Entrenamiento y comparativa de motores de TAE especializados en la localización de aplicaciones móviles

María Esperanza Fernández Ruíz
Pilar Sánchez-Gijón

revista.tradumàtica
tecnologies de la traducció

Rebuda: 7 de setembre de 2019 | Acceptació: 16 de desembre de 2019 | Publicació avançada: 23 de desembre de 2019

María Esperanza Fernández Ruíz
Universitat Autònoma de Barcelona
maria.esperanza.f.r@gmail.com;
ORCID:
0000-0000-0000-0000

Resumen

La calidad de la traducción automática estadística (TAE) está estrechamente relacionada con las características de los corpus utilizados, como son su volumen y su homogeneidad. Este artículo describe el proceso realizado durante el entrenamiento de tres motores de traducción automática estadística (TAE) a partir de diferentes combinaciones de colecciones de textos, todos ellos especializados en la traducción de aplicaciones móviles. Se compararán los tres motores con el fin de determinar qué composición de corpus es la ideal para entrenar motores especializados en la localización de aplicaciones y comprobar cómo abordan los motores entrenados ciertos aspectos problemáticos de este tipo de aplicaciones.

Palabras clave: traducción automática estadística, localización, aplicaciones móviles, MTradumàtica, corpus

Pilar Sánchez-Gijón
Grup Tradumàtica - Universitat Autònoma de Barcelona
pilar.sanchez.gijon@uab.cat;
ORCID:
0000-0001-5919-4629

Resum

La qualitat de la traducció automàtica estadística (TAE) està estretament relacionada amb les característiques dels corpus utilitzats, com ara el seu volum i la seva homogeneïtat. Aquest article descriu el procés realitzat durant l'entrenament de tres motors de traducció automàtica estadística a partir de diferents combinacions de col·leccions de textos, tots especialitzats en la traducció d'aplicacions mòbils. Es farà una comparació dels tres motors amb l'objectiu de determinar quina composició de corpus és la ideal per entrenar motors especialitzats en la localització d'aplicacions i comprovar com els motors entrenats aborden certs aspectes problemàtics de la traducció d'aplicacions.

Paraules clau: traducció automàtica estadística, localització, aplicacions mòbils, MTradumàtica, corpus

Abstract

The quality of statistical machine translation (SMT) is closely related to the characteristics of the corpus used, such as its volume and its homogeneity. This article describes the training process of three statistical machine translation engines with different collections of texts combi-

Revista Tradumàtica. Número 17 (desembre 2019).
Cita recomanada: FERNÁNDEZ RUÍZ, MARÍA ESPERANZA; SÁNCHEZ-GIJÓN, PILAR (2019). «Entrenamiento y comparativa de motores de TAE especializados en la localización de aplicaciones móviles.». Revista Tradumàtica. Tecnologies de la Traducció, 17, 162-183.
<https://doi.org/10.5565/rev/tradumatica.229>

http://revistes.uab.cat/tradumatica
ISSN 1578-7559

ned, all specialized in the translation of mobile apps. The three engines will be compared in order to determine which composition of corpus is the best for training engines specialized in app localisation, and to check how the trained engines deal with some problematic aspects related to this type of translation.

Keywords: statistical machine translation, localization, mobile apps, MTradumàtica, corpus

1. Introducción

Las aplicaciones y los dispositivos móviles están presentes en nuestro día a día. El número de descargas de aplicaciones a nivel mundial en 2017 fue de 178,1 miles de millones, y se prevé que, en 2019, el número de descargas ascienda a 258,2 miles de millones de aplicaciones (Fundación Telefónica, 2019). La mayor parte de ellas han tenido que localizarse previamente para que podamos usarlas en nuestra propia lengua, por lo que la localización de aplicaciones móviles tiene una gran relevancia y constituye una oportunidad de negocio dentro del ámbito de los servicios lingüísticos.

En cuanto a los plazos de entrega de los que se suele disponer para su traducción, estos son bastante cortos, debido en parte a que el ciclo de vida de algunas aplicaciones es también corto y a que hay actualizaciones para cada aplicación constantemente. Simón (2016) resume la situación del mercado de localización de aplicaciones de la siguiente manera: «When freelancers and small to medium-sized LSPs venture into this market niche, they are confronted with the challenge of making a profit from projects that are inevitably small, global and urgent». Por tanto, la prospección de la integración de la traducción automática en el proceso de localización resulta de especial interés, ya que, si la calidad de la traducción automática es la adecuada, nos permitirá reducir el tiempo requerido por este proceso y aumentar así la productividad.

El objetivo de este artículo es evaluar el rendimiento de diferentes motores de TAE en el ámbito de la localización de aplicaciones. Para ello, se hará una compilación de textos provenientes de *apps* y se crearán tres motores con diferentes composiciones de corpus. Los resultados del estudio se evaluarán desde la perspectiva de los problemas de traducción propios de la localización de aplicaciones, de carácter principalmente técnico e independientes de las lenguas de trabajo. Los resultados presentados en este artículo provienen del trabajo de fin de máster titulado «Entrenamiento y comparativa de motores de TAE especializados en la localización de aplicaciones móviles» (Fernández, 2019).[1]

2. Selección del corpus y creación de los motores

El objetivo de este estudio es comparar el rendimiento de tres motores TAE entrenados a partir de corpus específicos de aplicaciones, así como de corpus genéricos. El corpus monolingüe general se ha obtenido de la web OPUS (Tiedemann, 2012) y los corpus

[1] Màster Tradumàtica: Tecnologies de la Traducció. Universitat Autònoma de Barcelona, curso 2018-2019.

bilingües están formados por los archivos strings.xml originales y traducidos de distintas aplicaciones móviles disponibles en la Play Store de Google y en otros portales de descarga, así como por las cadenas de texto de las aplicaciones nativas de Android[2]. Nos hemos centrado en las aplicaciones de Android porque es el sistema operativo más usado a nivel mundial, con aproximadamente un 75 % de cuota de mercado según StatCounter (2019), y por la sencillez de su descarga. En la tabla 1 se resumen los corpus utilizados para la creación de los diferentes motores:

	Composición	N.º de líneas	N.º de palabras
Corpus genérico monolingüe en español	DGT: textos legislativos de la UE. EUROPARL v7: textos de las actas del Parlamento Europeo. GlobalVoices: noticias publicadas en el sitio web Global Voices. EUbookshop: documentos de la biblioteca de la UE.	14 534 589	333 619 923
Corpus bilingüe de aplicaciones	Archivos de texto originales en inglés y traducidos de todas las aplicaciones descargadas. A partir de este, se ha creado también un corpus monolingüe de *apps* en español.	211 768	en: 961 993 es: 1 096 769
Corpus bilingüe de aplicaciones de «comunicación»	Archivos de texto originales y traducidos de aplicaciones descargadas de la categoría «comunicación» (categorización propia según la temática de las *apps*; véase Fernández, 2019).	53 776	en: 239 331 es: 267 265

Tabla 1. Corpus utilizados en el entrenamiento de los motores

Las aplicaciones descargadas se han obtenido a través de portales de descarga de aplicaciones en formato APK (Apkpure[3] y Uptodown[4]). El formato APK se emplea para distribuir aplicaciones Android y consiste en un archivo comprimido que incluye todas las carpetas con los archivos compilados que forman una aplicación; por tanto, para poder acceder a los archivos y visualizar el contenido, hay que disponer de una

[2] Archivos disponibles en el repositorio de Google Git al que se puede acceder a través del siguiente enlace: https://android.googlesource.com/.
[3] APKPure: https://apkpure.com/es/.
[4] Uptodown: https://www.uptodown.com/android.

herramienta que nos permita descompilar los archivos APK. En este caso se ha empleado la herramienta Apktool (Wiśniewski & Tumbleson, 2019) para realizar esta tarea.

El siguiente paso ha sido la creación de los corpus paralelos en formato TMX, ya que este es compatible con la herramienta de entrenamiento MTradumàtica. La herramienta empleada para llevar a cabo la alineación ha sido la de Okapi Rainbow[5], ya que nos permite alinear gran cantidad de archivos al mismo tiempo mediante la opción de alineación basada en identificadores (ID-Based Alignment), lo que agiliza en gran medida el proceso, puesto que une las cadenas de texto de forma automática basándose en los identificadores. Además, esta herramienta cuenta con un filtro para archivos XML de Android.

Una vez creados los archivos TMX, se ha procedido a la limpieza de los mismos y se han eliminado todos los elementos innecesarios que no interesan en el entrenamiento de los motores, como segmentos vacíos, segmentos en otros idiomas, dobles espacios, espacios a comienzo y final de segmento, comillas que encierran algunos segmentos y etiquetas HTML. Para realizar esta limpieza, se han empleado las funciones de búsqueda y sustitución de la herramienta Okapi Olifant[6] y del editor de texto Notepad ++[7].

Para la realización de este estudio, se han entrenado tres motores de TAE con la herramienta MTradumàtica [8] (Martín-Mor y Piqué, 2017) a partir de los corpus mencionados anteriormente:

> **Motor 1.** Motor específico de aplicaciones, integrado por el corpus bilingüe completo que incluye todas las aplicaciones móviles que hemos recopilado y el corpus monolingüe en español de todas las aplicaciones.
>
> **Motor 2.** Este motor se centra en un tipo en concreto de aplicaciones; en este caso, hemos seleccionado la categoría «comunicación» para el entrenamiento mediante un corpus bilingüe formado por las aplicaciones de esta categoría repetido cuatro veces con el fin de aumentar su tamaño y el corpus monolingüe de todas las aplicaciones.
>
> **Motor 3**. Este motor está compuesto por el mismo corpus bilingüe que el primero, formado por todas las aplicaciones, más un corpus monolingüe general para la creación de un modelo de lengua genérico.

La interfaz de MTradumàtica es muy intuitiva y los pasos a seguir en el entrenamiento están detallados en la propia web. El proceso de entrenamiento se resume en los siguientes pasos:

1. Subida de los archivos
2. Creación de los monotextos (corpus monolingües)

[5] Okapi Rainbow está integrado en el paquete de aplicaciones de Okapi Framework. Se puede descargar en https://bintray.com/okapi/Distribution.
[6] Olifant es una herramienta de gestión de memorias de traducción. La última versión disponible (v3.0.8) se puede descargar aquí: http://okapi.sourceforge.net/downloads.html.
[7] Enlace de descarga: https://notepad-plus-plus.org/download/v7.7.1.html.
[8] MTradumàtica es un sistema de gestión de motores TAE desarrollado gracias a los proyectos de investigación ProjecTA (FFI2013-46041-R) y ProjecTA-U (FFI2016-78612-R), dirigidos por la Dra. Olga Torres Hostench y financiados por el Ministerio de Economía y Competitividad. Se puede acceder aquí: https://mtradumatica.uab.cat/.

3. Creación de los modelos de lengua
4. Creación de los bitextos (corpus bilingües)
5. Creación de los motores de traducción

A continuación, se ha seleccionado una aplicación para emplearla como texto original en el estudio. La aplicación elegida ha sido Signal[9], una *app* de mensajería instantánea y llamadas de código abierto. Para evaluar el rendimiento de los motores se ha seleccionado una muestra de segmentos representativa de los problemas habituales en la localización de aplicaciones. Finalmente, para obtener las traducciones de la muestra, hemos empleado la herramienta OmegaT, ya que cuenta con un filtro de archivos específico para archivos XML de aplicaciones Android y ofrece la posibilidad de añadir un motor de traducción automática de tipo Moses como el de MTradumàtica. Además, de este modo conseguimos integrar el estudio dentro del flujo real de trabajo y aplicar el entrenamiento de los motores a un escenario específico.

3. Análisis del rendimiento de los motores

Para comparar el rendimiento de los tres motores se han llevado a cabo dos tipos de análisis. Por un lado, se han extraído una serie de segmentos en relación a aspectos problemáticos de la localización de aplicaciones para Android y se ha evaluado su usabilidad, comparándolos con un motor existente, el de DeepL. Por otro lado, las traducciones de cada motor se han comparado a través de la función *ranking* de DQF[10] de TAUS. Se pueden consultar todos los segmentos y ejemplos del análisis en Fernández (2019). Cabe destacar que ambos análisis son métodos manuales de evaluación de la calidad, por lo que los resultados están sujetos a cierta subjetividad, ya que pueden variar en función de la persona que realiza la evaluación.

3.1 Usabilidad de las traducciones de cada motor

Los aspectos problemáticos de la localización de aplicaciones que hemos tenido en cuenta son i) la presencia de variables como %d o %1$d, ii) la presencia de apóstrofos escapados con contrabarra (don\'t), iii) la presencia de saltos de línea mediante la secuencia de escape \n, y iv) la traducción de segmentos de una cierta longitud (más de diez palabras). Se han comparado las traducciones de cada motor entre sí y también con la traducción que ofrece el motor de DeepL para analizar cómo aborda cada motor dichos aspectos problemáticos y qué problemas se presentan. En el caso de los problemas i), ii) y iii), se ha clasificado el resultado de cada motor en función de su usabilidad como punto de partida de la posedición. En concreto, se ha usado una escala de cuatro niveles para puntuarlos simulando las escalas de puntuación de la calidad de la traducción automática según el esfuerzo que requieren para ser editadas: 1. No aprovechable, 2. Poco aprovechable, 3. Bastante aprovechable, 4. Totalmente aprovechable. Para la traducción de oraciones de más de diez palabras se ha empleado el mismo baremo del 1 al 4, pero teniendo en cuenta adecuación y fluidez.

[9] Signal: https://signal.org/es/.
[10] DQF: https://dqf.taus.net/.

3.1.1 Segmentos con variables

Se han seleccionado un total de doce segmentos que incluyen variables para comprobar cómo aborda su traducción cada motor. En la siguiente tabla podemos ver los resultados:

Variables	Totalmente aprovechable	Bastante aprovechable	Poco aprovechable	No aprovechable
Motor especializado en aplicaciones de comunicación	0	8	3	1
Motor especializado en aplicaciones	0	10	1	1
Motor especializado en aplicaciones + genérico	0	8	2	2
DeepL	1	0	2	9

Tabla 1. Valoración de los segmentos seleccionados para el análisis de las variables

El motor que ha proporcionado unos mejores resultados ha sido el especializado en aplicaciones, mientras que el que peores resultados ha dado ha sido el de DeepL. Los dos restantes han obtenido también unos resultados bastante aceptables. El único problema recurrente que presentan los motores entrenados es que devuelven la variable con todos los elementos separados por espacios (% 1 $ d en lugar de %1$d), debido a que el motor no los reconoce como una unidad, sino como símbolos separados. Por su parte, el motor de DeepL sí que devuelve las variables sin espacios, pero tiende a realizar elecciones léxicas erróneas o a colocar las variables en lugares de la oración erróneos (p. ej., «This will permanently delete all %1$d selected messages» por «Esto eliminará permanentemente todos los mensajes seleccionados de %1$d»). Además, en un par de ocasiones, este motor ha omitido completamente las variables.

3.1.2 Segmentos con apóstrofos escapados

Se han seleccionado un total de cuatro segmentos para analizar y el motor con mejores resultados ha sido el de DeepL, cuyas traducciones de los cuatro segmentos han abordado perfectamente la presencia de apóstrofos escapados.

Apóstrofo escapado (\')	Totalmente aprovechable	Bastante aprovechable	Poco aprovechable	No aprovechable
Motor especializado en aplicaciones de comunicación	2	0	1	1

Motor especializado en aplicaciones	2	0	1	1
Motor especializado en aplicaciones + genérico	2	0	1	1
DeepL	4	0	0	0

Tabla 2. Valoración de los segmentos seleccionados para el análisis del apóstrofo escapado

Los tres motores especializados han obtenido las mismas puntuaciones. El resultado ha sido adecuado ante la presencia de palabras más comunes incluidas en los corpus como don\'t o doesn\'t. No obstante, cuando aparece el genitivo sajón (Signal\'s), los motores entrenados no han sabido identificarlo como tal, pues la contrabarra lo separa de la palabra a la que debería ir unido. Por su parte, el motor de DeepL sí que ha podido identificarlo y, en consecuencia, lo ha traducido de manera perfecta.

3.1.3 Segmentos con saltos de línea (\n)

Para analizar los saltos de línea hemos seleccionado cuatro segmentos. En este caso, también ha sido DeepL el motor que ha obtenido unos resultados más favorables.

Saltos de línea (\n)	Totalmente aprovechable	Bastante aprovechable	Poco aprovechable	No aprovechable
Motor especializado en aplicaciones de comunicación	0	0	0	4
Motor especializado en aplicaciones	0	0	0	4
Motor especializado en aplicaciones + genérico	0	0	0	4
DeepL	0	3	0	1

Tabla 3. Valoración de los segmentos seleccionados para los saltos de línea (\n)

Los malos resultados de los motores que hemos entrenado se deben a que no reconocen que \n es un elemento en sí mismo, pues se escribe pegado a la siguiente palabra. Por otro lado, DeepL tiene la desventaja de que ha eliminado las secuencias de salto de línea total o parcialmente y, en una ocasión, ha eliminado también la palabra posterior al salto de línea.

3.1.4 Oraciones largas

Para comprobar cómo aborda cada motor las oraciones de más de 10 palabras, hemos seleccionado cinco oraciones de mayor longitud. A continuación, podemos ver las tablas

con las puntuaciones de cada motor para adecuación y fluidez. En este caso, el acercamiento a estos dos aspectos se basa en la definición de adecuación y fluidez utilizada en la herramienta DQF de TAUS:

Oraciones largas (Fluidez)	Totalmente fluido	Bastante fluido	Poco fluido	Nada fluido
Motor especializado en aplicaciones de comunicación	1	0	1	3
Motor especializado en aplicaciones	2	0	1	2
Motor especializado en aplicaciones + genérico	1	2	2	0
DeepL	5	0	0	0

Tabla 4. Valoración de la fluidez en las oraciones de mayor extensión

Oraciones largas (Adecuación)	Totalmente adecuado	Bastante adecuado	Poco adecuado	Nada adecuado
Motor especializado en aplicaciones de comunicación	0	1	2	2
Motor especializado en aplicaciones	0	3	2	0
Motor especializado en aplicaciones + genérico	0	2	3	0
DeepL	4	1	0	0

Tabla 5. Valoración de la adecuación en las oraciones de mayor extensión

Aquí, como habíamos previsto, la presencia de oraciones de una cierta extensión hace que los resultados obtenidos en los tres motores entrenados con aplicaciones sean peores que los de DeepL, debido a que los corpus empleados en el entrenamiento están compuestos en gran medida de segmentos cortos, que en muchas ocasiones no constituyen siquiera una frase completa.

3.2 Clasificación de los motores entrenados

Una vez analizados los factores problemáticos, hemos llevado a cabo la clasificación de los motores mediante el método de comparación de motores (*ranking engine*), que consiste en mostrar a evaluadores humanos traducciones procedentes de distintos motores para que las clasifiquen de mejor a peor según un valor numérico. La

clasificación se ha realizado a través de la plataforma en línea DQF de TAUS con el fin de determinar qué composición de corpus es la más adecuada para el entrenamiento de un motor destinado a la traducción de aplicaciones móviles; es por este motivo que no se incluye el motor de DeepL en este análisis. Para ello hemos evaluado un total de 55 segmentos, a los que se ha otorgado una clasificación del 1 al 3, en la que 1 es el mejor resultado y 3, el peor. A partir de esto, la herramienta otorga una puntuación a cada segmento, suma todos los puntos y los transforma en porcentaje. Como podemos ver en la figura 1, el motor especializado en aplicaciones ha sido el que mejores resultados ha obtenido (36,8 %). En cuanto a los otros dos motores, ambos tienen un porcentaje similar.

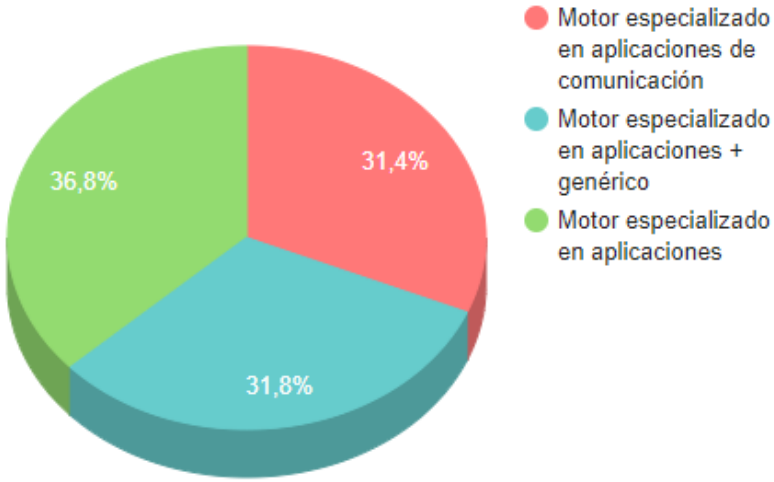

Figura 1. Gráfica de la clasificación de los motores entrenados

Además, como se puede observar en el siguiente diagrama de barras, el motor especializado en aplicaciones ha obtenido una clasificación de 1 (motor preferido) en más ocasiones, en torno a un 52 % de las valoraciones, y ha sido el que ha obtenido una clasificación de 3 en menos ocasiones (solo un 9 %).

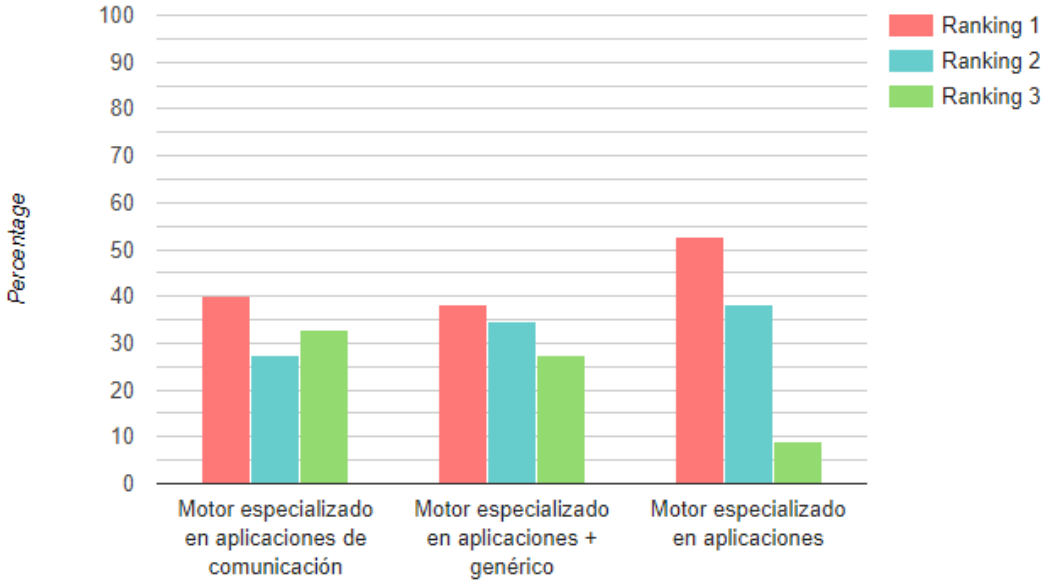

Figura 2. Gráfica con las valoraciones otorgadas a cada motor

3.3 Discusión de los resultados

Si bien es cierto que la cantidad de segmentos analizados con respecto a los aspectos problemáticos no ha sido muy elevada, este análisis nos permite obtener una imagen general de los motores entrenados y ver qué errores se nos pueden presentar, para así buscar soluciones y optimizarlos en el futuro.

Por otro lado, a la luz de los resultados del segundo análisis, podemos afirmar que el mejor rendimiento ha provenido de un motor entrenado con un corpus más voluminoso obtenido directamente de textos de todas las aplicaciones, a pesar de que se trate de aplicaciones sobre temas muy diversos.

4. Conclusiones

Mediante la realización de este estudio, hemos podido comprobar la utilidad del entrenamiento de motores de traducción personalizados según las necesidades específicas del usuario, ya sean empresas o traductores autónomos cuyo objetivo sea ampliar su abanico de recursos profesionales. Por ello, es de gran ayuda la existencia de herramientas de entrenamiento como MTradumàtica, que ofrecen la oportunidad a todo aquel que lo desee de entrenar su propio motor especializado, ya sea en un campo en concreto o en una tipología textual concreta, como ha sido el caso de este estudio. Estos sistemas facilitan la tarea de entrenamiento mediante una interfaz intuitiva y una serie de funcionalidades añadidas.

A la vista de los resultados, podemos afirmar que el motor especializado en aplicaciones sería la composición ideal de corpus para continuar con el entrenamiento de un motor especializado en la traducción de aplicaciones móviles si tuviera un volumen y una diversidad suficientes como para alcanzar un grado superior de fluidez. Por otro lado, a pesar de que los otros dos motores (especializado + genérico y especializado en aplicaciones de comunicación) han obtenido unos resultados similares, el motor especializado en aplicaciones con modelo de lengua genérico también ha arrojado unos resultados que no esperábamos, pues creíamos que, con la incorporación de un corpus genérico monolingüe al corpus especializado de aplicaciones, conseguiríamos aumentar la fluidez del texto y obtendríamos una mejor puntuación en la evaluación. La incorporación del corpus monolingüe al motor no solo no ha logrado mantener una calidad similar a la del motor especializado en aplicaciones, sino que incluso parece haber introducido ruido y ha acabado produciendo un rendimiento inferior. De este modo, lo que pensábamos que sería un componente que mejoraría el resultado del motor ha resultado ser un inconveniente.

Por su parte, el motor especializado en aplicaciones de comunicación ha obtenido peores resultados que el especializado en todo tipo de aplicaciones. Esto puede deberse al hecho de que el corpus empleado para su entrenamiento fuera más reducido y, aunque se ha intentado suplir esta falta de volumen replicándolo, esto no ha sido suficiente para conseguir unos mejores resultados al traducir una aplicación de la misma temática.

En cuanto a los elementos problemáticos, podemos afirmar que los motores entrenados han resuelto de manera adecuada la presencia de variables, pero presentan problemas a la hora de abordar los saltos de línea, los apóstrofos escapados y las oraciones largas. Por otro lado, hemos podido comprobar que el motor de DeepL sí que ha solucionado estos tres problemas de manera adecuada, resultados que han servido para analizar las carencias de los motores entrenados, o ventaja en el caso de las variables. Asimismo, ciertos problemas que presentan los motores entrenados podrían llegar a solucionarse añadiendo funcionalidades a la herramienta de entrenamiento o herramientas TAO empleadas: en el caso de los saltos de línea (\n), por ejemplo, sería interesante poder indicarle a la herramienta de entrenamiento que \n es un elemento independiente que debe ir separado de la palabra a la que va unida la letra *n*, para así evitar que dicha palabra quede sin traducir.

Referencias

Fernández-Ruiz, M. E. (2019). *Entrenamiento y comparativa de motores de TAE especializados en la localización de aplicaciones móviles* [Trabajo de fin de máster, supervisado por la Dra. Pilar Sánchez Gijón]. Universitat Autònoma de Barcelona. Bellaterra (Cerdanyola del Vallès). <https://ddd.uab.cat/pub/trerecpro/2019/hdl_2072_359326/Fernandez_Ruiz_MariaEsperanza_TFM.pdf>

Fundación Telefónica. (2019). *Sociedad Digital en España 2018*. Barcelona: Penguin Random House Grupo Editorial. <https://www.fundaciontelefonica.com/arte_cultura/publicaciones-listado/pagina-item-publicaciones/itempubli/655/>

Martín-Mor, A.; Piqué, R. (2017). MTradumàtica i la formació de traductors en Traducció Automàtica Estadística. *Revista Tradumàtica: tecnolgies de la traducció,* n. 15, pp. 97-115. <https://revistes.uab.cat/tradumatica/article/view/n15-martin-pique>. <https://doi.org/10.5565/rev/tradumatica.199>

Simón, E. (2016). A general view of the localization of apps for mobile devices: status, challenges and trends: Formats and customary processes in the translation of iOS and Android apps. *Revista Tradumàtica: tecnologies de la traducció,* n. 14, pp. 5-15. <https://ddd.uab.cat/pub/tradumatica/tradumatica_a2016n14/tradumatica_a2016n14p5.pdf>. <https://doi.org/10.5565/rev/tradumatica.174>

StatCounter. (2019). *Mobile Operating System Market Share Worldwide*. November 2019. <http://gs.statcounter.com/os-market-share/mobile/worldwide>

Tiedemann, J. (2012). Parallel Data, Tools and Interfaces in OPUS. *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012).* <http://opus.nlpl.eu/>

Wiśniewski, R.; Tumbleson, C. (2019). *Apktool (v2.4.0)* [Software]. https://ibotpeaches.github.io/Apktool/

Anexo 1. Aplicaciones descargadas

App	Versión	Descripción	Desarrollador	Contenido
Android app Settings	Android 9.0.0_r38	Ajustes	Open Handset Alliance / Google LLC	Herramientas
Android app AccountsAndSyncSettings	Android 2.2.3_r2	Sincronización de cuentas	Open Handset Alliance / Google LLC	Herramientas
Android app AlarmClock	Android 2.2.3_r2.1	Alarma	Open Handset Alliance / Google LLC	Servicios/estilo de vida
Android app Bluetooth	Android 9.0.0_r38	Bluetooth	Open Handset Alliance / Google LLC	Herramientas
Android app Browser	Android 6.0.1_r81	Buscador web	Open Handset Alliance / Google LLC	Herramientas
Android app Calendar	Android 9.0.0_r38	Calendario	Open Handset Alliance / Google LLC	Productividad
Android app Camera	Android 6.0.1_r81	Cámara	Open Handset Alliance / Google LLC	Multimedia
Android app Camera2	Android 9.0.0_r38	Cámara	Open Handset Alliance / Google LLC	Multimedia
Android app CertInstaller	Android 9.0.0_r38	Instalación de certificados	Open Handset	Herramientas

			Alliance / Google LLC	
Android app Contacts	Android 9.0.0_r38	Contactos	Open Handset Alliance / Google LLC	Comunicación
Android app ContactsCommon	Android 7.0.0_r36	Contactos	Open Handset Alliance / Google LLC	Comunicación
Android app DeskClock	Android 9.0.0_r38	Alarma	Open Handset Alliance / Google LLC	Servicios/estilo de vida
Android app DocumentsUI	Android 9.0.0_r38	Gestión de archivos	Open Handset Alliance / Google LLC	Herramientas
Android app Email	Android 9.0.0_r38	Correo electrónico	Open Handset Alliance / Google LLC	Comunicación
Android app EmergencyInfo	Android 9.0.0_r38	Información de emergencia	Open Handset Alliance / Google LLC	Servicios/estilo de vida
Android app ExactCalculator	Android 9.0.0_r38	Calculadora	Open Handset Alliance / Google LLC	Herramientas
Android app FMRadio	Android 6.0.1_r81	Radio	Open Handset Alliance / Google LLC	Multimedia
Android app Gallery	Android 9.0.0_r38	Galería multimedia	Open Handset Alliance / Google LLC	Multimedia

Android app Gallery2	Android 9.0.0_r38	Galería multimedia	Open Handset Alliance / Google LLC	Multimedia
Android app Gallery3D	Android 2.2.3_r2	Galería multimedia	Open Handset Alliance / Google LLC	Multimedia
Android app IM	Android 2.2.3_r2.1	Mensajes de texto	Open Handset Alliance / Google LLC	Comunicación
Android app InCallUI	Android 6.0.1_r81	Llamadas	Open Handset Alliance / Google LLC	Comunicación
Android app Launcher	Android 2.1_r2.1p2	Lanzador	Open Handset Alliance / Google LLC	Herramientas
Android app Launcher2	Android 9.0.0_r38	Lanzador	Open Handset Alliance / Google LLC	Herramientas
Android app Launcher3	Android 9.0.0_r38	Lanzador	Open Handset Alliance / Google LLC	Herramientas
Android app LegacyCamera	Android 9.0.0_r38	Cámara	Open Handset Alliance / Google LLC	Multimedia
Android app Messaging	Android 9.0.0_r38	Mensajes de texto	Open Handset Alliance / Google LLC	Comunicación
Android app Mms	Android 5.1.1_r38	Mensajes multimedia	Open Handset	Comunicación

			Alliance / Google LLC	
Android app Music	Android 9.0.0_r38	Reproducción de música	Open Handset Alliance / Google LLC	Multimedia
Android app Nfc	Android 9.0.0_r38	Conexión NFC	Open Handset Alliance / Google LLC	Herramientas
Android app PackageInstaller	Android 9.0.0_r38	Instalación de aplicaciones	Open Handset Alliance / Google LLC	Herramientas
Android app PhoneCommon	Android 9.0.0_r38	Llamadas	Open Handset Alliance / Google LLC	Comunicación
Android app SoundRecorder	Android 9.0.0_r38	Grabación de sonido	Open Handset Alliance / Google LLC	Multimedia
Android app StorageManager	Android 9.0.0_r38	Gestión de archivos	Open Handset Alliance / Google LLC	Herramientas
Android app TV	Android 9.0.0_r38	Control de televisión	Open Handset Alliance / Google LLC	Multimedia
Android app UnifiedEmail	Android 9.0.0_r38	Gestión de cuentas de correo electrónico	Open Handset Alliance / Google LLC	Comunicación
Android app VideoEditor	Android 4.4.4_r2.0.1	Editor de vídeos	Open Handset Alliance / Google LLC	Multimedia

YouTube	14.05.53	Reproducción y compartición de vídeos	Google LLC	Multimedia
WhatsApp Messenger	2.19.31	Servicio de mensajería y voz IP	WhatsApp Inc.	Comunicación
Telegram	5.3.0	Servicio de mensajería y voz IP	Telegram LLC	Comunicación
Snapseed	2.19.0.201907232	Edición de fotos	Google LLC	Multimedia
Calendario Menstrual - Fertilidad y ovulación	1.653.193	Calendario menstrual	Simple Design Ltd.	Servicios/estilo de vida
Nova Launcher	5.5.4	Lanzador	TeslaCoil Software	Herramientas
Instagram	79.0.0.21.101	Red social	Instagram	Comunicación
Google Fotos	4.8.0.229992926	Gestor de fotos	Google LLC	Multimedia
Google Docs	1.19.032.03.35	Creación y edición de documentos	Google LLC	Productividad
Google Calendar	6.0.20-231367491-release	Calendario/agenda	Google LLC	Productividad
Google App	9.5.3.21.arm	Buscador web	Google LLC	Herramientas
Gmail	8.12.30.230564275.release	Servicio de correo electrónico	Google LLC	Comunicación
Firefox	64.0.2	Navegador web	Mozilla Corporation	Herramientas
Aliexpress	7.1.0	Compras en línea	Alibaba Mobile	Servicios/estilo de vida
Avg Antivirus Free	6.16.4	Antivirus	AVG Mobile	Herramientas

Clean Master	7.0.1	Herramienta de limpieza	Cheetah Mobile	Herramientas
Discord-Chat for gamers	8.3.7	Servicio de mensajería para *gamers*	Discord Inc.	Comunicación
Amazon shopping	18.4.0.100	Compras	Amazon Mobile LLC	Servicios/estilo de vida
TripAdvisor	29.4.1	Viajes	TripAdvisor	Servicios/estilo de vida
Kindle	8.15.0.68	Lectura	Amazon Mobile LLC	Productividad
Spotify	8.4.94.817	Servicio de música en *streaming*	Spotify Ltd.	Multimedia
SwiftKey Keyboard	7.2.3.24	Teclado	SwiftKey	Herramientas
Trello	5.5.0.11966	Aplicación de notas colaborativa	Trello, Inc.	Productividad
Airdroid	4.2.1.12	Acceso remoto y transferencia de archivos	SAND STUDIO	Herramientas
Shareit	4.6.88_ww	Transferencia de archivos	SHAREit Technologies Co. Ltd	Herramientas
Booking.com	17.0	Viajes	Booking.com	Servicios/estilo de vida
Netflix	6.26.0 build 13 31685	Películas y series en *streaming*	Netflix, Inc.	Multimedia
Kinemaster	4.8.13.12545.GP	Edición de vídeos	NexStreaming Corp	Multimedia
Wish	4.26.5	Compras en línea	Wish Inc.	Servicios/estilo de vida

Google Drive	2.19.072.05.34	Almacenamiento en la nube	Google LLC	Productividad
Google Sheets	1.19.052.01.35	Creación y edición de hojas de cálculo	Google LLC	Productividad
WPS Office	11.4.4	Creación y edición de documentos	Kingsoft Office Software Corporation Limited	Productividad
TikTok	9.9.0	Red social de vídeos cortos	musical.ly	Comunicación
Fitbit	2.87	Deporte y vida sana	Fitbit, Inc.	Servicios/estilo de vida
Mx Player	1.10.44	Reproductor de vídeo	J2 Interactive	Multimedia
CamScanner - PDF creator	5.9.2.20190221	Escaneo y conversión de documentos a PDF	IntSig Information Co.,Ltd.	Productividad
Avast Cleanup	4.12.0	Herramienta de limpieza	Avast Software	Herramientas
Avast Battery Saber	2.8.1	Ahorro de batería	Avast Sofware	Herramientas
Adobe Illustrator Draw	3.5.1	Creación de diseños vectoriales	Adobe	Multimedia
File Commander	5.2.30001	Gestión de archivos	MobiSystems	Herramientas
Microsoft Outlook	3.0.14	Servicio de correo electrónico	Microsoft Corporation	Comunicación
Microsoft Powerpoint	16.0.11328.20080	Creación y edición de	Microsoft Corporation	Productividad

		presentaciones		
Evernote	8.8.1	Notas	Evernote Corporation	Productividad
Eventbrite	6.4.0	Gestión de eventos	Eventbrite	Servicios/estilo de vida
Deezer	6.0.6.79	Servicio de música en *streaming*	Deezer Mobile	Multimedia
Duolingo	4.3.1	Aprendizaje de idiomas	Duolingo	Productividad
Opera Browser	50.2.2426.136249	Navegador web	Opera	Herramientas
Pinterest	7.5.0	Red social para compartir imágenes	Pinterest	Comunicación
Adobe Photoshop Express	5.9.571	Edición de imágenes	Adobe	Multimedia
Flipboard	4.2.8	Noticias	Flipboard	Servicios/estilo de vida
Snapchat	10.51.0.0	Red social de mensajería multimedia	Snap Inc	Comunicación
FindNow	0.6.20	Ubicaciones en tiempo real	Ratech	Servicios/estilo de vida
Android Auto	4.0.590433-release	Asistente de viaje	Google LLC	Servicios/estilo de vida
Vlc for Android	3.0.13	Reproductor multimedia	Videolabs	Multimedia
Dropbox	130.2.6	Almacenamiento en la nube	Dopbox, Inc.	Productividad
Gboard	7.8.8.224901760	Teclado	Google LLC	Herramientas

Solid Explorer File Manager	2.6.1	Gestión de archivos	NeatBytes	Herramientas
Avast Mobile Antivirus	6.16.0	Antivirus	Avast Software	Herramientas
LastPass Password Manager	4.8.3632	Gestión de contraseñas	LogMeIn, Inc.	Herramientas
Twitter	7.83.0-release.33	Red social	Twitter, Inc.	Comunicación
Uber Eats	118.310.001	Comida a domicilio	Uber Technologies, Inc.	Servicios/estilo de vida
eBay	5.28.1.1	Compras	eBay Mobile	Servicios/estilo de vida
LinkedIn	4.1.236	Red social profesional	LinkedIn	Comunicación
Ikea Store	2.9.3	Compras	IKEA	Servicios/estilo de vida
Tinder	10.4.2	Aplicación de citas	Tinder	Comunicación
Sketch	8.4.A.3.2	Dibujo y edición de fotos	Sony Mobile Communications	Multimedia
Adobe Acrobat Reader	19.0.0.8512	Lector de PDF	Adobe	Productividad
Google Family Link	1.33.0.I.232598558	Control parental	Google LLC	Herramientas
YouTube Music	3.03.55	Servicio de música en *streaming*	Google LLC	Multimedia
TV Ttime	7.5.0.19021901	Seguimiento de series de TV	Toze Labs	Servicios/estilo de vida
Alarmy	4.7.1	Alarma	Alarmy	Servicios/estilo de vida

21 Buttons	4.7.3	Red social de moda	21 Buttons	Servicios/estilo de vida
Samsung Smart Switch Mobile	3.5.03.7	Transferencia de datos entre dispositivos	Samsung Electronics Co., Ltd.	Herramientas
Samsung Smart View	2.1.0.107	Transferencia de contenido multimedia a la TV	Samsung Electronics Co., Ltd.	Herramientas
Uber	425.010.001	Aplicación de transporte para particulares	Uber Technologies, Inc.	Servicios/estilo de vida
Huawei Health	9.0.4.332	Salud y deporte	Huawei Internet Service	Servicios/estilo de vida
Calls Blacklist	3.2.36	Bloqueo de llamadas y SMS	Vlad Lee	Comunicación
DU Recorder	1.7.9.7	Grabadora de pantalla	DU Recorder Team	Multimedia
Microsoft Edge	42.0.22.3333	Navegador web	Microsoft Corporation	Herramientas
Shazam	9.20.0.190215	Reconocimiento de canciones	Apple, Inc.	Multimedia
Amazon Prime Video	3.0.242.14741	Series y películas en *streaming*	Amazon Mobile LLC	Multimedia
Deliveroo	2.59.1	Comida a domicilio	Deliveroo	Servicios/estilo de vida
SmartThings	1.7.27-25	Control de dispositivos (domótica)	Samsung Electronics Co., Ltd.	Servicios/estilo de vida

Mi Remote	5.7.2	Mando para aparatos eléctricos	Xiaomi, Inc,	Herramientas
Google Duo	47.1.234325686.DR47_RC14	Videollamadas	Google LLC	Comunicación
File Manager - free and easily	V1-190128	Gestión de archivos	Xiaomi, Inc,	Herramientas
Google Allo	27.0.326_rc03 (armeabi-v7a_xxhdpi)	Servicio de mensajería	Google LLC	Comunicación
Google Home	2.9.40.16	Control de dispositivos (domótica)	Google LLC	Servicios/estilo de vida
McAfee-Mobile Security	5.2.0.152	Antivirus	McAfee LLC	Herramientas
Viber Messenger	10.1.0.1	Servicio de mensajería y llamadas	Viber Media SARL	Comunicación