Universitat Autònoma de Barcelona

Traducción automática del habla
Creación y evaluación de seis motores de TAE

Noelia Jiménez Molina

Tutora:
María Pilar Sánchez Gijón

Barcelona, 25 de junio de 2020

Trabajo de fin de máster

Máster en Tradumática: Tecnologías de la Traducción

Facultad de Traducción e Interpretación

Universitat Autònoma de Barcelona

Título

Traducción automática del habla: creación y evaluación de seis motores de TAE

Resumen

La traducción automática (TA) ha mejorado notablemente en los últimos años; sin embargo, la traducción del habla y el procesamiento del lenguaje natural siguen siendo todo un reto para los sistemas de TA. Este trabajo surge con la motivación de aportar una posible solución a la falta de naturalidad en la traducción automática del habla. Se parte de la hipótesis de que se puede mejorar la oralidad de las traducciones introduciendo corpus orales transcritos y optimizaciones en el entrenamiento de los sistemas de TA. Para probar esta hipótesis, se crean con KantanMT —tras probar MTradumàtica— seis motores de traducción automática estadística entrenados con distintos corpus orales transcritos y escritos y, después, se evalúan.

Palabras clave

traducción automática del habla, reconocimiento automático del habla, traducción automática estadística, MTradumàtica, KantanMT

Title

Spoken Language Translation: Creating and Evaluating Six SMT Engines

Abstract

Machine Translation (MT) has been greatly improved in recent years. Nevertheless, Spoken Language Translation (SLT) and natural language processing remain a major challenge for MT engines. The purpose of this work is to provide a possible solution to the lack of naturalness in SLT. The work is based on the hypothesis that it is possible to improve the orality of translations by introducing transcribed oral corpus and optimizations in the training process of MT systems. To test this hypothesis, six statistical machine translation engines, trained with different transcribed oral and written corpora, were created with KantanMT, after trying MTradumàtica, and then evaluated.

Keywords

spoken language translation, automatic speech recognition, statistical machine translation, MTradumàtica, KantanMT

Índice de contenido

Índice de tablas

Índice de figuras

Siglas

ASR	*automatic speech recognition*
IA	inteligencia artificial
LM	lengua meta
LO	lengua de origen
ML	modelo de lengua
MT	modelo de traducción
RAH	reconocimiento automático del habla
SLT	*spoken language translation*
SMTC	*statistical machine translation customisation*
SST	*speech-to-speech translation*
TA	traducción automática
TABR	traducción automática basada en reglas
TAE	traducción automática estadística
TAN	traducción automática neuronal
TH	traducción humana
TTS	*text-to-speech*

1 Introducción

Las nuevas tecnologías avanzan a un ritmo vertiginoso y ya forman parte del día a día de todos. Cada vez son más los sectores que las usan y se benefician de sus ventajas, y la traducción es uno de ellos. En los últimos años, se han digitalizado las tareas que envuelven el proceso de traducción, pero el objetivo principal de esta profesión sigue siendo el mismo: trasladar el contenido del texto de un idioma a otro (Martín-Mor, Sánchez-Gijón y Piqué 13).

En el sector de la traducción, como en cualquier otro sector, se pretende conseguir un gran número de productos en el menor tiempo posible; es decir, aumentar la productividad al traducir. En este contexto, el uso de la traducción automática (TA) es muy atractivo, puesto que agiliza el proceso de traducción y ayuda a reducir los costes. Aunque la realidad es que este aumento de productividad y abaratamiento del servicio solo ocurre cuando la calidad del resultado de la TA es lo suficientemente buena.

Desde la creación de los primeros sistemas de TA, se ha trabajado e investigado continuamente para conseguir mejorar la calidad y naturalidad de las traducciones. En la actualidad, existen sistemas de traducción automática que proporcionan unos resultados muy buenos y con una calidad equiparable a la humana, sobre todo en algunas especialidades como la traducción jurídica. Aun así, todavía quedan algunas áreas que mejorar.

El sueño de los primeros investigadores en TA era conseguir traducciones completamente automáticas producidas por sistemas informáticos y que no necesiten ser modificadas por seres humanos. Sin embargo, la realidad es distinta; normalmente, una persona revisa el resultado que proporciona el motor de TA y realiza los cambios pertinentes en el texto (posedita) para mejorar la calidad de la traducción y eliminar los errores que pueda proporcionar el sistema de TA. En este contexto, la mayoría de los esfuerzos de investigadores y profesionales de la traducción están dirigidos a conseguir que la traducción sea lo más natural posible con la mínima intervención humana; es decir, se pretende automatizar el proceso de traducción y posedición lo máximo posible (Freitag, Caswell y Roy 34).

Con la llegada de la IA, se abre un nuevo mundo en el que el sector de la traducción desempeña un papel muy importante. Las grandes empresas como Google y Amazon han diseñado sus propias inteligencias artificiales y las han puesto a disposición de todos en forma de altavoces inteligentes, asistentes en dispositivos portátiles, teléfonos móviles, etc. En el caso concreto de Google y Amazon, el idioma principal de sus IA es el inglés, por lo que utilizan la traducción automática para recitar información en los demás idiomas. A pesar de que la traducción que proporcionan es gramaticalmente correcta, aún no se ha conseguido la naturalidad del lenguaje oral humano.

Este trabajo surge con la motivación de aportar una posible solución al problema de la ausencia de naturalidad en la traducción automática del habla. Se pretende investigar cómo se puede mejorar la calidad y la oralidad de la TA con unos recursos básicos y concluir qué impacto podrían tener los resultados del presente trabajo si se llevase a cabo en una escala mayor y con más recursos.

2 Objetivos

El principal objetivo de este trabajo, como ya se ha mencionado, es comprobar si se puede mejorar la naturalidad y calidad de la traducción automática del habla, de diálogos y de textos que van a ser recitados, por ejemplo, por altavoces inteligentes y dispositivos similares, cambiando la configuración del entrenamiento de los motores de TA. Para conseguirlo, se parte de las siguientes hipótesis:

- Se puede conseguir una traducción más natural de un texto oral del inglés al español al introducir textos orales transcritos en el entrenamiento de un motor de TAE.
- Se pueden crear distintos motores de TAE con MTradumàtica entrenados con varios corpus provenientes tanto de textos orales como de escritos.

Para comprobar estas hipótesis y conseguir el objetivo principal del trabajo, se establecen varios objetivos específicos. En primer lugar, se recopilan corpus monolingües en español y bilingües de inglés y español formados por textos escritos y textos orales para entrenar los motores de TAE e introducir la optimización. Más adelante, se decide cuántos motores de TAE se crearán y qué configuración tendrán. Además, se prueban

distintas herramientas de transcripción de audio y se selecciona la que mejor funcione para el tipo de texto que se traducirá con los sistemas de TA.

Tras cumplir esos objetivos, el siguiente paso consiste en entrenar los motores de traducción automática siguiendo las características establecidas anteriormente y evaluar los resultados de dichos motores según la precisión y la fluidez de las traducciones, comparándolos entre ellos y con el resultado de una traducción humana. Además, se debe evaluar la presencia de marcas de oralidad en las traducciones. Por último, se concluye qué efectos podría tener el trabajo en una escala mayor, con más recursos, y se determina si se podría mejorar la calidad de la oralidad de la traducción de textos orales para el doblaje de películas y series o para altavoces inteligentes y otros dispositivos.

3 Marco teórico y antecedentes

La rápida evolución de la tecnología ha obligado a muchos sectores a actualizarse para aplicar el uso de las nuevas tecnologías en los procesos de producción. La traducción es uno de esos sectores que ha ido evolucionando con la digitalización de las tareas, pero sin cambiar su objetivo principal: trasladar el sentido de un texto en un idioma a otro (Martín-Mor, Sánchez-Gijón y Piqué 13).

En los siguientes apartados, se definen algunos de los conceptos más importantes de este trabajo y se contextualiza la situación de la traducción automática y de la traducción de textos orales.

3.1 Traducción automática

Como señalan Hutchnis y Somers (XI), se comenzó a especular sobre la traducción automática (TA) —definida como cualquier sistema informático que produzca traducciones de textos de un idioma a otro, con o sin intervención humana (Hutchnis y Somers 3)— hace ya muchos años, incluso antes de que aparecieran los primeros ordenadores. De hecho, se considera «uno de los primeros problemas … en el procesamiento del lenguaje natural (NLP) y de la inteligencia artificial (IA)» (Giménez Linares 1) que impulsó la creación de los primeros ordenadores y uno de los deseos más antiguos del ser humano. Además, como ya nos adelantaban Hutchnis y Somers en 1992 (2), la demanda de traducciones supera la capacidad de la profesión de la traducción, por

lo que la idea de automatizar el proceso de traducción a través de la informática atrajo a muchos investigadores de distintos ámbitos como la lingüística, las matemáticas y la informática, entre otros.

Aunque la aparición de la primera máquina de traducción data del siglo XVII, hasta finales de la década de 1940 no surgieron los que se consideran los primeros sistemas de TA. En 1933, se iniciaron distintos proyectos e intentos de automatizar la traducción. Uno de los más interesantes fue el que propuso Petr Smirnov-Troyanskii; imaginó que se podían crear tres fases, la primera para analizar el texto en el idioma de origen con un editor que conozca solo el idioma de origen, la segunda para trasladar partes de un idioma a partes equivalentes en otro idioma, y la tercera para conseguir que el resultado equivalente suene natural en el idioma de llegada con un editor que conozca solo este idioma (Hutchnis y Somers 5).

A pesar de que la idea de Troyanskii era muy interesante, no llegó a hacerse eco fuera de Rusia y se tardó varios años más en comenzar a investigar en traducción automática y en establecer las primeras líneas de investigación. En 1954 se mostraron públicamente los primeros resultados de un sistema de TA creado por Leon Dostert en la universidad de Georgetown en colaboración con IBM. A partir de ese momento, muchos grupos de investigación se involucraron en este ámbito hasta la actualidad. (Hutchnis y Somers 6).

En la década de 1990, aparece una nueva línea de investigación, la traducción de la lengua oral (s*poken language translation* o SLT por sus siglas en inglés), que supone un reto informático aún mayor, puesto que se debe comprender la voz en un idioma y traducir el contenido del discurso a otro idioma. A principios del siglo XXI, la situación de la SLT no permitía que el proceso de reconocer la voz y traducir un texto oral se consiguiera en menos de cinco segundos (Somers 7). Sin embargo, gracias a las grandes cantidades de datos disponibles y a la evolución de las nuevas tecnologías, la situación actual de la SLT es muy distinta.

A partir de este momento hasta la actualidad, investigadores de todo el mundo y de distintas disciplinas han centrado sus esfuerzos en mejorar la calidad de los resultados y en incluir las nuevas tecnologías, como la inteligencia artificial (IA), en los sistemas de

TA y viceversa. Gracias a estos avances, la TA ha dejado de ser un misterio indescifrable para convertirse en una realidad disponible para todas las personas.

A pesar de que el objetivo de las primeras investigaciones en TA era conseguir traducciones completamente automáticas y de buena calidad que no necesitasen de intervención humana alguna, la realidad es que, en muchos ámbitos, se sigue necesitando la supervisión de una persona para conseguir una calidad óptima. En el siglo XXI, uno de los objetivos principales de la investigación en TA se ha centrado en combinar aplicaciones de reconocimiento de voz con TA (Koehn 5), especialmente para dispositivos portátiles como los teléfonos y altavoces inteligentes, entre otros.

Según la tecnología usada, los sistemas de TA se pueden dividir en dos bloques: los sistemas de traducción automática basados en reglas (TABR) y los basados en corpus (Ping 162). Los sistemas de TABR se basan en diccionarios y en reglas semánticas, gramaticales y de transferencia, entre otras, que se deben escribir para cada uno de los idiomas del motor de TA, lo que implica la participación activa de recursos humanos. Sin embargo, los motores de TA basados en corpus —como son los sistemas de TA basados en ejemplos (EBMT por sus siglas en inglés), los estadísticos (TAE) y los neuronales (TAN)— necesitan grandes cantidades de textos originales y traducidos, pero no la participación humana no es tan necesaria como en la TABR (Martín-Mor 26). Los modelos de TA basados en corpus son capaces de traducir un texto desconocido por el motor a partir de información extraída de otros textos traducidos previamente por humanos (Hearne y Way 205). En los siguientes apartados, se explicará con más detalle en qué consisten estos sistemas.

3.1.1 Traducción automática estadística (TAE)

Los sistemas de traducción automática estadística (TAE) se diferencian de los basados en reglas en que «generan traducciones mediante modelos estadísticos cuyos parámetros se calculan gracias al análisis de grandes cantidades de corpus textuales bilingües» (Giménez Linares 85) y monolingües, mientras que la TABR funciona, como ya se ha mencionado, mediante reglas programadas. En este modelo de TA no se necesitan tantos recursos humanos, puesto que el ordenador realiza la mayor parte del trabajo. Es cierto que para entrenar estos sistemas se necesita un gran volumen de texto traducido por personas y alineado, pero si se dispone de este texto previamente, ya sea

usando corpus existentes o uniendo distintas memorias de traducción, la intervención humana en el proceso es mínima comparada con la TABR.

Los motores de TAE buscan las mejores traducciones para un segmento calculando probabilidades y ofrecen como resultado la traducción que obtenga la mejor puntuación —calculada por el mismo motor—, es decir, la probabilidad más alta de que esa traducción haya sido realizada por un ser humano. Para calcular estas probabilidades, el sistema de TAE consulta los distintos modelos entrenados con los corpus bilingües y monolingües que se hayan cargado para dicho motor. A continuación, se explican brevemente los modelos que se usan para entrenar un motor de TAE y cómo interactúa el motor con dichos modelos.

Como indican Hearne y Way (205), el proceso de la TAE se divide en dos fases: entrenamiento (*training*) y descodificación (*decoding*). La primera fase consiste en extraer modelos estadísticos de traducción a partir de corpus alineados bilingües y modelos estadísticos de la lengua meta a partir de corpus monolingües en dicho idioma, mientras que la segunda fase consiste en encontrar una posible traducción de un idioma A a un idioma B para una frase o texto buscando equivalencias entre ambos idiomas en el modelo de traducción y en reordenar la traducción más probable buscando posibles combinaciones en el modelo de lengua del idioma B.

El modelo de traducción (MT) de un sistema de TAE consta de un diccionario bilingüe probabilístico creado a través de la alineación de palabras de un bitexto (modelo de traducción de palabras) que se usa para extraer las probabilidades de traducción de segmentos (modelo de traducción de segmentos). El modelo de lengua (ML), por su parte, sirve para mejorar la fluidez e inteligibilidad de las traducciones obtenidas en el MT calculando la frecuencia en la que las palabras o los segmentos de palabras aparecen juntas en los monotextos del idioma meta. Además de estos modelos, a la hora de dar una puntuación final a la traducción, el sistema de TAE también tiene en cuenta la diferencia del número de palabras (*word penalty*) entre el original y la traducción para evitar producir traducciones muy cortas o largas y el número de segmentos bilingües usados en la traducción (*phrase penalty*) para impulsar el uso de segmentos largos. (Giménez Linares 86-89; Sánchez-Cartagena, Sánchez-Martínez y Pérez-Ortiz 90-91).

Los sistemas de TAE suelen funcionar muy bien en combinaciones de idiomas comunes, como es el caso del presente trabajo con inglés y español, puesto que hay una cantidad de texto mayor que en el caso de los idiomas minoritarios o de combinaciones de idiomas menos frecuentes. Además, se considera que los sistemas de TAE creados para un ámbito específico suelen dar mejores resultados que los sistemas genéricos (Martín-Mor 26).

3.2 Corpus escritos, de lengua oral y orales

Se conoce como corpus a un «conjunto estructurado y documentado de materiales recogidos en función de criterios explícitos» (Llisterri, «La lingüística de corpus»). Otros autores, como Sinclair, prefieren referirse a los corpus como «pieces of language» o fragmentos de lenguaje ordenados para una función concreta y que sirven como un ejemplo del lenguaje, de un idioma o de una variedad concreta. Ambas definiciones tienen en común el uso deliberado de términos ambiguos, «materiales» y «fragmentos», en lugar de «textos». Esto se debe a que, dependiendo de la finalidad del corpus, el material usado puede ser un conjunto de palabras, fonemas, textos, audios, vídeos, etc.

	Corpus de lengua oral (*spoken language corpora*) Lingüística de corpus	**Corpus orales (*speech corpora*) Fonética Tecnologías del habla**
Materiales	Habla espontánea *unelicited speech*	Corpus controlado *elicited speech*
Nivel de análisis	Discurso, diálogo	Enunciado
Obtención de los datos	Entorno natural	Entorno controlado
Transcripción	Transcripción ortográfica enriquecida	Transcripción fonética y ortográfica alineada con la señal sonora
Orientación	Representación simbólica, categorial	Señal sonora, representación temporal

Tabla 1. Diferencias entre los corpus de lengua oral y los corpus orales (Llisterri, «Los corpus de lengua oral»)

Llisterri («La lingüística de corpus») establece una división de corpus en tres tipos: corpus escritos o textuales, corpus de lengua oral y corpus orales. El primer tipo de corpus, el corpus escrito, está compuesto por textos escritos, como bien indica el nombre. En cuanto a los corpus orales, Llisterri distingue entre los corpus de lengua oral, definidos como «la transcripción en ortografía convencional (transliteración) de una grabación a partir de la cual se lleva a cabo el tratamiento y el análisis del corpus», y los corpus orales, en los que el tratamiento y análisis del corpus se lleva a cabo a partir de la señal sonora y no de la transcripción del texto.

Para crear un corpus de lengua oral, se debe establecer previamente una serie de características como el tema, el tipo de texto (oral espontáneo entre una o varias personas, oral leído o recitado, escrito para recitar, etc.), la situación comunicativa (monólogo, diálogo o entrevista…), la procedencia de las grabaciones, el estilo o registro del habla, etc. Una vez establecidas estas características, se debe llevar a cabo una selección de «informantes» o participantes y comenzar la grabación o recogida de datos (Llisterri, «Los corpus de lengua oral»).

El siguiente paso en la creación de corpus de lengua oral es la transcripción, es decir, el traslado del sonido a texto o, como la define Payrató, el «procedimiento de traslado o transposición a una forma gráfica (escrita) de una producción (lingüística, discursiva) originalmente oral» (45). La transcripción se puede realizar manualmente o de forma automática a través de herramientas de reconocimiento automático del habla. Este paso se considera la base sobre la que se puede ir añadiendo información adicional a través de etiquetas o anotaciones.

La transcripción debe ser un proceso neutral y fiel al material original, por lo que, si se quiere añadir algún tipo de información adicional para cumplir los objetivos del corpus, se hace a través de anotaciones o etiquetas, puesto que esta información se considera una interpretación de la persona que lo está creando (Llisterri, «La lingüística de corpus»). Algunas instituciones y grupos de investigación han establecido ciertas recomendaciones para la transcripción ortográfica de textos orales y para las anotaciones de los mismos; sin embargo, no existe ningún estándar que defina los elementos que se deben transcribir.

Las anotaciones y etiquetas del corpus se pueden mezclar con el contenido textual o separar de este a través de la codificación: «procedimiento de representación de los caracteres, de la estructura del texto y de la anotación, de modo que la estructura y la anotación se mantienen separadas del contenido del corpus» (Llisterri, «La lingüística de corpus»). La *Text Encoding Initiative*[1] ha creado un estándar con indicaciones para realizar este proceso de codificación de textos, aunque existen otras propuestas de codificación y no es obligatorio seguir ninguna de ellas.

3.3 Traducción automática del habla

Como ya se ha mencionado brevemente en las páginas anteriores, el interés por los sistemas de TA de textos orales nació en la década de 1990. Desde la creación del primer sistema de traducción de voz en 1989 hasta ahora, se han llevado a cabo grandes progresos en la investigación de este ámbito que han permitido conseguir una traducción del habla en menos de cinco segundos, a pesar de lo imposible que parecía en 2003, como indicaba Somers (7). Sin embargo, aunque la tecnología ha avanzado considerablemente, generar respuestas y traducciones fluidas de textos orales sigue siendo una tarea complicada para los sistemas de conversación (Balakrishnan et al. 1).

Los sistemas de traducción de voz —*spoken language translation* (SLT) o *speech-to-speech translation* (SST)— incluyen, además de la TA, dos componentes más: el reconocimiento automático de voz —*automatic speech recognition* o ASR—, que reconoce el habla y la convierte en texto, y la síntesis de voz —*text-to-speech* o TTS—, que convierte en voz el texto recibido de la TA (Waibel y Fugen 70).

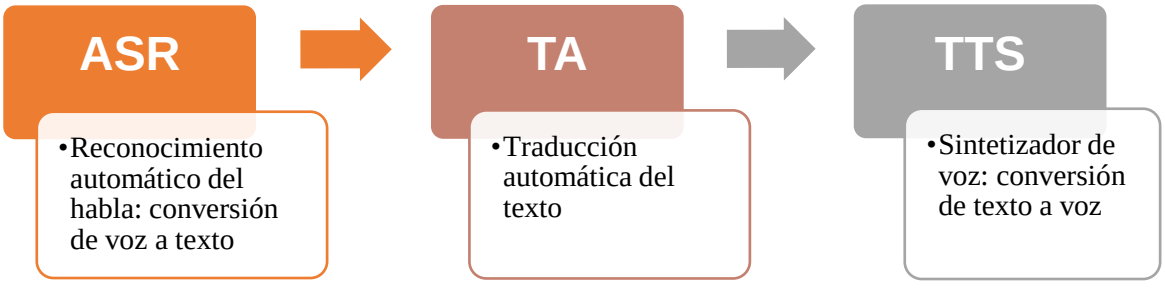


Figura 1. Proceso de los sistemas de traducción de textos orales. Fuente: elaboración propia de la autora

[1] Enlace a la página web de *Text Encoding Initiative*: https://tei-c.org/ [última consulta: junio de 2020].

Viendo el proceso que siguen los sistemas de SST, es fácil imaginar la enorme dificultad que puede suponer esta tarea para un sistema informático, puesto que carece del razonamiento del sentido común que tienen los seres humanos, es decir, carece del conocimiento de miles y miles de cosas que los seres humanos dan por hecho (Somers 121). Además, como se detallará más adelante, los textos orales (discursos, diálogos, etc.) cuentan con ciertas características, como los problemas de fluidez, que no suelen aparecer en los textos escritos y que dificultan la tarea de reconocimiento de voz y de traducción (Salesky, Sperber y Waibel 1).

3.3.1 Reconocimiento automático del habla

El reconocimiento automático del habla (RAH, o ASR por sus siglas en inglés) es uno de los ámbitos de investigación en los que más se ha trabajado en los últimos años. Investigadores de distintas disciplinas han unido sus esfuerzos para intentar conseguir que una máquina sea capaz de procesar el lenguaje natural de un ser humano, tanto la información lingüística como paralingüística, a partir de una señal de voz (Lleida 2).

Entre 1930 y 1950, se comienza a investigar en el reconocimiento de sonidos y en los años 50 se crea el primer sistema capaz de reconocer la voz de una persona. En esa década, el reconocimiento del habla se limitaba a identificar algunas palabras y se trataba de un sistema complejo y caro, pero capaz de dar buenos resultados, lo que impulsó la investigación en las siguientes décadas. En los años 70, cambió el rumbo de la investigación, que se dejó de centrar en el mero reconocimiento del habla y se dirigió hacia la comprensión del habla con la inclusión de elementos pragmáticos. En 1980, aparece el reconocimiento de habla estadístico, que permite reconocer un vocabulario mucho más amplio que lo que se había conseguido hasta el momento y que se entrena con grandes cantidades de datos. En la actualidad, este sigue siendo uno de los modelos más utilizado, junto con los sistemas de redes neuronales (Lleida 3-6).

3.4 Lenguaje coloquial y oralidad

Se entiende por lenguaje coloquial la ausencia de planificación en el habla; es decir, la espontaneidad o naturalidad de una conversación o discurso. El lenguaje coloquial es un registro que puede aparecer tanto en textos orales como en textos escritos, mientras que la oralidad, de acuerdo con la definición de Antonio Briz, consiste en la

aparición de características típicas de la lengua oral en textos escritos. Algunas de estas características son las oraciones coordinadas, la repetición de palabras y frases, el uso repetitivo de conectores, interjecciones, exclamaciones y frases hechas o expresiones, etc. (Barrio Arconada 439-442).

Según Níkleva (215), las tres características principales de los textos orales que los diferencian de los textos escritos son las «acumulaciones paradigmáticas, [las] idas y vueltas sobre el eje de los sintagmas y [la] introducción de oraciones incisas». Estas características —y algunas más (véase la tabla 2)— propias de la lengua oral son las que deben reconocer los sistemas de SLT en el idioma de origen y trasladar al idioma meta.

Modalidad oral	**Modalidad escrita**
elementos prosódicos	-
espontánea	planificada
dinámica	estática
informal	formal
dialogada	monologada
uso del lenguaje no verbal	ausencia del lenguaje no verbal
elementos paralingüísticos	-
-	posibilidad de corregir
interrupciones	-
solapamientos	-
mayor redundancia	menor redundancia
mayor uso de muletillas y onomatopeyas	menor uso de muletillas y onomatopeyas
mayor frecuencia de frases hechas y refranes	menor frecuencia de frases hechas y refranes

Tabla 2. Ejemplos de características de la lengua oral frente a la lengua escrita (Níkleva 222)

3.5 MTradumàtica

La creación de modelos de TAE para fines específicos se conoce como *Statistical Machine Translation Customisation* (SMTC). Con la ayuda de tecnologías como Moses[2], este proceso es tan sencillo como obtener un corpus bilingüe alineado en la LO y la LM y otro monolingüe en la LM para entrenar, respectivamente, el MT y el ML. Una vez seleccionados los corpus, Moses se encarga de preparar los textos y entrenar los distintos modelos con ellos. El proceso de preparación de los textos consiste en tres fases: *tokenising* —separar las palabras de los signos de puntuación—, *truecasing* —buscar la capitalización más frecuente de las palabras— y *cleaning*, el proceso en el que se eliminan segmentos erróneos o vacíos de los corpus (Martín-Mor 29). Tras preparar los textos, se comienza la fase de entrenamiento y la de descodificación detalladas en el apartado anterior. Una vez se han entrenado los traductores, se puede optimizar el sistema (fase de *tuning* o mejora) introduciendo un corpus paralelo nuevo y mucho más pequeño y específico que el usado para entrenar el modelo de traducción.

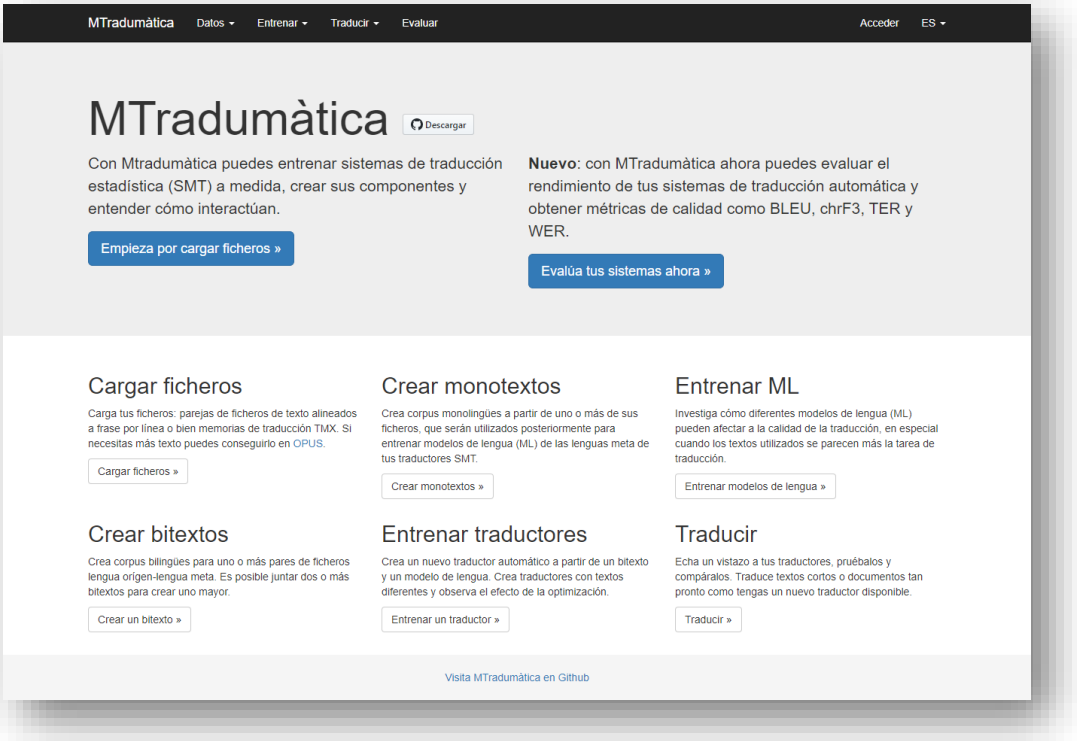


Figura 2. Página de inicio de MTradumàtica

[2] Para más información sobre Moses, ver http://www.statmt.org/moses/ [última consulta: abril de 2020].

Como se puede observar, el proceso de creación de un motor de TAE con Moses es bastante sencillo. No obstante, Moses carece de una interfaz gráfica que facilite la navegación a usuarios que no suelan tratar con el símbolo del sistema. De esta necesidad nace MTradumàtica, «una plataforma web basada en Moses con interfaz gráfica de usuario» cuyo principal objetivo es facilitar el proceso de creación de un motor de TAE a cualquier usuario (Martín-Mor 30-31).

Teniendo este objetivo en mente, la plataforma divide el proceso de creación de un sistema de TA en seis fases (véase la figura 2): cargar ficheros, crear monotextos, entrenar modelos de lengua, crear bitextos, entrenar traductores y traducir. Además de estas opciones, MTradumàtica cuenta con una función llamada `Evaluar` en la que se puede introducir una frase en un idioma A y ver el proceso que sigue el sistema de TA que se haya creado con las puntuaciones proporcionadas por cada modelo hasta conseguir la traducción más probable en un idioma B, pasando por los procesos de tokenización y *truecasing*.

3.6 KantanMT

KantanMT (O'Dowd) es una plataforma en la nube que permite crear sistemas de traducción automática estadística y de traducción automática neuronal personalizados (Shterionov 222). Al igual que MTradumàtica, uno de los principales objetivos de KantanMT consiste en facilitar la creación de motores de TA a través de una plataforma con una interfaz sencilla e intuitiva. Sin embargo, KantanMT se diferencia de MTradumàtica en que se trata de un *software* privativo y no de libre acceso, como es el caso del último.

Para entrenar un motor de TAE con esta plataforma, no es necesario crear o entrenar primero un modelo de lengua o de traducción; basta con subir los archivos de entrenamiento (bilingües y monolingües) a la sección `Training` con los nombres y el formato que se indica.

Además de crear sistemas de TA personalizado, KantanMT ofrece otras funciones como traducir con esos sistemas, analizar la calidad de los resultados, medir automáticamente la calidad con medidas como BLEU, TER y F-Measure e integrar el

motor de TA en herramientas de traducción asistida como memoQ y Memsource a través de un API.

Otra característica interesante de esta plataforma es que desglosa el proceso de cada tarea en distintos pasos para saber en todo momento en qué punto se encuentra esa tarea.

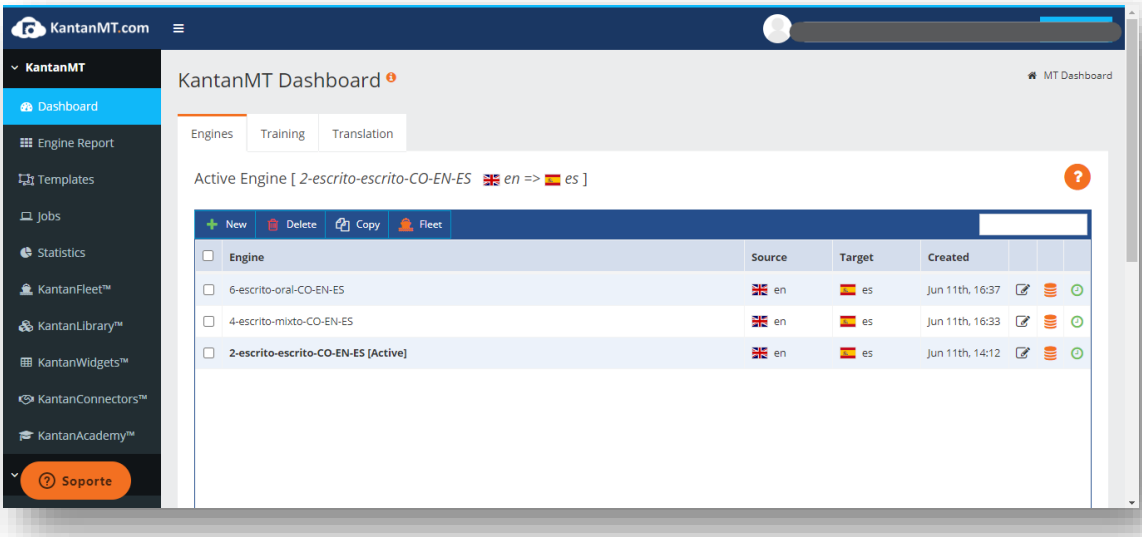


Figura 3. Vista de la pantalla principal de KantanMT

4 Metodología

Para conseguir los objetivos planteados y evaluar los sistemas de TA, se deben utilizar tres instrumentos: los sistemas de TA creados y los corpus que los conforman, el texto de evaluación y la escala de la evaluación. En los siguientes apartados, se describen los pasos seguidos para establecer estos tres instrumentos y se detalla el proceso creación y evaluación de los motores.

En primer lugar, se escoge un audio en inglés que contenga suficientes marcas de oralidad para funcionar como texto de evaluación de los distintos sistemas de TA creados y que, además, disponga de una traducción humana al español que también tenga marcas de oralidad (véase el apartado 4.1).

En el apartado 4.2, se buscan varios corpus monolingües en español y bilingües en inglés y español que se ajusten a las características del trabajo. Tras hacer una selección

de los corpus más relevantes, en el apartado 4.3, se plantean los diferentes escenarios y configuraciones de los sistemas de TA; es decir, se establece la forma en que se van a combinar los corpus seleccionados en el paso anterior teniendo en cuenta los objetivos del trabajo y las hipótesis planteadas.

El siguiente paso, detallado en el apartado 4.4, consiste en preparar los archivos de los corpus para después entrenar los motores de TA con las configuraciones establecidas (véanse los apartados 4.5 y 4.6).

Por último, en el apartado 4.7, se establece la escala en la que se evalúan los resultados; se realiza una evaluación de la calidad de las traducciones en cuanto a precisión y fluidez y se elabora una clasificación general comparándolas entre ellas. Además, se analiza la presencia o ausencia de marcas de oralidad comparando la traducción automática con la traducción humana y se estima el efecto que causa la introducción de corpus orales transcritos en los resultados.

4.1 Selección del texto de evaluación

El audio elegido para probar la calidad de los distintos motores se extrae de un diálogo de la película *Marriage Story* (2019) —título en español: *Historia de un matrimonio*—, dirigida por Noah Baumbach y nominada a mejor guion, entre otros premios y nominaciones, en los Premios Óscar y en los Globos de Oro. Esta película contiene unos diálogos naturales con suficientes marcas de oralidad en el texto en inglés que son fáciles de identificar y que se deberían trasladar a la traducción en español con los sistemas de TA creados. Además, la película ha sido doblada al español con una traducción humana, por lo que este doblaje sirve para identificar las marcas de oralidad de la traducción en español y comparar los resultados de la TA con la TH.

En concreto, se eligen dos minutos de un diálogo en una escena comprendida entre los minutos 91 y 100 de la película. En esta escena, los protagonistas, Nicole y Charlie, mantienen una discusión en la que prevalecen las interrupciones, los incisos, la repetición de palabras y frases… Es decir, el texto incluye muchas de las marcas de oralidad mencionadas en el apartado 3.4 y especificadas con más detalle en el apartado 4.1.2.

4.1.1 Transcripción del audio

La transcripción del diálogo original en inglés y del doblaje en español se lleva a cabo con Happy Scribe[3], una herramienta en línea que permite realizar transcripciones automáticamente a partir de un audio o vídeo y editar el texto generado en un editor que sincroniza el texto con audio palabra a palabra (véase la figura 4). Se elige Happy Scribe tras probar con un audio de un minuto y comprobar que la calidad de la transcripción es buena y no es necesario realizar muchos cambios.

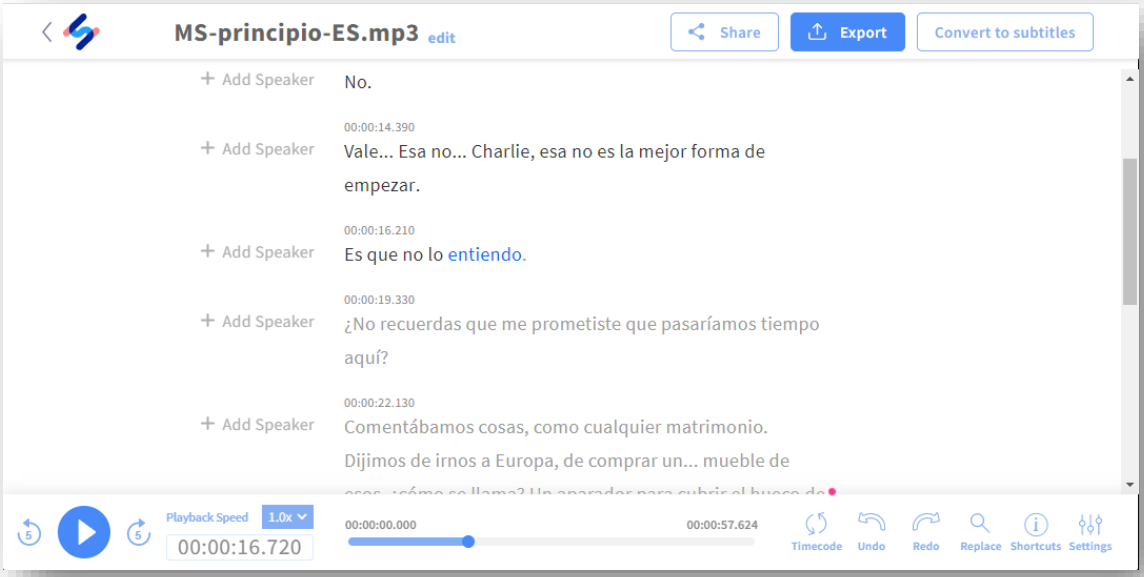


Figura 4. Vista del editor de Happy Scribe

Una vez comprobada la calidad de la herramienta de transcripción, se cargan los dos archivos de audio (el original y el doblaje) y se seleccionan los idiomas de cada uno de ellos. Tras unos minutos, se puede comenzar a editar la transcripción obtenida para corregir palabras o frases erróneas, cambiar la segmentación o incluir texto que no se haya detectado por entremezclarse con otra intervención o por el ruido del audio. El siguiente paso, consiste en exportar la transcripción en alguno de los formatos ofrecidos (TXT, DOCX, SRT, PDF, HTML, etc.).

[3] https://www.happyscribe.co/ [última consulta: junio de 2020].

4.1.2 Marcas de oralidad del texto

Como se menciona en el apartado 3.4, los textos orales contienen ciertas características que los diferencian de los textos escritos. Con el objetivo de medir la calidad de las traducciones de los sistemas de TA, se señalan, tanto en el texto original como en la traducción, las distintas marcas de oralidad. A continuación, se incluye una lista de los tipos de marcas de oralidad y los colores con los que se representarán dichas marcas:

- Interjección.
- Interrupción.
- Repetición de palabra o frase o redundancia.
- Inciso.
- Exclamación.
- Conectores, marcadores del discurso u oraciones coordinadas.
- Muletilla.

	Original (EN-US)	**Traducción humana (ES-ES)**
1.	I don't know how to start.	No sé cómo empezar.
2.	Do you understand why I want to stay in LA?	¿Entiendes por qué quiero quedarme en Los Ángeles?
3.	No.	No.
4.	Well, that's not… Charlie, that's not a useful way for us to start.	Vale… Esa no… Charlie, esa no es la mejor forma de empezar.
5.	I don't understand it.	Es que no lo entiendo.
6.	You don't remember promising that we could do time here?	¿No recuerdas que me prometiste que pasaríamos tiempo aquí?
7.	We discussed things; we were married. We said things. We talked about moving to Europe, about getting a… sideboard or what do you call it? A credenza, to fill the empty	Comentábamos cosas, como cualquier matrimonio. Dijimos de irnos a Europa, de comprar un… mueble de esos, ¿cómo se llama? Un aparador para cubrir el hueco de

	space behind the couch. We never did any of it.	detrás del sofá, y no hicimos nada de eso.
8.	You turn down the residency at the Geffen that would have brought us here for a… a year.	Rechazaste la residencia en el Greffen que nos habría permitido pasar aquí un año.
9.	It wasn't something I wanted. We had a great theater company and a great life where we were.	Porque no me apetecía. Teníamos una compañía de teatro y una vida increíbles.
10.	You call that a great life?	¿Aquello era una vida increíble?
11.	You know what I mean. I don't mean we had a great marriage. I mean, life in Brooklyn. Professionally. I don't know. Honestly, I never considered anything different.	Tú ya me entiendes. No digo que el matrimonio fuera increíble. Hablo de la vida en Brooklyn. Profesionalmente. No sé, la verdad es que nunca he querido hacer otra cosa.
12.	Well, that's the problem, isn't it? I mean, I was your wife. You should have considered my happiness, too.	Ya, y ese es el problema, Charlie. A ver, yo era tu mujer. Debiste pensar también en mi felicidad.
13.	Come on. You were happy. You've just decided you weren't now.	Venga ya, eras muy feliz y de repente decidiste que ya no lo eras.
14.	OK. Let's… let's not… M-my work is here now, my family's here…	Vale, vale… Va… Vamos a… Mi trabajo está aquí ahora, mi familia está aquí…
15.	And I agreed to put Henry in school here because your show went to series. I did that knowing that when you were done shooting, he would come back to New York.	Y yo acepté que Henry fuera al cole aquí, porque tu piloto se convirtió en serie, porque pensé que cuando acabaras volveríais a Nueva York.

16.	Honey, we never said that. That may have been your assumption, but we never expressly said that.	Cielo, nunca dijimos eso. Quizá lo supusieras, pero nunca dijimos eso.
17.	We did say it.	Claro que lo dijimos.
18.	When did we say it?	¿Y cuándo lo dijimos?
19.	I don't know when we said it, but we said it.	No sé cuándo lo dijimos, pero lo dijimos.
20.	I thought…	Yo creía que…
21.	We said it that time on the phone!	¡Aquella vez por teléfono!
22.	Honey! Let me finish! Sorry, I keep saying that. I thought that if Henry was happy here and my show continued, that we might do L.A. for a while.	¡Cielo, que me dejes acabar! Perdona por volver a decírtelo. Yo creía que, si Henry era feliz aquí y mi serie prosperaba, podríamos vivir aquí un tiempo.
23.	I was not privy to that thought process.	Nunca me informaste de ese proceso mental.
24.	The only reason we didn't live here is because you can't imagine desires other than your own. Unless they're forced on you.	La única razón de que no viviéramos aquí es que eres incapaz de pensar en otra cosa que no sean tus deseos si no se te obliga.
25.	Okay. You wish you hadn't married me. You wish you'd had a different life. But this is what happened. So, what do we do?	Vale, querrías no haberte casado conmigo y haber tenido otra vida. Pero esto es lo que hay. Entonces, ¿qué hacemos?
26.	I don't know.	No lo sé.

Tabla 3. Transcripción del diálogo original en inglés y del doblaje en español

4.2 Selección de corpus

El siguiente paso consiste en buscar distintos tipos de corpus con las características necesarias para resolver las hipótesis y estimar la cantidad de recursos lingüísticos accesibles de cada tipo para poder organizar la configuración de los motores de TA más adelante. En esta primera fase, se pretende encontrar el mayor número de corpus orales transcritos posible y uno o varios corpus escritos bilingües de temática variada para el modelo de traducción. Además, se intenta evitar el uso de corpus muy especializados o técnicos, puesto que el objetivo del trabajo consiste en obtener una traducción natural para el discurso oral y con terminología común.

Encontrar corpus de textos escritos bilingües en inglés y español y monolingües en español descargables y de libre acceso no es muy difícil, puesto que son dos de los idiomas más hablados del mundo y hay mucha información para esta combinación. Sin embargo, acceder a corpus bilingües y monolingües de textos orales transcritos que tengan una calidad y cantidad aceptables resulta una tarea más ardua. En la búsqueda de varios corpus de este tipo, se consultan varias fuentes que ofrecen recursos lingüísticos y corpus como ELRA[4] (European Language Resources Asociation), OPUS[5] y OpenSLR[6], entre otras.

ELRA es una asociación que dispone de un catálogo de recursos lingüísticos —sobre todo corpus, pero también diccionarios y recursos terminológicos— muy amplio para varios idiomas y en distintos formatos: texto, audio y vídeo. Algunos recursos se pueden descargar directamente desde la página web, sin embargo, para acceder a la mayoría de los recursos, se debe rellenar una solicitud de pedido y tan solo unos pocos recursos son gratuitos. Por este motivo, a pesar de que ELRA dispone de unos corpus muy interesantes para la elaboración de este trabajo, solo se consigue acceder a cinco corpus, de los cuales dos cumplen con las características necesarias para el presente trabajo:

- Glissando (Garrido et al. 945–971): este corpus oral de más de doce horas de grabaciones de voces en español y catalán contiene, además, las

[4] Página web del catálogo de ELRA: http://catalog.elra.info/en-us/ [última consulta: mayo de 2020].
[5] Página web de OPUS: http://opus.nlpl.eu/ [última consulta: mayo de 2020].
[6] Página web de OpenSLR: https://www.openslr.org/index.html [última consulta: mayo de 2020].

trascripciones de los diálogos —cerca de 130 000 palabras transcritas— con algunas etiquetas que marcan pausas, interrupciones, dudas, interjecciones, etc. Los textos son lecturas de noticias y diálogos formales e informales entre distintas personas. El corpus se ha desarrollado en el marco de dos proyectos en los que participan la Universitat Pompeu Fabra, la Universitat Autònoma de Barcelona y la Universidad de Valladolid: «Glissando, un corpus de habla anotado para estudios prosódicos en catalán y español» y «Modelización de los fenómenos prosódicos del español y catalán a partir del corpus GLISSANDO», financiados por el Plan Nacional de I+D del Gobierno español.

- European Parliament Interpretation Corpus (EPIC): corpus formado por audios y vídeos de discursos del Parlamento Europeo en inglés, español e italiano, audios de sus correspondientes interpretaciones a los otros idiomas y las transcripciones de todos ellos. Esto quiere decir que, por ejemplo, además del audio, vídeo y transcripción de un discurso en español, se incluye un audio y una transcripción de la interpretación al inglés y un audio y una transcripción de la interpretación al italiano. Este corpus es interesante porque puede servir como corpus monolingüe con las transcripciones de los discursos originales en español y como corpus bilingüe con los discursos originales en inglés y las interpretaciones en español de esos discursos. Contiene unas 42 000 palabras de discursos originales en inglés y unas 14 000 palabras originales en español.

El repositorio de OPUS ofrece numerosos corpus paralelos y monolingües en varios idiomas en distintos formatos: TMX, Moses, TXT, etc. (Tiedemann 2214-2218). Además, todos los recursos que se encuentran en esta página son de libre acceso y gratuitos. En OPUS se encuentran distintos corpus que pueden ser útiles para la elaboración del trabajo, puesto que cuenta con un gran número de palabras y corpus para la combinación de idiomas inglés-español. Los recursos más interesantes encontrados en OPUS son los siguientes:

- ParaCrawl[7] es un corpus escrito, paralelo y multilingüe en todos los idiomas oficiales de la Unión Europea —además de otros idiomas minoritarios no oficiales de la UE— alineados con el inglés (Esplà-Gomis 118). Los textos del corpus ParaCrawl se extraen de distintas páginas web bilingües o multilingües. El proceso de extracción del texto se lleva a cabo de forma automática a través de un robot que accede a las páginas web que se lo permitan mediante el archivo `robots.txt`, por lo que, aunque se desconoce el origen exacto de los textos, se entiende que la variedad temática es amplia. Para este trabajo, se descarga la versión 5 del corpus, que contiene 39 millones de segmentos y casi 900 millones de palabras para la combinación inglés-español.

- TildeMODEL: corpus multilingüe que dispone de casi cuatro millones de segmentos alineados en inglés y español y cerca de 130 millones de palabras originales en inglés. El corpus está formado por cinco subcorpus (EESC, RAPID, ECB, EMA y World Bank) y por los textos extraídos de seis páginas web, formando un corpus bastante completo y variado de 30 idiomas (Rozis y Skadiņš 263-265).

- Wikipedia: como indica el nombre, el corpus está formado por frases extraídas de distintos artículos de Wikipedia. Se trata de un corpus escrito y multilingüe —con 20 idiomas alineados con el inglés y el polaco— y contiene más de 38 millones de palabras en español (Wołk y Marasek 126).

- Tatoeba[8]: este recurso lingüístico recopila frases escritas en varios idiomas originales y sus traducciones al resto de idiomas. Además, para la mayoría de las frases se incluye también una grabación de voz. Se trata de un trabajo colaborativo, abierto y gratuito, lo que significa que tanto las frases originales como las traducciones son proporcionadas por personas voluntarias que desean colaborar con el proyecto. Tatoeba ofrece cerca de 300 000 frases en español (dos millones y medio de palabras). A pesar de

[7] Más información sobre ParaCrawl: https://www.paracrawl.eu/ [última consulta: mayo de 2020].
[8] Corpus extraído de https://tatoeba.org [última consulta: junio de 2020], publicado bajo la licencia CC-BY 2.0 FR.

que el corpus contiene grabaciones de voz de algunas frases, no se considera un corpus oral, sino escrito. Esto se debe a que el origen de las frases no es oral, es decir, no son transcripciones de discursos, más bien al contrario: se trata de lecturas de frases escritas. Por este motivo, se clasifica como corpus escrito.

- TED2013: este corpus se ha creado con las transcripciones de los discursos de TED Talks en inglés alineadas con las distintas traducciones proporcionadas por traductores voluntarios. Los archivos, proporcionados por WIT[3] (Cettolo, Girardi y Federico 261), incluyen dos millones y medio de palabras en español. En un principio, se descarta el uso de este recurso, puesto que el texto en español no es un texto oral transcrito, sino la traducción de un texto escrito. Sin embargo, tras analizar el contenido del corpus, se encuentran características propias del discurso oral en las traducciones como el uso repetitivo de conjunciones, muletillas, repeticiones de palabras, etc. Por este motivo, y por la cantidad de palabras del corpus, se decide tenerlo en cuenta para el trabajo.

OpenSLR es una página web dedicada a recopilar recursos lingüísticos orales que se puedan usar para el entrenamiento de programas de reconocimiento de voz y otros fines similares. Los recursos que se ofrecen en esta plataforma son gratuitos y variados en cuanto a formato e idioma. A continuación, se describen los dos corpus descargados de esta fuente:

- El corpus Heroico (Morgan) contiene grabaciones de voces y sus transcripciones de las academias militares de México (Heroico) y de Estados Unidos (USMA). En total, hay más de 50000 palabras transcritas en español.

- TEDx Spanish (Hernández-Mena): un corpus de unas 24 horas de grabaciones de distintos discursos de TED en español. Este corpus se ha elaborado en el marco del programa «Desarrollo de Tecnologías del Habla» de la Universidad Nacional Autónoma de México y del proyecto CIEMPIESS-UNAM[9].

[9] http://www.ciempiess.org/ [última consulta: junio de 2020].

Fuera de las tres fuentes mencionadas anteriormente, se encuentran más recursos lingüísticos interesantes para el trabajo:

- ACTIV-ES es un corpus de diálogos transcritos de películas en español de producciones de Argentina, México y España. Los textos se han extraído de repositorios de guiones de películas, de los subtítulos para personas sordas y contienen un total de casi cuatro millones de palabras.

- El corpus DiEspa (Diálogos en Español) está formado por 16 audios de diálogos en español grabados en Barcelona, Sevilla y Almería, aunque solo están disponibles las transcripciones de cuatro de ellos. El corpus ha sido elaborado por Parlare italiano y se puede descargar desde su página web[10].

- El corpus multilingüe PraTiD[11], preparado también por Parlare italiano, contiene doce diálogos, cuatro de ellos en español transcritos en TXT y anotados en XML. Los diálogos son espontáneos, pero la temática se ha planificado: cada participante tiene un dibujo y tiene que encontrar las diferencias de su dibujo con respecto al resto de dibujos a través del diálogo.

- AN.ANA.S.[12], al igual que el anterior, se trata de un corpus multilingüe de Parlare italiano. La parte en español del corpus está formada por un archivo extraído del corpus DiEspa, mencionado anteriormente, y un diálogo espontáneo transcrito de una emisión de radio. Solo se tiene en cuenta este último archivo, puesto que el primero ya aparece en el corpus DiEspa.

- El corpus MuST-C es un recurso multilingüe y paralelo que contiene transcripciones de discursos de TED en inglés y sus traducciones a varios idiomas, entre ellos, el español. Este corpus fue creado específicamente para usarlo como material de entrenamiento de sistemas de TA de textos

[10] Página web de Parlare italiano: http://www.parlaritaliano.it/ [última consulta: junio de 2020].
[11] http://www.parlaritaliano.it/index.php/it/corpora-di-parlato/672-corpus-pratid-nelle-lingue-europee [última consulta: junio de 2020].
[12] http://www.parlaritaliano.it/index.php/it/corpora-di-parlato/716-corpus-ananas-multilingue-ananasmt [última consulta: junio de 2020].

orales, por lo que resulta especialmente interesante. Se utiliza la versión 1.0, que incluye cerca de cinco millones de palabras en español.

- ESLORA[13] es un corpus oral transcrito formado por conversaciones de hablantes de Galicia y entrevistas en español y elaborado por el Grupo de Gramática del Español de la Universidad de Santiago de Compostela. El corpus contiene unas 650 000 palabras transcritas.

Todos estos recursos se seleccionan para formar parte del modelo de traducción y de los modelos de lengua con distintas configuraciones, como se detalla en el apartado 4.3 con el número exacto de líneas y palabras usadas de cada uno. Sin embargo, para la optimización de los motores, se decide crear un pequeño corpus bilingüe con textos seleccionados que se parezcan más al texto a traducir tanto en temática como en origen del texto y de la traducción.

4.2.1 Creación de corpus para la optimización

Como se menciona en el apartado anterior, debido a la falta de recursos libres con las características buscadas para optimizar los motores de TA, se opta por crear uno. Puesto que el audio seleccionado para probar los sistemas de TA proviene de una película en inglés que ha sido doblada al español, los textos para la optimización se extraen de películas y series en inglés que también dispongan de un doblaje en español. Además, la temática de la escena elegida es una discusión matrimonial de una pareja, por lo que el registro es informal. De esta manera, se descartan las escenas en las que haya un cierto grado de formalidad y se buscan escenas con parejas heterosexuales, como la pareja de *Marriage Story*.

Teniendo en cuenta estas características, se escoge una escena de las películas *The Notebook* (2004), *El diario de Noa* en español, *Friends With Benefits* (2011), *Con derecho a roce* en español, y *The Break Up* (2006), *Viviendo con mi ex* en español, y dos episodios de la serie *Friends* (2004). Los diálogos seleccionados de las tres películas son discusiones en pareja y, en el caso de *Friends*, se escogen dos episodios al azar («The One Where Monica Gets a Roommate (Pilot)» y «The One on the Last Night») de los que

[13] Corpus para el estudio del español oral (http://eslora.usc.es [última consulta: junio de 2020]), versión 1.2.2 de noviembre de 2018, ISSN: 2444-1430.

se extrae más texto con marcas de oralidad y diálogos informales entre personas que se conocen y parejas.

Para obtener los textos transcritos de los diálogos originales y de los doblajes, en primer lugar, se graban los diálogos de las películas, todas disponibles en la plataforma Netflix[14], con una extensión para Google Chrome llamada Chrome Audio Capture[15]. La transcripción de los audios obtenidos, al igual que con el diálogo de *Marriage Story*, se lleva a cabo de forma automática con Happy Scribe y se revisa con el editor del mismo programa. Una vez revisadas, se exportan las transcripciones del original y de la traducción en formato TXT y se segmentan y alinean de forma manual en una hoja de Excel.

En el caso de la serie *Friends*, no es necesario transcribir las distintas escenas, pues el diálogo original se descarga de las páginas web de fans de la serie en inglés[16] y en español[17]. Aun así, se comprueba que las transcripciones se correspondan con los diálogos y el doblaje y que no se hayan extraído de los subtítulos o se haya traducido literalmente el guion original.

Los diálogos de *Friends*, al contener mucho más texto que las escenas de las películas, se decide alinearlos de forma semiautomática con SDL Trados Studio 2019[18] (en adelante «Trados»). De esta manera, solo se deben modificar unos pocos segmentos y se agiliza el proceso. Además, la opción `Ir a` de Trados permite ir directamente a las alineaciones que la misma herramienta considera de mala calidad (véase la figura 5) y modificarlas. El editor de alineaciones de Trados también permite modificar el texto y dejar sueltos ciertos segmentos, ya sea porque no tienen traducción o porque no son de interés, por lo que se han podido corregir errores ortográficos de la transcripción y eliminar comentarios sobre las escenas que no eran diálogos. Tras confirmar todas las alineaciones, se deben importar a una memoria de traducción que se puede exportar en formato TMX.

[14] Página web de Netflix: https://www.netflix.com/ [última consulta: junio de 2020].
[15] Enlace de la extensión Chrome Audio Capture: https://chrome.google.com/webstore/detail/chrome-audio-capture/kfokdmfpdnokpmpbjhjbcabgligoelgp [última consulta: junio de 2020].
[16] http://www.livesinabox.com/friends/ [última consulta: junio de 2020].
[17] https://www.friendspeich.com/ [última consulta: junio de 2020].
[18] Página web de SDL Trados Studio: https://www.sdl.com/es/software-and-services/translation-software/sdl-trados-studio/ [última consulta: junio de 2020].

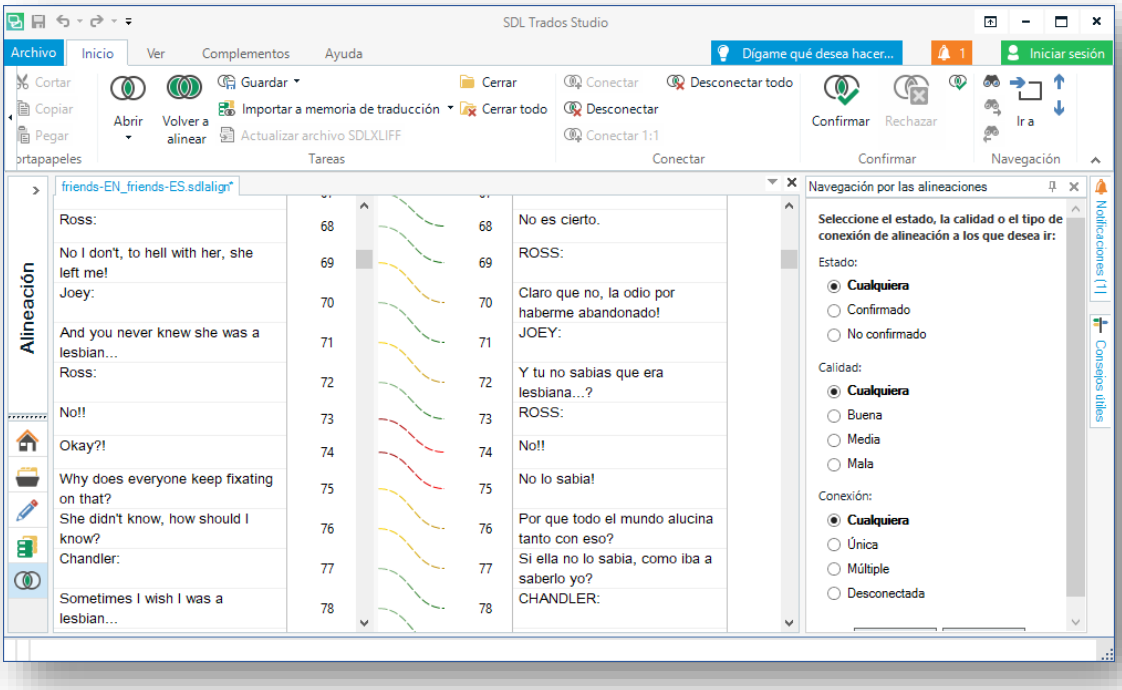


Figura 5. Vista del editor de alineaciones de SDL Trados Studio 2019 con las opciones de navegación por las alineaciones a la derecha

El resultado de todo este proceso es un pequeño corpus de 1272 líneas, 9154 palabras en inglés y 8615 palabras en español que se utilizará para optimizar algunos de los sistemas de TAE.

4.3 Configuración de los motores de TA

Para cumplir los objetivos del trabajo y resolver las hipótesis planteadas, se establecen seis escenarios de entrenamiento de motores de traducción automática estadística. En cada escenario, la configuración y las características del motor varían ligeramente para comprobar si se pueden obtener distintos resultados.

Tras comprobar en la búsqueda de recursos que apenas existen corpus orales transcritos y paralelos en inglés y español que sean de acceso libre, gratuitos y de buena calidad, se descarta la opción de insertar textos orales en el entrenamiento del modelo de traducción. En su lugar, se decide usar los corpus orales transcritos en español encontrados para modificar el modelo de lengua de los sistemas de TA. Esto significa que todos los motores tienen en común el modelo de traducción; simplemente se añaden o

eliminan ciertos corpus en el modelo de lengua y se introduce o no una optimización para observar la variación de los resultados de cada motor. En la tabla 4, se muestran las combinaciones de cada sistema de TA.

N.º de motor de TA	**Corpus del MT**	**Corpus del ML**	**Optimización**
1	Escrito	Escrito	No
2	Escrito	Escrito	Sí
3	Escrito	Mixto (escrito y oral)	No
4	Escrito	Mixto (escrito y oral)	Sí
5	Escrito	Oral transcrito	No
6	Escrito	Oral transcrito	Sí

Tabla 4. Configuración de los motores de TAE

El primer motor se considera la base sobre la que se realizan las modificaciones necesarias del resto de los motores. Se elabora con un modelo de traducción y un modelo de lengua entrenados con corpus escritos y sin ningún tipo de optimización. Para el modelo de traducción o bitexto de los seis sistemas de TA, se utiliza el corpus de ParaCrawl, puesto que contiene un gran número de palabras alineadas en inglés y español y las fuentes de los textos son variadas. Al tratarse de un corpus tan grande —casi 40 millones de líneas y 900 millones de palabras—, se decide utilizar la mitad del corpus, unos 20 millones de líneas, que será suficiente para obtener una traducción correcta del tipo de texto con el que se pretende trabajar.

En cuanto al modelo de lengua escrito que se utiliza en los motores 1 y 2 —y en los motores 3 y 4 combinado con el modelo de lengua oral transcrito—, se entrena con los corpus de Wikipedia y Tatoeba. En la tabla 5, se puede ver con detalle el número de líneas y de palabras de cada corpus y del total del ML escrito.

El modelo de lengua oral transcrito con el que se entrenan los motores 5 y 6 —y los motores 3 y 4 en combinación con el ML escrito— está formado por textos de once corpus distintos. La tabla 7 incluye el nombre, el número de líneas y el número de palabras de cada corpus y el número total de líneas y de palabras del ML oral transcrito.

Por último, el modelo de lengua mixto está formado por los tres corpus del ML escrito y los once corpus del ML oral transcrito, sumando un total de catorce corpus con más de tres millones y medio de líneas y setenta millones de palabras, como se puede observar en la tabla 6.

ML escrito		
Corpus	**Número de líneas**	**Número de palabras**
Tatoeba	318 320	2 250 336
Wikipedia	1 962 602	38 965 392
Total	**2 280 922**	**41 215 728**

Tabla 5. Corpus del ML escrito

ML mixto		
Corpus	**Número de líneas**	**Número de palabras**
Activ-ES	434 462	2 652 839
AN.ANA.S.	620	14 555
DiEspa	1138	8133
EPIC	1023	7016
ESLORA	283	2504
Glissando	11 071	129 809
Heroico	15 762	51 408
MuST-C	265 625	5 018 910
PraTiD	53 800	765 668
Tatoeba	318 320	2 250 336
TED2013	157 785	2 556 005
TEDx Spanish	11 243	244 915
Wikipedia	1 962 602	38 965 392

Total	**3 233 734**	**52 667 490**

Tabla 6. Corpus del ML mixto

ML oral transcrito		
Corpus	**Número de líneas**	**Número de palabras**
Activ-ES	434 462	2 652 839
AN.ANA.S.	283	2504
DiEspa	1023	7016
EPIC	620	14 555
ESLORA	53 800	765 668
Glissando	11 071	129 809
Heroico	15 762	51 408
MuST-C	265 625	5 018 910
PraTiD	1138	8133
TED2013	157 785	2 556 005
TEDx Spanish	11 243	244 915
Total	**952 812**	**11 451 762**

Tabla 7. Corpus del ML oral transcrito

4.4 Preparación de archivos

Al descargar un gran número de archivos de fuentes muy variadas, es necesario preparar algunos de ellos para unificarlos, facilitar la carga de archivos a MTradumàtica y evitar errores en el entrenamiento y en los resultados de los sistemas de TA.

El proceso de preparación de archivos se divide en varias fases: división y unión de archivos, limpieza de etiquetas, breve control de calidad, segmentación, conversión de formatos y de codificación, etc. En los siguientes apartados, se explica en qué consiste cada proceso y se enumeran los corpus que necesitan este tipo de preparación.

4.4.1 División de archivos

El primer paso de la preparación de archivos consiste en dividir los archivos de más de 500 MB en varios archivos más pequeños. El objetivo de este paso es facilitar la carga de archivos a MTradumàtica, la primera opción de la autora, puesto que la plataforma no admite archivos cuyo tamaño sea mayor a 500 MB.

El único corpus que se debe dividir por superar el límite de tamaño es ParaCrawl. El corpus contiene dos archivos, uno en inglés y otro en español, de unos 5 GB cada uno que contienen el mismo número de líneas, puesto que son paralelos y cada línea de un archivo se corresponde con el mismo número de línea del otro archivo.

Ante este problema, la primera solución que surge es la de dividir los archivos por tamaño en varios archivos de 400 o 500 MB cada uno. Sin embargo, con esta opción se podrían desalinear los segmentos en inglés y en español y, además, se podrían cortar las frases e incluso las palabras. Por este motivo, se busca otra solución que respete la alineación y mantenga las líneas completas.

`split` Nombre archivo entrada Prefijo archivo salida `-l` N.º de líneas `-a` N.º de dígitos del sufijo `-d`

Figura 6. Comando `split` para dividir archivos por número de líneas

En esa búsqueda, se descubre el comando `split` (véase la figura 6), que funciona en la terminal del sistema operativo de Linux y que permite dividir archivos tanto por tamaño como por líneas. Si se quiere dividir un archivo por líneas, como es el caso, se utiliza la siguiente estructura: `split -l` seguido del número de líneas y del nombre del archivo que se pretende dividir. Además, este comando tiene otras opciones para personalizar el nombre de los archivos generados, puesto que, si no se indica nada más, los archivos divididos se llamarán `xaa`, `xab`, `xac`, etc. Las opciones del comando que se usan en este trabajo se indican en la tabla 8. Para poder utilizar este comando en el sistema operativo de Windows 10, se descarga e instala la aplicación Git Bash[19].

[19] Página web de Git: https://git-scm.com/ [última consulta: abril de 2020].

Opciones del comando `split`	**Funciones**
`-l`	Divide el archivo en varios archivos con el número de líneas que se indique después.
`-d`	Utiliza números en el sufijo de los archivos generados en lugar de letras.
`-a`	Sirve para indicar el número de dígitos del sufijo.

Tabla 8. Opciones del comando `split` usadas en el trabajo

Antes de dividir el corpus de ParaCrawl, se realiza una prueba con un corpus más pequeño para comprobar si se consigue el resultado esperado. El corpus elegido para esta prueba es TED2013[20], que contiene 156 698 segmentos alineados en inglés y español y más de cinco millones de palabras. El corpus está formado por dos archivos, uno en inglés (`TED2013.en-es.en`) y uno en español (`TED2013.en-es.es`). Para la prueba, se decide dividir el archivo en inglés en 10 archivos más pequeños de 15 670 líneas cada uno que se denominen TED-EN_ seguido de un número de dos dígitos (por ejemplo, `TED-EN_00`, `TED-EN_01`, etc.). Para conseguir este objetivo, se utiliza la siguiente línea de comando: `split TED2013.en-es.en TED-EN_ -l 15670 -a 2 -d`. Al ejecutarlo, se crean diez archivos de tamaños similares y un undécimo archivo con las líneas sobrantes y notablemente más pequeño. En la figura 7 se puede observar el resultado de la división y el comando usado en la aplicación de Git Bash.

Antes de continuar con la división del corpus de ParaCrawl, se divide el archivo en español de TED2013 de la misma manera y se abren dos archivos en inglés y sus correspondientes en español con el editor de texto Notepad++[21] para comprobar que la división es correcta y que las líneas de los archivos en inglés se corresponden con su traducción en español. Tras comparar las primeras y las últimas líneas de estos dos archivos en inglés con las primeras y últimas líneas de los dos archivos en español, se concluye que la división es correcta y se procede a dividir el corpus de ParaCrawl.

[20] Más información sobre el corpus en el apartado 4.2 de este trabajo.
[21] Página web de Notepad++: https://notepad-plus-plus.org/ [última consulta: abril de 2020].

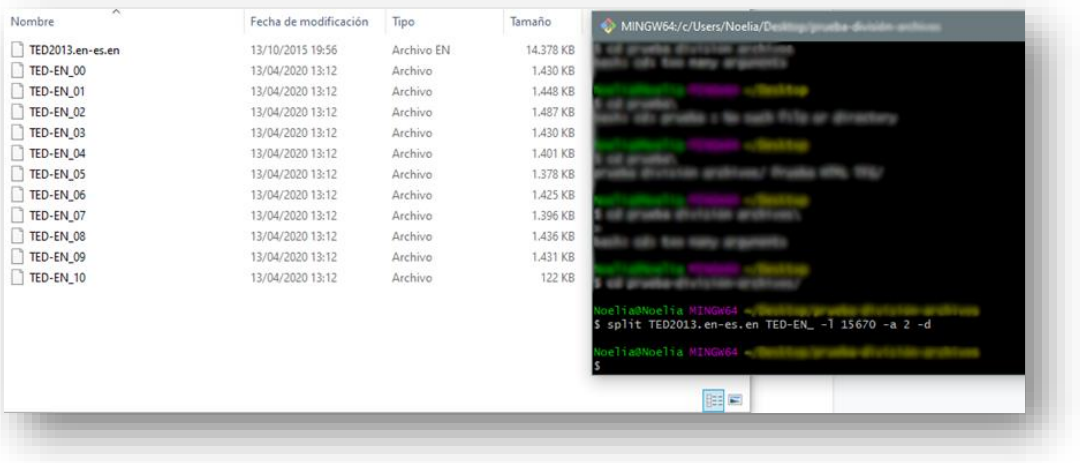
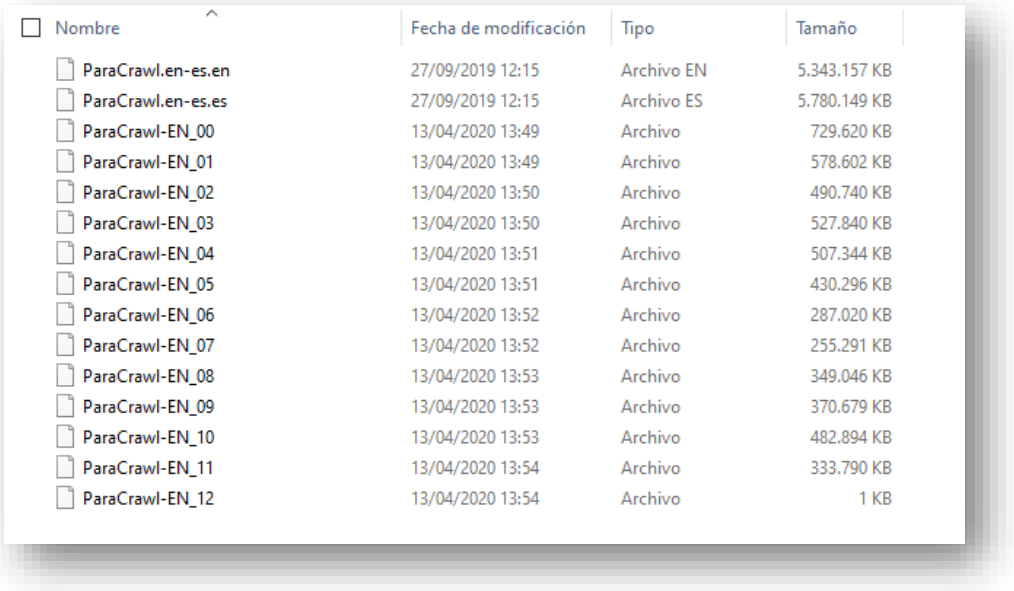


Figura 7. Resultado de la división del corpus de TED2013 (izquierda) y comando utilizado (derecha)

Como ya se ha indicado anteriormente, el corpus de ParaCrawl se compone de dos archivos (uno en inglés y otro en español) de casi 5 GB cada uno y formados por 38 971 348 segmentos alineados. Con el objetivo de conseguir archivos de un máximo de 500 MB, se decide dividir cada archivo en 12 partes de 3 247 616 líneas siguiendo los pasos anteriores, pero el resultado no es el esperado. Al contrario que en la prueba con el corpus de TED, los archivos obtenidos de la división son de tamaños muy variados a pesar de tener el mismo número de líneas, por lo que hay algunos archivos de hasta 700 MB y otros de tan solo 290 MB, como se puede observar en la figura 8. Para evitar problemas al cargar los archivos en MTradumàtica, se procede a dividirlos en 25 partes de 1 558 854 líneas; esta vez el archivo más grande ocupa unos 400 MB. El comando utilizado para realizar esta división es el siguiente: `split ParaCrawl.en-es.en ParaCrawl-EN_ -l 1558854 -a 2 -d`.

Al igual que en la prueba con el corpus de TED, se comparan las primeras y las últimas líneas de dos archivos al azar en inglés con sus correspondientes archivos en español para asegurar que la alineación es correcta.



Nombre	Fecha de modificación	Tipo	Tamaño
ParaCrawl.en-es.en	27/09/2019 12:15	Archivo EN	5.343.157 KB
ParaCrawl.en-es.es	27/09/2019 12:15	Archivo ES	5.780.149 KB
ParaCrawl-EN_00	13/04/2020 13:49	Archivo	729.620 KB
ParaCrawl-EN_01	13/04/2020 13:49	Archivo	578.602 KB
ParaCrawl-EN_02	13/04/2020 13:50	Archivo	490.740 KB
ParaCrawl-EN_03	13/04/2020 13:50	Archivo	527.840 KB
ParaCrawl-EN_04	13/04/2020 13:51	Archivo	507.344 KB
ParaCrawl-EN_05	13/04/2020 13:51	Archivo	430.296 KB
ParaCrawl-EN_06	13/04/2020 13:52	Archivo	287.020 KB
ParaCrawl-EN_07	13/04/2020 13:52	Archivo	255.291 KB
ParaCrawl-EN_08	13/04/2020 13:53	Archivo	349.046 KB
ParaCrawl-EN_09	13/04/2020 13:53	Archivo	370.679 KB
ParaCrawl-EN_10	13/04/2020 13:53	Archivo	482.894 KB
ParaCrawl-EN_11	13/04/2020 13:54	Archivo	333.790 KB
ParaCrawl-EN_12	13/04/2020 13:54	Archivo	1 KB

Figura 8. Resultado de la primera división del corpus de ParaCrawl

4.4.2 Unión de archivos

Al contrario que en el paso anterior, en aquellos corpus que contienen numerosos archivos con pocas líneas y de tamaño reducido es necesario llevar a cabo un proceso de fusión de todos ellos en un único archivo más grande. Esta necesidad surge porque, al crear los monotextos y el bitexto, se deben añadir todos los archivos uno a uno, y algunos corpus de los doce seleccionados contienen más de 400 archivos. Como se puede imaginar, se trataría de un proceso muy lento y arduo que se puede agilizar fácilmente mediante el uso de la tecnología.

Los corpus que necesitan esta fusión de archivos son Activ-ES, Glissando, Heroico, DiEspa, EPIC, ESLORA y PraTiD. Algunos de estos corpus solo contienen cuatro o cinco archivos, como DiEspa y PraTiD, pero otros, como Activ-ES y Glissando, están formados por más de 400 archivos de menos de 100 KB.

Tras encontrar el comando `split` que permite la división de archivos, se investiga la existencia de un comando similar que pueda realizar este proceso de unión de archivos. Entonces, se encuentra un comando que funciona en el sistema operativo de Windows 10 y que busca todos los archivos de un mismo formato en la ubicación desde donde se ejecute y los junta en un mismo archivo en la ubicación que se indique. En la figura 9 se incluye este comando con el ejemplo utilizado para los archivos XML del

corpus de Glissando. En el caso de este ejemplo, no solo se utiliza el comando para juntar los archivos, sino que, además, se aprovecha para convertir el formato a TXT.

```
for %f in (*.Extensión archivos) do type "%f" >> Ruta\Nombre archivo salida.Extensión archivo salida

for %f in (*.xml) do type "%f" >> C:\glissando\glissando.txt
```

Figura 9. Comando para unir archivos (primera línea) y ejemplo de uso con los archivos XML de Glissando (segunda línea)

Como ya se avanza en el párrafo anterior, la primera parte del comando —`for %f in (*.xml)`— sirve para buscar todos los archivos con la extensión que añadamos en el paréntesis. En el caso de la figura 9, se buscan las extensiones `(*.xml)`, pero se pueden buscar los archivos en TXT `(*.txt)` o en cualquier otro formato. En la segunda parte del comando —`do type "%f" >> C:\glissando\glissando.txt`—, se indica que se escriba el contenido de los archivos encontrados en un archivo nuevo en la ruta que se indique después de los signos >>. En el ejemplo de Glissando, el archivo `glissando.txt` que se indica al final en la línea de comandos no existe antes de ejecutarlo, sino que se crea una vez se realiza la orden.

Tras comprobar el funcionamiento del comando con los archivos XML de Glissando, se procede a realizar la misma acción con el resto de los corpus. En el caso de Activ-ES, que contiene archivos en español de Argentina, de México y de España, se decide juntarlos por estas variantes, obteniendo un total de tres archivos.

4.4.3 Limpieza de etiquetas e información

Muchos de los archivos de texto de los corpus orales transcritos contienen información insertada en el texto a través de etiquetas. Esto ocurre porque no todos los corpus descargados han sido creados para entrenar sistemas de TA, sino para investigaciones sobre el lenguaje oral o para entrenar sistemas de reconocimiento de voz, entre otras finalidades. Los corpus que requieren una limpieza de etiquetas e información innecesarias para el entrenamiento son AN.ANA.S, DiEspa, EPIC, ESLORA, Glissando, Heroico y PraTiD.

Cada corpus está etiquetado de una forma distinta y contiene información diferente, puesto que se crean en función del objetivo de cada corpus y según los criterios de sus autores. No obstante, la información de las etiquetas es bastante similar; se incluyen sonidos no lingüísticos que se puedan escuchar en los audios —como el sonido de una puerta o un golpe—, interjecciones que no forman parte del contenido textual, pero que aparecen en el discurso, palabras cortadas por interrupciones o cambios en el discurso, palabras acortadas, etc. En la figura 10 se pueden ver algunas de las etiquetas del corpus de Glissando.

Antes de comenzar a eliminar etiquetas, es necesario identificar cada una de ellas y decidir si el contenido de la etiqueta puede ser útil para el entrenamiento de los sistemas de TA. La herramienta que se utiliza para todo el proceso de búsqueda y reemplazo de etiquetas es el editor de texto Notepad++. Con este programa, se puede ver fácilmente el esquema de etiquetas para los archivos en XML, como algunos de los archivos del corpus Glissando, gracias a los colores de las etiquetas. No obstante, se utiliza una expresión regular para buscar todas las etiquetas del documento y se elabora una lista del tipo de etiquetas de cada corpus.

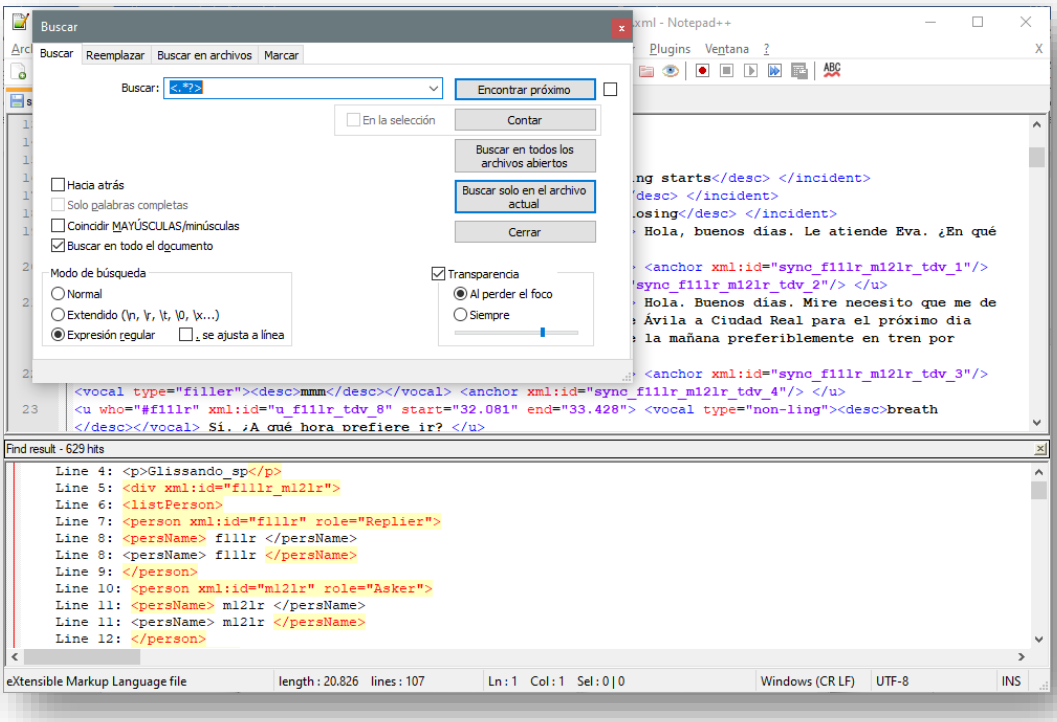


Figura 10. Expresión regular utilizada para buscar etiquetas del corpus Glissando con Notepad++

La mayoría de la información se encuentra en etiquetas, marcadas con el signo menor que (<) en el inicio y con el signo mayor que (>) en el final, salvo alguna información que se incluye entre corchetes [] o entre paréntesis (). Para buscar las etiquetas, se utiliza, por tanto, una expresión regular (véase la figura 10) para encontrar el texto que esté entre esas dos marcas de inicio y fin. La expresión regular está formada por un punto (`.`), que significa cualquier carácter (letra, número, signo, espacio, etc.), un asterisco (`*`), que indica que el carácter anterior puede aparecer de 0 a infinitas veces, y un signo de interrogación de cierre (`?`), que limita el fin de la búsqueda a lo más cercano. Es decir, la búsqueda `<.*?>` con las expresiones regulares activas encuentra cualquier letra, palabra, frase o incluso párrafo que se encuentre entre el signo de apertura (<) y el siguiente signo de cierre (>). Si no se incluye el signo de interrogación en la expresión regular y hay varias etiquetas en una frase, Notepad++ marcará desde el primer signo de apertura (<) hasta el último de cierre (>), incluyendo varias etiquetas en una misma búsqueda y texto que no forma parte de ninguna etiqueta.

Al seleccionar la opción de «Buscar en todo el documento» y hacer clic en «Buscar solo en el archivo actual», se abre una pestaña con todos los resultados del documento marcados con un color rojo y resaltados con amarillo (véase la figura 10). De esta manera, se pueden ver de un solo vistazo todas las etiquetas.

Una vez identificadas todas las etiquetas de cada corpus, se continúa eliminando aquellas que no son necesarias para el entrenamiento de los motores de TA y modificando las que sí contienen información útil sobre las características del lenguaje oral. Por ejemplo, la información sobre los hablantes o los sonidos del entorno no se consideran relevantes para el entrenamiento, pero las palabras y frases truncadas, las pausas largas y las interjecciones se dejan como parte del texto.

Se comienza el proceso de limpieza de etiquetas eliminando todas las innecesarias. Para ello, en el espacio para buscar, se incluye cada tipo de etiqueta y se deja en blanco el espacio para reemplazar. Para los archivos que contienen información sobre el corpus —nombre, participantes, referencias, etc.— (véase la figura 11) y etiquetas dobles con texto en medio (véase la figura 12), es necesario utilizar expresiones regulares para agilizar el proceso y eliminar el texto que se encuentra entre dos pares de etiquetas y que no forma parte del diálogo o discurso.

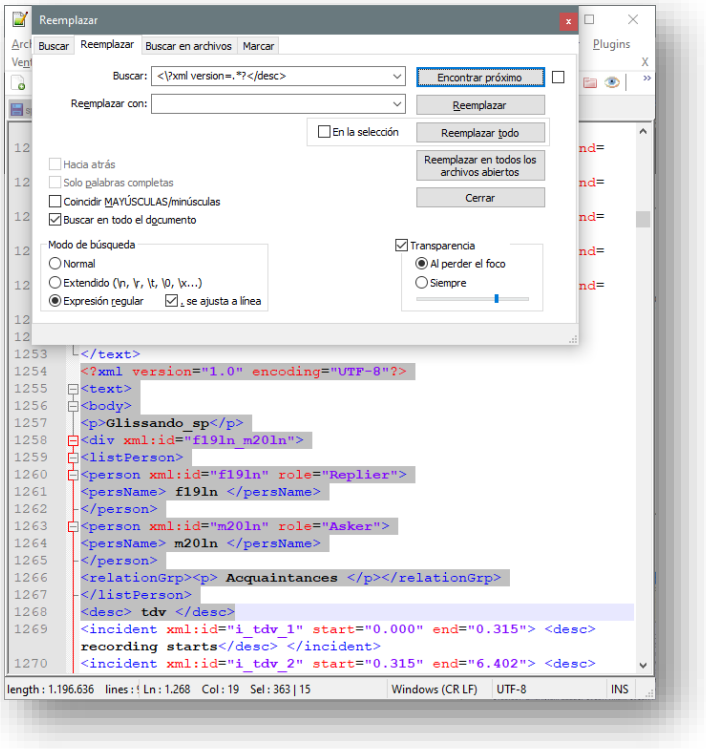


Figura 11. Ejemplo de búsqueda y reemplazo de información sobre el archivo con expresiones regulares

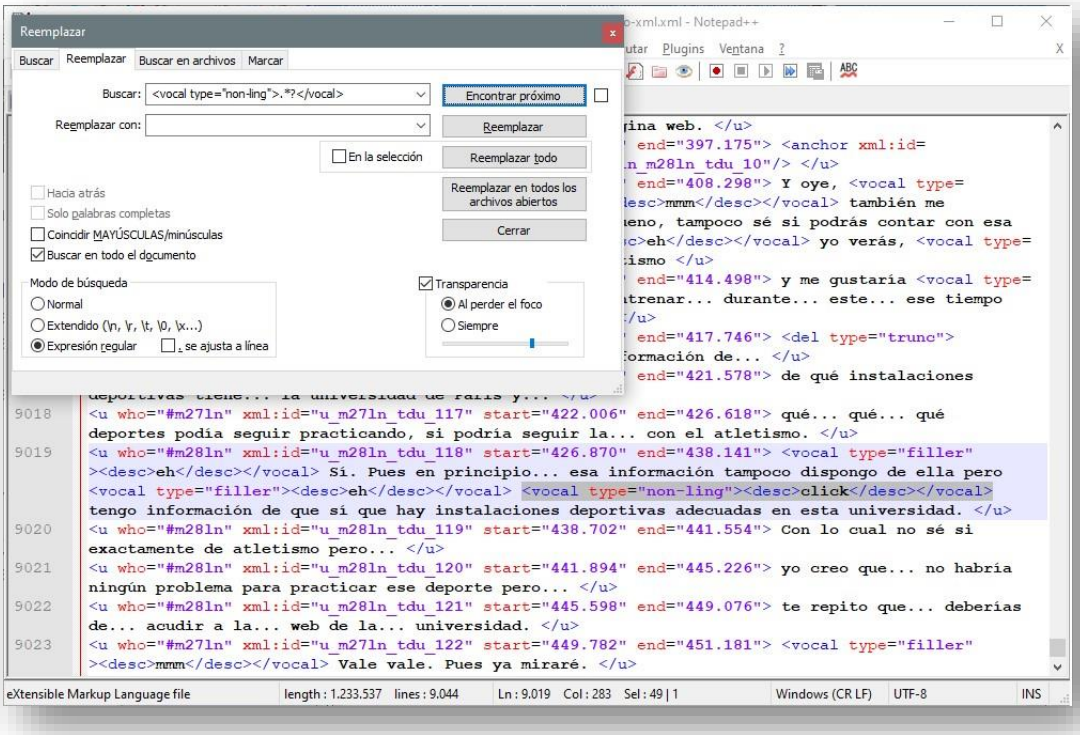


Figura 12. Ejemplo de búsqueda y reemplazo de etiquetas con elementos no lingüísticos con expresiones regulares

Tras quitar todas las etiquetas innecesarias, se modifican las etiquetas que incluyen información que se quiere mantener en el texto. Por ejemplo, los corpus de Parlare italiano (AN.ANA.S, DiEspa y PraTiD) insertan las interjecciones en etiquetas (<ehm>, <eeh>, <mhmh>, <ah>, etc.) y no como parte del texto. El corpus Glissando también incluye este tipo de interjecciones, pero con etiquetas dobles, de apertura y de cierre: <vocal type="filler"><desc>eh</desc></vocal>. Otros ejemplos de información etiquetada que se decide dejar son las frases o palabras truncadas, marcadas con el símbolo + en los corpus de Parlare italiano y con la etiqueta <del type="trunc"> en Glissando.

Para modificar este tipo de etiquetas, se buscan el texto de la etiqueta y se sustituye por lo que se quiere mostrar en el texto. Por ejemplo, en la figura 13 se ve cómo se sustituye la etiqueta <ah!> del corpus PraTiD por el texto «¡ah!,». Así pasa de ser mera información a formar parte del texto, del discurso. Se sigue el mismo procedimiento con el resto de las etiquetas con interjecciones.

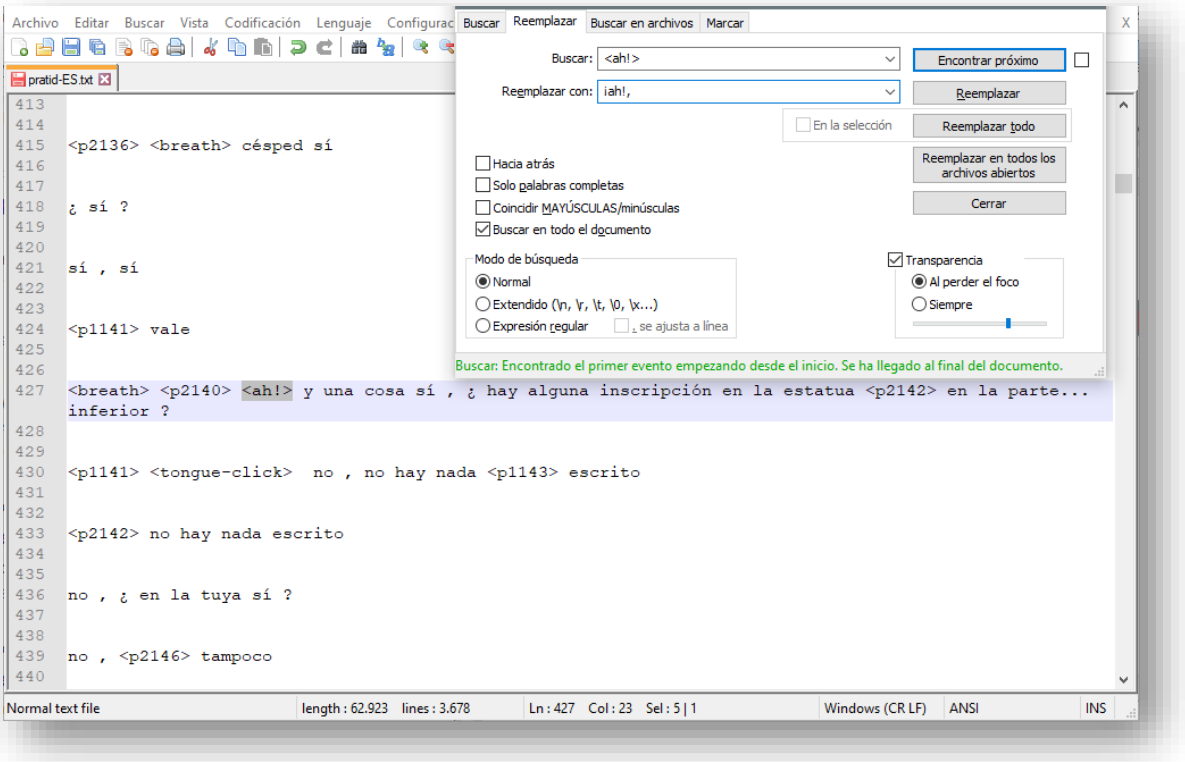


Figura 13. Ejemplo de búsqueda y reemplazo de etiquetas con interjecciones

En el caso de las palabras y frases truncadas, el proceso es bastante similar al anterior, sobre todo en el caso de los corpus de Parlare italiano, puesto que simplemente se sustituyen los símbolos + por tres puntos suspensivos (…) para indicar que la frase o palabra está incompleta. Sin embargo, para el corpus de Glissando se tienen que hacer dos búsquedas distintas porque la misma etiqueta contiene dos tipos de información: palabras cortadas por interrupciones o cambios de tema y palabras acortadas por la pronunciación de los hablantes. En el segundo caso, se incluyen las letras que no se han pronunciado entre paréntesis, por lo que es necesario utilizar expresiones regulares para incluir el texto sin que los paréntesis corten la palabra.

Una vez añadido todo el texto relevante de las etiquetas, se procede a eliminar el resto de las etiquetas e información para conseguir como resultado archivos de texto plano.

4.4.4 Control de calidad y segmentación

Algunos de los corpus elegidos para el entrenamiento de los motores de TA, a pesar de que son muy interesantes, no tienen una calidad perfecta. Esto ocurre, sobre todo, en los corpus orales transcritos, puesto que, en varias ocasiones, las transcripciones han sido realizadas por numerosas personas voluntarias, que pueden no ser profesionales de la lengua, o se han llevado a cabo de forma automática sin una revisión profunda. Aunque realizar un control de calidad a los corpus no es una prioridad ni uno de los objetivos de este trabajo, se considera necesario realizarlo para ciertos archivos, en concreto, para los corpus Activ-ES y TED2013.

Con solo abrir los archivos, se detectan algunos errores ortográficos y ortotipográficos que podrían dañar la calidad de los sistemas de TA. Por suerte, muchos de estos errores son repetitivos y se pueden solucionar con la función de búsqueda y reemplazo de Microsoft Word y de Notepad++ y con algunas macros generales para corregir errores de puntuación, como los dobles espacios, espacios antes de coma o punto, etc. Además de estas búsquedas, se pasa el corrector ortográfico de Word. Al tratarse de un proceso muy lento, se lleva a cabo solo en los archivos donde se considera realmente necesario.

Con el control de calidad realizado, se procede a segmentar algunos de los archivos. La mayoría de los corpus ya están preparados para el entrenamiento y segmentados por frases, pero algunos corpus como Activ-ES, que no está segmentado y contiene todo el texto en una sola línea, o Glissando, que tiene líneas con varias oraciones, necesitan una segmentación previa.

Para realizar esta segmentación, se utiliza el menú de búsqueda y reemplazo de Notepad++. En primer lugar, se sustituyen todos los puntos suspensivos (…) por una almohadilla (#), porque, si los puntos suspensivos marcan el truncamiento de una palabra o frase que continúa después, es interesante que se mantenga con la siguiente frase y que no se separen en dos segmentos distintos. A continuación, se reemplazan los puntos seguidos de espacio por un punto seguido de un salto de línea. De esta manera, las frases que tienen abreviaturas como «Sr.» o «Sra.» se cortan. Aunque no hay muchas abreviaturas en los documentos, es necesario realizar otra búsqueda para juntar estas frases: se buscan las abreviaturas identificadas en los documentos seguidas de un salto de línea y se reemplazan por esas mismas abreviaturas, pero seguidas de un espacio.

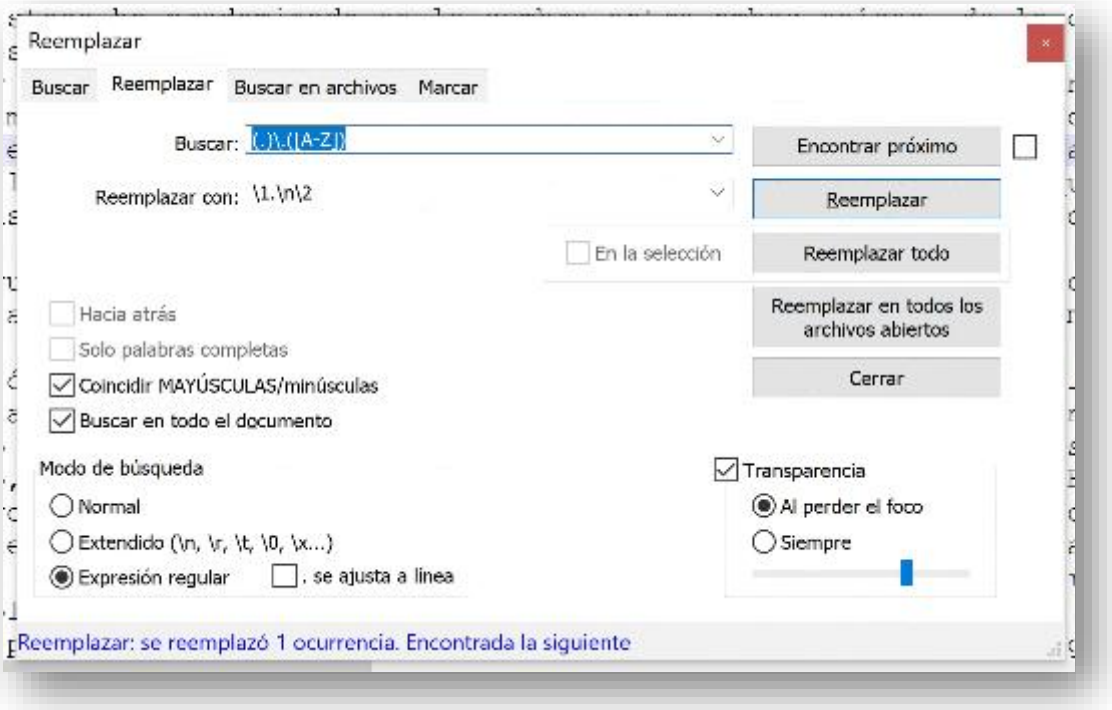


Figura 14. Expresión regular usada en Notepad++ para separar oraciones unidas por un punto

En el corpus de Glissando se encuentra un problema adicional: algunas oraciones no contienen un espacio detrás del punto, por lo que están unidas a las siguientes oraciones. Para solucionar este problema, se recurre a una expresión regular que encuentre cualquier carácter seguido por un punto y, a continuación, por una letra en mayúsculas (véase la figura 14). Esta búsqueda se reemplaza por ese primer carácter seguido de un salto de línea y de la letra en mayúsculas que se haya encontrado.

4.4.5 Conversión de formatos y codificaciones

A pesar de que muchos corpus contienen los archivos en formato TXT, algunos vienen en otros formatos como XML o RTF. Con el fin de unificar los formatos y evitar futuros problemas al cargarlos en MTradumàtica, se decide convertir todos los archivos a TXT, codificarlos en UTF-8 y convertir el fin de línea a formato UNIX.

En el caso de los XML, como se indica en el apartado 4.4.2 del trabajo, se aprovecha para convertir el formato a la vez que se juntan todos los archivos y se obtiene un solo archivo en TXT. Para el resto de los archivos que no es necesario unir, se cambia el formato manualmente, guardándolos como TXT o cambiando la extensión.

En cuanto a la codificación, la mayoría de los archivos ya están codificados en UTF-8, pero algunos necesitan una conversión. Al no ser muchos archivos, esta conversión también se realiza de forma manual mediante la herramienta Notepad++. Se abre el archivo con Notepad++ y se hace clic en `Codificación > Convertir a UTF-8`.

4.5 Entrenamiento de motores con MTradumàtica

4.5.1 Creación de los ML y del MT

Con todos los archivos preparados, el primer paso consiste en subir los archivos a la plataforma, donde se muestran algunos datos, como el número de líneas, palabras y caracteres por archivo.

Al tener disponibles todos los archivos en la plataforma, se puede comenzar a crear los monotextos con los corpus monolingües en español con los que se entrenarán los tres modelos de lengua a continuación. Para ello, se crea en `Datos > Monotexto`

tres nuevos monotextos denominados «1-escrito-ES», «2-mixto-ES» y «3-oral-transcrito-ES». Una vez creados, se añaden uno a uno los archivos del corpus monolingüe.

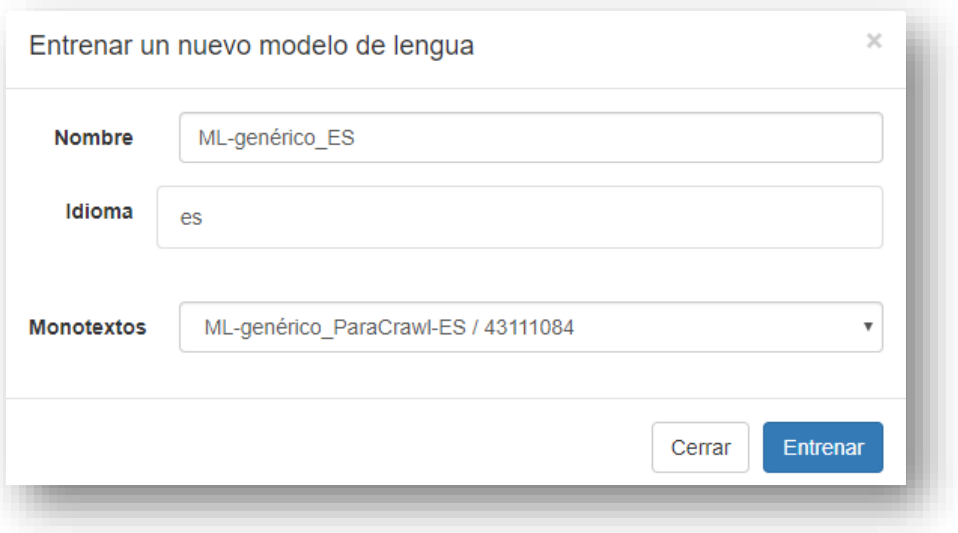


Figura 15. Diálogo de entrenamiento de un modelo de lengua con MTradumàtica

Cuando se haya agregado todo el contenido deseado a cada monotexto, se crean tres modelos de lengua en la sección `Entrenar > Modelo de lengua` con los nombres «ML-escrito-ES», «ML-mixto-ES» y «ML-oral-ES». En el diálogo de creación, se selecciona el idioma «español» y el monotexto correspondiente; por ejemplo, para el ML «ML-mixto-ES», se selecciona el monotexto «2-mixto-ES» (véase la figura 15). Al hacer clic en «Entrenar», MTradumàtica comienza a crear el modelo de lengua. El tiempo de entrenamiento de estos tres modelos de lengua oscila entre cuatro minutos y cuatro horas.

Mientras se entrenan los ML, se procede a crear el bitexto del MT común a los seis motores de TA. Para crearlo, se siguen los mismos pasos que para la creación del monotexto: en `Datos > Bitextos`, se crea un nuevo bitexto denominado «MT-escrito-ParaCrawl-EN-ES». Al igual que con los monotextos, hay que añadir uno a uno cada archivo en inglés con su correspondiente en español, como se observa en la figura 16. Por este motivo, es muy importante dividir los archivos por líneas y no por tamaño de archivo, para asegurarse de que, independientemente del tamaño de cada archivo, todos se corresponden con su traducción.

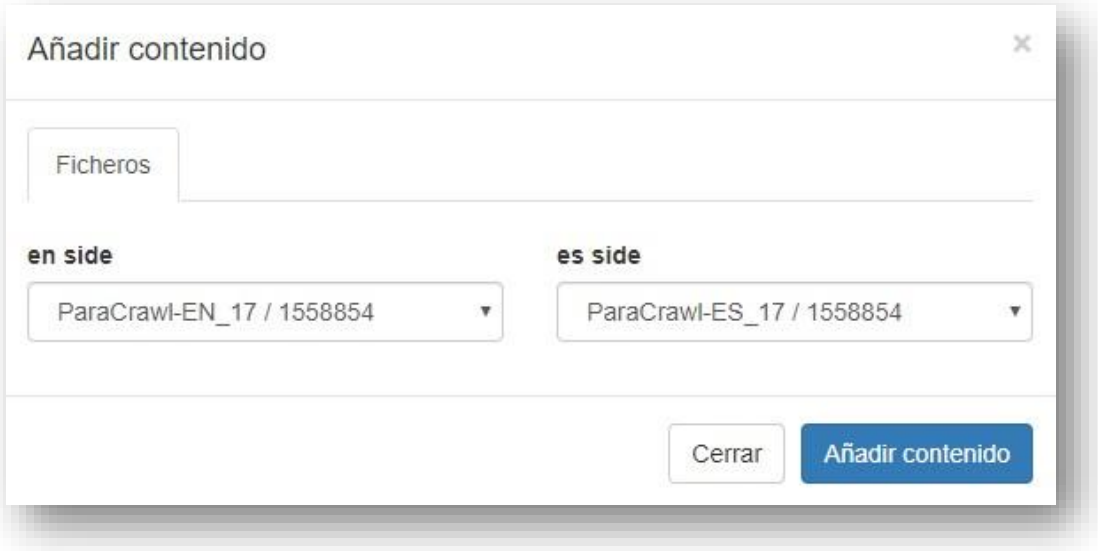


Figura 16. Diálogo para añadir contenido a un bitexto con MTradumàtica

Tras añadir todos los archivos al bitexto y con el ML escrito ya entrenado, se puede empezar a entrenar el primer motor de traducción automática. En la pestaña `Entrenar > Traductores`, se crea un nuevo traductor con el nombre «01-escrito-escrito-SO» y se configura con las características deseadas. En este caso, se selecciona la lengua de origen «inglés», la lengua meta «español», el bitexto creado en el paso anterior («MT-escrito-ParaCrawl-EN-ES») y, por último, el modelo de lengua escrito «ML-escrito-ES». Al hacer clic en «Añadir», MTradumàtica comienza a entrenar el sistema de traducción automática.

4.5.2 Error en MTradumàtica

Transcurrido más de un mes de entrenamiento, se descubre que los archivos del MT son demasiado grandes y que están bloqueando el entrenamiento. Por este motivo, se decide cambiar de corpus bilingüe y se elige el corpus TildeMODEL, que tiene cuatro millones de segmentos —16 millones de segmentos menos que ParaCrawl—, para formar el modelo de traducción escrito de los sistemas de TA.

Los archivos de texto plano de este corpus superan ligeramente el límite de 500 MB. No obstante, se intenta subirlos a la plataforma sin dividirlos antes y funciona. Tras cargar los nuevos archivos del MT y seguir el mismo proceso que en los pasos anteriores para crear el bitexto y entrenar los motores de TA, se observa que el

entrenamiento sigue bloqueándose. Al no conseguir ningún avance en varias semanas con MTradumàtica, se decide cambiar de herramienta y buscar otra alternativa.

4.6 Entrenamiento de motores con KantanMT

Después de descartar el uso de MTradumàtica, la primera herramienta elegida para elaborar el trabajo, se contacta con el equipo de KantanMT, quienes deciden colaborar con este proyecto y permiten a la autora del trabajo el acceso a su plataforma para entrenar y optimizar los distintos motores de TAE.

Al tratarse de una herramienta diferente, el procedimiento de entrenamiento varía al mencionado para MTradumàtica. Como principal diferencia cabe destacar que el usuario no debe entrenar un modelo de lengua o crear un bitexto previamente, sino que basta con subir todos los archivos necesarios en los formatos indicados y la herramienta se encarga de crear el ML y el MT de forma automática. Con MTradumàtica, el usuario debe intervenir en estos pasos; la herramienta fue desarrollada así por motivos didácticos y a fin de enseñar al usuario cada paso del entrenamiento de un sistema de TA. En los siguientes apartados, se detallan los pasos a seguir en el entrenamiento de los motores con KantanMT.

4.6.1 Preparación de archivos

Para llevar a cabo el entrenamiento de un sistema de TA con KantanMT, se deben cargar todos los archivos (el corpus bilingüe, el corpus monolingüe y el corpus de optimización) en el mismo apartado con un formato y nombre concreto para cada tipo de archivo. Por lo tanto, a pesar de que ya se había llevado a cabo un proceso de preparación de archivos anteriormente (véase el apartado 4.4), se deben volver a preparar para los requerimientos de KantanMT.

El corpus bilingüe se puede subir a la plataforma en un archivo TMX, XLIFF o XLSX bilingüe o también en dos archivos de texto plano codificados en UTF-8, uno en el idioma de origen y con el nombre `source.utf8.src` y otro archivo con la traducción denominado `source.utf8.trg`. En este caso, como ya se contaba con los archivos en texto plano y codificados en UTF-8, simplemente se cambian los nombres por los indicados.

Los corpus monolingües para los tres modelos de lengua, al igual que en el anterior caso, se pueden subir en un solo archivo codificado en UTF-8 y con el nombre `source.utf8.trg.mono`. Aunque ya se habían unido varios archivos en la fase de preparación de archivos anterior, aún se sigue teniendo un archivo por corpus, por lo que es necesario unir todos los archivos de cada ML en un solo archivo. Para ello, se recurre al comando usado en el apartado 4.4.2: `for %f in (*.txt) do type "%f" >> C:\ML-escrito\source.utf8.trg.mono`. Tras ejecutar este bucle con cada uno de los ML, se obtiene como resultado tres archivos con el mismo nombre, pero con distinto contenido.

En el caso del corpus para la optimización de motores, KantanMT ofrece dos opciones: la primera, cargar dos archivos de texto codificados en UTF-8, uno con el texto de origen llamado `source.tune.src` y otro con la traducción denominado `source.tune.trg` y, la segunda opción, subir un archivo de Excel con el texto original en la columna A y la traducción en la columna B y con el nombre `tune.reference.set.xlsx`.

Al tener parte del corpus de optimización en una hoja de Excel, se opta por esta segunda opción. Sin embargo, la parte de los episodios de Friends está guardada en formato TMX, puesto que es la única opción de exportación de las memorias de Trados, a parte del formato propio de la herramienta (SDLTM).

Para poder extraer el texto de origen y la traducción alineada del archivo TMX, se recurre a la herramienta Olifant[22]. Esta herramienta tiene numerosas funciones para editar memorias de traducción, pero se utiliza solo una, y probablemente la más simple: la exportación de la memoria. Con Olifant, una memoria de traducción se puede exportar en formato TMX o en formato TXT (Wordfast File). La estructura de este último formato es mucho más sencilla que la del TMX y con una expresión regular en Notepad++ se puede extraer el texto en inglés y después el texto en español.

[22] Olifant es una herramienta de Okapi Framework especial para editar memorias de traducción. Se puede descargar desde este enlace: http://okapi.sourceforge.net/downloads.html [última consulta: junio de 2020].

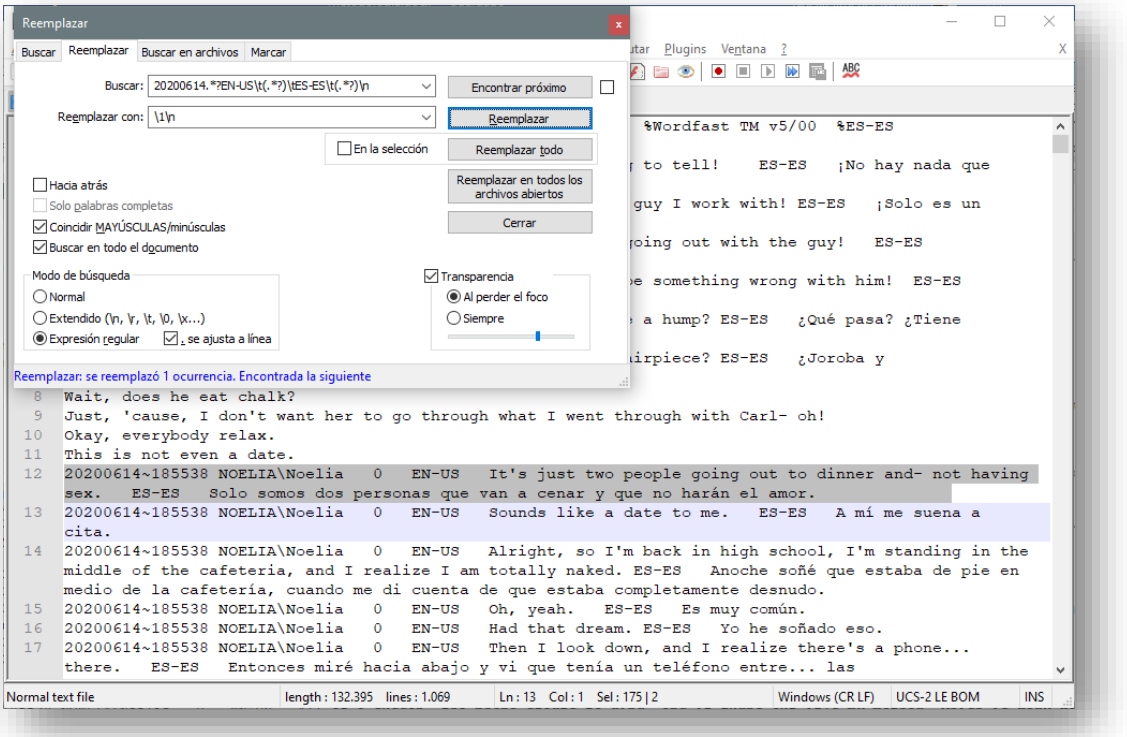


Figura 17. Expresión regular utilizada en Notepad++ para extraer el texto en inglés de la memoria en formato Wordfast File (.txt)

En la figura 17, se puede observar la expresión regular utilizada para extraer el texto en inglés y el resultado de la búsqueda y reemplazo en los segmentos del 8 al 11. Para extraer el texto en español se utiliza la misma expresión regular, pero en el recuadro de reemplazo se pone `\2\n` en lugar de `\1\n` para seleccionar el segundo grupo, marcado por los paréntesis. Tras guardar los dos archivos TXT con el texto plano, se copia el texto y se pega en la columna correspondiente de la hoja de Excel con el nombre `tune.reference.set.xlsx`.

4.6.2 Entrenamiento de los sistemas de TA

Con todos los archivos preparados, se puede comenzar a crear los sistemas de TA. En el panel de inicio de KantanMT, se hace clic en `New` para abrir el asistente de creación de un motor de TA (véase la figura 18) y se configuran las características del motor: el nombre, el tipo de motor (estadístico o neuronal), el idioma de origen y de destino e incluso se puede seleccionar alguno de los corpus que proporciona KantanMT en el apartado `Library`.

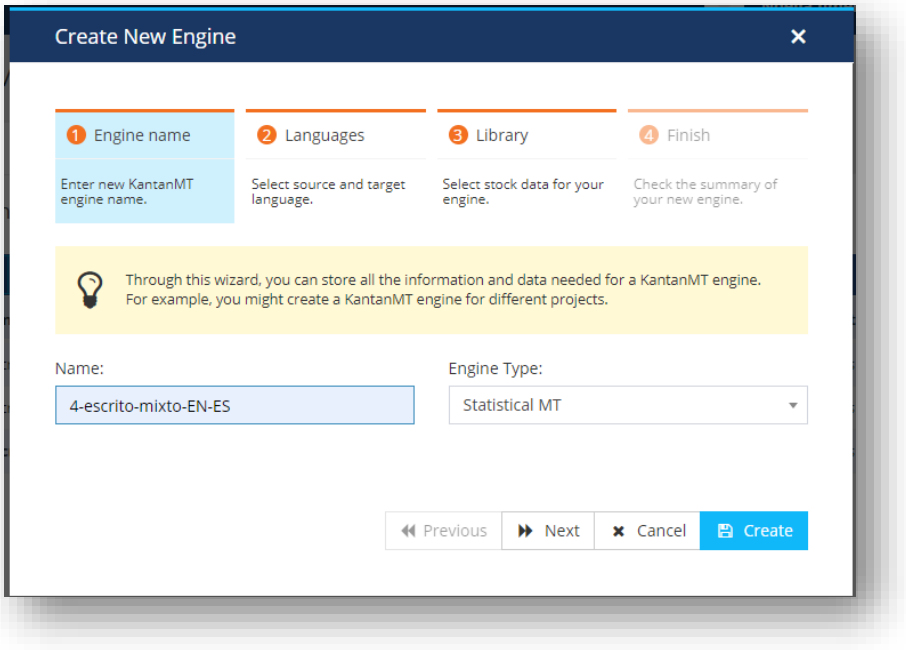


Figura 18. Asistente de creación de un sistema de TA de KantanMT

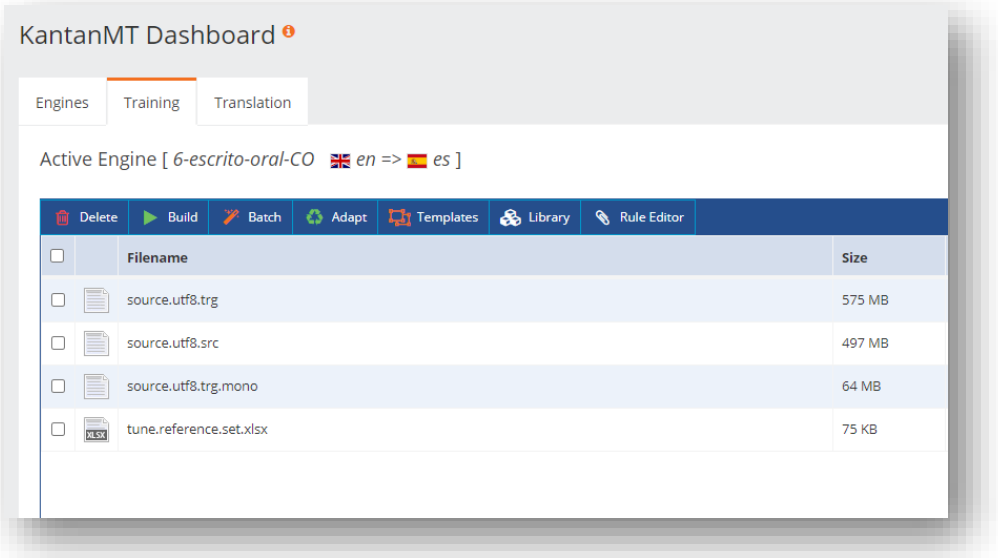


Figura 19. Pestaña `Training` del panel de inicio de KantanMT con el motor n.º 6 activo

Al hacer clic en `Create`, se crea el motor con las características seleccionadas y se puede comenzar a cargar los archivos con los que se quiere entrenar dicho motor. Para cargar los archivos, se abre la pestaña `Training` y se suben los archivos preparados en el paso anterior (véase la figura 19). El límite de tamaño de los archivos es de 1 GB, por lo que no se tiene ningún problema de este calibre al subirlos.

Una vez se han cargado los archivos en la plataforma, se puede comenzar a entrenar los motores haciendo clic en el botón `Build` (véase la figura 19). Los sistemas de TA se deben entrenar uno a uno con KantanMT, puesto que no se puede poner en marcha más de una tarea a la vez. El entrenamiento de los motores dura unas ocho horas cada uno.

4.6.3 Optimización de los sistemas de TA

Después de entrenar los motores que no se optimizan (1, 3 y 5) y de obtener las traducciones de dichos motores (véase el apartado 4.6.4), se procede a optimizarlos para crear los motores 2, 4 y 6. KantanMT solo permite crear tres motores de traducción y la base de los motores es la misma, por lo que se decide optimizar los primeros motores tras descargar la traducción y guardarla para la evaluación.

Para optimizar los sistemas de TA, se carga el archivo Excel de optimización en la pestaña `Training` y se abren las estadísticas del motor a optimizar. En la ventana de las estadísticas del motor, además de consultar los análisis BLEU, TER o F-Measure creados automáticamente por KantanMT al entrenar el motor, se puede optimizar el motor haciendo clic en `Tune` (véase la figura 20). Si ya se ha subido el archivo para la optimización, la tarea comienza automáticamente y, tras un par de horas, se completa el proceso de optimización del motor.

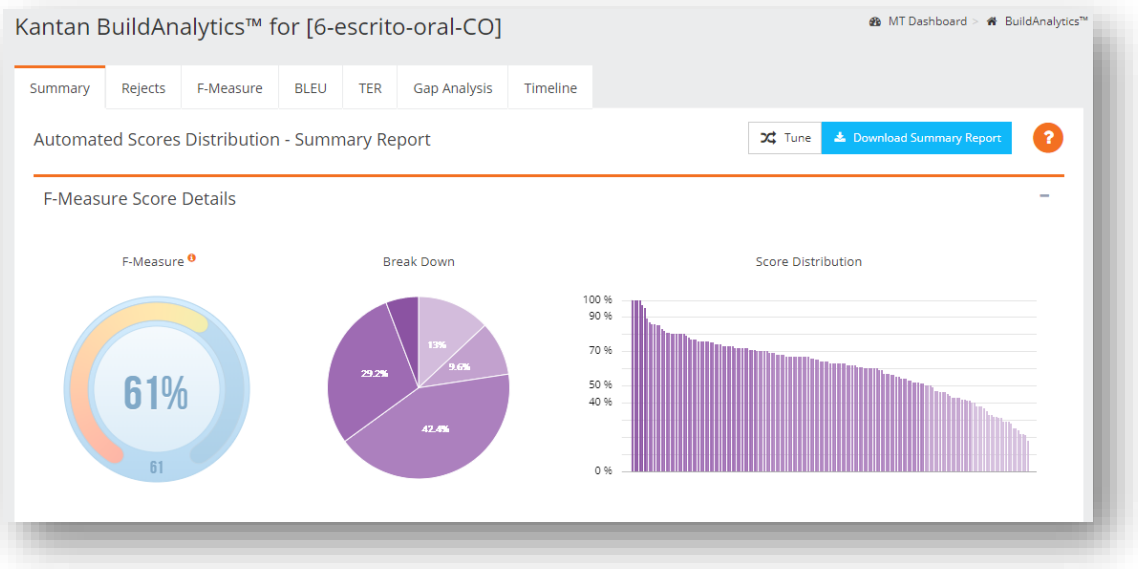


Figura 20. Vista de las estadísticas del motor n.º 6 y del botón de optimización (`Tune`, parte superior derecha)

4.6.4 Traducción del texto con los sistemas de TA

Con los motores ya entrenados, el último paso consiste en obtener las seis traducciones del texto seleccionado para probarlos. En la pestaña `Translation` del panel de inicio, se sube el archivo (o archivos) que se quiere traducir —KantanMT soporta diversos formatos que se pueden consultar en su página web de ayuda[23]— y se hace clic en el botón `Translate`. Entonces comienza una nueva tarea que dura cerca de 15 minutos, a diferencia de otras herramientas, como MTradumàtica, que muestran la traducción en unos segundos.

Una vez se ha completado la tarea de traducción, se puede descargar la traducción en el mismo formato de origen y, además, un conjunto de archivos con las palabras desconocidas que ha encontrado el sistema de TA al realizar la traducción.

4.7 Evaluación de las traducciones

La evaluación de la calidad de las traducciones producidas por los seis motores de TA creados en este trabajo se lleva a cabo teniendo en cuenta tres parámetros: la precisión, la fluidez y el estilo. Por último, se elabora una clasificación general mediante la comparación entre todos los motores. En las cuatro evaluaciones se le concede una puntuación a cada segmento y se hace una media de todos los segmentos para obtener la puntuación total de la traducción.

- **Precisión (véase el anexo 8.1)**. En esta primera evaluación, se compara cada traducción con el texto original y se tiene en cuenta solo si la traducción refleja el contenido del diálogo original. Se decide puntuar del 1 al 3, siendo 1 un resultado bueno, 2 un resultado con errores de poca importancia y 3 un resultado con muchos errores o errores graves. Los errores que se tienen en cuenta son la adición y la omisión de texto, las traducciones erróneas y las palabras o segmentos sin traducir.
- **Fluidez (véase el anexo 8.2)**. Para evaluar la fluidez del texto, se elimina el texto de partida para evitar calificar la traducción y observar solo la

[23] https://kantanmt.zendesk.com/hc/en-us/articles/200914076-What-file-formats-does-KantanMT-support- [última consulta: junio de 2020].

calidad del texto de llegada, independientemente de si traslada o no el significado del original. Al igual que con la evaluación de la precisión, se puntúan los segmentos del 1 al 3. Se observan, sobre todo, la gramática, las inconsistencias, la ortografía, la ortotipografía y la inteligibilidad de la traducción.

- **Estilo (véase el anexo 8.3)**. En cuanto al estilo, se decide evaluar solo la presencia de marcas de oralidad. Esta evaluación se realiza comparando las traducciones con el doblaje en español del diálogo y dando una puntuación de 1 si el segmento contiene alguna marca de oralidad y de 2 si no hay ninguna marca o característica propia de los textos orales. Puesto que no todos los segmentos del doblaje contienen marcas de oralidad, se puntúa también el doblaje para poder realizar una comparación más justa.
- **Clasificación general (véase el anexo 8.4)**. La última evaluación consiste en comparar los segmentos entre sí y clasificarlos de mejor a peor. La mejor traducción de cada segmento se puntúa con un 1, la segunda mejor con un 2 y así sucesivamente hasta el peor segmento, que se lleva un 6. En el caso de que haya dos o más segmentos iguales, se les concede la misma puntuación.

Las evaluaciones se realizan en un libro de Excel (véase el archivo adjunto al PDF «TAH-Creación y evaluación de 6 motores de TAE-evaluación-resultados.xlsx») con una hoja por cada tipo de evaluación. De esta manera, es más sencillo realizar los cálculos de las puntuaciones y elaborar gráficos que muestren los resultados.

Como se menciona anteriormente, KantanMT proporciona tres evaluaciones automáticas al crear cada sistema de TAE según las métricas BLEU, TER y F-Measure. Estas evaluaciones se realizan con los textos de prueba predeterminados de KantanMT, por lo que resultan útiles para compararlas con los resultados de la evaluación manual de la autora y comprobar si la oralidad del texto de origen influye en el rendimiento de los motores.

5 Resultados del estudio

Tras completar las cuatro evaluaciones mencionadas en el apartado anterior, se elaboran unos gráficos que sirvan para comparar la calidad de los resultados. Antes de presentar los datos de los resultados, cabe destacar que la calidad de las traducciones es bastante similar, puesto que las puntuaciones no varían demasiado. Las puntuaciones medias en precisión y fluidez no bajan de los dos puntos, por lo que se deduce que hay bastantes errores y pocos segmentos que tengan una traducción correcta.

En cuanto a la precisión, a la fluidez y a la clasificación general, la traducción con la mejor calidad, es decir, la puntuación más cercana al 1, es la del motor n.º 1 (2,38, 2,15 y 2,35 respectivamente), el motor entrenado con el modelo de lengua escrito y sin optimización (véanse la figura 21 y la figura 22). El motor n.º 2, con el ML escrito y con optimización, obtiene la peor puntuación de precisión (2,69), junto con el motor n.º 3 (ML mixto, sin optimización), y de fluidez (2,54), junto con el motor n.º 6 (ML oral, con optimización). Los motores n.º 4, con el ML mixto y optimización, y n.º 6, con el ML oral y optimización, quedan los últimos en la clasificación general con la puntuación más cercana al 4 (3,31).

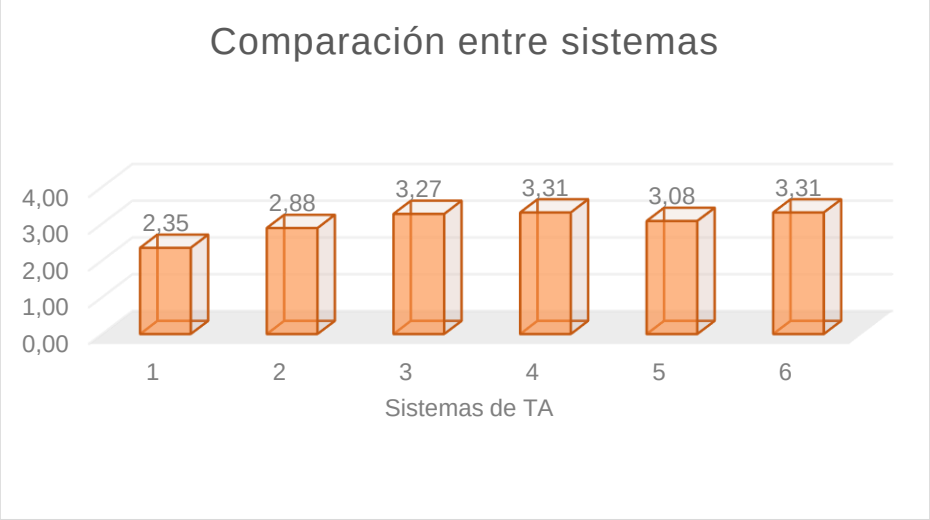
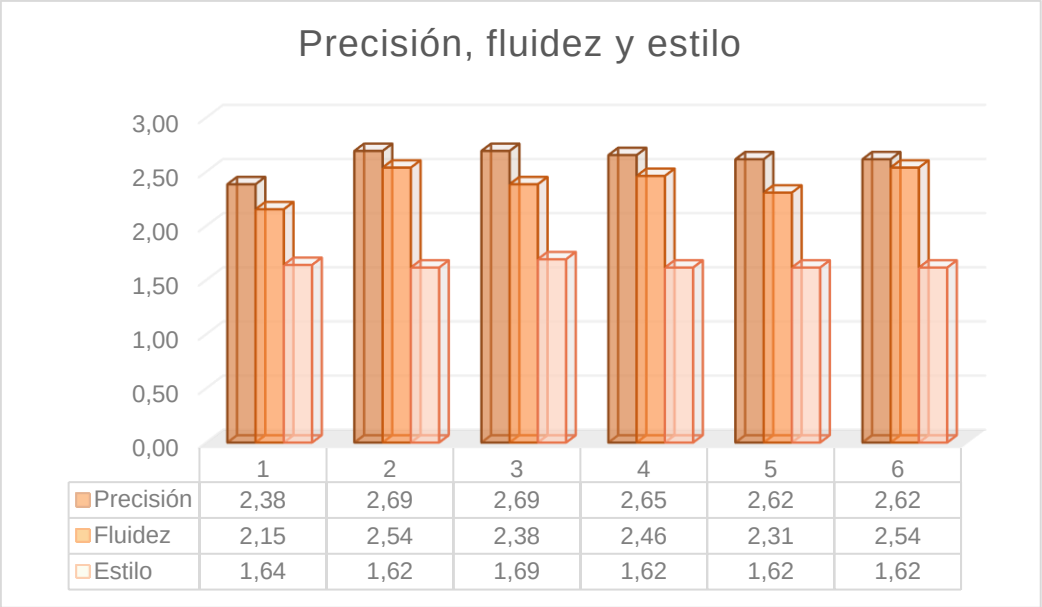


Figura 21. Gráfico con la media de las puntuaciones de la comparación entre los seis motores de TA



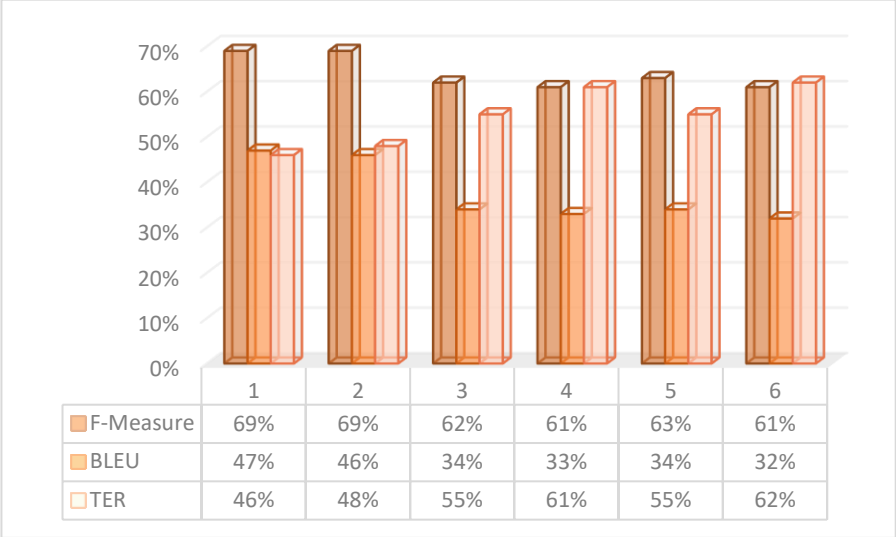
	1	2	3	4	5	6
Precisión	2,38	2,69	2,69	2,65	2,62	2,62
Fluidez	2,15	2,54	2,38	2,46	2,31	2,54
Estilo	1,64	1,62	1,69	1,62	1,62	1,62

Figura 22. Gráfico con la media de las puntuaciones de precisión, fluidez y estilo de los sistemas de TA

Los resultados de la evaluación del estilo —es decir, de la oralidad— de la traducción son bastante similares; la mejor puntuación se diferencia de la peor en tan solo siete centésimas. Como se indica en el apartado 4.7, el doblaje en español se puntúa de la misma forma que las traducciones de los sistemas de TA, con un 1 si hay marcas de oralidad y con un 2 si no las hay, puesto que no todos los segmentos contienen marcas de oralidad. La puntuación del doblaje es 1,35, mientras que la mejor puntuación de los sistemas creados es de 1,62, que corresponde a los motores 2 (ML escrito, con optimización), 4 (ML mixto, con optimización), 5 (ML oral, sin optimización) y 6 (ML oral, con optimización). En esta evaluación, el motor n.º 3 (ML mixto, sin optimización) obtiene la peor puntuación con 1,69.

Los resultados de la evaluación automática de KantanMT no se alejan mucho de los resultados mencionados en los párrafos anteriores. KantanMT evalúa el rendimiento de los sistemas de TA teniendo en cuenta tres métricas: F-Measure, que mide la precisión de las traducciones, BLEU, para medir la fluidez del texto meta, y TER, que indica el esfuerzo de posedición. Según las indicaciones de la propia plataforma, se considera que un motor de TA tiene un buen rendimiento si obtiene una puntuación mayor de 70 % en F-Measure, mayor de 50 % en BLEU y menor de 40 % en TER. Teniendo en cuenta estas indicaciones, el motor n.º 1, con el ML escrito y sin optimización, vuelve a situarse en la

primera posición y los motores 4 (ML mixto, con optimización) y 6 (ML oral, con optimización) reciben el peor puesto.



	1	2	3	4	5	6
F-Measure	69%	69%	62%	61%	63%	61%
BLEU	47%	46%	34%	33%	34%	32%
TER	46%	48%	55%	61%	55%	62%

Figura 23. Evaluación automática de los motores realizada por KantanMT

En general, se observa que la presencia de marcas orales aumenta en los motores entrenados con corpus orales transcritos, tanto en el ML como en la optimización. No obstante, las marcas de oralidad de las traducciones aparecen en los segmentos que originalmente ya tenían estas marcas. Esto ocurre porque las traducciones de los sistemas de TA creados son bastante literales y se ciñen mucho al texto original, sobre todo en cuanto a la puntuación. En la tabla 9, se incluyen algunos ejemplos de buenas prácticas de los motores entrenados con texto oral —el primer ejemplo se incluye porque se reconocen los puntos suspensivos, aunque la precisión y la fluidez no sean muy buenas, y el segundo muestra una mejor precisión de los textos orales— y en la tabla 10 se añade un ejemplo de buenas prácticas de los motores entrenados con texto escrito (para más ejemplos, véase el apartado 8).

Original EN	Traducción 1	Traducción 2	Traducción 3	Traducción 4	Traducción 5	Traducción 6
OK. Let's... let's not... M- my work is here now, my family's here...	OK. Vayamos. No debemos. M- mi trabajo está aquí ahora, mi familia de aquí.	RESPUESTA OK. Deje's... ¡Trabajemos not... M- mi trabajo está aquí ahora, mi familia here... de	OK. Hablemos de. No hay que. M- mi trabajo es aquí ahora, mi familia de aquí.	OK. Llamaremos's... Vayamos not... M- mi trabajo aquí ahora, mi familia here...	OK. Hablemos de. No hay que. M- mi trabajo es aquí ahora, mi familia de aquí.	OK. Llamaremos's... Vayamos not... M- mi trabajo es aquí ahora, mi familia here...

Do you understand why I want to stay in L.A.?	No entiende por qué quiero estancia en L.A.?	¿Le entender por qué quiero estancia en L.A.?	¿Sabe por qué me quieren permanecer en L.A.?	¿Sabe por qué quiero permanecer en L.A.?	Usted comprende por qué me quieren permanecer en L.A.?	¿Sabe por qué quiero permanecer en L.A.?

Tabla 9. Ejemplos de buenas traducciones de los motores entrenados con textos orales

Original EN	Traducción 1	Traducción 2	Traducción 3	Traducción 4	Traducción 5	Traducción 6
I was not privy to that thought process.	Yo no estaba enterada de que el proceso de reflexión.	No estaba enterada de ese pensamiento proceso.	Yo no es que creía conocer.	Yo no conocer que pensaba.	Yo no era conocer que pensaba.	Yo no era conocer que pensaba.

Tabla 10. Ejemplos de buenas traducciones de los motores entrenados con corpus escritos

Las principales dificultades para los sistemas de TA que han desembocado en problemas en las traducciones se pueden clasificar en tres tipos: (i) la oralidad del texto original, puesto que no se reconocen algunas expresiones y contracciones del inglés como *that's*, *you've* o *you'd*; (ii) la traducción del sentido, que se pierde en algunas ocasiones al verse entorpecida por la oralidad del texto original, las palabras desconocidas por el motor y los signos de puntuación —por ejemplo, los motores sin optimización no reconocen los puntos suspensivos—, y, por último, (iii) la oralidad y la fluidez del texto final, que no se consigue transmitir a la traducción si no aparece una marca clara y equivalente en el original porque no se reconocen las marcas orales típicas del idioma de destino.

6 Conclusiones

Después de evaluar las traducciones de los seis motores de TA, se pueden sacar varias conclusiones. En primer lugar, se descubre una falta de recursos orales transcritos aptos para el entrenamiento de sistemas de TA que se adecúen al tema del trabajo y que sean de libre acceso. Además, los pocos recursos que se encuentran con las características perfectas para el trabajo —como Activ-ES, que contiene diálogos de películas en español— no tienen una calidad óptima y, a pesar de que se realiza un control de calidad, se mantienen algunos errores que se trasladan a las traducciones. La mayoría de los recursos orales a los que se ha podido acceder están pensados para mejorar el

reconocimiento de voz y no para el entrenamiento de sistemas de TA, por lo que contienen anotaciones e información adicional al texto en sí. Esto condiciona bastante la elaboración del trabajo, ya que se debe invertir más tiempo en preparar los archivos y realizar un control de calidad antes de entrenar los motores.

Con respecto a la primera hipótesis planteada al inicio del trabajo, se comprueba que, al introducir corpus orales transcritos en el entrenamiento, las traducciones contienen más marcas de oralidad, pero la calidad de la precisión y de la fluidez empeora notablemente. No obstante, los resultados no se consideran totalmente concluyentes, puesto que la calidad general de las traducciones es mala y las puntuaciones son bastante parecidas. Por otro lado, las traducciones se ciñen mucho al texto original, sobre todo con la puntuación ortográfica, lo que crea textos demasiado literales y poco naturales para ser recitados.

En cuanto a la segunda hipótesis, se consigue crear varios motores de TAE con corpus de textos escritos y de textos orales transcritos, pero se descarta el uso de MTradumàtica, la primera opción. Si bien MTradumàtica puede ser una herramienta muy útil para distintos contextos, en el presente trabajo, KantanMT resulta ser más eficiente. A pesar de que comparar las herramientas no era uno de los objetivos iniciales del trabajo, parece interesante aprovechar las circunstancias y añadir algunos apuntes sobre las ventajas y desventajas de cada una de ellas:

1. Con MTradumàtica no se puede hacer un seguimiento del proceso de entrenamiento, por lo que, si hay algún error, como ha sido el caso, no se puede saber el origen de este sin ayuda del soporte técnico.

2. KantanMT, por el contrario, muestra el paso a paso de cada tarea —el entrenamiento de los motores, la traducción de textos, la optimización, etc.— además del porcentaje de entrenamiento, haciendo mucho más sencilla la planificación del proyecto.

3. MTradumàtica permite identificar los archivos, los monotextos y los bitextos con mayor facilidad, puesto que no tienen todos los mismos nombres como en KantanMT, y, además, se detalla el número de líneas, palabras, palabras únicas y caracteres de cada archivo.

4. KantanMT ofrece unas estadísticas muy detalladas sobre el rendimiento del sistema según las métricas BLEU, TER y F-Measure con tan solo entrenar el motor de TA, incluyendo ejemplos de los segmentos que mejor y peor ha traducido del texto de prueba. También muestra los segmentos del corpus que se excluyen del entrenamiento por cuestiones de calidad o longitud y los segmentos del corpus que podrían no estar traducidos. De esta manera, se pueden editar esos segmentos y adaptar el motor para que tenga una mejor calidad.

5. MTradumàtica es una herramienta de código abierto y libre, mientras que KantanMT es una plataforma privativa a la que no todo el mundo puede tener acceso.

En general, se puede concluir que, aunque se cumplen todos los objetivos planteados al principio del trabajo, los resultados no son los esperados. Si bien es cierto que la introducción de textos orales en el entrenamiento de los sistemas de TA ayuda a mejorar la presencia de marcas de oralidad y de signos de puntuación en las traducciones, la precisión y la fluidez de los textos empeora con respecto a los motores entrenados con textos escritos. La causa de esta pérdida de calidad se puede ver influida por el carácter oral de los textos de entrenamiento, sobre todo por las interrupciones y los cambios de tema repentinos, y por la literalidad de las traducciones, puesto que las marcas de oralidad trasladadas a la traducción son las típicas del idioma de origen y no del idioma de destino.

Se estima que en una escala mayor con más recursos específicos y con las características del texto de evaluación, se puede mejorar el rendimiento de los sistemas de TA, puesto que la mejor y la peor puntuación de estos solo varía en un 10 %. No obstante, según los resultados de este trabajo, conseguir una traducción automática del habla que suene natural y contenga marcas de oralidad típicas del idioma de llegada sigue siendo un reto para los sistemas de TAE.

7 Bibliografía

Balakrishnan, Anusha, et al. «Constrained decoding for neural NLG from compositional representations in task-oriented dialogue». *arXiv preprint arXiv:1906.07220*, 2019, https://arxiv.org/abs/1906.07220 [última consulta: abril de 2020].

Barrio Arconada, María Luisa. «El español coloquial y las marcas de oralidad en los textos y en la clase de E/LE». *Actas del III Simposio Internacional de la Lengua Española del Instituto Cervantes de São Paulo (2010)*, Instituto Cervantes de São Paulo, 2010, pp. 439-445, https://cvc.cervantes.es/ensenanza/biblioteca_ele/publicaciones_centros/PDF/sao paulo_2010/38_barrio.pdf [última consulta: abril de 2020].

Cettolo, Mauro, Christian Girardi y Marcello Federico. «Wit3: Web inventory of transcribed and translated talks». *Conference of european association for machine translation,* 2012, pp. 261-268, http://hdl.handle.net/11582/104409 [última consulta: junio de 2020].

Di Gangi, Mattia A., et al. "MuST-C: a multilingual speech translation corpus." *2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics, 2019, http://hdl.handle.net/11582/319646 [última consulta: junio de 2020].

Esplà-Gomis, Miquel, et al. "ParaCrawl: Web-scale parallel corpora for the languages of the EU." *Proceedings of Machine Translation Summit XVII Volume 2: Translator, Project and User Tracks*, 2019, pp. 118-119, https://www.aclweb.org/anthology/W19-6721.pdf [última consulta: mayo de 2020].

Francom, Jerid, Mans Hulden, and Adam Ussishkin. "ACTIV-ES: a comparable, cross-dialect corpus of'everyday'Spanish from Argentina, Mexico, and Spain." *LREC*. 2014, https://www.researchgate.net/profile/Jerid_Francom/publication/263028684_ACTIV-ES_a_comparable_cross-dialect_corpus_of_'everyday_'_Spanish_from/links/00463539998029c34f00000

0/ACTIV-ES-a-comparable-cross-dialect-corpus-of-everyday-Spanish-from.pdf [última consulta: mayo de 2020].

Freitag, Markus, Isaac Caswell y Scott Roy. «APE at Scale and its Implications on MT Evaluation Biases». *Proceedings of the Fourth Conference on Machine Translation (Volume 1: Research Papers)*, 2019, https://www.aclweb.org/anthology/W19-5204.pdf [última consulta: mayo de 2020].

Friends With Benefits. Dirigida por Will Gluck. Reparto: Justin Timberlake, Mila Kunis. 2011. Screen Gems y Castle Rock Entertainment. *Netflix*, https://www.netflix.com/watch/70167075 [última consulta: junio de 2020].

Garrido, Juan María, et al. "Glissando: a corpus for multidisciplinary prosodic studies in Spanish and Catalan." *Language resources and evaluation* 47.4, 2013: 945-971, https://link.springer.com/article/10.1007/s10579-012-9213-0 [última consulta: mayo de 2020].

Graham, Yvette, Barry Haddow y Philipp Koehn. «Translationese in machine translation evaluation». *arXiv preprint arXiv:1906.09833*, 2019, https://arxiv.org/abs/1906.09833 [última consulta: mayo de 2020].

Hearne, Mary y Andy Way. «Statistical machine translation: a guide for linguists and translators». *Language and Linguistics Compass* 5.5, 2011, pp. 205-226, https://doi.org/10.1111/j.1749-818X.2011.00274.x [última consulta: abril de 2020].

Hernández-Menz, Carlos D. *TEDx Spanish Corpus. Audio and transcripts in Spanish taken from the TEDx Talks; shared under the CC BY-NC-ND 4.0 license*. Universidad Nacional Autonoma de Mexico, 2019, https://www.openslr.org/67/ [última consulta: junio de 2020].

Hutchins, William John y Harold L. Somers. *An introduction to machine translation*. Vol. 362. London: Academic Press, 1992.

Koehn, Philipp. *Statistical machine translation*. Cambridge University Press, 2009.

Linares, Jesús Angel Giménez. *Empirical Machine Translation and its Evaluation*. Sociedad Española para el Procesamiento del Lenguaje Natural, 2009.

Lleida, Eduardo y Alfonso Ortega. «Reconocimiento del lenguaje hablado». Editado por Ángel Luis Gonzalo, *Tecnologías del lenguaje en España. Comunicación inteligente entre personas y máquinas*. Madrid - Barcelona: Fundación Telefónica – Editorial Ariel., 2016, pp. 1-18, https://www.fundaciontelefonica.com/arte_cultura/publicaciones-listado/pagina-item-publicaciones/itempubli/565/ [última consulta: junio de 2020].

Llisterri, Joaquim. «La lingüística de corpus». *Grup de Fonètica*. Departament de Filologia Espanyola, Universitat Autònoma de Barcelona, 11 de marzo de 2020, http://liceu.uab.es/~joaquim/language_resources/lang_res/linguistica_corpus.html [última consulta: junio de 2020].

---. «Los corpus de lengua oral». *Grup de Fonètica*. Departament de Filologia Espanyola, Universitat Autònoma de Barcelona, 11 de marzo de 2020, http://liceu.uab.es/~joaquim/language_resources/spoken_res/Corpus_lengua_oral.html [última consulta: junio de 2020].

Marriage Story. Dirigida por Noah Baumbach. Reparto: Scarlett Johansson, Adam Driver. 2019. Heyday Films, Netflix. *Netflix*, https://www.netflix.com/watch/80223779 [última consulta: junio de 2020].

Martín-Mor, Adrià, Pilar Sánchez-Gijón y Ramon Piqué i Huerta. *Tradumàtica: Tecnologies de la traducció*. Eumo Editioral, 2016.

Martín-Mor, Adrià. «MTradumàtica: Statistical machine translation customisation for translators». *Skase journal of translation and interpretation* 11.1, 2017, pp. 25-40, http://www.skase.sk/Volumes/JTI12/pdf_doc/02.pdf [última consulta: mayo de 2020].

Morgan, John. *West point heroico Spanish speech*. Tech. Rep., LDC, Philadelphia, Pennsylvania, 2006, https://catalog.ldc.upenn.edu/LDC2006S37 [última consulta: junio de 2020].

Níkleva, Dimitrinka Georgíeva. «La oposición oral/escrito: consideraciones terminológicas, históricas y pedagógicas». *Didáctica. Lengua y Literatura*, vol.

20, Universidad Complutense de Madrid, 2008, pp. 211-227, https://digibug.ugr.es/handle/10481/17407 [última consulta: abril de 2020].

O'Dowd, Tony. "Cloud-Based Machine Translation Platform." *KantanMT*, adquirido por Keywords Studio, 2013-2020, kantanmt.com/ [última consulta: junio de 2020].

Payrató, Lluís. «Transcripción del discurso coloquial». *El español coloquial: actas del I Simposio sobre análisis del discurso oral: Almería*, 1995, pp. 43-70.

Ping, Ke. «Machine translation». *Routledge Encyclopedia of Translation Studies. 2nd edition. New York: Routledge*, 2009, pp. 162-169.

Rozis, Roberts y Raivis Skadiņš. «Tilde MODEL-multilingual open data for EU languages». *Proceedings of the 21st Nordic Conference on Computational Linguistics*, 2017, https://www.aclweb.org/anthology/W17-0235.pdf [última consulta: junio de 2020].

Salesky, Elizabeth, Matthias Sperber y Alex Waibel. «Fluent translations from disfluent speech in end-to-end speech translation». *arXiv preprint arXiv:1906.00556*, 2019, https://arxiv.org/abs/1906.00556 [última consulta: abril de 2020].

Sánchez-Cartagena, Víctor M., Felipe Sánchez-Martínez y Juan Antonio Pérez-Ortiz. «Enriching a statistical machine translation system trained on small parallel corpora with rule-based bilingual phrases». *Proceedings of the International Conference Recent Advances in Natural Language Processing 2011*, 2011, pp. 90-96, https://www.aclweb.org/anthology/R11-1013.pdf [última consulta: abril de 2020].

Shterionov, Dimitar, et al. «Improving KantanMT training efficiency with fast align». *AMTA*, 2016, http://doras.dcu.ie/23348/ [última consulta: junio de 2020].

Sinclair, John. «Preliminary recommendations on corpus typology». *EAGLES Document TCWG-CTYP/P*, mayo de 1996, http://www.ilc.cnr.it/EAGLES96/corpustyp/corpustyp.html [última consulta: junio de 2020].

Somers, Harold, ed. *Computers and translation: a translator's guide*. Vol. 35. John Benjamins Publishing, 2003.

Tiedemann, Jörg. «Parallel Data, Tools and Interfaces in OPUS». *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC'2012)*. European Language Resources Association (ELRA), 2012, pp. 2214-2218 http://www.lrec-conf.org/proceedings/lrec2012/pdf/463_Paper.pdf [última consulta: mayo de 2020].

The Break-Up. Dirigida por Peyton Reed. Reparto: Vince Vaugh, Jennifer Aniston. 2006. Universal Pictures. *Netflix*, https://www.netflix.com/watch/70042688 [última consulta: junio de 2020].

The Notebook. Dirigida por Nick Cassavetes. Reparto: Ryan Gosling, Rachel McAdams. 2004. New Line Cinema. *Netflix*, https://www.netflix.com/watch/60036227 [última consulta: junio de 2020].

«The One on the Last Night». *Friends*, creada por Marta Kauffman y David Crane, temporada 6, episodio 6, Warner Brothers, 2004. *Netflix*, https://www.netflix.com/watch/70274123 [última consulta: junio de 2020].

«The One Where Monica Gets a Roommate (Pilot)». *Friends*, creada por Marta Kauffman y David Crane, temporada 1, episodio 1, Warner Brothers, 2004. *Netflix*, https://www.netflix.com/watch/70273997 [última consulta: junio de 2020].

Waibel, Alex y Christian Fugen. «Spoken language translation». *IEEE Signal Processing Magazine* 25.3, 2008, pp. 70-79, https://doi.org/10.1109/MSP.2008.918415 [última consulta: abril de 2020].

Wołk, Krzysztof y Krzysztof Marasek. «Building subject-aligned comparable corpora and mining it for truly parallel sentence pairs». *Procedia Technology*, 18, 2014, pp. 126-132, https://doi.org/10.1016/j.protcy.2014.11.024 [última consulta: mayo de 2020].

Wu, Yonghui, et al. «Google's neural machine translation system: Bridging the gap between human and machine translation». *arXiv preprint arXiv:1609.08144*, 2016, https://arxiv.org/abs/1609.08144 [última consulta: abril de 2020].

8 Anexos: evaluación de los resultados

A continuación, se incluyen las traducciones de cada motor con sus evaluaciones en cuanto a precisión (anexo 8.1), fluidez (anexo 8.2) y estilo u oralidad (anexo 8.3). También se añaden la clasificación general de los sistemas de TA (anexo 8.4) y la evaluación automática de KantanMT (anexo 8.5). Por último, en el anexo 8.6, se incluyen los gráficos creados a partir de dichas evaluaciones. Además, se puede consultar el libro de Excel creado con todas estas evaluaciones y gráficos (véase el archivo adjunto al PDF «TAH-Creación y evaluación de 6 motores de TAE-evaluación-resultados.xlsx»).

8.1 Evaluación de la precisión

Original EN	Trad. 1	Punt. 1	Trad. 2	Punt. 2	Trad. 3	Punt. 3	Trad. 4	Punt. 4	Trad. 5	Punt. 5	Trad. 6	Punt. 6
I don't know how to start.	I no sabe cómo comenzar.	3	I no sabemos cómo comenzar.	3	No sé si la manera de empezar.	2	No sé si cómo comenzar.	2	No sé si la manera de empezar.	2	No sé si cómo empezar.	2
Do you understand why I want to stay in L.A.?	No entiende por qué quiero estancia en L.A.?	2	¿Le entender por qué quiero estancia en L.A.?	2	¿Sabe por qué me quieren permanecer en L.A.?	2	¿Sabe por qué quiero permanecer en L.A.?	1	Usted comprende por qué me quieren permanecer en L.A.?	2	¿Sabe por qué quiero permanecer en L.A.?	1
No.	Nº.	3	Nº.	3	Nº.	3	Nº.	3	Nº.	3	Nº.	3
Well, that's not… Charlie, that's not a useful way for us to start.	Bien, eso no. Charlie, que no de una manera útil para nosotros, para empezar.	2	Bien, eso not… Charlie, eso no una manera útil para nosotros a comenzar.	2	Así, que de no. Charlie, que no de una manera útil para nosotros de empezar a.	2	Bien, ¡es not… Charlie, que no hay una manera útil a nosotros empezar.	3	Así, que de no. Charlie, que no de una manera útil para nosotros, para empezar.	3	Bien, ¡es not… Charlie, que no hay una manera útil para nosotros empezar.	3
I don't understand it.	I no entenderlo.	3	I no entenderlo.	3	No comunicaré comprender.	3	No comunicaré comprender.	3	No comunicaré comprender.	3	No comunicaré comprender.	3

You don't remember promising that we could do time here?	No te recuerde prometedoras que podríamos hacer tiempo aquí?	2	No te recuerde prometedoras que podríamos hacer tiempo aquí?	2	¿No vas a recordar prometedoras que podemos hacer aquí?	3	Usted No comunicaré recordar prometedoras que podríamos hacer aquí?	3	¿No vas a recordar prometedoras que podemos hacer aquí?	2	Usted No comunicaré recordar prometedoras que podríamos hacer aquí?	3
We discussed things; we were married. We said things. We talked about moving to Europe, about getting a… sideboard or what do you call it? A credenza, to fill the empty space behind the couch. We never did any of it.	Hemos debatido comportamientos. Estuvimos casados. Dijo las cosas. Hemos hablado de que se desplazan a Europa, sobre la posibilidad de padecer un. sideboard o qué ustedes? Un credenza, para llenar el espacio vacío detrás del sofá. Que nunca tuvo ninguna de ellas.	3	Debatimos comportamientos. Estuvimos casados. Dijo cosas. Hablamos acerca hacia Europa, parecería salir a… sideboard o qué ustedes? Un credenza, para llenar el espacio vacío detrás del sofá. Nunca nos hicimos todo de ella.	3	Hemos debatido comportamientos. estábamos casados. Decíamos cosas. Hemos hablado acerca de trasladarse a Europa, alrededor de conseguir un. sideboard o qué hacen ustedes? Un credenza, para llenar el espacio vacío detrás del couch. Nosotros nunca hizo ningún de ella.	2	Hemos debatido comportamientos . estábamos casados. Decíamos cosas. Hemos hablado acerca hacia Europa, parecería salir a… sideboard o qué hacen ustedes? Un credenza, para llenar el espacio vacío detrás couch. Nosotros nunca hizo ningún de ella.	3	Hemos debatido comportamientos . estábamos casados. Decíamos cosas. Hemos hablado acerca de la transición hacia Europa, alrededor de conseguir un. sideboard o qué hacen ustedes? Un credenza, colmar el vacío espacio detrás del couch. Nunca nos hizo ningún de ella.	2	Hemos debatido comportamientos. estábamos casados. Decíamos cosas. Hemos hablado sobre hacia Europa, parecería salir a… sideboard o qué hacen ustedes? Un credenza, colmar el espacio vacío detrás couch. Nosotros nunca hizo ningún de ella.	2
You turn down the residency at the Geffen that would have brought us here for a… a year.	A su vez, por la residencia a Geffen que nos han traído aquí para un año.	3	Le rechazan La residency a Geffen que habría permitido aquí para a… un año.	3	Usted, a su vez, en la residencia a Geffen que nos han traído aquí para una. Un año.	3	Usted rechazar la residencia a Geffen que nos han traído aquí para a… un año.	2	Usted, a su vez, en la residencia a Geffen que nos han traído aquí para una. Un año.	3	Usted rechazar la residencia al Geffen que hubiera traído nosotros aquí para a… un año.	2

It wasn't something I wanted. We had a great theater company and a great life where we were.	No es algo que Yo quería. Hemos tenido un gran teatro sociedad y una gran vida de dónde venimos.	2	No era algo que Yo quería. Tuvimos un gran teatro empresa y una gran vida cuando éramos.	3	No es algo que yo quería. Tuvimos un gran teatro de la sociedad y una gran vida donde estábamos.	2	No era algo me quería. Tuvimos un gran teatro sociedad y una gran vida donde estábamos.	2	No es algo que yo quería. Hemos tenido una gran teatro sociedad y una gran vida donde estábamos.	1	Era algo no me querían. Tuvimos un gran teatro sociedad y una gran vida donde estábamos.	2
You call that a great life?	Llame a que una gran vida?	3	Ustedes que una gran vida?	3	Ustedes que una gran vida?	3	Ustedes que una gran vida?	3	Ustedes que una gran vida?	3	Ustedes que una gran vida?	3
You know what I mean. I don't mean we had a great marriage. I mean, life in Brooklyn. Professionally. I don't know. Honestly, I never considered anything different.	Necesita saber a qué me refiero. I no significa que tuvo un gran matrimonio. Me refiero, vida en Brooklyn. Profesionalmente. Yo no lo sé. Honestamente, yo no consideran algo diferente.	2	Saben qué me refiero. I no significa tuvimos un gran matrimonio. Me refiero, vida en Brooklyn. Es profesional. Yo no contesta. Honestamente, yo nunca considerado algo diferente.	3	Saben lo que me hacen. No comunicaré significa que tuvimos un gran matrimonio. I significa, vida en Brooklyn. Profesional. No sé si. Honestamente, yo nunca considerado algo diferente.	3	Sabes qué me hacen. No comunicaré significa tuvimos un gran matrimonio. I significa, vida en Brooklyn. Es profesional. No sé si. Honestamente, yo nunca considerado algo diferente.	3	Saben lo que me hacen. No comunicaré significa que tuvimos una gran ceremonia del matrimonio. Me refiero a una vida, en Brooklyn. El punto de vista profesional. No sé si. Honestamente, yo nunca como algo diferente.	3	Saben qué me hacen. No comunicaré significa tuvimos una gran matrimonio. I significa, vida en Brooklyn. Vista profesional. No sé si. Honestamente, yo nunca consideró algo diferente.	2
Well, that's the problem, isn't it? I mean, I was your wife. You should have considered my happiness, too.	Bien, eso el problema, no es él? Me refiero, era su esposa. Debe tener en cuenta mi felicidad, demasiado.	2	Bien, eso el problema, no es él? Me refiero, yo era su esposa. Debe han considerado mi felicidad, demasiado.	3	Así que el problema, no es él? I significa, yo era su esposa. Usted tener que mi felicidad, también.	3	Así que el problema, ¿no es esto? I significa fui tu esposa. Deberías haber estudiado mi felicidad, también.	2	Así que el problema, ¿no es él? Me hacen, yo estaba tu esposa. Usted tener que mi felicidad, también.	3	Bien, ¡es el problema, ¿no es esto? Me significa fui tu esposa. Deberías haber considerado mi felicidad, también.	2

Come on. You were happy. You've just decided you weren't now.	Llegado. Le fueron felices. Solo le hemos decidido que no estaban ahora.	3	Llegado sobre. Le fueron felices. Solo le hemos decidido no le fueron ahora.	3	A. Ustedes son felices. Le've sólo se decidió usted no ahora.	3	A el. Ustedes son felices. Usted've sólo decidió usted eran desconocidas actualmente.	3	A. Ustedes estaban felices. Usted've sólo se decidió usted no ahora.	3	A el. Ustedes estaban felices. Usted've sólo decidió usted no eran ahora.	3
OK. Let's… let's not… M- my work is here now, my family's here…	OK. Vayamos. No debemos. M- mi trabajo está aquí ahora, mi familia de aquí.	2	RESPUESTA OK. Deje's… ¡Trabajemos not… M- mi trabajo está aquí ahora, mi familia here… de	3	OK. Hablemos de. No hay que. M- mi trabajo es aquí ahora, mi familia de aquí.	3	OK. Llamaremos' s… Vayamos not… M- mi trabajo aquí ahora, mi familia here…	3	OK. Hablemos de. No hay que. M- mi trabajo es aquí ahora, mi familia de aquí.	2	OK. Llamaremos' s… Vayamos not… M- mi trabajo es aquí ahora, mi familia here…	3
And I agreed to put Henry in school here because your show went to series. I did that knowing that when you were done shooting, he would come back to New York.	Y estoy de acuerdo a Henry a la escuela porque su show fue a la serie. Hice que saber que cuando se le han hecho tiro, volver a Nueva York.	2	Y yo acordaron poner Henry en la escuela aquí porque su show fue a serie. Hice que sabiendo que cuando se realizaron tiro, él volverá a Nueva York.	3	Y me acuerdo a Henry en la escuela, porque Aquí te muestran fue para la serie. Hice que saber que cuando se realizaron tiro, que quiere volver a Nueva York.	3	Y me acordaron poner Henry escolar aquí porque tu muestran fue a la serie. Hice que saber que cuando se realizaron tiro, debería volver a Nueva York.	3	Y me acordaron poner Henry en la escuela aquí porque va a dar tu serie. Hice que saber que cuando se realizaron tiro, va a volver a Nueva York.	3	Y me acordaron poner Henry escolar aquí porque tu muestran fue a serie. Hice que saber que cuando se realizaron tiro, debería volver a Nueva York.	3
Honey, we never said that. That may have been your assumption, but we never expressly said that.	La miel, dijo que nunca. Que puede haber sido su hipótesis, pero nunca explícitamente declara que.	3	Miel, dijo que nunca. Que puede haber sido su suposición, pero nunca dijo que expresamente.	2	La miel, dijo que nunca. Que puede haber sido su supuesto, pero nunca explícitamente declara que.	3	Miel, dijo que nunca. Que puede haber sido tu suposición, pero nunca explícitamente declara que.	3	Miel, nos dijo que nunca. Que puede haber sido su hipótesis, pero dijo que nunca expresamente.	3	Miel, nunca dijo que. Que puede haber sido tu supuesto, pero nunca expresamente que.	3

We did say it.	Podemos decir que no.	3	Hicimos decirlo.	3	Decir que no.	3	Hicimos decir.	3	Hicimos decir.	3	Hicimos decir.	3
When did we say it?	Cuando decimos que sí?	2	Hizo Cuando decimos que?	3	Cuando no podemos decir?	3	Cuando no podemos decir?	3	Cuando no podemos decir?	3	Cuando no podemos decir?	3
I don't know when we said it, but we said it.	No sé si cuando nos lo dijo, pero nos lo dijo.	2	No sé si cuando nos lo dijo, pero nos dijo.	2	No sé si se dice que cuando él, pero se dice que es.	3	No sé si cuando decíamos, pero dijo ella.	3	No sé si cuando decíamos, pero nos dijo.	3	No sé si cuando decíamos, pero dijo ella.	3
I thought…	Yo pensaba.	1	I thought…	3	I pensaba.	2	Thought… I	3	I pensaba.	2	Thought… I	3
We said it that time on the phone!	Ella dijo que el tiempo en el teléfono.	3	Dijo que ese tiempo en el teléfono.	3	Se dice que ese tiempo en el teléfono.	3	Decíamos que entonces el teléfono!	3	Dijo que ese momento, sobre el teléfono!	3	Decíamos que entonces el teléfono!	3
Honey! Let me finish! Sorry, I keep saying that. I thought that if Henry was happy here and my show continued, that we might do L.A. for a while.	La miel. Permítanme terminar. Siento, estoy diciendo que mantener. Pensé que si Henry fue feliz aquí y mi muestran continuación, que podría hacer L.a. Para un tiempo.	3	La miel! Permítanme terminar! Siento, estoy diciendo que mantener. Pensé que si Henry fue feliz aquí y mi muestran continuó que podríamos hacer L.a. Durante un tiempo.	2	La miel. Permítanme terminar. Siento, tengo que seguir diciendo. I pensaba que cuando Henry era feliz aquí y mi muestran continuación, que se puede hacer L.a. Para un tiempo.	3	La miel! Permítanme terminar! Siento, me seguir diciendo que. I pensaba que cuando Henry era feliz aquí y mi espectáculo continuó, que podríamos hacer L.a. Un rato.	2	Miel! Permítanme terminar. Siento, tengo que seguir diciendo. I pensaba que si Henry es feliz aquí y mi muestra de continuidad, que se puede hacer L.a. Un rato.	3	La miel! Permítanme terminar! Siento, me seguir diciendo que. I pensaba que si Henry es feliz aquí y mi muestran continuación, que podríamos hacer L.a. Un rato.	3
I was not privy to that thought process.	Yo no estaba enterada de que el proceso de reflexión.	2	No estaba enterada de ese pensamiento proceso.	2	Yo no es que creía conocer.	3	Yo no conocer que pensaba.	3	Yo no era conocer que pensaba.	3	Yo no era conocer que pensaba.	3
The only reason we didn't live here is because you	La única razón de que no la vida aquí es debido a que no se puede imaginar desea	2	El único motivo tampoco lo hemos conseguido vivir aquí es porque	2	La única razón por la que no la vida aquí es porque usted no podrán imaginarse desea	3	La única razón tuviéramos vivir aquí es porque usted no podrán imaginarse desea	3	La única razón, no estábamos viviendo aquí es porque usted no podrán	3	La única razón tuviéramos vivir aquí es porque usted no podrán imaginarse	3

can't imagine desires other than your own. Unless they're forced on you.	que su propia. Si no estamos obligados a usted.		usted no puede imaginarse desea excepto tu propio. Salvo que estamos obligados a usted.		que no sean tus propios. Si le obligó a la obra».		excepto tu propio. Salvo obra» obligado a usted.		imaginarse desea que no sean tus propios. Si le obligó a la obra».		desea excepto tu propia. Salvo obra» forzoso sobre usted.	
Okay. You wish you hadn't married me. You wish you'd had a different life. But this is what happened. So, what do we do?	Normal. Le deseo de que no había casado conmigo. Le deseo «d tenido una vida diferente. Pero esto es lo que ocurrió. Así pues, ¿qué podemos hacer?	3	Importa. Le deseo no había casado conmigo. Le deseo «d tenido una vida diferente. Pero esto es lo que ocurrió. Afirmativo, ¿qué podemos hacer?	3	Okay. Te deseo no le había casado conmigo. Desea usted «d tenido una vida diferente. Pero esto es lo que ocurrió. Así, ¿Qué podemos hacer?	2	Okay. Desea usted no había casado conmigo. Desea usted «d tenido una vida diferente. Pero es lo que ocurrió. Así, ¿Qué podemos hacer?	2	Okay. Desea usted no me había casado. Desea usted «d tenido una vida diferente. Pero esto es lo que ocurrió. Así, ¿Qué podemos hacer?	2	Okay. Desea usted no había casado conmigo. Desea usted «d tenido una vida diferente. Pero es lo que ocurrió. Así, ¿Qué podemos hacer?	2
I don't know.	Yo no lo sé.	1	Yo no contesta.	3	No sé si.	2	No sé si.	2	No sé si.	2	No sé si.	2
	Puntuación media	**2,38**	**Puntuación media**	**2,69**	**Puntuación media**	**2,69**	**Puntuación media**	**2,65**	**Puntuación media**	**2,62**	**Puntuación media**	**2,62**

8.2 Evaluación de la fluidez

Trad. 1	Punt. 1	Trad. 2	Punt. 2	Trad. 3	Punt. 3	Trad. 4	Punt. 4	Trad. 5	Punt. 5	Trad. 6	Punt. 6
I no sabe cómo comenzar.	2	I no sabemos cómo comenzar.	2	No sé si la manera de empezar.	2	No sé si cómo comenzar.	2	No sé si la manera de empezar.	2	No sé si cómo empezar.	2
No entiende por qué quiero estancia en L.A.?	2	¿Le entender por qué quiero estancia en L.A.?	3	¿Sabe por qué me quieren permanecer en L.A.?	2	¿Sabe por qué quiero permanecer en L.A.?	1	Usted comprende por qué me quieren	2	¿Sabe por qué quiero permanecer en L.A.?	1

								permanecer en L.A.?		
Nº.	2	Nº.	2	Nº.	2	Nº.	2	Nº.	2	Nº.	2
Bien, eso no. Charlie, que no de una manera útil para nosotros, para empezar.	2	Bien, eso not… Charlie, eso no una manera útil para nosotros a comenzar.	2	Así, que de no. Charlie, que no de una manera útil para nosotros de empezar a.	3	Bien, ¡es not… Charlie, que no hay una manera útil a nosotros empezar.	3	Así, que de no. Charlie, que no de una manera útil para nosotros, para empezar.	2	Bien, ¡es not… Charlie, que no hay una manera útil para nosotros empezar.	3
I no entenderlo.	2	I no entenderlo.	2	No comunicaré comprender.	2	No comunicaré comprender.	2	No comunicaré comprender.	2	No comunicaré comprender.	2
No te recuerde prometedoras que podríamos hacer tiempo aquí?	3	No te recuerde prometedoras que podríamos hacer tiempo aquí?	3	¿No vas a recordar prometedoras que podemos hacer aquí?	2	Usted No comunicaré recordar prometedoras que podríamos hacer aquí?	3	¿No vas a recordar prometedoras que podemos hacer aquí?	2	Usted No comunicaré recordar prometedoras que podríamos hacer aquí?	3
Hemos debatido comportamientos. Estuvimos casados. Dijo las cosas. Hemos hablado de que se desplazan a Europa, sobre la posibilidad de padecer un. sideboard o qué ustedes? Un credenza, para llenar el espacio vacío detrás del sofá. Que nunca tuvo ninguna de ellas.	2	Debatimos comportamientos. Estuvimos casados. Dijo cosas. Hablamos acerca hacia Europa, parecería salir a… sideboard o qué ustedes? Un credenza, para llenar el espacio vacío detrás del sofá. Nunca nos hicimos todo de ella.	3	Hemos debatido comportamientos. estábamos casados. Decíamos cosas. Hemos hablado acerca de trasladarse a Europa, alrededor de conseguir un. sideboard o qué hacen ustedes? Un credenza, para llenar el espacio vacío detrás del couch. Nosotros nunca hizo ningún de ella.	2	Hemos debatido comportamientos. estábamos casados. Decíamos cosas. Hemos hablado acerca hacia Europa, parecería salir a… sideboard o qué hacen ustedes? Un credenza, para llenar el espacio vacío detrás couch. Nosotros nunca hizo ningún de ella.	3	Hemos debatido comportamientos. estábamos casados. Decíamos cosas. Hemos hablado acerca de la transición hacia Europa, alrededor de conseguir un. sideboard o qué hacen ustedes? Un credenza, colmar el vacío espacio detrás del couch. Nunca nos hizo ningún de ella.	2	Hemos debatido comportamientos. estábamos casados. Decíamos cosas. Hemos hablado sobre hacia Europa, parecería salir a… sideboard o qué hacen ustedes? Un credenza, colmar el espacio vacío detrás couch. Nosotros nunca hizo ningún de ella.	3

A su vez, por la residencia a Geffen que nos han traído aquí para un año.	2	Le rechazan La residency a Geffen que habría permitido aquí para a… un año.	3	Usted, a su vez, en la residencia a Geffen que nos han traído aquí para una. Un año.	2	Usted rechazar la residencia a Geffen que nos han traído aquí para a… un año.	2	Usted, a su vez, en la residencia a Geffen que nos han traído aquí para una. Un año.	2	Usted rechazar la residencia al Geffen que hubiera traído nosotros aquí para a… un año.	2
No es algo que Yo quería. Hemos tenido un gran teatro sociedad y una gran vida de dónde venimos.	2	No era algo que Yo quería. Tuvimos un gran teatro empresa y una gran vida cuando éramos.	2	No es algo que yo quería. Tuvimos un gran teatro de la sociedad y una gran vida donde estábamos.	1	No era algo me quería. Tuvimos un gran teatro sociedad y una gran vida donde estábamos.	2	No es algo que yo quería. Hemos tenido una gran teatro sociedad y una gran vida donde estábamos.	2	Era algo no me querían. Tuvimos un gran teatro sociedad y una gran vida donde estábamos.	2
Llame a que una gran vida?	3	Ustedes que una gran vida?	3	Ustedes que una gran vida?	3	Ustedes que una gran vida?	3	Ustedes que una gran vida?	3	Ustedes que una gran vida?	3
Necesita saber a qué me refiero. I no significa que tuvo un gran matrimonio. Me refiero, vida en Brooklyn. Profesionalmente. Yo no lo sé. Honestamente, yo no consideran algo diferente.	2	Saben qué me refiero. I no significa tuvimos un gran matrimonio. Me refiero, vida en Brooklyn. Es profesional. Yo no contesta. Honestamente, yo nunca considerado algo diferente.	3	Saben lo que me hacen. No comunicaré significa que tuvimos un gran matrimonio. I significa, vida en Brooklyn. Profesional. No sé si. Honestamente, yo nunca considerado algo diferente.	3	Sabes qué me hacen. No comunicaré significa tuvimos un gran matrimonio. I significa, vida en Brooklyn. Es profesional. No sé si. Honestamente, yo nunca considerado algo diferente.	3	Saben lo que me hacen. No comunicaré significa que tuvimos una gran ceremonia del matrimonio. Me refiero a una vida, en Brooklyn. El punto de vista profesional. No sé si. Honestamente, yo nunca como algo diferente.	3	Saben qué me hacen. No comunicaré significa tuvimos una gran matrimonio. I significa, vida en Brooklyn. Vista profesional. No sé si. Honestamente, yo nunca consideró algo diferente.	3
Bien, eso el problema, no es él? Me refiero, era su esposa. Debe tener en cuenta mi	2	Bien, eso el problema, no es él? Me refiero, yo era su esposa. Debe han considerado mi	2	Así que el problema, no es él? I significa, yo era su esposa. Usted tener que mi felicidad, también.	3	Así que el problema, ¿no es esto? I significa fui tu esposa. Deberías haber estudiado mi	2	Así que el problema, ¿no es él? Me hacen, yo estaba tu esposa. Usted tener que mi felicidad, también.	3	Bien, ¡es el problema, ¿no es esto? Me significa fui tu esposa. Deberías haber considerado mi	2

felicidad, demasiado.		felicidad, demasiado.				felicidad, también.				felicidad, también.	
Llegado. Le fueron felices. Solo le hemos decidido que no estaban ahora.	3	Llegado sobre. Le fueron felices. Solo le hemos decidido no le fueron ahora.	3	A. Ustedes son felices. Le've sólo se decidió usted no ahora.	3	A el. Ustedes son felices. Usted've sólo decidió usted eran desconocidas actualmente.	3	A. Ustedes estaban felices. Usted've sólo se decidió usted no ahora.	3	A el. Ustedes estaban felices. Usted've sólo decidió usted no eran ahora.	3
OK. Vayamos. No debemos. M- mi trabajo está aquí ahora, mi familia de aquí.	2	RESPUESTA OK. Deje's… ¡Trabajemos not… M- mi trabajo está aquí ahora, mi familia here… de	3	OK. Hablemos de. No hay que. M- mi trabajo es aquí ahora, mi familia de aquí.	2	OK. Llamaremos' s… Vayamos not… M- mi trabajo aquí ahora, mi familia here…	3	OK. Hablemos de. No hay que. M- mi trabajo es aquí ahora, mi familia de aquí.	2	OK. Llamaremos' s… Vayamos not… M- mi trabajo es aquí ahora, mi familia here…	3
Y estoy de acuerdo a Henry a la escuela porque su show fue a la serie. Hice que saber que cuando se le han hecho tiro, volver a Nueva York.	3	Y yo acordaron poner Henry en la escuela aquí porque su show fue a serie. Hice que sabiendo que cuando se realizaron tiro, él volverá a Nueva York.	2	Y me acuerdo a Henry en la escuela, porque Aquí te muestran fue para la serie. Hice que saber que cuando se realizaron tiro, que quiere volver a Nueva York.	3	Y me acordaron poner Henry escolar aquí porque tu muestran fue a la serie. Hice que saber que cuando se realizaron tiro, debería volver a Nueva York.	3	Y me acordaron poner Henry en la escuela aquí porque va a dar tu serie. Hice que saber que cuando se realizaron tiro, va a volver a Nueva York.	3	Y me acordaron poner Henry escolar aquí porque tu muestran fue a serie. Hice que saber que cuando se realizaron tiro, debería volver a Nueva York.	3
La miel, dijo que nunca. Que puede haber sido su hipótesis, pero nunca explícitamente declara que.	3	Miel, dijo que nunca. Que puede haber sido su suposición, pero nunca dijo que expresamente.	2	La miel, dijo que nunca. Que puede haber sido su supuesto, pero nunca explícitamente declara que.	3	Miel, dijo que nunca. Que puede haber sido tu suposición, pero nunca explícitamente declara que.	2	Miel, nos dijo que nunca. Que puede haber sido su hipótesis, pero dijo que nunca expresamente.	1	Miel, nunca dijo que. Que puede haber sido tu supuesto, pero nunca expresamente que.	3
Podemos decir que no.	1	Hicimos decirlo.	2	Decir que no.	1	Hicimos decir.	3	Hicimos decir.	3	Hicimos decir.	3

Cuando decimos que sí?	2	Hizo Cuando decimos que?	3	Cuando no podemos decir?	2	Cuando no podemos decir?	2	Cuando no podemos decir?	2	Cuando no podemos decir?	2
No sé si cuando nos lo dijo, pero nos lo dijo.	2	No sé si cuando nos lo dijo, pero nos dijo.	2	No sé si se dice que cuando él, pero se dice que es.	3	No sé si cuando decíamos, pero dijo ella.	2	No sé si cuando decíamos, pero nos dijo.	2	No sé si cuando decíamos, pero dijo ella.	3
Yo pensaba.	1	I thought…	3	I pensaba.	2	Thought… I	3	I pensaba.	2	Thought… I	3
Ella dijo que el tiempo en el teléfono.	2	Dijo que ese tiempo en el teléfono.	2	Se dice que ese tiempo en el teléfono.	2	Decíamos que entonces el teléfono!	2	Dijo que ese momento, sobre el teléfono!	2	Decíamos que entonces el teléfono!	2
La miel. Permítanme terminar. Siento, estoy diciendo que mantener. Pensé que si Henry fue feliz aquí y mi muestran continuación, que podría hacer L.a. Para un tiempo.	3	La miel! Permítanme terminar! Siento, estoy diciendo que mantener. Pensé que si Henry fue feliz aquí y mi muestran continuó que podríamos hacer L.a. Durante un tiempo.	3	La miel. Permítanme terminar. Siento, tengo que seguir diciendo. I pensaba que cuando Henry era feliz aquí y mi muestran continuación, que se puede hacer L.a. Para un tiempo.	3	La miel! Permítanme terminar! Siento, me seguir diciendo que. I pensaba que cuando Henry era feliz aquí y mi espectáculo continuó, que podríamos hacer L.a. Un rato.	3	Miel! Permítanme terminar. Siento, tengo que seguir diciendo. I pensaba que si Henry es feliz aquí y mi muestra de continuidad, que se puede hacer L.a. Un rato.	3	La miel! Permítanme terminar! Siento, me seguir diciendo que. I pensaba que si Henry es feliz aquí y mi muestran continuación, que podríamos hacer L.a. Un rato.	3
Yo no estaba enterada de que el proceso de reflexión.	2	No estaba enterada de ese pensamiento proceso.	2	Yo no es que creía conocer.	3	Yo no conocer que pensaba.	3	Yo no era conocer que pensaba.	3	Yo no era conocer que pensaba.	3
La única razón de que no la vida aquí es debido a que no se puede imaginar desea que su propia. Si	3	El único motivo tampoco lo hemos conseguido vivir aquí es porque usted no puede imaginarse desea	3	La única razón por la que no la vida aquí es porque usted no podrán imaginarse desea que no sean tus	3	La única razón tuviéramos vivir aquí es porque usted no podrán imaginarse desea excepto tu propio.	3	La única razón, no estábamos viviendo aquí es porque usted no podrán imaginarse desea que no sean tus	3	La única razón tuviéramos vivir aquí es porque usted no podrán imaginarse desea excepto tu propia. Salvo obra»	3

no estamos obligados a usted.		excepto tu propio. Salvo que estamos obligados a usted.		propios. Si le obligó a la obra».		Salvo obra» obligado a usted.		propios. Si le obligó a la obra».		forzoso sobre usted.	
Normal. Le deseo de que no había casado conmigo. Le deseo «d tenido una vida diferente. Pero esto es lo que ocurrió. Así pues, ¿qué podemos hacer?	2	Importa. Le deseo no había casado conmigo. Le deseo «d tenido una vida diferente. Pero esto es lo que ocurrió. Afirmativo, ¿qué podemos hacer?	3	Okay. Te deseo no le había casado conmigo. Desea usted «d tenido una vida diferente. Pero esto es lo que ocurrió. Así, ¿Qué podemos hacer?	3	Okay. Desea usted no había casado conmigo. Desea usted «d tenido una vida diferente. Pero es lo que ocurrió. Así, ¿Qué podemos hacer?	2	Okay. Desea usted no me había casado. Desea usted «d tenido una vida diferente. Pero esto es lo que ocurrió. Así, ¿Qué podemos hacer?	2	Okay. Desea usted no había casado conmigo. Desea usted «d tenido una vida diferente. Pero es lo que ocurrió. Así, ¿Qué podemos hacer?	2
Yo no lo sé.	1	Yo no contesta.	3	No sé si.	2	No sé si.	2	No sé si.	2	No sé si.	2
Puntuación media	**2,15**	**Puntuación media**	**2,54**	**Puntuación media**	**2,38**	**Puntuación media**	**2,46**	**Puntuación media**	**2,31**	**Puntuación media**	**2,54**

8.3 Evaluación del estilo (oralidad)

Doblaje ES	Punt.	Trad. 1	Punt. 1	Trad. 2	Punt. 2	Trad. 3	Punt. 3	Trad. 4	Punt. 4	Trad. 5	Punt. 5	Trad. 6	Punt. 6
No sé cómo empezar.	2	I no sabe cómo comenzar.	2	I no sabemos cómo comenzar.	2	No sé si la manera de empezar.	2	No sé si cómo comenzar.	2	No sé si la manera de empezar.	2	No sé si cómo empezar.	2
¿Entiendes por qué quiero quedarme en Los Ángeles?	2	No entiende por qué quiero estancia en L.A.?	2	¿Le entender por qué quiero estancia en L.A.?	2	¿Sabe por qué me quieren permanecer en L.A.?	2	¿Sabe por qué quiero permanecer en L.A.?	2	Usted comprende por qué me quieren permanecer en L.A.?	2	¿Sabe por qué quiero permanecer en L.A.?	2
No.	2	Nº.	2	Nº.	2	Nº.	2	Nº.	2	Nº.	2	Nº.	2
Vale... Esa no... Charlie, esa no	1	Bien, eso no. Charlie, que no	1	Bien, eso not… Charlie, eso no	1	Así, que de no. Charlie, que no	1	Bien, ¡es not… Charlie, que no	1	Así, que de no. Charlie, que no	1	Bien, ¡es not… Charlie, que no	1

es la mejor forma de empezar.		de una manera útil para nosotros, para empezar.		una manera útil para nosotros a comenzar.		de una manera útil para nosotros de empezar a.		hay una manera útil a nosotros empezar.		de una manera útil para nosotros, para empezar.		hay una manera útil para nosotros empezar.	
Es que no lo entiendo.	1	I no entenderlo.	2	I no entenderlo.	2	No comunicaré comprender.	2	No comunicaré comprender.	2	No comunicaré comprender.	2	No comunicaré comprender.	2
¿No recuerdas que me prometiste que pasaríamos tiempo aquí?	1	No te recuerde prometedoras que podríamos hacer tiempo aquí?	2	No te recuerde prometedoras que podríamos hacer tiempo aquí?	2	¿No vas a recordar prometedoras que podemos hacer aquí?	2	Usted No comunicaré recordar prometedoras que podríamos hacer aquí?	2	¿No vas a recordar prometedoras que podemos hacer aquí?	2	Usted No comunicaré recordar prometedoras que podríamos hacer aquí?	2
Comentábamos cosas, como cualquier matrimonio. Dijimos de irnos a Europa, de comprar un... mueble de esos, ¿cómo se llama? Un aparador para cubrir el hueco de detrás del sofá, y no hicimos nada de eso.	1	Hemos debatido comportamientos. Estuvimos casados. Dijo las cosas. Hemos hablado de que se desplazan a Europa, sobre la posibilidad de padecer un. sideboard o qué ustedes? Un credenza, para llenar el espacio vacío detrás del sofá. Que nunca tuvo ninguna de ellas.	1	Debatimos comportamientos. Estuvimos casados. Dijo cosas. Hablamos acerca hacia Europa, parecería salir a… sideboard o qué ustedes? Un credenza, para llenar el espacio vacío detrás del sofá. Nunca nos hicimos todo de ella.	1	Hemos debatido comportamientos. estábamos casados. Decíamos cosas. Hemos hablado acerca de trasladarse a Europa, alrededor de conseguir un. sideboard o qué hacen ustedes? Un credenza, para llenar el espacio vacío detrás del couch. Nosotros nunca hizo ningún de ella.	1	Hemos debatido comportamientos. estábamos casados. Decíamos cosas. Hemos hablado acerca hacia Europa, parecería salir a… sideboard o qué hacen ustedes? Un credenza, para llenar el espacio vacío detrás couch. Nosotros nunca hizo ningún de ella.	1	Hemos debatido comportamientos. estábamos casados. Decíamos cosas. Hemos hablado acerca de la transición hacia Europa, alrededor de conseguir un. sideboard o qué hacen ustedes? Un credenza, colmar el vacío espacio detrás del couch. Nunca nos hizo ningún de ella.	1	Hemos debatido comportamientos. estábamos casados. Decíamos cosas. Hemos hablado sobre hacia Europa, parecería salir a… sideboard o qué hacen ustedes? Un credenza, colmar el espacio vacío detrás couch. Nosotros nunca hizo ningún de ella.	1

Rechazaste la residencia en el Greffen que nos habría permitido pasar aquí un año.	2	A su vez, por la residencia a Geffen que nos han traído aquí para un año.	2	Le rechazan La residency a Geffen que habría permitido aquí para a… un año.	1	Usted, a su vez, en la residencia a Geffen que nos han traído aquí para una. Un año.	1	Usted rechazar la residencia a Geffen que nos han traído aquí para a… un año.	1	Usted, a su vez, en la residencia a Geffen que nos han traído aquí para una. Un año.	1	Usted rechazar la residencia al Geffen que hubiera traído nosotros aquí para a… un año.	1
Porque no me apetecía. Teníamos una compañía de teatro y una vida increíbles.	1	No es algo que Yo quería. Hemos tenido un gran teatro sociedad y una gran vida de dónde venimos.	2	No era algo que Yo quería. Tuvimos un gran teatro empresa y una gran vida cuando éramos.	2	No es algo que yo quería. Tuvimos un gran teatro de la sociedad y una gran vida donde estábamos.	2	No era algo me quería. Tuvimos un gran teatro sociedad y una gran vida donde estábamos.	2	No es algo que yo quería. Hemos tenido una gran teatro sociedad y una gran vida donde estábamos.	2	Era algo no me querían. Tuvimos un gran teatro sociedad y una gran vida donde estábamos.	2
¿Aquello era una vida increíble?	2	Llame a que una gran vida?	2	Ustedes que una gran vida?	2	Ustedes que una gran vida?	2	Ustedes que una gran vida?	2	Ustedes que una gran vida?	2	Ustedes que una gran vida?	2
Tú ya me entiendes. No digo que el matrimonio fuera increíble. Hablo de la vida en Brooklyn. Profesionalmente. No sé, la verdad es que nunca he querido hacer otra cosa.	1	Necesita saber a qué me refiero. I no significa que tuvo un gran matrimonio. Me refiero, vida en Brooklyn. Profesionalmente. Yo no lo sé. Honestamente, yo no consideran algo diferente.	2	Saben qué me refiero. I no significa tuvimos un gran matrimonio. Me refiero, vida en Brooklyn. Es profesional. Yo no contesta. Honestamente, yo nunca considerado algo diferente.	2	Saben lo que me hacen. No comunicaré significa que tuvimos un gran matrimonio. I significa, vida en Brooklyn. Profesional. No sé si. Honestamente, yo nunca considerado algo diferente.	2	Sabes qué me hacen. No comunicaré significa tuvimos un gran matrimonio. I significa, vida en Brooklyn. Es profesional. No sé si. Honestamente, yo nunca considerado algo diferente.	2	Saben lo que me hacen. No comunicaré significa que tuvimos una gran ceremonia del matrimonio. Me refiero a una vida, en Brooklyn. El punto de vista profesional. No sé si. Honestamente, yo nunca como algo diferente.	2	Saben qué me hacen. No comunicaré significa tuvimos una gran matrimonio. I significa, vida en Brooklyn. Vista profesional. No sé si. Honestamente, yo nunca consideró algo diferente.	2

Ya, y ese es el problema, Charlie. A ver, yo era tu mujer. Debiste pensar también en mi felicidad.	1	Bien, eso el problema, no es él? Me refiero, era su esposa. Debe tener en cuenta mi felicidad, demasiado.	1	Bien, eso el problema, no es él? Me refiero, yo era su esposa. Debe han considerado mi felicidad, demasiado.	1	Así que el problema, no es él? I significa, yo era su esposa. Usted tener que mi felicidad, también.	2	Así que el problema, ¿no es esto? I significa fui tu esposa. Deberías haber estudiado mi felicidad, también.	1	Así que el problema, ¿no es él? Me hacen, yo estaba tu esposa. Usted tener que mi felicidad, también.	1	Bien, ¡es el problema, ¿no es esto? Me significa fui tu esposa. Deberías haber considerado mi felicidad, también.	1
Venga ya, eras muy feliz y de repente decidiste que ya no lo eras.	1	Llegado. Le fueron felices. Solo le hemos decidido que no estaban ahora.	2	Llegado sobre. Le fueron felices. Solo le hemos decidido no le fueron ahora.	2	A. Ustedes son felices. Le've sólo se decidió usted no ahora.	2	A el. Ustedes son felices. Usted've sólo decidió usted eran desconocidas actualmente.	2	A. Ustedes estaban felices. Usted've sólo se decidió usted no ahora.	2	A el. Ustedes estaban felices. Usted've sólo decidió usted no eran ahora.	2
Vale, vale... Va... Vamos a... Mi trabajo está aquí ahora, mi familia está aquí...	1	OK. Vayamos. No debemos. M- mi trabajo está aquí ahora, mi familia de aquí.	1	RESPUESTA OK. Deje's… ¡Trabajemos not… M- mi trabajo está aquí ahora, mi familia here… de	1	OK. Hablemos de. No hay que. M- mi trabajo es aquí ahora, mi familia de aquí.	1	OK. Llamaremos' s… Vayamos not… M- mi trabajo aquí ahora, mi familia here…	1	OK. Hablemos de. No hay que. M- mi trabajo es aquí ahora, mi familia de aquí.	1	OK. Llamaremos' s… Vayamos not… M- mi trabajo es aquí ahora, mi familia here…	1
Y yo acepté que Henry fuera al cole aquí, porque tu piloto se convirtió en serie, porque pensé que cuando acabaras volveríais a Nueva York.	1	Y estoy de acuerdo a Henry a la escuela porque su show fue a la serie. Hice que saber que cuando se le han hecho tiro, volver a Nueva York.	1	Y yo acordaron poner Henry en la escuela aquí porque su show fue a serie. Hice que sabiendo que cuando se realizaron tiro, él volverá a Nueva York.	1	Y me acuerdo a Henry en la escuela, porque Aquí te muestran fue para la serie. Hice que saber que cuando se realizaron tiro, que quiere volver a Nueva York.	1	Y me acordaron poner Henry escolar aquí porque tu muestran fue a la serie. Hice que saber que cuando se realizaron tiro, debería volver a Nueva York.	1	Y me acordaron poner Henry en la escuela aquí porque va a dar tu serie. Hice que saber que cuando se realizaron tiro, va a volver a Nueva York.	1	Y me acordaron poner Henry escolar aquí porque tu muestran fue a serie. Hice que saber que cuando se realizaron tiro, debería volver a Nueva York.	1

Cielo, nunca dijimos eso. Quizá lo supusieras, pero nunca dijimos eso.	1	La miel, dijo que nunca. Que puede haber sido su hipótesis, pero nunca explícitamente declara que.	1	Miel, dijo que nunca. Que puede haber sido su suposición, pero nunca dijo que expresamente.	1	La miel, dijo que nunca. Que puede haber sido su supuesto, pero nunca explícitamente declara que.	1	Miel, dijo que nunca. Que puede haber sido tu suposición, pero nunca explícitamente declara que.	1	Miel, nos dijo que nunca. Que puede haber sido su hipótesis, pero dijo que nunca expresamente.	1	Miel, nunca dijo que. Que puede haber sido tu supuesto, pero nunca expresamente que.	1
Claro que lo dijimos.	2	Podemos decir que no.	2	Hicimos decirlo.	2	Decir que no.	2	Hicimos decir.	2	Hicimos decir.	2	Hicimos decir.	2
¿Y cuándo lo dijimos?	1	Cuando decimos que sí?	2	Hizo Cuando decimos que?	2	Cuando no podemos decir?	2	Cuando no podemos decir?	2	Cuando no podemos decir?	2	Cuando no podemos decir?	2
No sé cuándo lo dijimos, pero lo dijimos.	1	No sé si cuando nos lo dijo, pero nos lo dijo.	1	No sé si cuando nos lo dijo, pero nos dijo.	1	No sé si se dice que cuando él, pero se dice que es.	1	No sé si cuando decíamos, pero dijo ella.	2	No sé si cuando decíamos, pero nos dijo.	2	No sé si cuando decíamos, pero dijo ella.	2
Yo creía que...	1	Yo pensaba.	1	I thought…	1	I pensaba.	2	Thought… I	2	I pensaba.	1	Thought… I	2
¡Aquella vez por teléfono!	1	Ella dijo que el tiempo en el teléfono.		Dijo que ese tiempo en el teléfono.	2	Se dice que ese tiempo en el teléfono.	2	Decíamos que entonces el teléfono!	1	Dijo que ese momento, sobre el teléfono!	1	Decíamos que entonces el teléfono!	1
¡Cielo, que me dejes acabar! Perdona por volver a decírtelo.Yo creía que, si Henry era feliz aquí y mi serie prosperaba, podríamos vivir aquí un tiempo.	1	La miel. Permítanme terminar. Siento, estoy diciendo que mantener. Pensé que si Henry fue feliz aquí y mi muestran continuación, que podría hacer L.a. Para un tiempo.	2	La miel! Permítanme terminar! Siento, estoy diciendo que mantener. Pensé que si Henry fue feliz aquí y mi muestran continuó que podríamos hacer L.a.	1	La miel. Permítanme terminar. Siento, tengo que seguir diciendo. I pensaba que cuando Henry era feliz aquí y mi muestran continuación, que se puede	2	La miel! Permítanme terminar! Siento, me seguir diciendo que. I pensaba que cuando Henry era feliz aquí y mi espectáculo continuó, que podríamos hacer L.a. Un rato.	1	Miel! Permítanme terminar. Siento, tengo que seguir diciendo. I pensaba que si Henry es feliz aquí y mi muestra de continuidad, que se puede	1	La miel! Permítanme terminar! Siento, me seguir diciendo que. I pensaba que si Henry es feliz aquí y mi muestran continuación, que podríamos hacer L.a. Un rato.	1

				Durante un tiempo.		hacer L.a. Para un tiempo.				hacer L.a. Un rato.		
Nunca me informaste de ese proceso mental.	2	Yo no estaba enterada de que el proceso de reflexión.	2	No estaba enterada de ese pensamiento proceso.	2	Yo no es que creía conocer.	2	Yo no conocer que pensaba.	2	Yo no era conocer que pensaba.	2	Yo no era conocer que pensaba.	2
La única razón de que no viviéramos aquí es que eres incapaz de pensar en otra cosa que no sean tus deseos si no se te obliga.	2	La única razón de que no la vida aquí es debido a que no se puede imaginar desea que su propia. Si no estamos obligados a usted.	2	El único motivo tampoco lo hemos conseguido vivir aquí es porque usted no puede imaginarse desea excepto tu propio. Salvo que estamos obligados a usted.	2	La única razón por la que no la vida aquí es porque usted no podrán imaginarse desea que no sean tus propios. Si le obligó a la obra».	2	La única razón tuviéramos vivir aquí es porque usted no podrán imaginarse desea excepto tu propio. Salvo obra» obligado a usted.	2	La única razón, no estábamos viviendo aquí es porque usted no podrán imaginarse desea que no sean tus propios. Si le obligó a la obra».	2	La única razón tuviéramos vivir aquí es porque usted no podrán imaginarse desea excepto tu propia. Salvo obra» forzoso sobre usted.	2
Vale, querrías no haberte casado conmigo y haber tenido otra vida. Pero esto es lo que hay. Entonces, ¿qué hacemos?	1	Normal. Le deseo de que no había casado conmigo. Le deseo «d tenido una vida diferente. Pero esto es lo que ocurrió. Así pues, ¿qué podemos hacer?	1	Importa. Le deseo no había casado conmigo. Le deseo «d tenido una vida diferente. Pero esto es lo que ocurrió. Afirmativo, ¿qué podemos hacer?	2	Okay. Te deseo no le había casado conmigo. Desea usted «d tenido una vida diferente. Pero esto es lo que ocurrió. Así, ¿Qué podemos hacer?	1	Okay. Desea usted no había casado conmigo. Desea usted «d tenido una vida diferente. Pero es lo que ocurrió. Así, ¿Qué podemos hacer?	1	Okay. Desea usted no me había casado. Desea usted «d tenido una vida diferente. Pero esto es lo que ocurrió. Así, ¿Qué podemos hacer?	2	Okay. Desea usted no había casado conmigo. Desea usted «d tenido una vida diferente. Pero es lo que ocurrió. Así, ¿Qué podemos hacer?	1
No lo sé.	2	Yo no lo sé.	2	Yo no contesta.	2	No sé si.	2	No sé si.	2	No sé si.	2	No sé si.	2

Puntuación media	1,35	Puntuación media	1,64	Puntuación media	1,62	Puntuación media	1,69	Puntuación media	1,62	Puntuación media	1,62	Puntuación media	1,62

8.4 Clasificación general

Original EN	Trad. 1	Punt. 1	Trad. 2	Punt. 2	Trad. 3	Punt. 3	Trad. 4	Punt. 4	Trad. 5	Punt. 5	Trad. 6	Punt. 6
I don't know how to start.	I no sabe cómo comenzar.	4	I no sabemos cómo comenzar.	5	No sé si la manera de empezar.	3	No sé si cómo comenzar.	2	No sé si la manera de empezar.	3	No sé si cómo empezar.	1
Do you understand why I want to stay in L.A.?	No entiende por qué quiero estancia en L.A.?	4	¿Le entender por qué quiero estancia en L.A.?	5	¿Sabe por qué me quieren permanecer en L.A.?	2	¿Sabe por qué quiero permanecer en L.A.?	1	Usted comprende por qué me quieren permanecer en L.A.?	3	¿Sabe por qué quiero permanecer en L.A.?	1
No.	Nº.	3	Nº.	3	Nº.	3	Nº.	3	Nº.	3	Nº.	3
Well, that's not… Charlie, that's not a useful way for us to start.	Bien, eso no. Charlie, que no de una manera útil para nosotros, para empezar.	1	Bien, eso not… Charlie, eso no una manera útil para nosotros a comenzar.	3	Así, que de no. Charlie, que no de una manera útil para nosotros de empezar a.	5	Bien, ¡es not… Charlie, que no hay una manera útil a nosotros empezar.	6	Así, que de no. Charlie, que no de una manera útil para nosotros, para empezar.	4	Bien, ¡es not… Charlie, que no hay una manera útil para nosotros empezar.	2
I don't understand it.	I no entenderlo.	4	I no entenderlo.	4	No comunicaré comprender.	3	No comunicaré comprender.	3	No comunicaré comprender.	3	No comunicaré comprender.	3
You don't remember promising that we could do time here?	No te recuerde prometedoras que podríamos hacer tiempo aquí?	2	No te recuerde prometedoras que podríamos hacer tiempo aquí?	2	¿No vas a recordar prometedoras que podemos hacer aquí?	1	Usted No comunicaré recordar prometedoras que podríamos hacer aquí?	3	¿No vas a recordar prometedoras que podemos hacer aquí?	1	Usted No comunicaré recordar prometedoras que podríamos hacer aquí?	3

We discussed things; we were married. We said things. We talked about moving to Europe, about getting a… sideboard or what do you call it? A credenza, to fill the empty space behind the couch. We never did any of it.	Hemos debatido comportamientos. Estuvimos casados. Dijo las cosas. Hemos hablado de que se desplazan a Europa, sobre la posibilidad de padecer un. sideboard o qué ustedes? Un credenza, para llenar el espacio vacío detrás del sofá. Que nunca tuvo ninguna de ellas.	3	Debatimos comportamientos. Estuvimos casados. Dijo cosas. Hablamos acerca hacia Europa, parecería salir a… sideboard o qué ustedes? Un credenza, para llenar el espacio vacío detrás del sofá. Nunca nos hicimos todo de ella.	4	Hemos debatido comportamientos. estábamos casados. Decíamos cosas. Hemos hablado acerca de trasladarse a Europa, alrededor de conseguir un. sideboard o qué hacen ustedes? Un credenza, para llenar el espacio vacío detrás del couch. Nosotros nunca hizo ningún de ella.	1	Hemos debatido comportamientos. estábamos casados. Decíamos cosas. Hemos hablado acerca hacia Europa, parecería salir a… sideboard o qué hacen ustedes? Un credenza, para llenar el espacio vacío detrás couch. Nosotros nunca hizo ningún de ella.	5	Hemos debatido comportamientos. estábamos casados. Decíamos cosas. Hemos hablado acerca de la transición hacia Europa, alrededor de conseguir un. sideboard o qué hacen ustedes? Un credenza, colmar el vacío espacio detrás del couch. Nunca nos hizo ningún de ella.	2	Hemos debatido comportamientos. estábamos casados. Decíamos cosas. Hemos hablado sobre hacia Europa, parecería salir a… sideboard o qué hacen ustedes? Un credenza, colmar el espacio vacío detrás couch. Nosotros nunca hizo ningún de ella.	6
You turn down the residency at the Geffen that would have brought us here for a… a year.	A su vez, por la residencia a Geffen que nos han traído aquí para un año.	2	Le rechazan La residency a Geffen que habría permitido aquí para a… un año.	1	Usted, a su vez, en la residencia a Geffen que nos han traído aquí para una. Un año.	5	Usted rechazar la residencia a Geffen que nos han traído aquí para a… un año.	4	Usted, a su vez, en la residencia a Geffen que nos han traído aquí para una. Un año.	5	Usted rechazar la residencia al Geffen que hubiera traído nosotros aquí para a… un año.	3

It wasn't something I wanted. We had a great theater company and a great life where we were.	No es algo que Yo quería. Hemos tenido un gran teatro sociedad y una gran vida de dónde venimos.	3	No era algo que Yo quería. Tuvimos un gran teatro empresa y una gran vida cuando éramos.	5	No es algo que yo quería. Tuvimos un gran teatro de la sociedad y una gran vida donde estábamos.	1	No era algo me quería. Tuvimos un gran teatro sociedad y una gran vida donde estábamos.	4	No es algo que yo quería. Hemos tenido una gran teatro sociedad y una gran vida donde estábamos.	2	Era algo no me querían. Tuvimos un gran teatro sociedad y una gran vida donde estábamos.	6
You call that a great life?	Llame a que una gran vida?	1	Ustedes que una gran vida?	2	Ustedes que una gran vida?	2	Ustedes que una gran vida?	2	Ustedes que una gran vida?	2	Ustedes que una gran vida?	2
You know what I mean. I don't mean we had a great marriage. I mean, life in Brooklyn. Professionally. I don't know. Honestly, I never considered anything different.	Necesita saber a qué me refiero. I no significa que tuvo un gran matrimonio. Me refiero, vida en Brooklyn. Profesionalmente. Yo no lo sé. Honestamente, yo no consideran algo diferente.	1	Saben qué me refiero. I no significa tuvimos un gran matrimonio. Me refiero, vida en Brooklyn. Es profesional. Yo no contesta. Honestamente, yo nunca considerado algo diferente.	2	Saben lo que me hacen. No comunicaré significa que tuvimos un gran matrimonio. I significa, vida en Brooklyn. Profesional. No sé si. Honestamente, yo nunca considerado algo diferente.	3	Sabes qué me hacen. No comunicaré significa tuvimos un gran matrimonio. I significa, vida en Brooklyn. Es profesional. No sé si. Honestamente, yo nunca considerado algo diferente.	4	Saben lo que me hacen. No comunicaré significa que tuvimos una gran ceremonia del matrimonio. Me refiero a una vida, en Brooklyn. El punto de vista profesional. No sé si. Honestamente, yo nunca como algo diferente.	6	Saben qué me hacen. No comunicaré significa tuvimos una gran matrimonio. I significa, vida en Brooklyn. Vista profesional. No sé si. Honestamente, yo nunca consideró algo diferente.	5
Well, that's the problem, isn't it? I mean, I was your wife. You should have considered	Bien, eso el problema, no es él? Me refiero, era su esposa. Debe tener en cuenta mi felicidad, demasiado.	1	Bien, eso el problema, no es él? Me refiero, yo era su esposa. Debe han considerado mi felicidad, demasiado.	3	Así que el problema, no es él? I significa, yo era su esposa. Usted tener que mi felicidad, también.	5	Así que el problema, ¿no es esto? I significa fui tu esposa. Deberías haber estudiado mi	2	Así que el problema, ¿no es él? Me hacen, yo estaba tu esposa. Usted tener que mi	6	Bien, ¡es el problema, ¿no es esto? Me significa fui tu esposa. Deberías haber considerado mi	4

my happiness, too.						felicidad, también.		felicidad, también.		felicidad, también.	
Come on. You were happy. You've just decided you weren't now.	Llegado. Le fueron felices. Solo le hemos decidido que no estaban ahora.	1	Llegado sobre. Le fueron felices. Solo le hemos decidido no le fueron ahora.	2	A. Ustedes son felices. Le've sólo se decidió usted no ahora.	4	A el. Ustedes son felices. Usted've sólo decidió usted eran desconocidas actualmente.	6	A. Ustedes estaban felices. Usted've sólo se decidió usted no ahora.	5	A el. Ustedes estaban felices. Usted've sólo decidió usted no eran ahora.	3
OK. Let's… let's not… M- my work is here now, my family's here…	OK. Vayamos. No debemos. M- mi trabajo está aquí ahora, mi familia de aquí.	1	RESPUESTA OK. Deje's… ¡Trabajemos not… M- mi trabajo está aquí ahora, mi familia here… de	6	OK. Hablemos de. No hay que. M- mi trabajo es aquí ahora, mi familia de aquí.	2	OK. Llamaremos' s… Vayamos not… M- mi trabajo aquí ahora, mi familia here…	5	OK. Hablemos de. No hay que. M- mi trabajo es aquí ahora, mi familia de aquí.	2	OK. Llamaremos' s… Vayamos not… M- mi trabajo es aquí ahora, mi familia here…	4
And I agreed to put Henry in school here because your show went to series. I did that knowing that when you were done shooting, he would come	Y estoy de acuerdo a Henry a la escuela porque su show fue a la serie. Hice que saber que cuando se le han hecho tiro, volver a Nueva York.	2	Y yo acordaron poner Henry en la escuela aquí porque su show fue a serie. Hice que sabiendo que cuando se realizaron tiro, él volverá a Nueva York.	1	Y me acuerdo a Henry en la escuela, porque Aquí te muestran fue para la serie. Hice que saber que cuando se realizaron tiro, que quiere volver a Nueva York.	5	Y me acordaron poner Henry escolar aquí porque tu muestran fue a la serie. Hice que saber que cuando se realizaron tiro, debería volver a Nueva York.	4	Y me acordaron poner Henry en la escuela aquí porque va a dar tu serie. Hice que saber que cuando se realizaron tiro, va a volver a Nueva York.	3	Y me acordaron poner Henry escolar aquí porque tu muestran fue a serie. Hice que saber que cuando se realizaron tiro, debería volver a Nueva York.	4

back to New York.											
Honey, we never said that. That may have been your assumption, but we never expressly said that.	La miel, dijo que nunca. Que puede haber sido su hipótesis, pero nunca explícitamente declara que.	6	Miel, dijo que nunca. Que puede haber sido su suposición, pero nunca dijo que expresamente.	2	La miel, dijo que nunca. Que puede haber sido su supuesto, pero nunca explícitamente declara que.	5	Miel, dijo que nunca. Que puede haber sido tu suposición, pero nunca explícitamente declara que.	3	Miel, nos dijo que nunca. Que puede haber sido su hipótesis, pero dijo que nunca expresamente.	1	Miel, nunca dijo que. Que puede haber sido tu supuesto, pero nunca expresamente que.	4
We did say it.	Podemos decir que no.	3	Hicimos decirlo.	1	Decir que no.	4	Hicimos decir.	2	Hicimos decir.	2	Hicimos decir.	2
When did we say it?	Cuando decimos que sí?	2	Hizo Cuando decimos que?	4	Cuando no podemos decir?	3	Cuando no podemos decir?	3	Cuando no podemos decir?	3	Cuando no podemos decir?	3
I don't know when we said it, but we said it.	No sé si cuando nos lo dijo, pero nos lo dijo.	1	No sé si cuando nos lo dijo, pero nos dijo.	2	No sé si se dice que cuando él, pero se dice que es.	4	No sé si cuando decíamos, pero dijo ella.	5	No sé si cuando decíamos, pero nos dijo.	3	No sé si cuando decíamos, pero dijo ella.	5
I thought…	Yo pensaba.	1	I thought…	3	I pensaba.	2	Thought… I	4	I pensaba.	2	Thought… I	4
We said it that time on the phone!	Ella dijo que el tiempo en el teléfono.	5	Dijo que ese tiempo en el teléfono.	2	Se dice que ese tiempo en el teléfono.	4	Decíamos que entonces el teléfono!	1	Dijo que ese momento, sobre el teléfono!	3	Decíamos que entonces el teléfono!	1

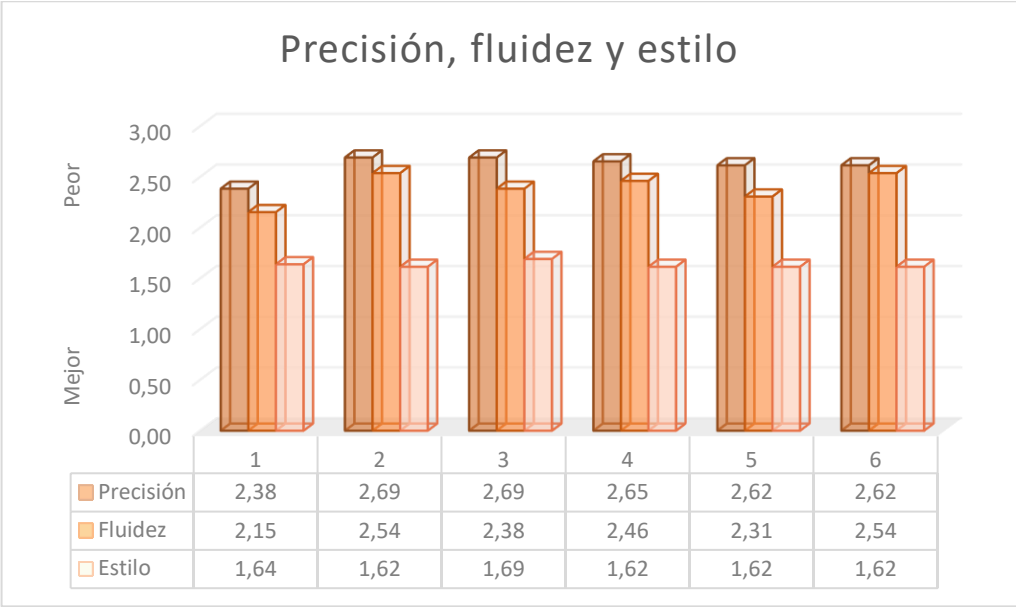
Honey! Let me finish! Sorry, I keep saying that. I thought that if Henry was happy here and my show continued, that we might do L.A. for a while.	La miel. Permítanme terminar. Siento, estoy diciendo que mantener. Pensé que si Henry fue feliz aquí y mi muestran continuación, que podría hacer L.a. Para un tiempo.	4	La miel! Permítanme terminar! Siento, estoy diciendo que mantener. Pensé que si Henry fue feliz aquí y mi muestran continuó que podríamos hacer L.a. Durante un tiempo.	2	La miel. Permítanme terminar. Siento, tengo que seguir diciendo. I pensaba que cuando Henry era feliz aquí y mi muestran continuación, que se puede hacer L.a. Para un tiempo.	3	La miel! Permítanme terminar! Siento, me seguir diciendo que. I pensaba que cuando Henry era feliz aquí y mi espectáculo continuó, que podríamos hacer L.a. Un rato.	1	Miel! Permítanme terminar. Siento, tengo que seguir diciendo. I pensaba que si Henry es feliz aquí y mi muestra de continuidad, que se puede hacer L.a. Un rato.	5	La miel! Permítanme terminar! Siento, me seguir diciendo que. I pensaba que si Henry es feliz aquí y mi muestran continuación, que podríamos hacer L.a. Un rato.	6
I was not privy to that thought process.	Yo no estaba enterada de que el proceso de reflexión.	1	No estaba enterada de ese pensamiento proceso.	2	Yo no es que creía conocer.	5	Yo no conocer que pensaba.	4	Yo no era conocer que pensaba.	3	Yo no era conocer que pensaba.	3
The only reason we didn't live here is because you can't imagine desires other than your own. Unless they're forced on you.	La única razón de que no la vida aquí es debido a que no se puede imaginar desea que su propia. Si no estamos obligados a usted.	2	El único motivo tampoco lo hemos conseguido vivir aquí es porque usted no puede imaginarse desea excepto tu propio. Salvo que estamos obligados a usted.	1	La única razón por la que no la vida aquí es porque usted no podrán imaginarse desea que no sean tus propios. Si le obligó a la obra».	4	La única razón tuviéramos vivir aquí es porque usted no podrán imaginarse desea excepto tu propio. Salvo obra» obligado a usted.	6	La única razón, no estábamos viviendo aquí es porque usted no podrán imaginarse desea que no sean tus propios. Si le obligó a la obra».	3	La única razón tuviéramos vivir aquí es porque usted no podrán imaginarse desea excepto tu propia. Salvo obra» forzoso sobre usted.	5

Okay. You wish you hadn't married me. You wish you'd had a different life. But this is what happened. So, what do we do?	Normal. Le deseo de que no había casado conmigo. Le deseo «d tenido una vida diferente. Pero esto es lo que ocurrió. Así pues, ¿qué podemos hacer?	2	Importa. Le deseo no había casado conmigo. Le deseo «d tenido una vida diferente. Pero esto es lo que ocurrió. Afirmativo, ¿qué podemos hacer?	5	Okay. Te deseo no le había casado conmigo. Desea usted «d tenido una vida diferente. Pero esto es lo que ocurrió. Así, ¿Qué podemos hacer?	4	Okay. Desea usted no había casado conmigo. Desea usted «d tenido una vida diferente. Pero es lo que ocurrió. Así, ¿Qué podemos hacer?	1	Okay. Desea usted no me había casado. Desea usted «d tenido una vida diferente. Pero esto es lo que ocurrió. Así, ¿Qué podemos hacer?	3	Okay. Desea usted no había casado conmigo. Desea usted «d tenido una vida diferente. Pero es lo que ocurrió. Así, ¿Qué podemos hacer?	1
I don't know.	Yo no lo sé.	1	Yo no contesta.	3	No sé si.	2	No sé si.	2	No sé si.	2	No sé si.	2
	Puntuación media	**2,35**	**Puntuación media**	**2,88**	**Puntuación media**	**3,27**	**Puntuación media**	**3,31**	**Puntuación media**	**3,08**	**Puntuación media**	**3,31**

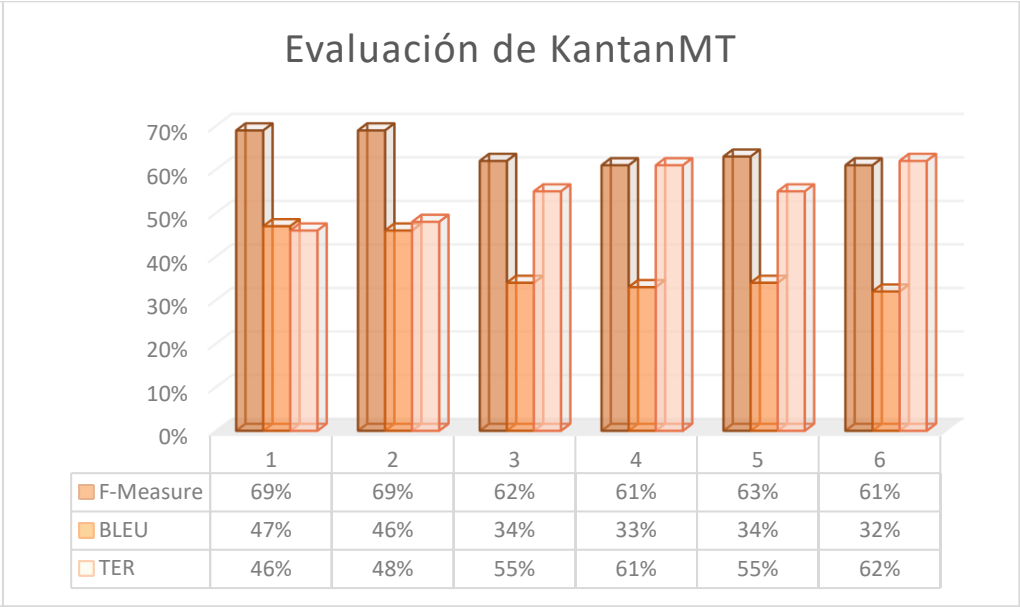
8.5 Evaluación automática de KantanMT

Engine	Source	Target	Word count	Monolingual word count	Unique word count	F-Measure	BLEU	TER
1-escrito-escrito-SO	en	es	80.044.130	25.417.814	844,376	69%	47%	46%
2-escrito-escrito-CO	en	es	80.044.130	25.417.814	844,376	69%	46%	48%
3-escrito-mixto-SO	en	es	80.044.130	33.020.097	844,377	62%	34%	55%
4-escrito-mixto-CO	en	es	80.044.130	33.020.097	844,377	61%	33%	61%
5-escrito-oral-SO	en	es	80.044.130	7.602.283	844,377	63%	34%	55%
6-escrito-oral-CO	en	es	80.044.130	7.602.283	844,377	61%	32%	62%

8.6 Gráficos



	1	2	3	4	5	6
Precisión	2,38	2,69	2,69	2,65	2,62	2,62
Fluidez	2,15	2,54	2,38	2,46	2,31	2,54
Estilo	1,64	1,62	1,69	1,62	1,62	1,62



	1	2	3	4	5	6
F-Measure	69%	69%	62%	61%	63%	61%
BLEU	47%	46%	34%	33%	34%	32%
TER	46%	48%	55%	61%	55%	62%