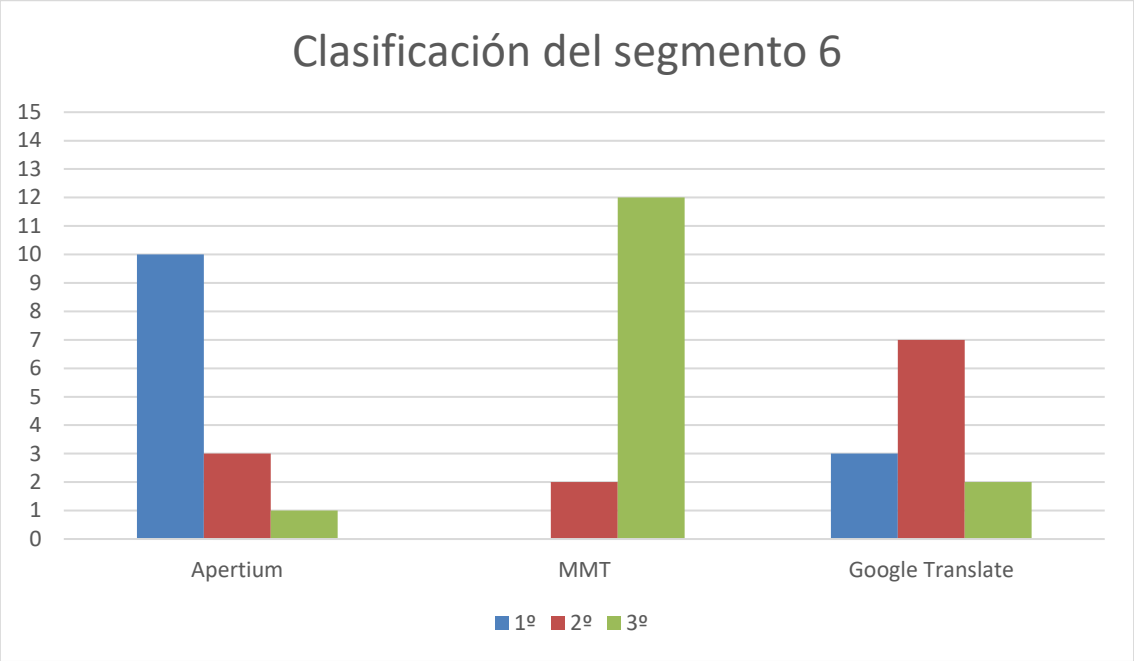


Ilustración 29. Clasificación del segmento 6

En el segmento 6, volvemos a encontrarnos con discrepancias a la hora de clasificar los segmentos, aunque sí se aprecia claramente que la opción de MMT es la peor clasificada. Apertium cuenta con más votos como primera opción, aunque 1 persona también lo ha seleccionado como 3ª opción y Google, por su parte, ha sido seleccionado por más personas como segunda opción. Estos resultados se replican en la pregunta sobre aprovechamiento: solo han indicado que se puede aprovechar la opción de Apertium y la de Google. Al observar el tipo de errores cometidos por cada motor en estos segmentos, se aprecia que la falsa traducción de MMT al traducir la preposición *a* por *con* junto con la omisión del pronombre de objeto directo se penalizan mucho más, que el cambio de registro y error gramatical en la conjunción *ou* que comete Google o la falsa traducción de Apertium (*una/unha*). Se observa una clara tendencia a descartar el resultado de TA si se dan errores de fluidez.

Tabla 11. Aprovechamiento para posedición del segmento 6

	Apertium	**MMT**	**Google Translate**
Porcentaje de respuesta afirmativas	93, 3%	40%	73,3%
Aprovechamiento	Apto	No apto	Apto

- *Segmento 7*

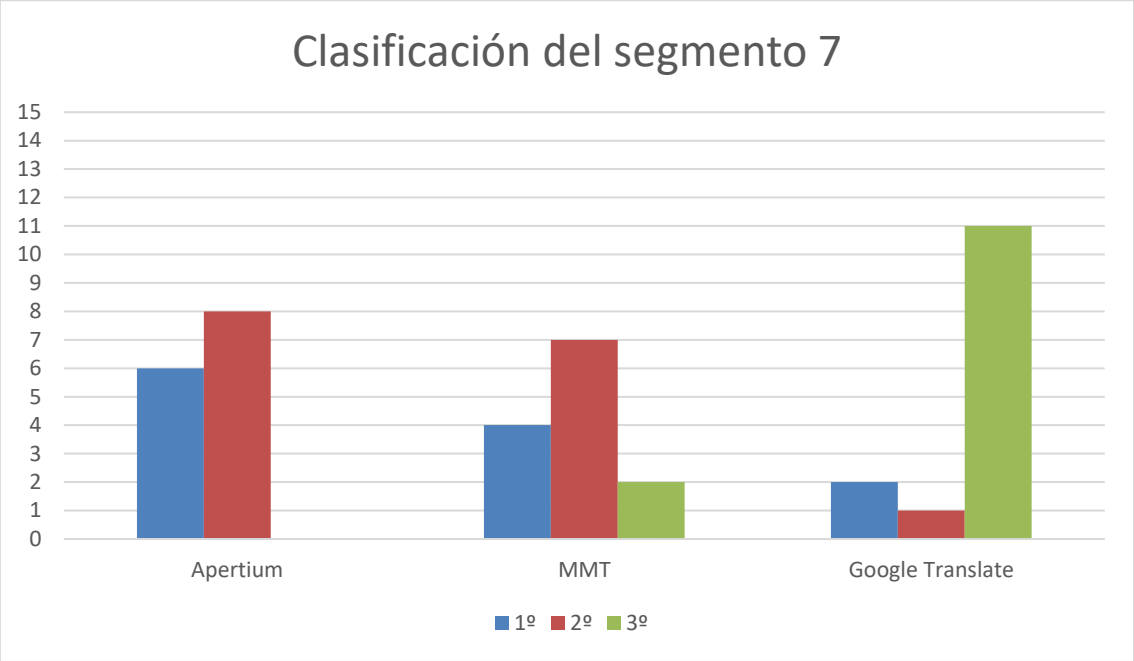


Ilustración 30. Clasificación del segmento 7

En este segmento, tanto la opción de Apertium como la de MMT se sitúan en la parte alta de la clasificación. Apertium se sitúa como primera opción, seguido de cerca por MMT y finalmente, la tercera opción es Google. A pesar de esto, MMT obtiene un mayor porcentaje de respuestas afirmativas sobre su aprovechamiento, lo que sorprende ya que dos personas la han seleccionado como tercera opción. Por su parte, Google roza casi el 50% de respuestas afirmativas, aunque la mayoría lo sitúan en el fondo de la clasificación. Este segmento es uno de los segmentos largos evaluados del documento y, por tanto, esta discrepancia de criterio se puede explicar debido a la mayor presencia de errores y a la subjetividad inherente a la hora de considerar su gravedad. Sin embargo, se observa una tendencia clara a penalizar la traducción más despegada del original y, a veces, con más errores gramaticales y terminológicos del motor neuronal.

Tabla 12. Aprovechamiento para posedición del segmento 7

	Apertium	**MMT**	**Google Translate**
Porcentaje de respuesta afirmativas	66,7	73,3%	46,7%
Aprovechamiento	Apto	Apto	No apto

- *Segmento 8*

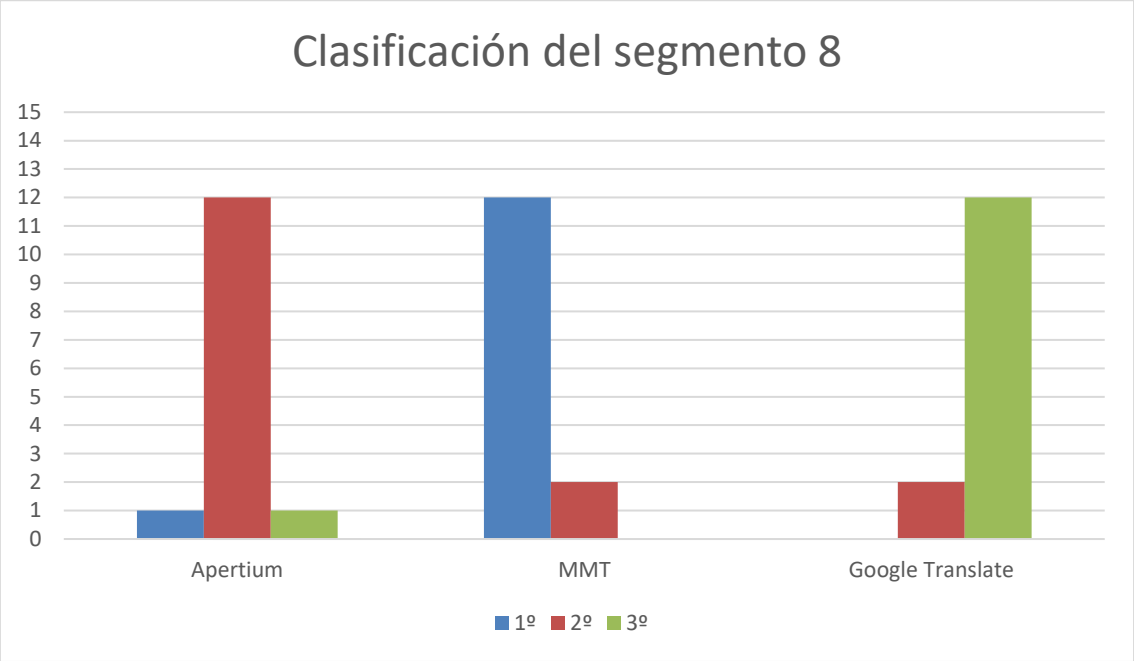


Ilustración 31. Clasificación del segmento 8

Tabla 13. Aprovechamiento para posedición del segmento 8

	Apertium	**MMT**	**Google Translate**
Porcentaje de respuesta afirmativas	46,7%	100%	26,7%
Aprovechamiento	No apto	Apto	No apto

Las opiniones sobre el segmento 8 claramente clasifican la opción de MMT como la mejor, seguida de Apertium y de Google. No obstante, la única que aprovecharían sería la opción de MMT por unanimidad, ya que tanto Apertium (46,7 %) como Google Translate (26,7 %) no consiguen más de la mitad de votos a favor de aprovechar el segmento para poseditar. Una posible explicación de por qué se han respondido negativamente al aprovechamiento del segundo clasificado podría deberse a que Apertium comete dos errores muy llamativos: falsa traducción debido a una mala interpretación del analizado morfológico, y que se mencionan en el apartado 3.1. Por su parte, Google cambia de nuevo el registro, comete una falsa traducción del verbo *pasar* y cambia dos complementos circunstanciales que provocan un cambio de estilo en la oración.

- *Segmento 9*

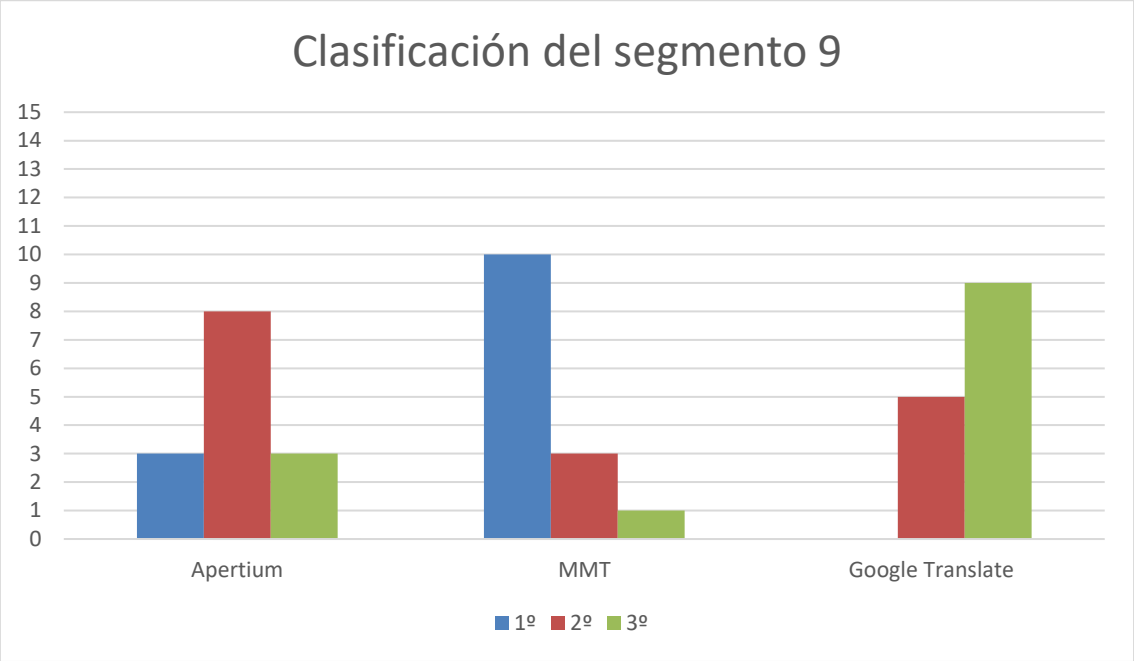


Ilustración 32. Clasificación del segmento 9

A pesar de que no todas las opiniones coinciden, MMT se sitúa en primera posición, seguido de Apertium y de Google. En los tres casos, se aprovecharía la traducción para poseditar, lo que apunta a que las traducciones ofrecidas por los distintos motores son similares y la clasificación de uno y otro dependerá de cómo prefieran trabajar los poseditores. Este segmento puede servir como un buen ejemplo para identificar pequeños matices diferentes en casa sistema con el objetivo de mejorarlos: por ejemplo, MMT no comete ningún error y se ajusta totalmente al registro administrativo de la oración principal, fruto del corpus de entrenamiento con el que se trabajó (uso del infinitivo conjugado). Por su parte, Apertium falla a la hora de trasvasar *en caso que*, debido a que en el original está mal construido (*en caso de que*) y no rige el artículo determinado en la construcción y calca la estructura del castellano en la construcción *en caso de no cumplirse*. Por su parte, Google también tiene problemas a la hora de resolver las mismas construcciones que Apertium y comete errores de tiempo/modo verbal.

Tabla 14. Aprovechamiento para posedición del segmento 9

	Apertium	**MMT**	**Google Translate**
Porcentaje de respuesta afirmativas	66,7%	100%	66,7%
Aprovechamiento	Apto	Apto	Apto

- *Segmento 10*

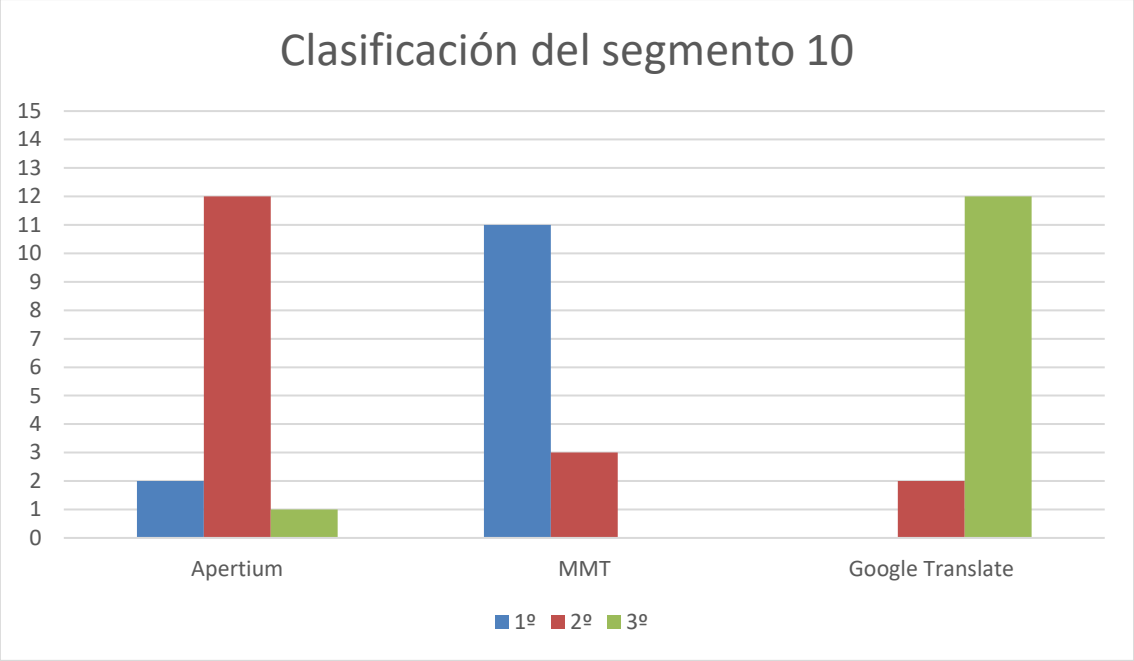


Ilustración 33. Clasificación del segmento 10

El segmento 10 es un claro ejemplo del tipo de oraciones que la TA resuelve bien, ya que los participantes concuerdan en que se podrían aprovechar las tres opciones, incluso si sitúan al segmento en una posición inferior. Esto es un claro indicativo de que los tres motores son capaces de resolver de forma aceptable la traducción, de hecho, solo Google comete un error de tiempos verbales. Si analizamos en profundidad la frase, vemos que se trata de una oración típica del ámbito administrativo, con una estructura sencilla y bien redactada en el original. No sorprende, por tanto, que la primera opción sea MMT ya que el corpus de base de entrenamiento contiene oraciones similares.

Tabla 15. Aprovechamiento para posedición del segmento 10

	Apertium	**MMT**	**Google Translate**
Porcentaje de respuesta afirmativas	93,3%	100%	60%
Aprovechamiento	Apto	Apto	Apto

- *Segmento 11*

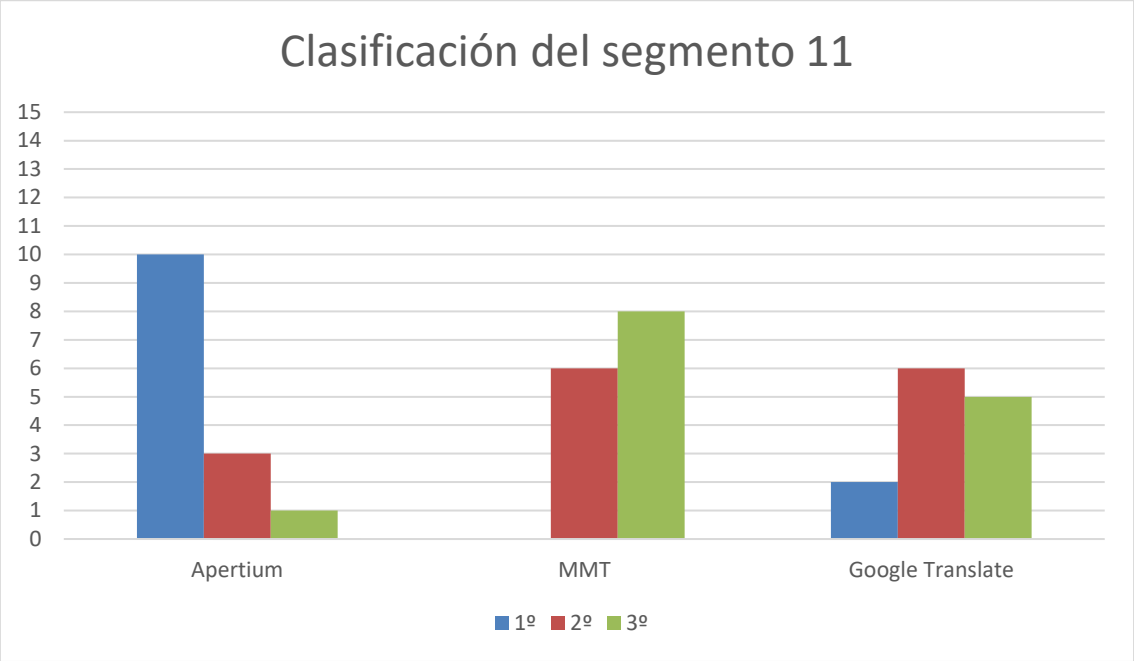


Ilustración 34. Clasificación del segmento 11

La mejor opción para el segmento 11 es la de Apertium. MMT y Google Translate reciben los mismos votos en segunda posición, sin embargo, MMT no tiene ningún voto para la primera clasificación y obtiene más votos como 3ª opción.

En línea con la clasificación obtenida, los participares aprovecharían la opción de Apertium. Si se compara Google con MMT, podemos observar que, aunque Google haya sido seleccionado como tercera opción al igual que MMT, los errores que comete no se consideran lo suficientemente graves como para no aprovechar la traducción, a diferencia del escenario que se plantea con MMT (fallos en la concordancia, adición de palabras, etc.). Este segmento, además, evidencia un posible problema de MMT a la hora de gestionar palabras en mayúsculas y que convendría solventar. En el segmento original, aparece en mayúscula «LA EMPRESA» y MMT lo traduce por «DA EMPRESA» (de la empresa, en castellano) como si se tratara de una estructura independiente sin tener en cuenta la preposición anterior al sintagma nominal en las dos veces que aparece el segmento.

Tabla 16. Aprovechamiento para posedición del segmento 11

	Apertium	**MMT**	**Google Translate**
Porcentaje de respuesta afirmativas	93,3%	43,7%	66,7%

Aprovechamiento	Apto	No apto	Apto

- *Segmento 12*

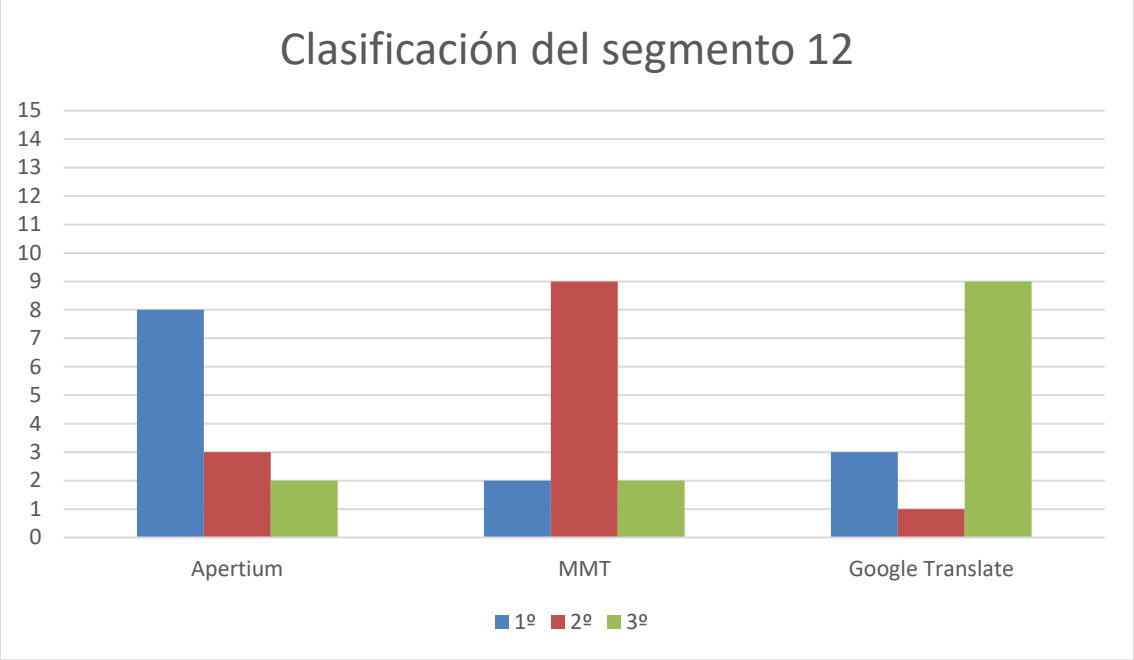


Ilustración 35. Clasificación del segmento 12

En este segmento, llama la atención la variabilidad de las respuestas, ya que todos los poseditores han elegido las tres opciones para las distintas traducciones. Podemos aventurar que un posible motivo de este resultado se debe a que el segmento 12 es muy corto y las diferencias entre las distintas opciones son muy sutiles. Esto hace que la elección de la clasificación sea muy subjetiva dependiendo de las preferencias de cada poseditor, especialmente si comparamos este resultado con la homogeneidad de las opiniones en los segmentos 4 y 10 donde una de las opciones destaca por encima de las otras con una mayor claridad. A pesar de ello, se puede afirmar que los 3 motores resuelven adecuadamente la traducción, ya que en los tres casos los poseditores coinciden en que se podría aprovechar.

Tabla 17. Aprovechamiento para posedición del segmento 12

	Apertium	**MMT**	**Google Translate**
Porcentaje de respuesta afirmativas	93,3%	93,3%	60%
Aprovechamiento	Apto	Apto	Apto

- *Segmento 13*

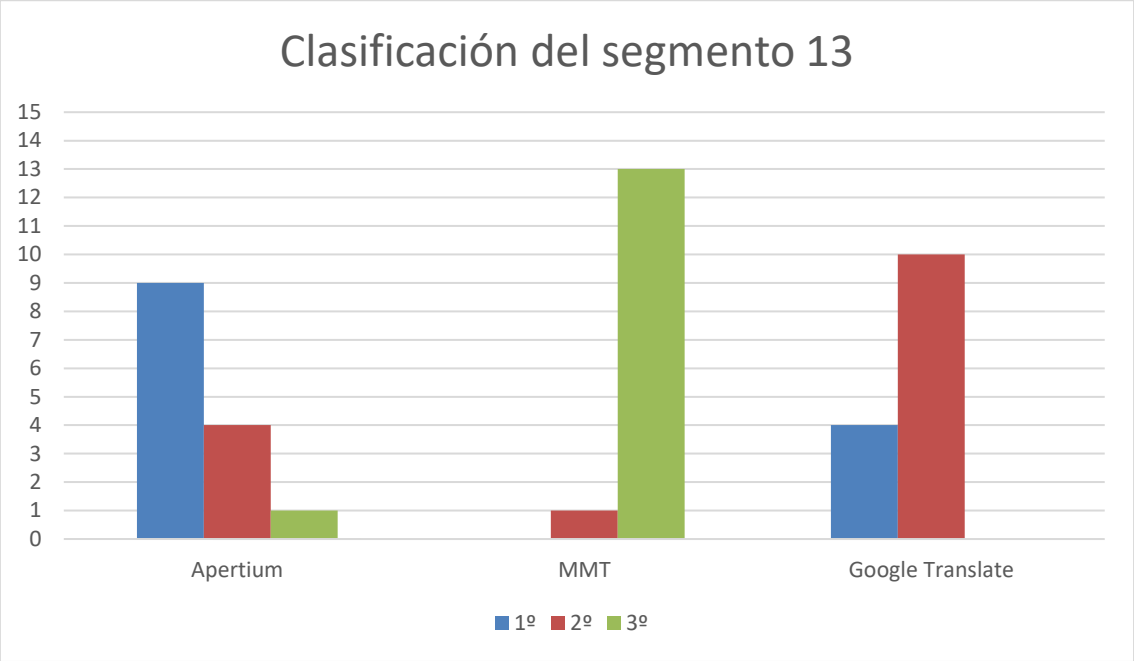


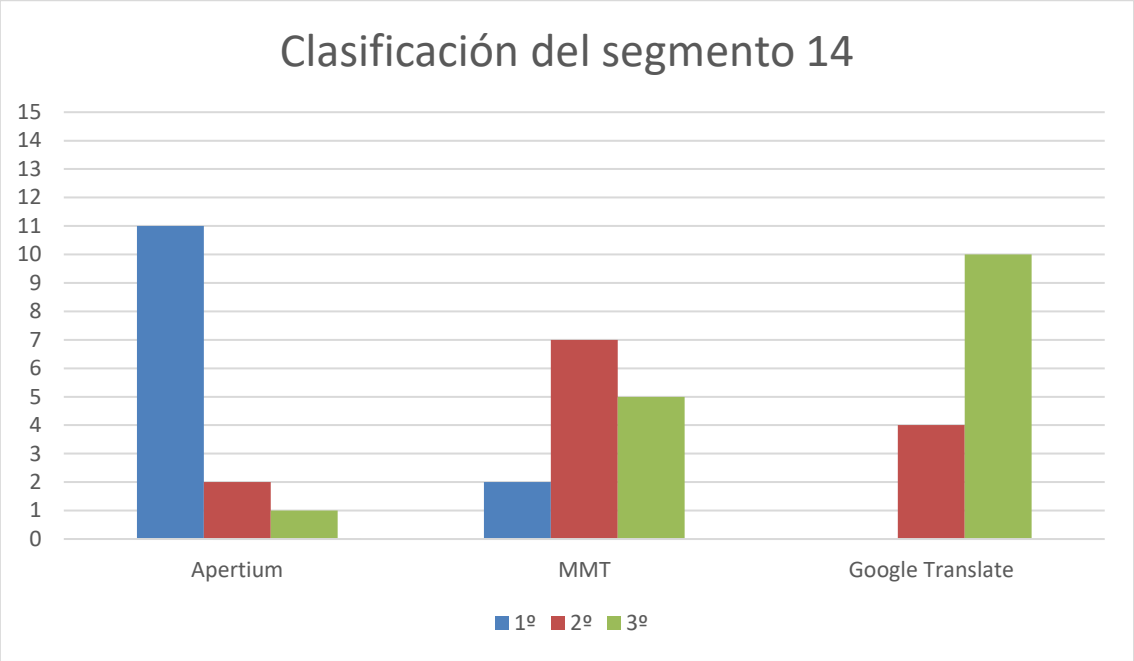
Ilustración 36. Clasificación del segmento 13

En el segmento 13, destaca la unanimidad para clasificar a MMT como la tercera opción y para indicar que no consideran que se puede aprovechar para poseditar. Los errores que no pasan el filtro de los poseditores son una palabra sin traducir y la omisión del artículo determinado. En los otros dos casos, se utilizaría la traducción en bruto. Vuelve a destacar que a pesar de que una persona haya seleccionado Apertium como la 3ª opción, el 100 % de los encuestados afirmen que se puede aprovechar la traducción. De hecho, ninguno de los dos motores comete errores, lo que hace que esta elección sea puramente estilística, por ejemplo, Google opta por una construcción pasiva.

Tabla 18. Aprovechamiento para posedición del segmento 13

	Apertium	**MMT**	**Google Translate**
Porcentaje de respuesta afirmativas	100%	33,3%	86,7%
Aprovechamiento	Apto	No apto	Apto

- *Segmento 14*



De nuevo, la clasificación de las traducciones coincide con el porcentaje de aprovechamiento. Los participantes clasifican Apertium en primer lugar y el total de los encuestados considera que se puede aprovechar. Asimismo, se repite la situación del segmento anterior, en el que a pesar de que se ha seleccionado como tercera opción en una ocasión, esto no repercute en el aprovechamiento. No es el caso de los otros dos motores, que registran más votos en contra de usar la traducción en bruto para poseditar. Aquí sorprende que MMT a pesar de haber sido seleccionado por dos personas como primera opción, no sea considerado como apto. Si se analizan los motivos del no aprovechamiento de MMT y Google, vemos que MMT deja palabras sin traducir, comete errores de flexión de género y número y de orden de palabras. Por su parte, Google produce errores de tiempo y modo verbal, falsas traducciones y cambios de registro.

Tabla 19. Aprovechamiento para posedición del segmento 14

	Apertium	**MMT**	**Google Translate**
Porcentaje de respuesta afirmativas	100%	40%	26,7%
Aprovechamiento	Apto	No apto	No apto

- *Análisis total*

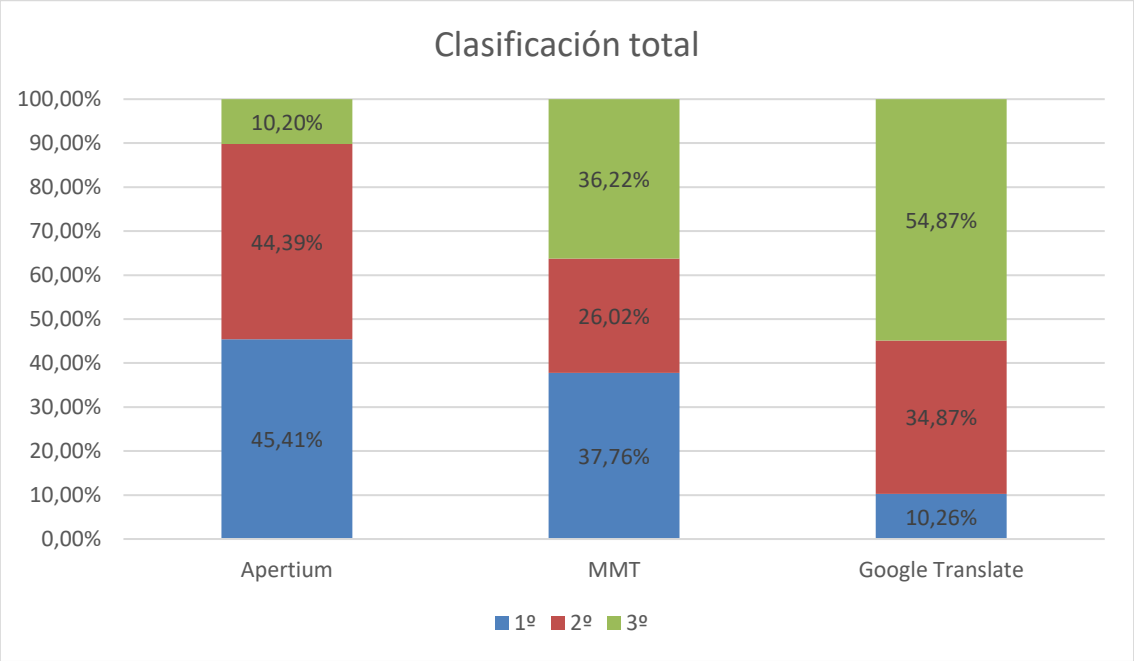


Ilustración 37. Clasificación total de la mediana obtenida en los segmentos

El gráfico anterior muestra en qué posición se ha elegido cada motor. Mayoritariamente, Apertium es la primera opción, dato que coincide con el porcentaje de segmentos marcados como aceptables para poseditar de la Ilustración 38. Por su parte, MMT tiene unos resultados bastante inconstantes, ya que cuenta con segmentos clasificados tanto como primera opción, como segunda y como tercera en una proporción bastante similar. El análisis posterior de errores del apartado 3.2 arrojará luz sobre los problemas de MMT para resolver correctamente los segmentos peor valorados.

En cuanto al aprovechamiento de las traducciones, los participantes solo han indicado en cinco segmentos que se podrían aprovechar todas las traducciones en bruto (35% del total) lo que claramente indica que no todos los motores funcionan igual de bien para esta combinación de idiomas y este contexto. En el resto de los casos, han rechazado una (15%) o dos (50%) de las opciones. No se registraron casos en los que los traductores rechacen la primera opción, lo que respalda los altos índices obtenidos en la evaluación automática. El siguiente gráfico muestra las veces en las que el motor se ha seleccionado como apto en los 14 segmentos evaluados.

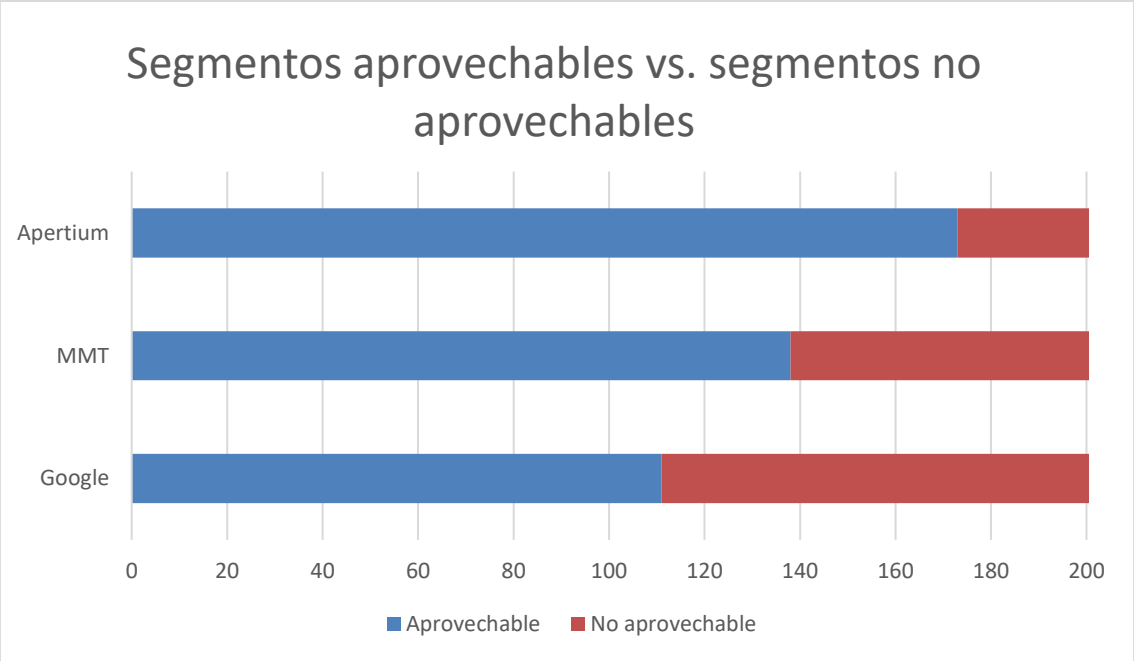


Ilustración 38. Aprovechamiento global de los motores

Para calcular la validez de los datos, se ha aplicado la prueba Q de Cochran (prueba estadística no paramétrica).

Tabla 20. Prueba Q de Cochran del documento entero

Variable	Categorías	Frecuencias	%
Apertium	0	37	17,619
	1	173	82,381
MmT	0	72	34,286
	1	138	65,714
Google	0	99	47,143
	1	111	52,857

Tabla 21. P-value del documento entero

C (Valor observado)	42,014
C (Valor crítico)	5,991
FD	2
p-value	< 0,0001
alfa	0,05

Las diferencias son significativas (p-value < 0,0001) y las proporciones entre los tres grupos son estadísticamente significativas (procedimiento Marascuilo):

Tabla 22. Proporciones entre los tres grupos para el documento entero

Pares	Valor	Valor crítico	¿Significativo?
|p(Apertium) - p(MmT)|	0,167	0,103	Sí
|p(Apertium) - p(Google)|	0,295	0,106	Sí
|p(Mmt) - p(Google)|	0,129	0,116	Sí

Por último, las proporciones muestran que los tres grupos son diferentes:

Muestra	Proporción	Grupos		
Google	0,529	A		
MmT	0,657		B	
Apertium	0,824			C

- *Análisis separado de los segmentos muy cortos*

En la encuesta, solo se evaluó un segmento corto, el segmento 12 (ver apartado Segmento 12) que se corresponde con el segmento 22 del documento entero.

- *Análisis separado de los segmentos muy largos*

Se evaluaron 4 de los cinco segmentos largos del documento (segmentos 1, 7, 9 y 11 de la encuesta que se corresponden con los segmentos 2, 17, 18 y 19 del documento entero respectivamente). Este es el resultado:

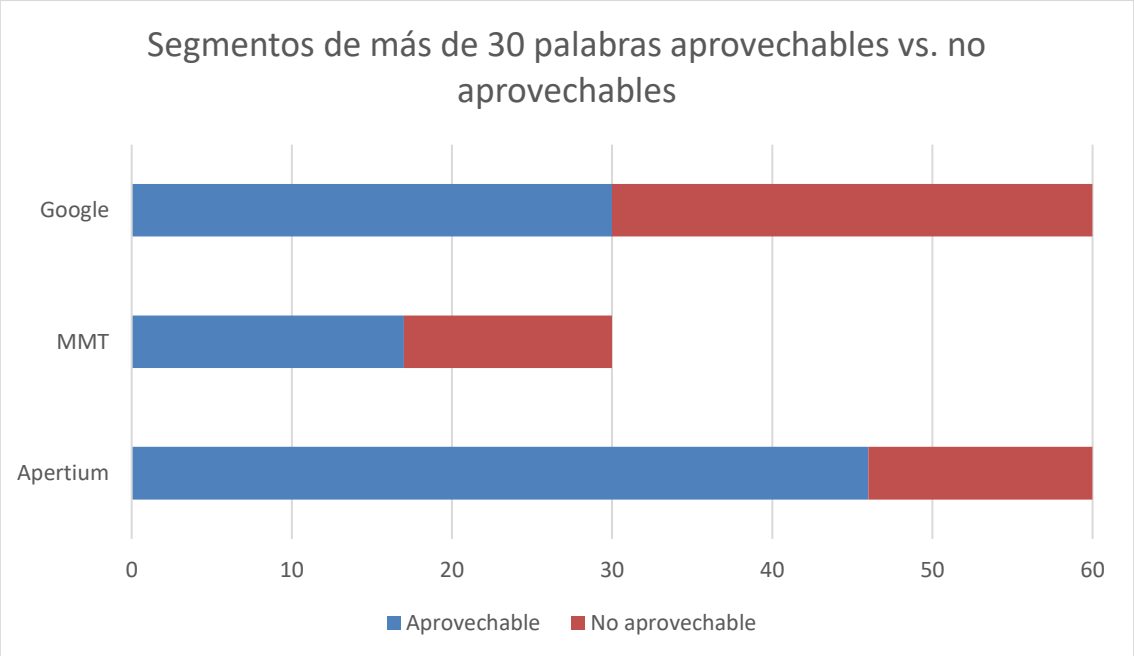


Ilustración 39. Aprovechamiento de los segmentos muy largos

Para calcular la validez de los datos, se ha aplicado la prueba Q de Cochran (prueba estadística no paramétrica).

Tabla 23. Prueba Q de Cochran en segmentos muy largos

Variable	Categorías	Frecuencias	%
RBMT	0	14	23,333
	1	46	76,667
PBMT	0	13	21,667
	1	47	78,333
NMT	0	30	50,000
	1	30	50,000

Tabla 24. P-value en segmentos muy largos

C (Valor observado)	15,167
C (Valor crítico)	5,991
FD	2
p-value	0,001
alfa	0,05

Las diferencias son significativas (p-value < 0,0001) y las proporciones son estadísticamente significativas, pero no entre los tres grupos (procedimiento Marascuilo):

Tabla 25. Proporciones en los segmentos muy largos

Pares	Valor	Valor crítico	¿Significativo?
|p(Apertium) - p(MmT)|	0,017	0,187	No
|p(Apertium) - p(Google)|	0,267	0,207	Sí
|p(Mmt) - p(Google)|	0,283	0,205	Sí

Por último, las proporciones muestran que existen diferencias entre Google y los otros dos sistemas de TA:

Muestra	Proporción	Grupos	
Google	0,500	A	
MmT	0,767		B
Apertium	0,783		B

En cuanto a la clasificación de los segmentos, la siguiente figura muestra en qué posición se ha seleccionado cada motor:

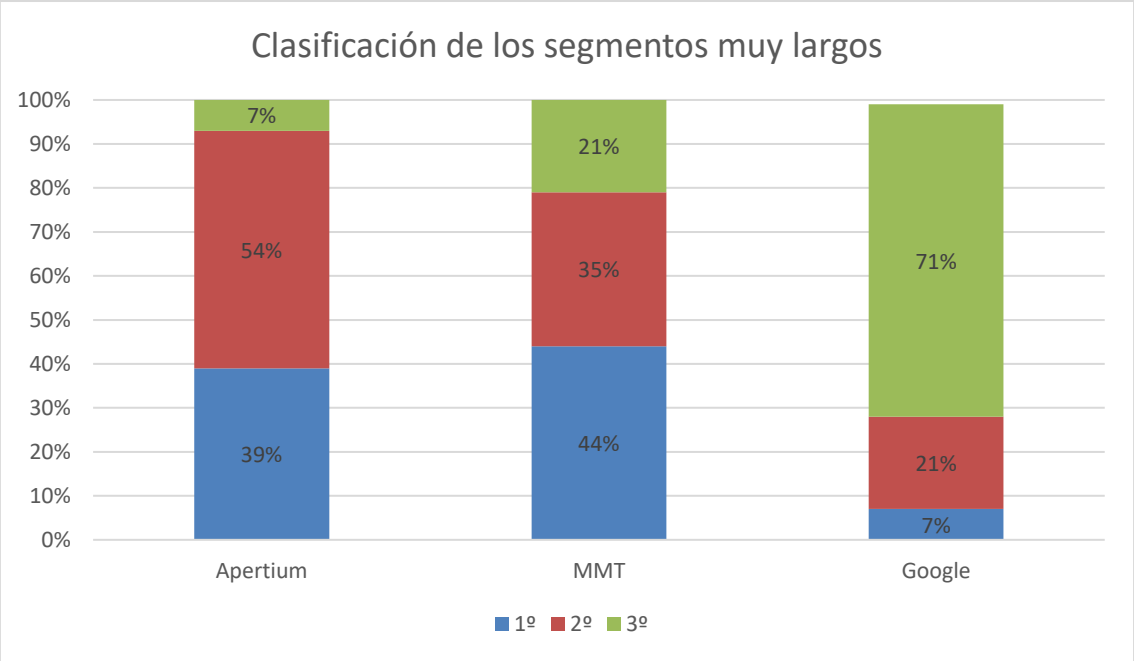


Ilustración 40. Clasificación de los segmentos muy largos

Se puede afirmar que tanto Apertium como MMT resuelven razonablemente bien la complejidad de los segmentos muy largos, fruto del entrenamiento previo efectuado y de la buena redacción del original. Por su parte, Google es el motor que se clasifica más veces como peor opción, probablemente por la mayor presencia de errores cometidos (ver apartado 3.5).

3. Datos del análisis de errores

Para el análisis de errores se ha diseñado un marco conforme a MQM que permita centrarse especialmente en la precisión y fluidez de las traducciones en bruto, ya que consideramos que son los aspectos básicos de la lengua que posibilitarían el uso de los motores en el ámbito empresarial.

En primer lugar, se presentarán los resultados individuales de cada motor y, posteriormente, se compararán los resultados para saber cuál de ellos obtiene mejores resultados en este tipo de evaluación.

3.1. Análisis de errores en Apertium

El análisis de errores se realizó en los 30 segmentos y este es el resultado:

Tabla 26. Análisis de errores en Apertium

Accuracy		
	Addition	
	Mistranslation	10
	Omission	
	Untranslated	1
Subtotal		11
Fluency		
	Ambiguity	
	Unintelligible	
	Spelling	
	Typography	
	Grammar	
	Word-form	
	Part-of-speech	
	Agreement	2
	Tense-mood-aspect	2
	Word-order	1
	Function-words	4
Subtotal		9
Terminology		
	Inconsistent with termbase	
	Inconsistent with domain	
Subtotal		0
Style		
	Register	
	Awkward	1
	Unidiomatic	2
Subtotal		3
Total		**23**

Como se puede apreciar, dentro de la categoría de precisión el error más frecuente es la falsa traducción, mientras que, en la categoría gramatical, los errores más comunes son de concordancia, de tiempo, modo y aspecto de los verbos y de las palabras de función (artículo, preposiciones…). No se han registrado errores terminológicos, ya que se ha velado porque la terminología sea la correcta, a diferencia de Google Translate en donde no era posible. Por el contrario, sí se han producido errores de estilo, principalmente, con expresiones que, aunque correctas gramaticales, no sonaban idiomáticas (consultar ejemplos de errores en el anexo IV).

Tras este análisis cualitativo, se pueden extraer los siguientes patrones de errores:

- Mala gestión del desambiguador léxico en la traducción del artículo indeterminado uno/una y de la preposición para. En ambos casos, el

motor interpreta el artículo como una forma verbal, por lo que traduce ambas palabras por la primera persona del presente de subjuntivo del verbo unir y del verbo parir.

- Mala gestión de las contracciones en gallego. Cuando en gallego la preposición rige el uso del artículo determinado antepuesto al sustantivo y no coincide con el castellano, el motor no lo traduce bien. Por ejemplo, «en caso de», se debe traducir por «no caso de».
- No se produce la flexión de género y número según la palabra traducida, sino según el género y número del original.
- Mala gestión de las perífrasis verbales en gallego. Por ejemplo, hay casos en los que no rigen una preposición, que, sin embargo, el motor si mantiene porque en español lleva preposición.
- Al convertir palabra a palabra, la traducción es bastante literal, lo que provoca que en ocasiones se produzcan expresiones poco idiomáticas.

3.2. Análisis de errores en MMT v2.5

El análisis de errores se realizó en los 30 segmentos y este es el resultado:

Tabla 27. Análisis de errores en MmT

Accuracy			
	Addition		2
	Mistranslation		1
	Omission		2
	Untranslated		5
Subtotal			10
Fluency			
	Ambiguity		
	Unintelligible		
	Spelling		1
	Typography		1
	Grammar		
		Word-form	
			Part-of-speech 2
			Agreement 5
			Tense-mood-aspect
		Word-order	2
		Function-words	6
Subtotal			17
Terminology			
	Inconsistent with termbase		
	Inconsistent with domain		1

Subtotal		1
Style		
	Register	
	Awkward	
	Unidiomatic	2
Subtotal		2
Total		**30**

La mayoría de errores se concentran en el apartado de precisión y de fluidez. Dentro de precisión, destaca el elevado número de palabras no traducidas. Por otro lado, buena parte de los errores de fluidez se concentran en la categoría gramatical, especialmente en cuanto a concordancia y palabras de función. Además, se ha detectado un error de inconsistencia con la terminología del dominio (predio/finca) y errores de estilo dentro de la subcategoría de no idiomático.

Este análisis cualitativo permite identificar los siguientes patrones:

- Mayor presencia de palabras sin traducir que en el caso anterior.
- Elevada presencia de adiciones y omisiones.
- Mala gestión de los artículos, las preposiciones y las contracciones entre preposición y artículo, que provoca desde errores gramaticales (falta de preposición, no contracción entre preposición y artículo) a incluso adiciones (en dos ocasiones «la empresa» se traduce como «da empresa» y no «a empresa»).
- Mala gestión del orden de palabras en construcciones complejas, lo que produce errores gramaticales y expresiones muy poco idiomáticas.
- Falta de concordancia de género y de número.
- Mala gestión de las formas verbales y de la colación de los pronombres, especialmente cuando los verbos pronominales no coinciden entre español y gallego.
- Mayor tendencia a la inconsistencia terminológica.

3.3. Análisis de errores en Neural Google Translate

Estos son los resultados del análisis de errores realizado en los 30 segmentos del documento:

Tabla 28. Análisis de errores en Google

Accuracy		
	Addition	1
	Mistranslation	6
	Omission	1
	Untranslated	4
Subtotal		12
Fluency		
	Ambiguity	
	Unintelligible	
	Spelling	1
	Typography	2
	Grammar	
	Word-form	
	Part-of-speech	1
	Agreement	3
	Tense-mood-aspect	1
	Word-order	1
	Function-words	2
Subtotal		10
Terminology		
	Inconsistent with termbase	
	Inconsistent with domain	4
Subtotal		4
Style		
	Register	11
	Awkward	2
	Unidiomatic	1
Subtotal		14
Total		**41**

Al contrario que en los resultados anteriores para los otros motores, se obtienen errores en todas las categorías, especialmente en estilo, precisión y fluidez. El principal error en estilo es el registro, ya que cuando en el original se opta por el tratamiento de tú, se traduce como usted y al revés. Dentro de precisión, destaca el elevado número de errores de falsas traducciones y de palabras no traducidas. Por otra parte, dentro del apartado de fluidez, hay una mayor presencia de errores relativos a la concordancia de género y número y las palabras de función. Finalmente, se han detectado inconsistencias dentro del mismo texto relativas a la terminología propia del dominio, como, por ejemplo, la palabra «confort» se ha traducido como «comodidade» en un segmento y como «confort» en otro; o «finca» que se ha traducido como «granxa» y como «facenda», lo que a su vez también son falsas traducciones en ambos casos.

Este análisis cualitativo permite identificar los siguientes patrones:

- Cambios de registro constante, incluso dentro del mismo segmento, en el tratamiento de tú o usted.
- Hay una tendencia a omitir palabras, especialmente artículos y preposiciones
- Elevada presencia de palabras sin traducir. Resulta especialmente llamativo que en un caso el término «homogéneo» no se traduzca la primera vez que aparece en el texto, pero sí las siguientes.
- Excesiva presencia de falsas traducciones. Por ejemplo, se ha traducido «regalamos» por «nos dan».
- En general, las construcciones tienden a separarse más del original, lo que en algunos casos da buenos resultados, pero en otros lleva a falsas traducciones, omisiones, adiciones, construcciones no idiomáticas y raras.
- Errores tipográficos llamativos como empleo de guiones para separar la forma verbal del pronombre, como, por ejemplo, «chama-nos» en lugar de «chámanos».

3.4. Análisis de errores en segmentos muy cortos

Este es el desglose de errores encontrados en los 10 segmentos muy cortos:

Tabla 29. Errores en los segmentos cortos

Accuracy	**Apertium**	**MMT**	**Google**
Addition			
Mistranslation			1
Omission			
Untranslated			1
Fluency			
Ambiguity			
Unintelligible			
Spelling			
Typography			
Grammar			
Word-form			
Part-of-speech	1	1	1
Agreement			
Tense-mood-aspect			
Word-order			

Function-words			
Terminology			
Inconsistent with termbase			
Inconsistent with domain			
Style			
Register			1
Awkward			
Unidiomatic			1
Total	1	1	5

Como se puede apreciar, se produce el mismo error gramatical en MMT v2.5 y Apertium. Por su parte, se registran errores de precisión, fluidez y estilo en Google Translate. Por ejemplo, si analizamos el segmento 22:

TO: Les informamos que a causa de:

TT de Google: Informámos**che** que **debido a**:

TT de referencia: Informámolos de que por mor de:

En este caso, Google opta por la segunda persona del singular y emplea una locución preposicional menos idiomática, más infrecuente y pegada al español que en el caso de la traducción de referencia, lo que produce un cambio de registro y estilo más pobre.

Con todo, cabe decir que, en general, el porcentaje de errores registrados en los segmentos cortos no impide la compresión ni el aprovechamiento de los segmentos para poseditar.

3.5. Análisis de errores en segmentos muy largos

Este es el desglose de errores encontrados en los 5 segmentos muy largos:

Tabla 30. Errores en los segmentos largos

Accuracy	Apertium	MMT	Google
Addition		2	1
Mistranslation	6		2
Omission		1	
Untranslated		1	1
Fluency			
Ambiguity			
Unintelligible			
Spelling		1	1
Typography			
Grammar			

Word-form		
Part-of-speech	1	1
Agreement		2
Tense-mood-aspect		1
Word-order		
Function-words	2	1
Terminology		
Inconsistent with termbase		
Inconsistent with domain		1
Style		
Register		4
Awkward		1
Unidiomatic		1
Total	9	9	16

En este caso, se puede apreciar el aumento exponencial de errores derivados de la complejidad estructural de las frases más largas. Si comparamos el resultado obtenido en la encuesta a los poseditores con esta gráfica, vemos que se puede establecer una relación directa entre mayor índice de errores y no apto para poseditar, especialmente, si se trata de errores de estilo y fluidez que obligan a realizar más cambios en el segmento. Un claro ejemplo de esto lo vemos en el segmento 1. Apertium y MMT registran una falsa traducción, mientras que Google comete errores de registro, palabras no traducidas y construcciones raras. En los dos primeros casos, el poseditor solo tendría que realizar un cambio, mientras que, en el tercer caso, sería necesaria la reformulación de la frase, por lo que los poseditores optan por no clasificarlo como no apto para poseditar.

Algo similar sucede con el segmento 11, en este caso, los errores gramaticales (palabras de función y concordancia) que registra MMT hacen que los poseditores lo consideren lo suficientemente graves como para no poder aprovechar la traducción.

3.6. Análisis conjunto de errores

Una vez analizadas todas las traducciones en bruto dentro de nuestro marco personalizado, se presentarán los resultados obtenidos por cada motor para comprobar cuál es el que menos errores tiene en general y cuáles son los errores comunes a ambos.

En cuanto al rendimiento general, la siguiente gráfica muestra los puntos totales obtenidos por cada motor:

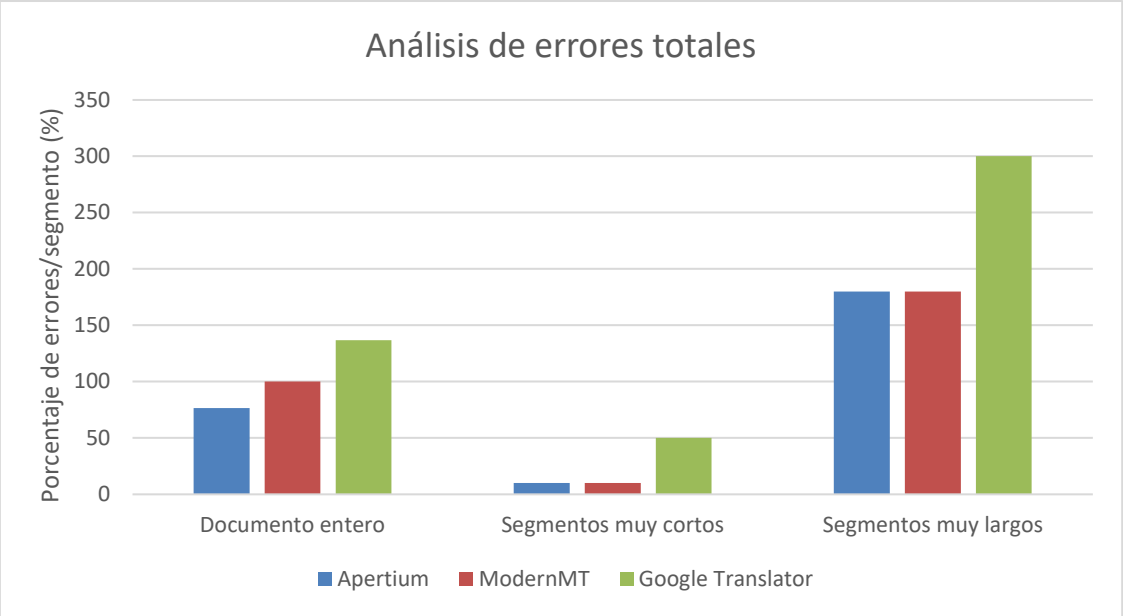


Ilustración 41. Comparación del análisis de errores en los tres motores

Como se puede apreciar en la gráfica, Google Translator es el motor que más porcentaje de errores tiene. Apertium obtiene menos errores en la evaluación general, pero empata con MMT en los segmentos cortos y queda por detrás de este en los segmentos muy largos. Podemos decir que tanto uno como el otro son los que mejor resultado obtienen en esta evaluación.

En cuanto al número de errores generales, Google obtiene más errores que número de segmentos totales, a diferencia de Apertium y ModernMT, que han resuelto correctamente varios segmentos.

Por otra parte, hay una mayor presencia de errores en los segmentos muy largos, por lo que se puede decir que los tres motores tienen dificultades con este tipo de segmentos. Sin embargo, tanto Apertium como MMT gestionan bien los segmentos cortos, al contrario que Google que produce un error en cada uno de los 5 segmentos.

Con respecto al tipo de error, la siguiente gráfica muestra el reparto de errores entre los motores. Para facilitar la interpretación del gráfico, solo se han incluido aquellos errores que se dan al menos en algún motor. Por tanto, se han excluido los errores que no están presentes en ningún motor.

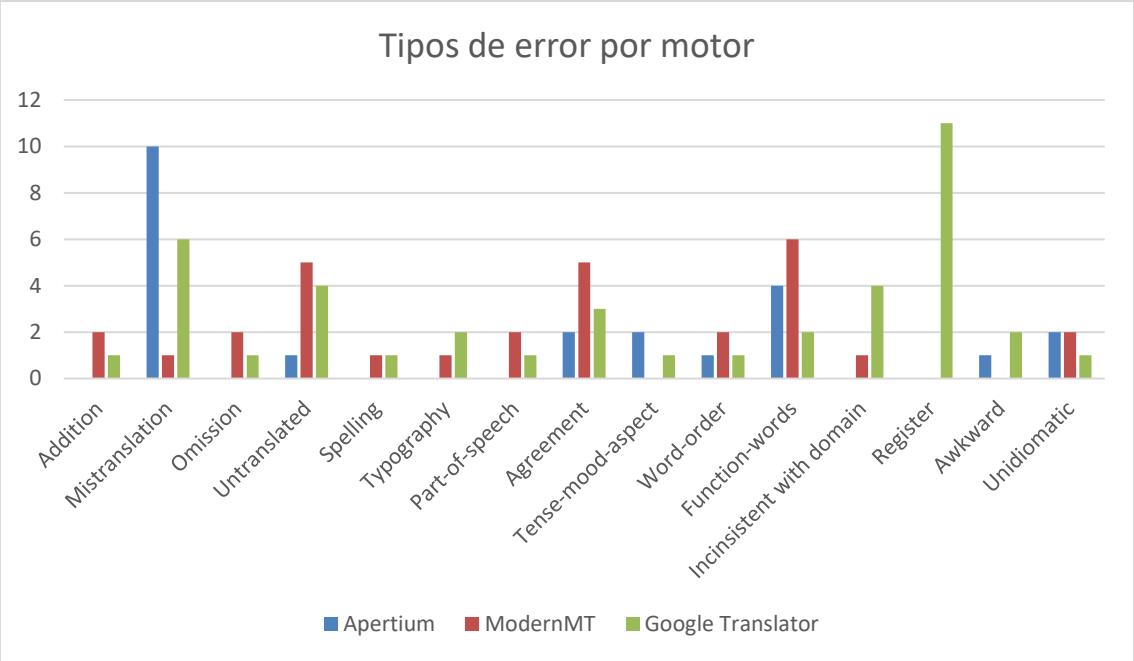


Ilustración 42. Comparación de los tipos de error en los tres motores

Los errores más frecuentes en todos los motores son las falsas traducciones, las palabras sin traducir, la falta de concordancia de género y número, los errores relativos a las palabras de función, el orden de palabras y las construcciones poco idiomáticas.

ModernMT y Google Translate coinciden en errores como adiciones, omisiones, ortografía, tipografía, partes del discurso e inconsistencias con la terminología del dominio.

Por otra parte, Apertium y Google Translate registran errores de concordancia verbal y construcciones raras.

Finalmente, el único motor con errores de registro es Google Translate. Así, como podemos ver, no hay ningún error que Google no cometa y los otros motores sí.

4. Análisis global de los resultados

Una vez realizado las evaluaciones por separado, consideramos que es necesario triangular los resultados para poder extraer conclusiones de relevancia.

4.1. Análisis global de los resultados de todo el texto

Las tres evaluaciones muestran resultados ajustados y ningún motor destaca especialmente por encima del otro, aunque sí se aprecia una diferencia evidente de

MMT y Apertium con respecto a Google. En la siguiente tabla comparativa se puede ver una comparación de los resultados obtenidos en cada una de las evaluaciones. A partir de los datos, se ha ordenado del 1 al 3 según el mejor rendimiento del motor.

Tabla 31. Comparación global de las tres evaluaciones

	Apertium	ModernMT	Google Translate
BLEU	1º	2º	3º
Encuesta	1º	2º	3º
Análisis de errores	1º	2º	3º

4.2. Análisis global de los segmentos muy cortos

De forma similar al análisis global del texto, este es el resultado de triangular los datos obtenidos en las evaluaciones realizadas sobre los segmentos muy cortos.

Tabla 32. Comparación de las tres evaluaciones en los segmentos muy cortos

	Apertium	ModernMT	Google Translate
BLEU	1º	2º	3º
Encuesta	1º	2º	3º
Análisis de errores	1º (empatado)	1º (empatado)	3º

4.3. Análisis global de los segmentos muy largos

Por último, cabe realizar la misma comparación en los segmentos muy largos para comprobar si se obtiene o no el mismo resultado.

Tabla 33. Comparación de las tres evaluaciones en los segmentos muy largos

	Apertium	ModernMT	Google Translate
BLEU	2º	1º	3º
Encuesta	1º	2º	3º
Análisis de errores	1º(empatado)	1º(empatado)	3º

CAPÍTULO IV

CONCLUSIONES

A lo largo del presente trabajo se han expuesto los conceptos teóricos y la relación existente entre traducción automática y lenguas minorizadas; se han trazado las estrategias existentes para optimizar los recursos de los que disponen y se ha conceptualizado el panorama actual de la traducción automática en lengua gallega. Esta contextualización ha sido necesaria para poder llevar a cabo el tipo de evaluación realizada en los distintos motores de traducción.

Dicho esto, y sin lugar a duda, podemos afirmar que la hipótesis de partida no se cumple, ya que tal y como se ha puesto de manifiesto en los resultados de las distintas evaluaciones, la traducción en bruto ofrecida por un motor neuronal general como Google Translate no supera, por el momento, las traducciones ofrecidas por Apertium y por MMT. En este sentido, Apertium y ModernMT obtienen unos resultados bastante próximos en la métrica BLEU, menos errores en MQM y son percibidos como mejores por los profesionales. Si bien, Apertium se sitúa unos puntos por encima de ModernMT. Asimismo, los resultados de las tres evaluaciones coinciden, lo que refuta todavía más el no cumplimiento de la hipótesis.

Si bien es cierto, se observa en cada motor una tendencia diferente en lo que respecta a errores de traducción, tal y como se adelantaba en la subhipótesis de partida. Mientras que en Apertium el error más común son las falsas traducciones, la falta de concordancia y errores en las construcciones verbales, en ModernMT se registra una mayor presencia de adiciones, omisiones, falsas traducciones y, finalmente, en Google, además de los errores presentes en ModernMT, se han detectado construcciones extrañas a la lengua, errores de registro, inconsistencias terminológicas y errores gramaticales. Por ejemplo, el principal error que comete Google son errores de terminología y de registro. Este tipo de error se ha podido evitar en los otros sistemas gracias al entrenamiento previo con la terminología y corpus relevantes para el tipo de texto evaluado. En cuanto a los errores de registro en Google, se podría decir que los datos empleados para entrenar su motor neuronal no están filtrados adecuadamente y mezclan textos de distintos registros, lo que puede llegar a confundir al motor y

provocar que, como hemos visto, en una misma frase se cambie el registro indiscriminadamente.

Se puede afirmar, por tanto, que cada motor tiene su punto fuerte y que la elección de uno u otro dependerá de los recursos textuales y económicos disponibles y del tipo de documento. Mientras Apertium construye frases con un mayor sentido gramatical gracias a la buena definición de las estructuras en la lengua de destino, ModernMT ofrece una mayor adecuación al contexto y un mejor estilo gracias al corpus específico con el que se entrenado y, tanto este como Google Translate consiguen una mayor fluidez en el texto, aunque cometen más falsas traducciones al alejarse más del original. La balanza dependerá, por tanto, de la calidad del corpus disponible para entrenar al motor estadístico o neuronal y de su tamaño.

La segunda hipótesis planteaba que la calidad de los motores sería peor en segmentos muy cortos y muy largos. Esta se cumple especialmente en los segmentos muy largos, ya que registran un mayor número de errores y un rendimiento inferior al resto de segmentos en la métrica BLEU y en la encuesta. Estos resultados ponen de manifiesto la necesidad de preedición del texto original para intentar solventar frases demasiado largas o con una construcción pobre que impidan el buen rendimiento de los sistemas.

Por todo lo expresado anteriormente, la conclusión principal sobre la evaluación de la traducción automática en el par de idiomas minorizados español-gallego es que, a pesar de que a priori la tecnología neuronal pueda partir con ventaja sobre los otros dos motores, esto no se cumple en una lengua minorizada porque no dispone del corpus especializado lo suficientemente grande como para poder arrojar un buen resultado. En el caso de Google, la posibilidad de partir de una lengua pivote (*zero shot translation*) no resuelve adecuadamente la falta de datos necesarios en la lengua minorizada y claramente queda un buen camino por recorrer para que el motor neuronal español-gallego ofrezca resultados aceptables. En este sentido, se abre la puerta a la creación de motores personalizados español-gallego de Google u otro sistema de TA neuronal entrenados específicamente para un campo de especialidad. Si bien, serían necesario un mayor volumen de corpus que en el caso de un motor estadístico.

A la par de esta conclusión, se pueden extraer dos más. En primer lugar, el factor determinante de éxito de un motor de traducción, independientemente de su naturaleza, es una buena planificación y un entrenamiento adecuado a las necesidades textuales del

documento. En segundo lugar, queda patente la necesidad de acceso a los recursos necesarios para poder llevar a cabo el entrenamiento necesario en los tres sistemas. En este sentido, cabe resaltar el carácter primordial de los esfuerzos de la comunidad universitaria para seguir avanzado en la investigación de la TA en las lenguas minorizadas.

Parece claro, por tanto, concluir que cualquier línea de investigación futura sobre TA y lenguas minorizadas debe centrarse en la búsqueda y optimización de recursos, tanto tecnológicos como textuales. A pesar de que ya se ha iniciado el camino, las tecnologías de procesamiento del lenguaje natural todavía no están al servicio de las lenguas minorizadas y las futuras investigaciones deben centrar sus esfuerzos en ampliar la democratización de estas, en proponer soluciones para romper con la escasez de recursos y en aprovechar los nuevos descubrimientos para ponerlos al servicio de la TA.

Para finalizar, cabe mencionar que los datos extraídos en la encuesta realizada abren también la puerta a la investigación sobre el perfil de los poseditores para determinar con qué segmentos prefieren trabajar y cuáles son los tipos de errores que rechazan. No se ha profundizado en este análisis porque estaba fuera de los objetivos de este trabajo, pero, desde luego, se trataría de una investigación muy útil para ayudar a las empresas de traducción a adelantarse a posibles rechazos de encargos de posedición y a utilizar los análisis de errores para corregir y entrenar los motores según sus necesidades específicas.

CAPÍTULO V

BIBLIOGRAFÍA

Agerri, R., Gómez Guinovart, X., Rigau, G. & Solla Portela, M. A. (2018). Developing New Linguistic Resources and Tools for the Galician Language. *Proceedings of the 11th Language Resources and Evaluation Conference (LREC'18)*: 2322-2325.

Agerri, R.; Bel, N.; Rigau, G. & Saggion, H. (2018). TUNER: Multifaceted Domain Adaptation for Advanced Textual Semantic Processing. First Results Available. *Procesamiento del Lenguaje Natural*, 61: 163-166

Agirre, E., Aldezabal, I., Alegria, I., Arregi, X., Arriola, J. M., Artola, X. et al. (2001). Developing Language Technology for a Minority Language: Progress and Strategy. *ELSNews 10.1. Special Issue on Minority Languages*.

ALPAC. (1966). *Language and Machines. Computers in Translation and Linguistics.* National Academy of Sciences, National Research Council. Washington, D.C. < http://www.mt-archive.info/ALPAC-1966.pdf> [Consulta: 20 de diciembre de 2018]

Alegria, I., Artola, X., Diaz de Harraza, A. & Sarasola, K. (2011). Strategies to develop Language Technologies for Less-Resourced Languages based on the case of Basque. *The 5th Language & Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics*. 25-27 de noviembre de 2011, Poznań, Polonia.

Armentano-Oller, C., Corbí-Bellot, A.M., Forcada, M.L., Ginestí-Rosell, M., Bonev, B., Ortiz-Rojas, S. et al. (2005) An open-source shallow-transfer machine translation toolbox: consequences of its release and availability. En *Proceedings of OSMaTran: Open-Source Machine Translation, A workshop at Machine Translation Summit X*, 12-16 de septiembre, Phuket, Tailandia.

Armentano-Oller, C & Forcada, M. L. (2006). Open-source machine translation between small languages: Catalan and Aranese Occitan. *Strategies for developing*

machine translation for minority languages (5th SALTMIL workshop on Minority Languages). 22-28 de mayo, p. 51-54.

Armentano-Oller, C., Carrasco, R. C., Corbí-Bellot, A. M., Forcada, M. L., Ginestí-Rosell, M., Ortiz-Rojas, S. et al. (2006). Open-source Portuguese-Spanish machine translation. En *Computational Processing of the Portuguese Language: 7th Workshop on Computational Processing of Written and Spoken Portuguese,* PROPOR. Lecture Notes in Artificial Intelligence 3960. Springer-Verlag, 50–59

Armentano-Oller, C., Corbí-Bellot, A. M., Forcada, M. L., Ginestí-Rosell, M., Ortiz-Rojas, S. et al. (2007). Proceedings of FLOSS (Free/Libre/Open Source Systems) International Conference. 7-9 de marzo, Jerez de la Frontera, 5-20.

Arnold, D., Balkan, L., Meijer, S., Humphreys, R. L. & Sadler, L. (1994). *Machine Translation: An Introductory Guide.* Londres: Blackwells-NCC.

Austin, P. K. & Sallabank, J. (2011). *The Cambridge Handbook of Endangered Languages.* Cambridge: Cambridge University Press.

Austin, P. K. & Sallabank, J. (2013). Endangered languages: an introduction. *Journal of Multilingual and Multicultural Development,* 34 (4): 313-316.

Aydin, B. & Özgür, A. (2014). Expanding Machine Translation Training Data with an Out-of-Domain Corpus using Language Modeling based Vocabulary Saturation. En Yaser Al-Onaizan & Michel Simard (Eds.). *Proceedings of AMTA 2014: MT Researches,* 1: 180-192.

Babych, B. & Hartley, A. (2004). Extending the BLEU MT Evaluation Metric with Frequency Ratings. *Proceedings of ACL 2004* (42nd Annual Meeting of the Association of Computational Linguistics). Barcelona, España, 621-628.

Bengio, Y., Ducharme, R., Vincent, P. & Janvin, C. (2003). A neuronal probabilístic Language model. *Journal of Machine Learning Research*, vol. 3, 1137-1155.

Bentivogli, L., Bisazza, A., Cettolo, M. & Federico, M. (2016). Neural versus Phrase-Based Machine Translation Quality: a Case Study. *Proceedings of Conference on Empirical Methods in Natural Language Processing* (EMNLP), 257-267.

Berner, S. (2003). “Lost in Translation”: cross-lingual communication, and virtual academic communities. *5th Annual Conference on World Wide Web Applications*, 10-12 de septiembre de 2003, Durban, Sudáfrica.

Bertoldi, N., Cattoni, R., Cettolo, M., Farajian, A. & Federico, M. (2017). MMT: New Open Source MT for the Translation Industry. *The 20th Annual Conference of the European Association for Machine Translation (EAMT)*.

Bowker, L. (2002). *Computer-Aided Translation Technology*. Ottawa: University of Ottawa Press.

Bowker, L. (2008). Official Language Minority Communities, Machine Translation, and Translator. TTR: 21(2), pp. 15-61.

Bowker, L. (2015). Computer-aided translation. Translator training. En Chan Sin-Wai (Ed.). *The Routledge Encyclopedia of Translation Technology* (pp. 88-104). Londres: Routledge.

Branchadell, A. & West, L. M. (Eds.) (2005). *Less Translated Languages*. Ámsterdam/Filadelfia: John Benjamins.

Branchadell, A. (2011). Minority languages and translation. En Yves Gambier & Luc Van Doorslaer. *Handbook of Translation Studies*, vol. 2 (pp. 97-101). Ámsterdam: John Benjamins.

Burchardt, A. & Lommel, A. (2014). Practical Guidelines for the Use of MQM in Scientific Research on Translation Quality. < http://www.qt21.eu/downloads/MQM-usage-guidelines.pdf > [Descargado el 18 de mayo de 2019].

Burschardt, A., Macketanz, V., Dehdari, J., Higold, G., Peter, J-T. & Williams, P. (2017). A Linguistic Evaluation of Rule-Based, Phrase-Based, and Neural MT Engines. *The Prague Bulletin of Mathematical Linguistics*, 108: 159-170.

Casacuberta, F. & Peris, A. (2017). Traducción automática neuronal. *Revista Tradumàtica*, 15: 66-74.

Callison-Burch, C., Osborne, M., Koehn, P. (2006). Re-evaluating the role of BLEU in Machine Translation Research". *Proceedings of EACL 2006* (11th Conference of the European Chapter of the Association of Computational Linguistics). Trento, Italia, 249-246.

Castilho, S., Moorkens, J., Gaspari, F., Sennrich, R., Sosoni, V., Georgakopoulou, Y. … & Gialama, M. (2017a). A Comparative Quality Evaluation of PBSMT and NMT using Professional Translators. *Proceedings of the MT Summit XVI, 1*: 116-131.

Castilho, S., Moorkens, J., Gaspari, F., Calixto, I., Tinsley, J. & Way, A. (2017b). Is Neural Machine Translation the New State-of-the-Art? *The Prague Bulletin of Mathematical Linguistics*, 108: 109-120.

Coughlin, D. (2003). Correlating Automated and Human Assessments of Machine Translation Evaluation. *Online Proceedings of the MT Summit IX* (2003). Nueva Orleans.

Cronin, M. (1995). Altered States: Translation and Minority Languages. *TTR* 8, 85-103.

Díaz Fouces, O. (2005). Translation policy for minority languages in the European Union: Globalisation and resistance. En Albert Branchadell & Lovell Margaret West (Eds.). *Less Translated Languages*. Ámsterdam/Filadelfia: John Benjamins.

Diz Gamallo, I. (2001). The importance of MT for the survival of minority languages: Spanish-Galician MT system. *Proceedings of MT Summit VIII*, Noviembre 2001, España.

Doddington, G. (2002). Automatic Evaluation of Machine Translation Quality Using N-gram Cooccurrence Statistics. *Proceedings of Human Language Technology Conference* (HLT-02), San Diego, CA, 138−145.

Doğru, G., Martín-Mor, A. & Aguilar-Amat, A. (2018). Parallel Corpora Preparation for Machine Translation of Low-Resource Languages: Turkish to English Cardiology Corpora. *MultilingualBIO: Multilingual Biomedical Text Processing*: 12-16.

Doherty, S., O'Brien, S &. Carl, M. (2010). Eye tracking as an MT evaluation technique. *Machine translation*, 24 (1): 1-13.

Fiederer, R., & O'Brien, S. (2009). Quality and Machine Translation: A realistic objective? *JoSTrans. The Journal for Specialised Translation*, 52–74.

Fields, P, Hague, D., Koby, G.S., Lommel, A & Melby, A. (2014). "What Is Quality? A Management Discipline and the Translation Industry Get Acquainted". *Revista Tradumàtica, Translation and Quality*,12: 404-412.

Flanagan, M. (1994) Error Classification for MT Evaluation. *Proceedings of the 1st Conference of the Association for Machine Translation in the Americas: Technology Partnerships for Crossing the Language Barrier* (AMTA-94), Columbia, MD, 65−72.

Folaron, D. (2015). Translation and minority, lesser-used and lesser-translated languages and cultures. *JOSTRANS: The Journal of Specialised Translation*, 24: 16-27.

Forcada, M. (2009). Apertium: traducció automàtica de codi obert per a les llengües romàniques. *Linguamática*, *1*(1): 13-23.

Forcada, M. (2015). Open-source Machine Translation Technology. En Chan Sin-Wai (Ed.). *The Routledge Encyclopedia of Translation Technology* (pp. 152–166). Abingdon, Nueva York: Routledge.

Forcada, M. (2017). Making sense of neural machine translation. *Translation Spaces*, 6: 291-309.

Gamallo, P. & Garcia, M. (2017). Linguakit: a multilingual tool for linguistic analysis and information extraction. *Linguamática*, 9(1): 19–28.

García Mateo, C. & Arza Rodríguez, M. (2012). *O idioma galego na era dixital*. Berlín: Springer Heidelberg.

Germann, U. (2015). Sampling phrase tables for the Moses statistical machine translation system. *The Prague Bulletin of Mathematical Linguistics*, 104: 39–50.

Gómez Guinovart, X. (2009). Tecnologies lingüístiques de la llengua gallega. *Llengua, Societat i Comunicació*, 7: 27-34.

Gómez Guinovart, X. & López Fernández, S. (2009). Anotación morfosintáctica do Corpus Técnico do Galego. *Linguamática*, 1(1): 61–71.

Gómez Guinovart, X. (2011). Galego 3.0: Novas oportunidades e desafíos para a investigación lingüística. *Grial*, XLIX/191: 28-33.

Gómez Guinovart, X. & Solla Portela, M. A. (2017). Building the Galician wordnet: methods and applications. *Language Resources and Evaluation,* 52 (1): 317–339.

Gornostay, T., Gojun, A., Weller, M. et al. (2012). Terminology Extraction, Translation Tools and Comparable Corpora: TTC concept, midterm progress and achieved results. *Workshop on Creating Cross-language Resources for Disconnected Languages and Styles*: 35-38.

Hamon, O. et al. (2007). Assessing Human and Automated Quality Judgments in the French MT-Evaluation Campaign CESTA. *Proceedings of The Journal of Specialised Translation Issue 11 - January 2009 71 Machine Translation Summit XI*. 10-14 de septiembre. Copenhage, Dinamarca. 231-238.

Johnson, M. Schuster, M. Le, Q. V., Krikun, M. Wu, Y. et al. (2017). Google's Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation. *Transactions of the Association for Computational Linguistics*, 5 (1): 339-351.

Hutchins, W. J. (2014). The history of machine translation in a nutshell. Recuperado de: http://www.mt-archive.info/10/Hutchins-2014.pdf.

Hutchins, W. J. & Somers, H. L. (1992). *An introduction to machine translation*. Londres: Academic Press.

Iglesias Iglesias, G., Rodríguez Liñares, L., Rodríguez Banga, E., Campillo Díaz, F. L. & Méndez Pazó, F. (2010). Perspectivas de la traducción automática castellano-gallego mediante técnicas estadísticas y por transferencia. *IV Jornadas en Tecnología del Habla*, 8-10 de noviembre de 2006, Zaragoza. pp. 111-116.

Keegan, T. T. & Manuirirang, H. (2011). Minority languages & translation technologies case study: te reo Māori & Google Translator Toolkit. *Proceedings of the Thirty-third International Conference on Translating and the Computer*, 17-18 de noviembre de 2011, Londres.

King, M., Popescu-Belis, A. & Hovy, E. (2003). FEMTI: Creating and Using a Framework for MT Evaluation. *Proceedings of Machine Translation Summit IX*, 23-27 de septiembre, Nueva Orleans. 224-231.

Kit, C. & Wong, B. (2015). Evaluation in Machine Translation and Computer Aided Translation. En Chan Sin-Wai (Ed.). *The Routledge Encyclopedia of Translation Technology* (pp. 213–236). Abingdon, Nueva York: Routledge.

Koby, G., Fields, P., Hague, D., Lommel, A. & Melby, A. (2014). "Defining Translation Quality". *Revista Tradumàtica, Translation and Quality*, 12: 413-420.

Lauscher, S. (2000). Translation Quality Assessment. *The Translator*, *6*(2): 149–168.

Linguistics Data Consortium [LDC]. (2002). *Linguistic data annotation specification: Assessment of fluency and adequacy in translations.* <https://catalog.ldc.upenn.edu/docs/LDC2003T17/TransAssess02.pdf> [Consultado el 27 marzo de 2019].

Loinaz, I., Arantzabal, I., Forcada, M., Gómez Guinovart, X., Padró, L., Pichel Campos, J. R., & Waliño, J. (2006). OpenTrad: Traducción automática de código abierto para las lenguas del Estado español. *Procesamiento del Lenguaje Natural*, *37*, 357–360.

Lommel, A. & Melby, A. K. (2014). *Multidimensional Quality Metrics (MQM) Issue types. Version 0.2.0 (2014-08-19)* <http://www.qt21.eu/mqm-definition/issues-list-2014-08-19.html> [Consultado el 27 marzo de 2019].

Lommel, A. et al. (2014). Multidimensional Quality Metrics (MQM): A Framework for Declaring and Describing Translation Quality Metrics. *Revista Tradumàtica: tecnologies de la traducció*, 12, 455-463

Lommel, A. et al. (2015a). *Multidimensional Quality Metrics (MQM) Definition. Version 0.3.0 (2015-01-20)* <http://www.qt21.eu/mqm-definition/definition-2015-01-20.html> [Consultado el 27 marzo de 2019].

Lommel, A.et al. (2015b). *Multidimensional Quality Metrics (MQM) Definition. Version 0.9.3 (2015-06-16)* <http://www.qt21.eu/mqm-definition/definition-2015-06-16.html> [Consultado el 27 marzo de 2019].

Lommel, A. et al. (2015c). Multidimensional Quality Metrics (MQM) Issue types. Version 0.5.1 (2015-06-16) <http://www.qt21.eu/mqm-definition/issues-list-2015-05-27.html> [Consultado el 27 marzo de 2019].

López Pereira, A. (2018). Determining translators' perception, productivity and post-editing effort when using SMT and NMT systems. En: Juan Antonio Pérez-Ortiz et al. (Eds.). *Proceedings of the 21st Annual Conference of the European Association for Machine Translation: 28-30 May 201* (p. 327). Alicante: Universitat d'Alacant.

Malvar Fernández, P., Pichel Campos, J. R., Senra Gómez, O., Gamallo Otero, P. & García, A. (2010) Vencendo a escassez de recursos computacionais. Carvalho: Tradutor Automático Estatístico Inglês-Galego a partir do corpus paralelo Europarl Inglês-Português. *Linguamatica*, 2(5): 31–38.

Marimon, M.; Vivaldi, J & Bel, N. (2017). Annotation of negation in the IULA Spanish clinical record corpus. *Proceedings of the Workshop SemBEaR* 2017. ACL. p. 43-52.

Martí, M. A. (2009). Introducció. Les tecnologies de la llengua i les llengües minoritzades. *Llengua, Societat i Comunicació, 7:* 1–2.

Martín-Mor, A. (2017). Technologies for endangered languages: The languages of Sardinia as a case in point. *mTm journal*, 9: 365-386.

Martinez Calvo, M. L. (2002). O ES-GA, o valor dun sistema de traducción automática nunha lingua minorizada. En María Xesús Bugarín López et al. *Actas da VIII Conferencia Internacional de Linguas Minoritarias: Santiago de Compostela, 22, 23, 24 de novembro de 2001 (pp.* 171-174). Santiago de Compostela: Dirección Xeral de Política Lingüística.

Melby, A., Fields, P., Hague, D., Koby, G. S. & Lommel, A. (2014). "Defining the Landscape of Translation". *Revista Tradumàtica, Translation and Quality*, 12: 392-403.

Mercader-Alarcón, J. & Sánchez-Martínez, F. (2016). Analysis of translation errors and evaluation of preediting rules for the translation of English news texts into Spanish with Lucy LT. *Revista Tradumàtica: tecnologies de la traducció*, 14: 172-186.

Millán-Varela, C. (2000). Translation, Normalisation and Identitity in Galicia(n). Target, 12 (2): 267-282.

Moorkens, J. & Way, A. (2016). Comparing Translators Acceptability of TM and SMT Outputs. Baltic J. *Modern Computing*, 4(2): 141-151.

O'Brien, S. (2012). Towards a dynamic Quality Evaluation Model for Translation. *The Journal of Specialised Translation*, 17: 55–77.

Otegi, A.; Imaz, O.; Díaz de Ilarraza, A.; Iruskieta, M. & Uria, L. (2017). ANALHITZA: a tool to extract linguistic information from large corpora in Humanities research. *Procesamiento del Lenguaje Natural*, 58: 77-84.

Padró, L. & Stanilovsky, E. (2012). Freeling 3.0: Towards wider multilinguality. In Nicoletta Calzolari et al. (Eds.). *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC-2012)*, pp. 2473–2479, Estámbul, Turquía.

Papineni, K., Roukos, S., Ward, T. & Zhu, W. (2001). *BLEU: A Method for Automatic Evaluation of Machine Translation*, IBM Research Report RC22176 (W0109−022).

Pérez, N., Cuadros, M. & Rigau, G. (2018). Biomedical term normalization of EHRs with UMLS. *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018).*

Pichel Campos, J. R., Malvar Fernández, P., Senra Gómez, O., Gamallo Otero, P. & García González, A. (2009). Carvalho: English-Galician SMT system from EuroParl English-Portuguese parallel corpus. *Procesamiento del Lenguaje Natural*, 23: 379-381.

Popović, M. & Ney, H. (2011). Towards automatic error analysis of machine translation output. *Computational Linguistics*, 37(4): 657–688.

Popović, M. (2017). Comparing Language Related Issues for NMT and PBMT between German and English. *The Prague Bulletin of Mathematical Linguistics*, 108: 209-220.

Saggion, H.; Ronzano, F.; Accuosto, P. & Ferrés, D. (2017). MultiScien: a Bi-Lingual Natural Language Processing System for Mining and Enrichment of Scientific Collections. BIRNDL@SIGIR (1) 2017: 26-40

Saldanha, G., & O'Brien, S. (2014). *Research methodologies in translation studies*. Londres: Routledge.

Sánchez-Gijón, P., Moorkens, J. & Way, A. (2019). Perception vs. Acceptability of TM and SMT Output: What do translators prefer? En: Juan Antonio Pérez-Ortiz et al. (Eds.). *Proceedings of the 21st Annual Conference of the European Association for Machine Translation: 28-30 May 2018* (p. 331). Alicante: Universitat d'Alacant.

Sánchez Martínez, F. (2008). *Empleo de métodos no supervisados basados en corpus para construir traductores automáticos basados en reglas* (Tesis). Alicante: Universitat d'Alacant.

Sánchez-Martínez, F., Armentano Oller, C., Pérez-Ortiz, J. A. & Forcada Zubizarreta, M. (2007). Training part-of-speech taggers to build machine translation systems for less-resourced language pairs. En: *Procesamiento del lenguaje natural, 39: 257-264.*

Scannell, K. P. (2007). The Crúbadán Project: Corpus building for underresourced languages. *Building and Exploring Web Corpora: Proceedings of the 3rd Web as Corpus Workshop*, 4: 5–15.

Snover, M., Dorr, B. J., Schwartz, R. Micciulla, L.& Makhoul, J. (2006). A Study of Translation Edit Rate with Targeted Human Annotation. *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas: Visions for the Future of Machine Translation* (AMTA-06), Cambridge, MA, 223–231.

Solla Portela, M. A. & Gómez Guinovart, X. (2015). Termonet: Construcción de terminologías a partir de WordNet y corpus especializados. *Procesamiento del Lenguaje Natural*, 55:165–168.

Solla Portela, M. A. & Gómez Guinovart, X. (2016). Dbpedia del gallego: recursos y aplicaciones en procesamiento del lenguaje. *Procesamiento del Lenguaje Natural*, 57:139–142.

Solla Portela, M. A. & Gómez Guinovart, X. (2017). Diseño y elaboración del corpus SemCor del gallego anotado semánticamente con Wordnet 3.0. *Procesamiento del Lenguaje Natural*, 59:137–140.

Somers, H. L. (1997). Machine Translation and Minority Languages. *19th ASLIB Translating & the Computer Conference*, 1–13.

Somers, H. L. (2003). *Computers and Translation: A translator's guide*. Ámsterdam, John Benjamins.

Somers, H. L. (2003). Translation technologies and minority languages. En Harold Somers (Ed.). *Computers and translation: A translator's guide* (pp. 87-103). Ámsterdam/Filadelfia, PA: John Benjamins.

Somers, H. L. (2006). Machine translation. En Mona Baker & Kirsten Malmkjær (eds.). *Routledge Encyclopedia of Translation Studies* (pp. 136-149). Londres: Routledge.

Toral, A. & Sánchez-Cartagena, V. M. (2017). A Multifaceted Evaluation of Neural versus Statistical Machine Translation for 9 Language Directions. En Mirella Lapata et al. (eds.). *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 1063-1073). Valencia: Association for Computational Linguistics.

Tyers, F. M., Alòs i Font, H., Fronteddu, G. & Martín-Mor, A. (2017). Rule-Based Machine Translation for the Italian-Sardinian Language Pair. *The Prague Bulletin of Mathematical Linguistics*, 108 (1): 221–232.

Vacalopoulou, A., Giouli, V., Efthimiou, E. & Giagkou, M. (2012). Bridging the gap between disconnected languages: the eMiLang multi-lingual database. *Workshop on Creating Cross-language Resources for Disconnected Languages and Styles*: 1-6.

White, J. S. (2003). How to Evaluate Machine Translation. En Harold L. Somers (ed.). *Computers and Translation: A Translator's Guide*. Ámsterdam/Filadelfia: John Benjamins Publishing Company, 211–244.

Williams, M. (2009). Translation quality assessment. *Mutatis Mutandis*, 2(1), 3–23.

Wu, Y., Schuster, M., Chen, Z., Le, Q. V. & Norouzi, M. (2016). Google's Neural Machine Translation System: Bridging. *ArXiv*:1609.08144.

Zapata, I. (1995). Un recorrido por algunos principios de Traducción Automática. En Germán Ruipérez. *Enseñanza de lenguas y traducción con ordenadores* (pp. 94-97). Madrid: Ediciones pedagógicas.

HERRAMIENTAS

Analthitza (http://ixa2.si.ehu.es/clarink/analhitza.php)

AsisTerm (http://scientmin.taln.upf.edu/scielo)

CLUVI (http://sli.uvigo.es/CLUVI/)

Galnet (http://sli.uvigo.gal/galnet/)

IULA (http://eines.iula.upf.edu/brat/#/NegationOnCR_IULA)

Lingaliza (http://sli.uvigo.gal/lingaliza/)

PDFIngest (http://taln.upf.edu/pdfdigest)

Proyecto Tuner (http://ixa2.si.ehu.es/tuner/)

QTLaunchPad (http://www.qt21.eu/launchpad/content/new-goal-quality-translation.html)

SensoGal (http://sli.uvigo.gal/SensoGal/)

SLI. Seminario de Lingüística Informática. http://sli.uvigo.gal/recursos.html

SouceForge (sourceforge.net)

TALG. Tecnoloxías e Aplicacións da Lingua Galega. http://talga.webs.uvigo.es/

TILDE. Tilde Custom Machine Translation. https://www.letsmt.eu/Bleu.aspx

UML Mapper (http://demosv2.vicomtech.org/umlsmapper)

CAPÍTULO VI

Anexo I: texto original

Añade a estos momentos todo el bienestar que se merecen

El confort siempre es bienvenido. Instala ahora el gas y disfruta de todas sus ventajas desde el primer momento: calefacción homogénea en todos los rincones de tu casa, duchas de agua caliente para todos y la tranquilidad de usar la energía con los precios más estables del mercado. Además, la instalación es más rápida y fácil de lo que imaginas y enseguida notarás todos sus beneficios. Tenemos una oferta que se adapta a tus necesidades para que puedas disfrutar desde este momento de todas las ventajas del gas. Por eso, próximamente uno de nuestros comerciales pasará por tu vivienda para facilitarte toda la información necesaria. Si prevés no estar en casa, puedes llamarnos y solicitar un presupuesto sin compromiso.

Instala ahora el gas

y te regalamos 100 €

Instala el gas

y disfruta de calor homogéneo en toda la casa

Hemos venido a explicarte las ventajas que tiene instalar el gas en tu hogar.

Instalar el gas ahora tiene más ventajas y además, disfrutarás de todo el bienestar y confort en tu hogar con una energía muy económica.

• El gas es actualmente una de las energías más baratas del mercado que te garantiza el máximo confort en tu hogar y además se paga una vez la has consumido.

• La instalación será rápida y sencilla. En pocos días y prácticamente sin obras lo tendrás instalado.

• Disfrutarás de la calefacción que proporciona calor homogéneo en toda la casa, sin contraste de temperaturas entre habitaciones.

• Dispondrás de agua caliente al instante, siempre que la necesites y al mejor precio.

• Al ser una energía de suministro continuo, no ocupa espacio en tu hogar porque no hay que almacenarla ni necesita reposición.

Consulta a tu instalador y descubre todas las ventajas de disfrutar del gas.

Oferta sujeta a disponibilidad de gas en la zona y en la finca, para contrataciones de nuevos puntos de suministro en fincas con más de 5 años de antigüedad y de puntos de suministro existentes inactivos o cesados durante más de 2 años, que se conecten a la red de distribución de las empresas distribuidoras de gas pertenecientes al grupo (ver www.páginaweb.com), válida para Solicitudes de Conexión a Red hechas entre el 01/10 y el 31/12 de 2012 y puestas en servicio antes del 30/06/2013. Los 100 € de regalo se ingresarán directamente por transferencia en la cuenta bancaria indicada por el cliente, tras la puesta en servicio del gas.

Las condiciones económicas anteriormente mencionadas solo serán efectivas en caso que el punto de suministro de gas y/o electricidad cumpla con los requisitos de tarifa establecidos para la promoción. En caso de no cumplirse los mencionados requisitos se aplicarán los precios sin descuentos.

¿Confirma su voluntad de contratar con LA EMPRESA, según las condiciones de la oferta realizada, así como su voluntad de causar baja en los suministros anteriores y autoriza a LA EMPRESA a realizar a tal efecto las gestiones necesarias con su empresa distribuidora?

GAS CERRADO

Señores Clientes:

Les informamos que a causa de:

Fuga de gas.

Defecto en la instalación común.

Otros.

el suministro de gas a esta finca queda momentáneamente interrumpido.

La empresa está gestionando la reparación de esta instalación para que pueda restablecerse el suministro en el plazo más breve posible.

Entretanto, mantengan cerradas todas las llaves de paso de gas, tanto la general de sus viviendas como las de sus aparatos de consumo.

Les agradecemos su colaboración, a la vez que rogamos disculpen las molestias que por este motivo podamos ocasionar.

Muchas gracias

Engade a estes momentos

todo o benestar que se merecen

O confort sempre é benvido. Instala agora o gas e goza de todas as súas vantaxes dende o primeiro momento: calefacción homoxénea en todos os recunchos da túa casa, duchas de auga quente para todos e a tranquilidade de usar a enerxía cos prezos máis estables do mercado.

Ademais, a instalación é máis rápida e doada do que imaxinas e deseguido notarás todos os seus beneficios.

Temos unha oferta que se adapta ás túas necesidades para que poidas gozar dende este momento de todas as vantaxes do gas. Por iso, proximamente un dos nosos comerciais pasará pola túa vivenda para facilitarche toda a información necesaria. Se prevés non estar na casa, podes chamarnos e solicitar un orzamento sen compromiso.

Instala agora gas

e regalámosche 100 €

Instala gas

e goza de calor homoxénea en toda a casa

Viñemos explicarche as vantaxes que ten instalar o gas natural no teu fogar.

Instalar o gas natural agora ten máis vantaxes e ademais, gozarás de todo o benestar e confort no teu fogar cunha enerxía moi económica.

- O gas natural é actualmente unha das enerxías máis baratas do mercado que che garante o máximo confort no teu fogar e ademais págase unha vez consumido.
- A instalación será rápida e sinxela. En poucos días e practicamente sen obras teralo instalado.
- Gozarás da calefacción que proporciona calor homoxénea en toda a casa, sen contraste de temperaturas entre habitacións.

- Disporás de auga quente ao instante, sempre que a necesites e ao mellor prezo.
- Ao ser unha enerxía de subministración continua, non ocupa espazo no teu fogar porque non a hai que almacenar nin precisa reposición.

Consulta o teu instalador e descubre todas as vantaxes de gozar do gas natural.

Oferta suxeita a dispoñibilidade de gas na zona e no edificio, para contratacións de novos puntos de subministración en edificios con máis de 5 anos de antigüidade e de puntos de subministración existentes inactivos ou cesados durante máis de 2 anos, que se conecten á rede de distribución das empresas distribuidoras de gas pertencentes ao grupo (ver www.páxinaweb.com), válida para Solicitudes de Conexión a Rede feitas entre o 01/10 e 31/12 de 2012 e postas en servizo antes do 30/06/13. Os 100 € de agasallo ingresaranse directamente por transferencia na conta bancaria indicada polo cliente, coincidindo coa posta en servizo do gas.

As condicións económicas anteriormente mencionadas só serán efectivas no caso de que o punto de subministración de gas e/ou electricidade cumpra cos requisitos de tarifa establecidos para a promoción. No caso de non cumprirse os mencionados requisitos, aplicaranse os prezos sen os descontos.

Confirma a súa vontade de contratar con A EMPRESA (segundo contratase o cliente), segundo as condicións da oferta realizada, así como a súa vontade de causar baixa nas subministracións anteriores e autoriza A EMPRESA a realizar a tal efecto as xestións necesarias coa súa empresa distribuidora?

GAS PECHADO

Estimados clientes:

Informámolos de que por mor de:

Fuga de gas.

Defecto na instalación común:

Outros.

o abastecemento de gas deste predio queda momentaneamente interrompido.

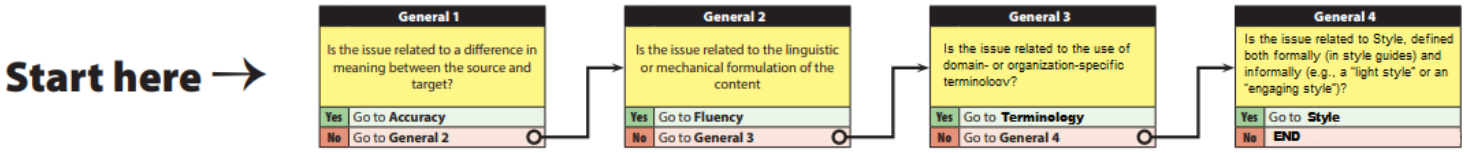
A empresa está a xestiona-la reparación desta instalación para que se poida restablece-lo abastecemento o antes posible.

Entrementres, manteñan pechadas tódalas chaves de paso de gas, tanto a xeral das súas vivendas como as dos seus aparellos de consumo.

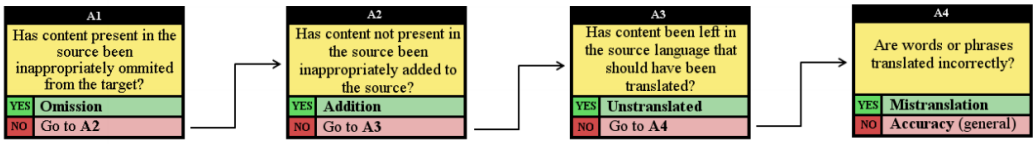
Agradecémoslles a súa colaboración, ó tempo que lles rogamos que desculpen as molestias que por este motivo lles poidamos ocasionar.

Moitas grazas.

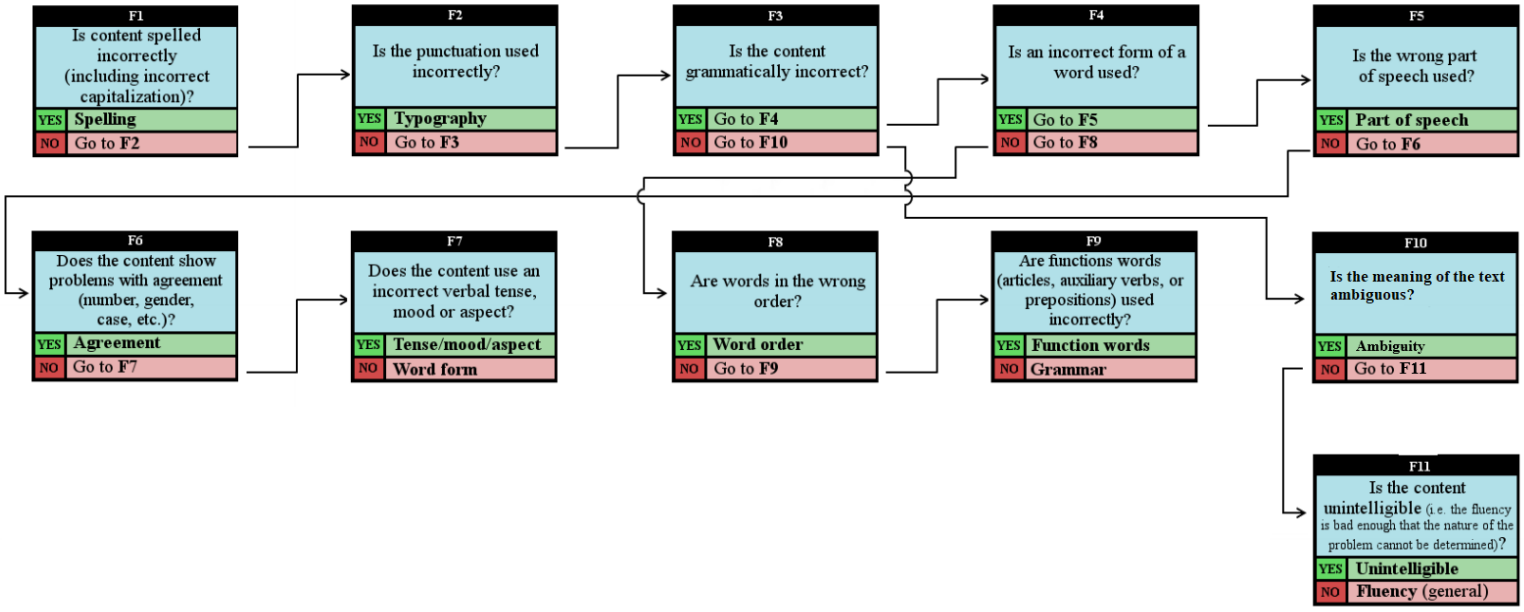
Anexo III: diagrama de decisiones (MQM)



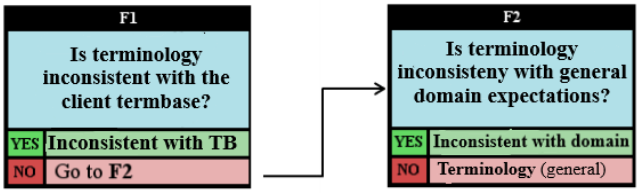
Accuracy



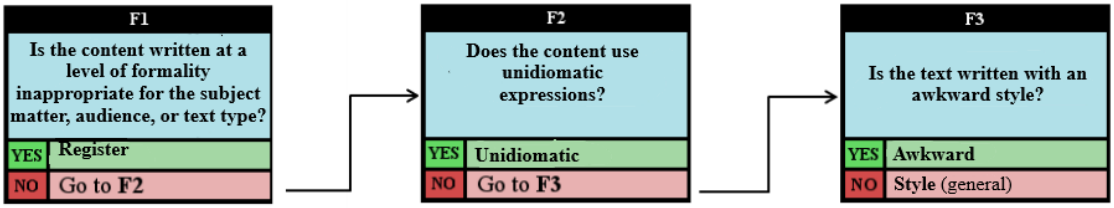
Fluency



Terminology



Style



Anexo IV: ejemplos de errores detectados

TO: Las condiciones económicas anteriormente mencionadas solo serán efectivas en caso que el punto de suministro de gas y/o electricidad cumpla con los requisitos de tarifa establecidos para la promoción. En caso de no cumplirse los mencionados requisitos se aplicarán los precios sin descuentos.

TT Apertium: As condicións económicas anteriormente mencionadas só serán efectivas en caso que o punto de subministración de gas e / ou electricidade cumpra cos requisitos de tarifa establecidos paira a promoción. En caso de non cumprirse os mencionados requisitos aplicaranse os prezos sen descontos.

- *En caso que* se traduce al gallego como *No caso de que*. En este caso, falta el artículo determinado necesario para formar la contracción preposición en + o y la preposición de.
- A lo largo de todo el texto, se confunde la preposición para con una forma verbal.

TT Google: As condicións económicas antes mencionadas só serán efectivas se o punto de subministración de gas e / ou electricidade cumpre cos requisitos de tarifas establecidos para a promoción. Se non se cumpran os requisitos anteriormente mencionados, aplicaranse os prezos sen descontos.

- Forma verbal y construcción gramatical incorrecta. Lo correcto hubiese sido *de non se cumpriren*.

TO: ¿Confirma su voluntad de contratar con LA EMPRESA, según las condiciones de la oferta realizada, así como su voluntad de causar baja en los suministros anteriores y autoriza a LA EMPRESA a realizar a tal efecto las gestiones necesarias con su empresa distribuidora?

TT MMT: Confirma a súa vontade de contratar con DA EMPRESA, segundo as condicións da oferta realizada, así como a súa vontade de causar baixa na subministracións anteriores e autoriza DA EMPRESA a realizar para ese efecto as xestións necesarias coa súa empresa distribuidora?

- Adición innecesaria de la preposición de en el primer caso y preposición incorrecta en el segundo caso. La preposición correcta que rige autoriza es *a* que contraería con el artículo y formaría *á*.

TO: El gas es actualmente una de las energías más baratas del mercado que te garantiza el máximo confort en tu hogar y además se paga una vez la has consumido

TT Apertium: o gas é actualmente una das enerxías máis baratas do mercado que che garante o máximo confort no teu fogar e ademais págase una vez consumíchela.

- Al igual que en el caso de la preposición para, Apertium identifica la forma femenina del artículo indeterminado una como forma verbal por lo que no lo resuelve bien.
- Incorrecta colocación del pronombre de objeto directo que debería ir en posición antepuesta.

TO: Al ser una energía de suministro continuo, no ocupa espacio en tu hogar porque no hay que almacenarla ni necesita reposición.

TT MMT: Co ser unha enerxía de subministración continua, non ocupa espazo no teu fogar porque non hai que almacenar nin necesita reposición.

- Preposición incorrecta con. La preposición correcta es *a*.
- Omisión del pronombre de objeto directo *a*.

TT Google: Como é unha enerxía de subministración continua, non ocupa espazo na súa casa porque non ten que ser almacenada ou necesita ser reemplazada.

- Cambio de registro de segunda persona del singular a segunda personal del plural.
- Construcción gramatical incorrecta, debería haberse usado la conjunción *nin*.