
This is the **published version** of the article:

Fernández Ruiz, María Esperanza; Sánchez-Gijón, Pilar, tut. Entrenamiento y comparación de motores de TAE especializados en la localización de aplicaciones móviles. 2019. 85 p.

This version is available at <https://ddd.uab.cat/record/210869>

under the terms of the  license

Máster en Tradumática: Tecnologías de la Traducción

**ENTRENAMIENTO Y COMPARATIVA
DE MOTORES DE TAE ESPECIALIZADOS
EN LA LOCALIZACIÓN DE
APLICACIONES MÓVILES**

TRABAJO DE FIN DE MÁSTER

Curso 2018-2019

María Esperanza Fernández Ruiz

Tutora: Pilar Sánchez Gijón

Facultad de Traducción e Interpretación, UAB

Resumen:

En este trabajo se entrenan, mediante la plataforma MTradumàtica, tres motores de traducción automática estadística con corpus de aplicaciones móviles de distinta composición para la combinación lingüística inglés-español. Los motores creados son los siguientes: uno especializado en aplicaciones, otro especializado en aplicaciones de la categoría «comunicación» y otro especializado en aplicaciones con un corpus monolingüe general para entrenar el modelo de lengua. De este modo, se exponen las cuestiones teóricas consideradas relevantes para el estudio como son la localización de aplicaciones móviles, el entrenamiento de motores de traducción automática mediante corpus, la posesición y la evaluación de la calidad de la traducción automática. Asimismo, se explica detalladamente el proceso que se ha llevado a cabo durante el entrenamiento de los tres motores. Una vez hecho esto, se ponen a prueba los motores para evaluar los resultados: por un lado, se presta atención a una serie de aspectos problemáticos específicos de la localización de aplicaciones móviles y se comparan los resultados con los de otro motor existente; y, por otro lado, se comparan entre sí los tres motores para establecer qué composición de corpus sería la más adecuada.

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

Abstract:

In this dissertation, three statistical machine translation engines with corpus of mobile applications of different compositions are trained for the English-Spanish language pair using the MTradumàtica platform. The engines created are the following: one engine specialized in all apps, another specialized in “communication” apps, and another one specialized in all apps with a general monolingual corpus in order to train the language model. Therefore, the theoretical bases considered relevant for this research are presented: the localization of mobile applications, the training process of corpus-based machine translation engines, post-editing and the quality evaluation of machine translation. In addition, the process that has been carried out during the training of these three engines is explained in detail. Following this, the engines are tested to evaluate their results: on the one hand, attention is paid to a series of problematic aspects specific to mobile apps localisation and the results are compared with those of an existing engine; and, on the other hand, the three engines are compared in order to establish which corpus composition would be the most appropriate.

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

Resum:

En aquest treball s'entrenen, mitjançant la plataforma MTradumàtica, tres motors de traducció automàtica estadística amb corpus d'aplicacions mòbils de diferent composició per la combinació lingüística anglès-espanyol. Els motors creats són els següents: un motor especialitzat en aplicacions, un altre especialitzat en aplicacions de la categoria «comunicació» i un altre especialitzat en aplicacions amb un corpus monolingüe general per entrenar el model de llengua. D'aquesta manera, s'exposen les qüestions teòriques considerades rellevants per a l'estudi com són la localització d'aplicacions mòbils, l'entrenament de motors de traducció automàtiques mitjançant corpus, la posició i l'avaluació de la qualitat de la traducció automàtica. Així mateix, s'explica detalladament el procés que s'ha dut a terme durant l'entrenament dels tres motors. Un cop fet això, es posen a prova els motors per avaluar els resultats: d'una banda, es presta atenció a una sèrie d'aspectes problemàtics específics de la localització d'aplicacions mòbils i es comparen els resultats amb els d'un altre motor existent; i, d'altra banda, es comparen entre si els tres motors per establir quina composició de corpus seria la més adequada.

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

Índice de contenidos

1.	Introducción	1
2.	Objetivos: preguntas e hipótesis.....	2
3.	Marco teórico y antecedentes	3
3.1.	La localización de aplicaciones móviles.....	4
3.2.	Aspectos técnicos de la localización de aplicaciones	7
3.2.1.	Las aplicaciones para Android	10
3.2.2.	Las aplicaciones para iOS	13
3.3.	Aspectos lingüísticos y culturales de la localización de aplicaciones ..	14
3.4.	Entrenamiento de motores de TAE.....	17
3.5.	Posedición	21
3.6.	La calidad de la traducción automática.....	24
4.	Metodología	27
4.1.	Motores de TAE creados	28
4.2.	Búsqueda de recursos.....	30
4.3.	Creación de los corpus paralelos	34
4.4.	Limpieza de las memorias de traducción.....	36
4.5.	Creación de los motores mediante MTradumàtica	39
4.6.	Elección de la muestra para realizar el análisis	43
5.	Análisis de los resultados	47
5.1.	Aspectos problemáticos	47
5.2.	Clasificación de los motores entrenados.....	52
6.	Conclusiones	54
7.	Bibliografía.....	57
8.	Anexos.....	62
	Anexo 1. Composición de corpus	62

Anexo 2. Segmentos extraídos de la aplicación procesada por los motores .. 74

Índice de figuras

FIGURA 1. ECOSISTEMA DE UNA APLICACIÓN SEGÚN ROTURIER (2015, PÁG. 2).....	5
FIGURA 2. ARCHIVO STRINGS.XML DE LA APLICACIÓN DE TELEGRAM PARA ANDROID.....	9
FIGURA 3. EJEMPLO DE ARCHIVO ARRAYS.XML EXTRAÍDO DE LA GUÍA DE LOCALIZACIÓN DE ANDROID.....	11
FIGURA 4. ARCHIVO LOCALIZABLE.STRINGS DE LA APLICACIÓN TELEGRAM PARA IOS.....	14
FIGURA 5. EJEMPLO DE USO DE VARIABLE EN UNA CADENA DE TEXTO DE LA APLICACIÓN TELEGRAM.....	15
FIGURA 6. ARCHIVO RESULTANTE DE LA DESCOMPRESIÓN DE UNA APLICACIÓN EN FORMATO APK.....	32
FIGURA 7. EJEMPLO DE LA ESTRUCTURA DE CARPETAS QUE SE CREA AL DESCOMPILAR EL ARCHIVO APK DE UNA APLICACIÓN.....	33
FIGURA 8. INTERFAZ DE OKAPI RAINBOW CON LOS ARCHIVOS DE LAS APLICACIONES DE «COMUNICACIÓN».....	35
FIGURA 9. MEMORIAS DE TRADUCCIÓN EN FORMATO TMX CREADAS TRAS LA ALINEACIÓN.....	36
FIGURA 10. VISTA DEL PROGRAMA OKAPI OLIFANT AL ABRIR UNA MEMORIA TMX.....	37
FIGURA 11. SEPARACIÓN DE LOS ARCHIVOS TMX EN EL APARTADO FICHEROS DE MTRADUMÀTICA.....	40
FIGURA 12. MONOTEXTO CREADO PARA EL MOTOR DE TRADUCCIÓN ESPECÍFICO DE LA TIPOLOGÍA TEXTUAL.....	41
FIGURA 13. MOTORES DE TAE CREADOS CON MTRADUMÀTICA.....	42
FIGURA 14. CONFIGURACIÓN DE MTRADUMÀTICA EN OMEGAT.....	46
FIGURA 15. GRÁFICA DE LA CLASIFICACIÓN DE LOS MOTORES ENTRENADOS.....	52
FIGURA 16 . GRÁFICA CON LAS VALORACIONES OTORGADAS A CADA MOTOR.....	53

Índice de tablas

TABLA 1. CARACTERES QUE DEBEN ESCAPARSE EN LOS ARCHIVOS .XML DE ANDROID EXTRAÍDOS DE LA GUÍA DE LOCALIZACIÓN DE ANDROID.....	12
TABLA 2. TABLA CON LA CLASIFICACIÓN DE LAS APLICACIONES QUE INTEGRAN EL CORPUS.....	29
TABLA 3. TABLA DE LA COMPOSICIÓN DE CADA UNO DE LOS CORPUS TRAS EL PROCESO DE LIMPIEZA.....	39
TABLA 4. TABLA DESCRIPTIVA DEL MOTOR ESPECIALIZADO EN APLICACIONES + GENÉRICO.....	43
TABLA 5. TABLA DESCRIPTIVA DEL MOTOR ESPECIALIZADO EN APLICACIONES.....	43
TABLA 6. TABLA DESCRIPTIVA DEL MOTOR ESPECIALIZADO EN APLICACIONES DE COMUNICACIÓN.....	43
TABLA 7. PUNTUACIÓN DE LOS SEGMENTOS SELECCIONADOS PARA EL ANÁLISIS DE LAS VARIABLES.....	47
TABLA 8. PUNTUACIÓN DE LOS SEGMENTOS SELECCIONADOS PARA EL ANÁLISIS DEL APÓSTROFO ESCAPADO.....	49
TABLA 9. PUNTUACIÓN DE LOS SEGMENTOS SELECCIONADOS PARA LOS SALTOS DE LÍNEA (\N).....	50
TABLA 10. PUNTUACIÓN DE LA FLUIDEZ EN LAS ORACIONES DE MAYOR EXTENSIÓN.....	51
TABLA 11. PUNTUACIÓN DE LA ADECUACIÓN EN LAS ORACIONES DE MAYOR EXTENSIÓN.....	51

1. Introducción

Hoy en día, pocas son las personas que no cuentan con un teléfono móvil inteligente, una tableta o cualquier otro dispositivo que integre aplicaciones móviles o, como también se las conoce, *apps*. De hecho, para realizar muchas de nuestras actividades diarias hacemos uso de estas aplicaciones, por ejemplo, para poner una alarma, enviar un mensaje a un conocido o intercambiar dinero con amigos. Los dispositivos móviles cada vez van ganando más fuerza, ya que se pueden usar en cualquier lugar y, por lo general, siempre los llevamos encima, de ahí que el uso de este tipo de dispositivos sea incluso mayor que el de los ordenadores de sobremesa o los portátiles, algo que no ocurría en el pasado, cuando los teléfonos móviles ni siquiera disponían de conexión a internet. Pero no solo se hace un uso personal de las aplicaciones, sino que también instituciones pertenecientes a campos como la educación, la salud o el deporte las emplean con diversos fines.

Para hacernos una idea del volumen e importancia que esto implica, 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 estas aplicaciones han tenido que localizarse previamente para que podamos usarlas en nuestra propia lengua, pues el idioma por excelencia empleado para su desarrollo es el inglés, al igual que ocurre con la localización de software. Por este motivo, la localización de estas aplicaciones dentro del mercado de la traducción tiene una gran relevancia y constituye una oportunidad de negocio tanto para traductores como para empresas proveedoras de servicios lingüísticos.

El número de aplicaciones diseñadas para dispositivos móviles que existen en el mercado está en constante cambio, pues hay aplicaciones que nacen y otras que desaparecen. Algunas de estas aplicaciones ya vienen preinstaladas en los dispositivos que compramos y otras las vamos añadiendo o borrando según nuestras necesidades y preferencias con el paso del tiempo. En cuanto a los plazos de entrega de los que suele disponer para su traducción 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. Desde el punto de vista del desarrollador de la aplicación, lo que se pretende es que la traducción esté lista lo antes posible, por tanto, la prospección de la

integración de la traducción automática al proceso de localización resulta de especial interés, ya que, si la calidad de la traducción automática es la adecuada, nos permitirá reducir el tiempo necesario para llevar a cabo este proceso y aumentar así la productividad.

2. Objetivos: preguntas e hipótesis

Mediante la elaboración de este trabajo se pretende evaluar la calidad de distintos motores de traducción automática estadística de distinta composición para la traducción de textos perteneciente al género textual de las aplicaciones móviles, centrándonos en la forma en la que soluciona cada uno de ellos una serie de aspectos problemáticos establecidos con anterioridad y comparándolos entre sí. Al tratarse de un género tan heterogéneo que puede incluir desde texto genérico hasta texto especializado en función del tipo de aplicación, se nos plantean una serie de preguntas a la hora de abordar el entrenamiento de un motor de traducción que ofrezca un resultado adecuado, independientemente del tipo de aplicación que se pretenda traducir.

El objetivo en este caso será averiguar si mediante el entrenamiento de motores de traducción automática se pueden lograr unos resultados aceptables a través del entrenamiento de un motor de traducción automática estadística inglés>español, en base a un corpus creado con aplicaciones de distinto tipo al alcance del traductor, con el fin de llevar a cabo la traducción de un tipo concreto de aplicaciones móviles. Para ello, gracias a la herramienta de entrenamiento de motores de traducción automática MTradumàtica se crearán varios motores de traducción: un motor especializado en la traducción de aplicaciones de cualquier tipo, otro especializado en la traducción de aplicaciones de la categoría comunicación y otro especializado en la traducción de aplicaciones que también incluye un corpus genérico para entrenar el motor de lengua. Mediante la elección de una aplicación de muestra y la selección de las cadenas que se utilizarán en la evaluación en función de los parámetros establecidos, se evaluará y comparará el resultado que ofrezca cada uno de ellos.

La hipótesis de la que partimos en este estudio es que el entrenamiento de un motor de traducción automática estadística a partir de un corpus formado con aplicaciones móviles de composición heterogénea no será suficiente para alcanzar el nivel de calidad esperado y que, desde un punto de vista productivo, la carga de posesición será mayor;

todo esto teniendo en cuenta las restricciones y especificidades que comporta la localización de aplicaciones móviles. Por este motivo, también se entrenará un motor específico de un tipo de aplicaciones y otro también específico, pero con el añadido de un corpus genérico monolingüe para observar si ayuda a mejorar la fluidez de la traducción.

Además, las aplicaciones móviles tienen una serie de características propias de esta tipología textual que plantean problemas a la hora de abordar su traducción mediante traducción automática, por ello, se establecerán aspectos problemáticos para ver cómo aborda cada motor la aparición de dichos elementos en la traducción y se compararán con los resultados de otro motor de traducción, el de DeepL, con el fin de determinar si un motor entrenado específicamente con textos de aplicaciones puede otorgar mejores resultados para estos elementos problemáticos característicos de la tipología. Asimismo, se valorará qué composición de corpus de las seleccionadas sería la más adecuada para entrenar un motor destinado a la traducción de aplicaciones móviles.

3. Marco teórico y antecedentes

En esta sección, nos adentraremos en la parte más teórica de la localización de aplicaciones móviles y de la traducción automática.

En primer lugar, en el apartado 3.1, con el fin de tener claras las particularidades de la localización de aplicaciones móviles, explicaremos la situación de la misma dentro del mercado de la localización y las diferencias que existen entre aplicaciones móviles y de escritorio, también conocidas como software o programas informáticos. Asimismo, en los apartados 3.2 y 3.3, indagaremos en los aspectos técnicos de este tipo de localización y los aspectos lingüísticos que debemos tener en cuenta a la hora de llevar a cabo la localización de una aplicación móvil.

A continuación, en el apartado 3.4, nos centraremos en la creación y entrenamiento de motores de traducción automática estadística (de aquí en adelante TAE) con corpus específicos para la traducción de ciertos tipos de textos. En el apartado 3.5 introduciremos el concepto de posesición y su relevancia dentro de la industria actual de la traducción. Y, finalmente en el apartado 3.6, hablaremos sobre el concepto de calidad en la traducción automática y los distintos métodos de medición que se pueden utilizar para evaluarla.

3.1. La localización de aplicaciones móviles

El crecimiento en el uso de teléfonos móviles y, con ello, en el uso de aplicaciones creadas para estos dispositivos ha hecho que la traducción y la localización de aplicaciones haya crecido durante los últimos años hasta convertirse en una oportunidad de mercado para proveedores de servicios lingüísticos y traductores que pretenden encontrar un nicho en este sector. Este aumento se debe a la globalización del mercado de las aplicaciones móviles, ya que cualquier persona tiene acceso a millones de aplicaciones creadas por desarrolladores de cualquier parte del mundo y desarrolladas en un idioma que no es el suyo propio. Esta diferencia lingüística y cultural entre el desarrollador y el usuario hace necesaria la localización de las aplicaciones, es decir, se necesita adaptar el producto a un mercado en concreto para que pueda llegar a un público mayor.

Pero este mercado de las aplicaciones no afecta solo a los teléfonos inteligentes o tabletas, a pesar de que sean los dispositivos más empleados y extendidos entre el público, pues el abanico de productos a los que tenemos acceso y que usan también aplicaciones de este tipo es muy amplio. De hecho, en los últimos años han ganado bastante popularidad relojes, pulseras, televisiones, altavoces o incluso objetos cotidianos inteligentes que han contribuido significativamente al aumento en el uso de estas aplicaciones que precisan de un proceso de localización. No obstante, el dispositivo más usado por el público sigue siendo el *smartphone*. Según el informe de Sociedad Digital en España 2018 de la Fundación Telefónica (2019), el dispositivo más utilizado para cualquier propósito por los españoles es el teléfono móvil inteligente, independientemente de la franja de edad de los usuarios: «El acceso a internet se ha hecho progresivamente más móvil y es posible hacerlo desde diferentes lugares. En consecuencia, el *smartphone* se mantiene como el dispositivo preferido para cualquier tipo de uso (un 91,9 % de la población así lo indica), seguido del ordenador (62,9 %), y a cierta distancia permanece la tableta (29,7 %)».

Por localización entendemos la adaptación cultural y lingüística de contenido digital a las necesidades de un mercado extranjero (Schäler, 2011), pero dentro la localización, podemos encontrarnos con distintos tipos de productos digitales que se pueden localizar: programas informáticos, sitios web, videojuegos o aplicaciones móviles. La aparición de las aplicaciones móviles y su localización es más reciente que

la de programas informáticos o sitios web, lo que convierte a esta actividad en algo relativamente nuevo dentro de la industria de la localización, pero con una gran perspectiva de futuro.

A pesar del hecho de que distinguir entre cualquier tipo de localización puede resultar confuso, pues todas comparten características comunes en el proceso independientemente del tipo de producto que se vaya a localizar (Sánchez Ramos, 2017), debemos tener claro en primer lugar a qué nos referimos con aplicaciones móviles y qué características las diferencian de las aplicaciones de escritorio, diferencias que ciertamente influyen en la localización de ambos productos. Así pues, podemos decir que «se denomina aplicación móvil o *app* a toda aplicación informática diseñada para ser ejecutada en teléfonos inteligentes, tabletas y otros dispositivos móviles» (Santiago, Trbaldo, Kamijo, & Fernández, 2015, pág. 7).

No obstante, la línea que divide una aplicación móvil de una aplicación para ordenador es muy delgada, de hecho, algunos autores las incluyen dentro del mismo grupo. En este sentido, para Roturier (2015) todo sería una aplicación, ya sean sitios web, aplicaciones móviles o programas de escritorio, y estas aplicaciones estarían formadas por un ecosistema digital. Dentro de este ecosistema, la interfaz sería el elemento más visible para el usuario final, pues incluye las cadenas de texto que se muestran al usuario en pantalla y otros contenidos como pueden ser las imágenes, pero también están dentro de este ecosistema elementos como el contenido de ayuda y el contenido de *marketing*



Figura 1. Ecosistema de una aplicación según Roturier (2015, pág. 2)

creado para una aplicación, y que también hay que traducir. Desde esta perspectiva, aplicaciones móviles y softwares no difieren en gran medida.

Una de las diferencias que presenta la localización de aplicaciones móviles con respecto a la de aplicaciones de escritorio radica en los idiomas a los que se localiza finalmente el producto para su salida en el mercado, ya que las combinaciones lingüísticas típicas pierden fuerza en la localización de aplicaciones (es decir, las traducciones al francés, alemán, japonés, español o italiano, a pesar de que se mantienen, no son las únicas). Esto se debe a la aparición en escena de mercados secundarios que hacen que un proyecto de localización de una aplicación móvil pueda llegar a incluir hasta 25 idiomas (Simón, 2016). De hecho, según un análisis de las tendencias del mercado de la localización de aplicaciones móviles, juegos y sitios web en 2017 realizado por OneSky (empresa del sector de la localización), los mercados emergentes como países de África, del sureste de Asia y países de Europa del Este, tradicionalmente minoritarios en la localización de software, han incrementado en gran medida la demanda de localización de los productos antes mencionados (Li, 2018).

También hay que tener presente que, por lo general, el volumen de palabras de las aplicaciones móviles que se traducen suele ser bastante bajo, ya que las interfaces de usuario tienden a usar más elementos gráficos o de audio y pocos textuales con el fin de facilitar su uso. Asimismo, en el caso de las aplicaciones móviles, los plazos de entrega para un proyecto de localización se reducen considerablemente. 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».

Además, dentro del mercado actual de la localización de aplicaciones móviles, podemos encontrarnos con distintos tipos de desarrolladores. Por un lado, estarían las grandes empresas que son capaces de permitirse un coste elevado en la localización con tal del llegar a un mayor número de usuarios y, por otro lado, empresas desarrolladoras de tamaño medio o incluso desarrolladores independientes que disponen de unos recursos mucho más limitados para invertir en la localización de sus aplicaciones. Hay que tener en cuenta que no todos los desarrolladores pueden permitirse un elevado coste de localización, por lo que la calidad lingüística puede variar dependiendo de cada

aplicación. Tampoco debemos olvidar la existencia de aplicaciones de código abierto para móviles que se traducen en su mayoría gracias a personas que hacen este trabajo de forma voluntaria, lo que no nos permite tampoco tener garantías de la calidad de dichas traducciones.

3.2. Aspectos técnicos de la localización de aplicaciones

Cuando una empresa o un traductor pretende hacerse hueco en el mundo de la localización de aplicaciones móviles, debe tener en cuenta los aspectos técnicos asociados a esta actividad. Las aplicaciones móviles pueden estar diseñadas para distintos sistemas operativos específicos para dispositivos móviles, como pueden ser iOS, Android o Windows Phone entre otros, los cuales llevan asociados unos formatos específicos. Sin embargo, los dos sistemas más utilizados por la mayoría de usuarios de todo el mundo son Android y iOS, podemos decir que son los líderes del mercado. Según StatCounter (2019), el sistema operativo con mayor cuota de mercado en todo el mundo con diferencia es Android (74,85 % en abril de 2019), seguido de iOS con un 22,94 % de cuota de mercado, mientras que el resto de sistemas operativos no llegan siquiera al uno por ciento.

Por tanto, los profesionales que se dedican a la localización deben tener un conocimiento profundo de los distintos formatos a los que se pueden enfrentar o tener los conocimientos necesarios para poder hacer frente a nuevos formatos. Pero este proceso, además de estar relacionado con los aspectos técnicos, es un proceso dinámico, como señala Sánchez Ramos, «el proceso de trasvase en la localización está directamente imbricado con los aspectos técnicos, además de caracterizarse por un flujo de trabajo en constante cambio, que lo dota de dinamismo entre los distintos actores implicados (traductores, gestores de proyectos, ingenieros, etc.) y que se conoce como *agile localization*» (2017, pág. 123). Antes de que se empleara este modelo de trabajo dinámico de colaboración por el que el producto se traduce a la vez que está siendo desarrollado (*agile localization*), se utilizaba el modelo tradicional conocido como *waterfall localization*, que consiste en que cada etapa del desarrollo tiene que completarse antes de pasar a la siguiente, es decir, la localización comienza una vez ha terminado la fase de desarrollo (Tabacchi, 2018).

Además, en el sector del desarrollo de aplicaciones móviles hay innovaciones y avances continuos, por lo que un traductor que se especializa en localización,

independientemente del tipo de producto con el que suele trabajar, debe adaptarse y seguir formándose a lo largo de toda su carrera profesional, ya que se trata de un sector en continua evolución.

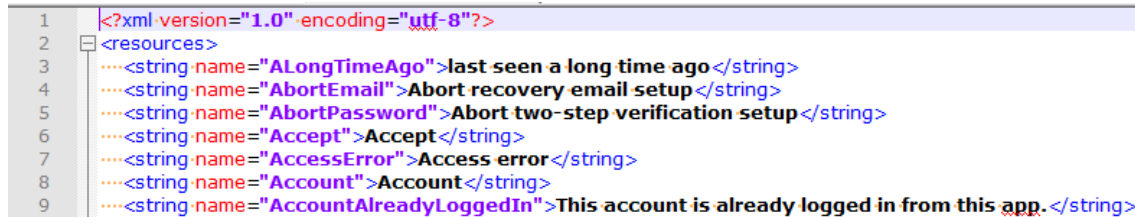
Cuando un proveedor de servicios lingüísticos recibe los archivos que hay que localizar, puede recibir distintos tipos de archivos, que a su vez pueden variar en función del encargo. Estos archivos pueden ser archivos ya procesados por el cliente y que simplemente habrá que traducir (archivos de texto simple o en plataformas en línea) los cuales no comportan un desafío técnico para el localizador; archivos de texto de aplicaciones en formatos estándar (.strings, .xml, .po, etc.) que requieren cierto grado de procesamiento; o aplicaciones completas (.app, .apk, etc.) (Simón, 2016). Este último formato es el que plantea más problemas desde un punto de vista técnico.

Como señala Simón (2016), cuando hablamos de archivos de texto en formatos estándar nos referimos a «any format based on simple text, tagging language or a format for data interchange that uses a systematic syntactic pattern of a “key/value” pair (or, in other words, an “identifier/translatable text” pair)». Es decir, se trata de archivos que incluyen las cadenas de texto de una aplicación móvil, y cada una de esas cadenas va acompañada de un identificador único o ID en toda la aplicación. Estos identificadores son el resultado del proceso de internacionalización de la aplicación, proceso por el cual se prepara una aplicación para ser localizada, y durante el que se separa el código de la aplicación (todo aquello que no se debe modificar durante el proceso de localización) del texto que hay que traducir. De este modo, los identificadores permiten llamar a cada cadena de texto desde el resto de archivos que forman parte de la aplicación. Esto hace más fácil el proceso de localización e impide que el traductor pueda provocar un error en el código de la aplicación al realizar cambios en partes que no debería.

Por otro lado, hoy en día disponemos de herramientas de traducción asistida especializadas en localización (como Alchemy CATALYST o SDL Passolo) que nos permiten introducir la mayoría de archivos estándar y separar aquello que es traducible de lo que no se debe traducir. De este modo, facilitan en gran medida el proceso de traducción, pues permiten la integración de memorias de traducción y glosarios en el proceso, a la vez que ayudan a mantener la consistencia interna de los textos. Pero la función más poderosa de este tipo de herramientas es la posibilidad que ofrecen de crear filtros personalizados para procesar formatos que la herramienta no es capaz de reconocer,

sin la necesidad de tener conocimientos de programación; así, cada localizador puede hacer uso de esta utilidad cuando se enfrente a un nuevo formato (Simón, 2016).

En la siguiente imagen podemos ver el aspecto de este tipo de archivos con identificadores, donde lo que aparece entre comillas y de color lila es el identificador de la cadena que nos aporta cierta información sobre el contenido de dicha cadena, y lo que aparece en negro es el texto que hay que traducir.



```
1 <?xml version="1.0" encoding="utf-8"?>
2 <resources>
3 <string name="ALongTimeAgo">last seen a long time ago</string>
4 <string name="AbortEmail">Abort recovery email setup</string>
5 <string name="AbortPassword">Abort two-step verification setup</string>
6 <string name="Accept">Accept</string>
7 <string name="AccessError">Access error</string>
8 <string name="Account">Account</string>
9 <string name="AccountAlreadyLoggedIn">This account is already logged in from this app.</string>
```

Figura 2. Archivo strings.xml de la aplicación de Telegram para Android

Los formatos estándar de texto más empleados son los que emplean los dos sistemas operativos más comunes, archivos .xml para Android y archivos .strings para iOS, no obstante, en este trabajo nos centraremos en los archivos de Android, ya que son los que utilizaremos más adelante para llevar a cabo el estudio.

Como hemos mencionado, también existe la posibilidad de que el cliente envíe la aplicación completa compilada en formato APP para iOS o APK para Android. Para hacer frente a este tipo de encargos es necesario tener un conocimiento más avanzado de las herramientas de desarrollo de aplicaciones usadas para cada plataforma o de herramientas externas que nos permitan descompilar las aplicaciones para así poder acceder al contenido. Además, hay que tener claro dónde se encuentran los archivos que hay que traducir, dentro de todo entramado de carpetas de una aplicación. Sin embargo, hay herramientas como AppleGlott, iLocalize, Alchemy CATALYST o SDL Passolo que ayudan a identificar los archivos traducibles dentro de las carpetas de recursos de forma automática (Simón, 2016).

Un aspecto técnico que ha provocado problemas en el pasado en la localización de aplicaciones móviles ha sido la restricción de caracteres con motivo del pequeño tamaño de las pantallas de los teléfonos móviles y la falta de adaptabilidad del diseño. Sin embargo, actualmente esto ya no supone un gran problema, pues los desarrolladores de aplicaciones lo tienen en cuenta a la hora de crearlas y tienen la opción de hacer que el

diseño sea flexible y se adapte en función de la longitud del texto traducido (Android Open Source Project, s.f.).

3.2.1. Las aplicaciones para Android

Las aplicaciones de Android utilizan el formato estándar XML para crear los archivos de cadenas de texto o recursos de *strings*. Pero si lo que recibimos es una aplicación Android en formato APK, necesitaremos utilizar herramientas para desarrolladores como Android Studio u otras herramientas que nos permitan descompilar el archivo como Apktool. Al acceder al contenido, lo que nos encontramos es un entramado de carpetas y archivos que forman la aplicación, pero los archivos que nos interesan con respecto al proceso de localización se encuentran, por lo general, dentro de una carpeta llamada *res*, que en su interior incluye otras carpetas. Dentro de esta carpeta *res*, todos los recursos se encuentran ordenados en función del tipo de recurso. Las carpetas que pueden aparecer en su interior son las siguientes (Android Open Source Project, s.f.):

- § *animator*: archivos XML que definen animaciones de propiedades.
- § *anim*: archivos XML que definen animaciones de interpolación de movimientos.
- § *color*: archivos XML que definen una lista de estados de color (define una lista de colores para objetos que cambian de color según el estado).
- § *drawable*: elementos de diseño que incluyen archivos de mapas de bits (.png, .9.png, .jpg, .gif) o archivos XML con distintas funciones de diseño.
- § *mipmap*: archivos de diseño para los iconos de la aplicación de diferentes densidades para lanzadores.
- § *layout*: archivos XML que definen el diseño de la interfaz de usuario.
- § *menu*: archivos XML que incluyen los menús de las aplicaciones.
- § *raw*: archivos aleatorios que no encajan en el resto de carpetas como vídeos, archivos de audio o archivos .json.
- § *values*: archivos de valores simples como *strings*, valores enteros y colores.
- § *xml*: archivos XML aleatorios.

Dentro de las carpetas *values* es donde se encuentran los archivos de tipo *string* que se deben traducir. La carpeta que no cuenta con un código de idioma es la carpeta de los

recursos originales, a partir de los que se realizan el resto de traducciones que se incluyen en carpetas del tipo *values-es* (el sufijo *es* corresponde con el código del español). Estos códigos de idioma se definen mediante «un código de idioma ISO 639-1 de dos letras, opcionalmente seguido de un código de región ISO 3166-1-alfa-2 de dos letras (precedido por una “r” en minúscula)» (Android Open Source Project, s.f.). Así pues, la carpeta destinada a las cadenas en español de España sería *values-es-rES*. A pesar de que los archivos de tipo *string* son aquellos más fácilmente identificables como archivos que hay que traducir, no debemos olvidarnos de que puede haber imágenes, audios o vídeos en otras carpetas de recursos que también se deben localizar.

Dentro de estos archivos de recursos tipo *string*, podemos encontrarnos con tres tipos de archivos que debemos tener en cuenta a la hora localizar una aplicación. El primero es el archivo *strings.xml*, archivo de recursos básico, este aparece en cualquier aplicación y contiene el grueso de las cadenas de texto de la misma. Los otros dos tipos de archivos son los archivos *arrays.xml* y *plurals.xml*, y no siempre aparecen. El primero incluye listas ordenadas de objetos, mientras el segundo permite la pluralización de las cadenas de la aplicación. Como podemos ver en el siguiente ejemplo de un archivo *arrays.xml*, estos archivos incluyen un mismo identificador para diversos elementos:

```
<?xml version="1.0" encoding="utf-8"?>
<resources>
  <string-array name="planets_array">
    <item>Mercury</item>
    <item>Venus</item>
    <item>Earth</item>
    <item>Mars</item>
  </string-array>
</resources>
```

Figura 3. Ejemplo de archivo *arrays.xml* extraído de la guía de localización de Android

Teniendo en cuenta que cada idioma tiene una manera propia de marcar los plurales, dentro de los archivos *plurals.xml*, Android aconseja a los desarrolladores de aplicaciones para esta plataforma que incluyan como mínimo en el original una cadena para *one* (<item quantity="one">) y otra cadena para el resto cuyo valor es *other* (<item quantity="other">). Todas las palabras clave que se pueden usar en Android para marcar la forma de plural son las siguientes: *zero*, *one*, *two*, *few*, *many* y *other*. No obstante, es el traductor quien debe elegir cuáles son necesarias en su idioma, y colocar en el archivo traducido únicamente aquellas que respondan a la necesidad gramatical del idioma, es

decir, si en español solo distinguimos entre un objeto (*one*) o más de uno (*other*), no hace falta que se incluyan el resto de opciones.

Además, hay que tener en cuenta que Android permite utilizar etiquetas de formato HTML, por lo que ciertos caracteres, que cuentan con un uso especial en Android o en HTML (@, ?, <, ', "), pueden provocar problemas a la hora de mostrar las cadenas de texto y, por este motivo, deben escaparse. Algunos se pueden escapar usando una contrabarra (\), mientras que para otros se necesita sustituirlos por su entidad HTML. El caso de los apóstrofes, por ejemplo, se puede solucionar rodeando la cadena con comillas dobles. En la siguiente tabla extraída de la guía para desarrolladores de Android (Android Open Source Project, s.f.) podemos ver de forma detallada cómo se debe proceder con cada uno de estos caracteres:

Characters	Escaped form(s)
@	\@
?	\?
<	<
&	&
Single quote (')	Any of the following: <ul style="list-style-type: none">< &apos;< \'< Enclose the entire string in double quotes ("This'll work", for example)
Double quote (")	Any of the following: <ul style="list-style-type: none">< &quot;< \" <p>Note that you must escape double quotes. Surrounding the string with single quotes does not work.</p>

Tabla 1. Caracteres que deben escaparse en los archivos .xml de Android extraídos de la guía de localización de Android

3.2.2. Las aplicaciones para iOS

En el caso de las aplicaciones para iOS, podemos enfrentarnos a los archivos localizables originales en formato .strings o a la aplicación completa en formato APP. Para poder ver el contenido de la aplicación APP necesitaremos disponer de la herramienta para desarrolladores de Apple, Xcode que funciona únicamente en MacOS y que se puede descargar de forma gratuita desde la Mac App Store si dispones de un ID de Apple. Además, Xcode permite exportar textos en formatos estándar XLIFF para que la persona encargada de la localización pueda traducirla de manera sencilla (Apple Inc., 2019). Al igual que Android, Apple también dispone de documentación para desarrolladores que incluye información sobre el proceso de localización para hacer más sencillo el proceso.

En cuanto a los archivos localizables de cadenas de texto, estos se encuentran dentro de una carpeta [ID del idioma original].lproj, que en el caso de que se hubiera desarrollado en inglés sería *en.lproj*, mientras los archivos traducidos al español se encontrarían en la carpeta *es.lproj* (se emplean los mismos códigos ISO de dos letras para definir idioma y región que en Android). En caso de que se hayan exportado e importado los archivos XLIFF con Xcode, el programa creará la carpeta específica del idioma e incluirá los archivos .strings en su correspondiente carpeta. Además, si se importan versiones más recientes de la traducción, el archivo nuevo y el existente se combinan para aplicar los cambios (Apple Inc., 2019).

El grueso de las cadenas de texto de las aplicaciones iOS se almacena en archivos Localizable.strings situados dentro de la carpeta *.lproj* del idioma original, y cuya estructura difiere un poco de la de los archivos XML de Android, pero en los que se puede reconocer con facilidad el identificador de la cadena y el texto a traducir. Como podemos ver en la siguiente imagen, ambos van entre comillas, primero aparece el identificador, seguido de un símbolo de igual y, a continuación, la cadena de texto traducible.

```
"PUSH_MESSAGE_TEXT" = "%1$@|%2$@";  
"PUSH_MESSAGE_NOTEXT" = "%1$@|sent you a message";  
"PUSH_MESSAGE_PHOTO" = "%1$@|sent you a photo";  
"PUSH_MESSAGE_PHOTO_SECRET" = "%1$@|sent you a self-destructing photo";  
"PUSH_MESSAGE_VIDEO" = "%1$@|sent you a video";  
"PUSH_MESSAGE_VIDEO_SECRET" = "%1$@|sent you a self-destructing video";  
"PUSH_MESSAGE_ROUND" = "%1$@|sent you a video message";  
"PUSH_MESSAGE_CONTACT" = "%1$@|shared a contact %2$@ with you";  
"PUSH_MESSAGE_GEO" = "%1$@|sent you a map";  
"PUSH_MESSAGE_GEOLIVE" = "%1$@|started sharing their live location";  
"PUSH_MESSAGE_DOC" = "%1$@|sent you a file";  
"PUSH_MESSAGE_AUDIO" = "%1$@|sent you a voice message";
```

Figura 4. Archivo Localizable.strings de la aplicación Telegram para iOS

Otro tipo de archivos de cadenas de texto en iOS es InfoPlist.strings, que ofrece información sobre la aplicación y cómo está configurada en el sistema, es decir, aquello que tiene impacto fuera de la propia aplicación como, por ejemplo, los mensajes que informan de los permisos que necesita la aplicación para realizar determinadas acciones o los avisos de copyright. Estos archivos también están situados dentro de la carpeta *.lproj* de cada idioma, y las cadenas de las que están compuestos se muestran en pantalla al usuario, por lo que también hay que traducirlas.

Por otro lado, el tipo de archivo que recoge las diferentes formas de pluralización en las aplicaciones iOS es un archivo con extensión *.stringsdict*. En este tipo de archivo, al igual que ocurre con los archivos *plurals.xml* de Android, solo hay que incluir aquellas formas de plural que tengan relevancia para el idioma al que se traduce y se usan las mismas palabras clave: *zero*, *one*, *two*, *few*, *many* y *other*.

3.3. Aspectos lingüísticos y culturales de la localización de aplicaciones

A parte de los aspectos técnicos que conlleva la localización de aplicaciones, también hay otros aspectos que se deben tener en cuenta a la hora de localizar una aplicación. En este apartado comentaremos ciertos aspectos problemáticos de este tipo de localización como pueden ser la presencia de variables; la terminología y especialización del texto; el registro del lenguaje; la falta de contexto de las cadenas, y la adaptación de elementos culturales.

a) Variables:

Al igual que en la localización de cualquier producto, las aplicaciones están plagadas de variables que se sustituirán por otro elemento, y se deben traducir las cadenas de texto teniendo esto en cuenta. Las variables son elementos empleados por los desarrolladores que consisten en caracteres que se reemplazan por una palabra, un valor o una cadena de texto en el momento en que se ejecuta la aplicación (Esselink, 2000). A continuación, podemos ver un ejemplo de una cadena de texto de la aplicación de Telegram para Android, donde *un1* sería la variable que en la aplicación se sustituirá por el nombre del usuario que ha cambiado la foto del grupo.

```
<string name="ActionChangedPhoto">un1 changed the group  
photo</string>
```

Figura 5. Ejemplo de uso de variable en una cadena de texto de la aplicación Telegram

Desde el punto de vista de la traducción, uno de los grandes problemas que plantean las variables es el hecho de que no sabemos por qué elemento se van a sustituir después, lo cual puede dar lugar a una traducción donde el género o el número no concuerden. Pero, en ocasiones, los desarrolladores de aplicaciones móviles tienen estos aspectos en cuenta y crean archivos específicos donde incluyen las cadenas que pueden ser problemáticas en cuanto a la pluralización o, en ocasiones, duplican las cadenas. Por ejemplo, en el archivo strings.xml de la aplicación de Firefox, primero aparece una cadena con variable y el sustantivo en plural (`<string name="home_closed_tabs_number">%d tabs</string>`) y, a continuación, la misma cadena en la que se sustituye la variable por el número uno para el singular (`<string name="home_closed_tabs_one">1 tab</string>`). Además, el identificador de la cadena también da una pista sobre el contenido y, en este caso, el número. También podemos encontrarnos con otras aplicaciones en las que se tiene en cuenta el género de la variable y se crean variables complejas, como en la de Google Family Link, donde a la variable {CHILD}, que se sustituye por el nombre del hijo del usuario, se le añaden atributos de género, a pesar de que en inglés no supone ninguna diferencia: {GENDER, select, male {Install for {CHILD}} female {Install for {CHILD}} other {Install for {CHILD}}}

Por otro lado, también se debe tener cuidado con el lugar de la frase en el que se coloca la variable en la traducción. Esto se debe a las diferencias de colocación de los elementos de la frase en función del idioma, por tanto, hay que intentar intuir si la variable se sustituirá por un número o una palabra y, en tal caso, de qué tipo de palabra se va a tratar y cuál será su función en la frase.

b) Terminología:

La terminología y la especialización del texto también pueden complicar la tarea de localización. Las aplicaciones móviles actuales son muy heterogéneas en lo que se refiere a la temática y pueden estar enfocadas a distintos campos de especialidad como la medicina, las finanzas, la arquitectura, etc. Por ello, es necesario que el profesional que se encargue de la localización controle la especialidad o, por el contrario, sepa documentarse de forma adecuada para poder localizar la aplicación de forma correcta y cumplir con las expectativas del cliente.

Además, las aplicaciones se usan en dispositivos de distintas plataformas, como Android o iOS, y cada una de estas plataformas usa una terminología diferente. Por este motivo, como indica la guía de localización de Android (Android Open Source Project, s.f.), hay que intentar mantener la consistencia terminológica, sobre todo cuando se hace referencia a elementos de la interfaz del sistema externos a la aplicación. Apple, por su parte, ofrece la opción de descargar glosarios de los sistemas iOS, watchOS, tvOS y macOS en distintos idiomas para ayudar a la localización de estos elementos; estos glosarios se pueden descargar desde la sección para desarrolladores de su sitio web, una vez inicias sesión con tu ID de Apple (Apple Inc., 2019).

c) Registro:

Otro aspecto a tener en cuenta es el registro, es decir, el traductor debe considerar a quién va a estar dirigida la aplicación para adaptar su lenguaje a los usuarios y, en caso de que las haya, debe seguir las indicaciones del cliente al respecto. Por ejemplo, si se trata de una aplicación para niños habrá que emplear un vocabulario más sencillo, totalmente distinto al que se utilizará en una aplicación para la gestión de finanzas o de citas.

d) Falta de contexto:

Otro problema en la localización de aplicaciones móviles es la falta de contexto en las cadenas a traducir. A pesar de que la guía para desarrolladores de Android aconseja proporcionar contexto suficiente para cada cadena «As you declare strings in your strings.xml file, make sure to describe the context in which the string is used. This information is invaluable to the translator and result in better quality translation. It also helps you manage your strings more effectively» (Android Open Source Project, s.f.), la realidad es que la mayoría de aplicaciones (excepto las aplicaciones nativas de Android) no incluyen información adicional a modo de comentario que supla la falta de contexto. No obstante, el nombre identificativo de cada cadena (p. ej., name="NotificationMessagePhoto") suele servir de ayuda para identificar a qué hace referencia el texto incluido en esa cadena. Con motivo de esta falta de contexto, es de especial relevancia la fase de *testing* lingüístico que se lleva a cabo tras la traducción, que consiste en verificar cómo se ve el texto traducido en pantalla para detectar posibles problemas y que constituye una de las fases más importantes del proceso de localización (Muñoz-Sánchez & Lopez-Sánchez, 2016).

e) Elementos culturales

Tampoco debemos olvidar que se deben tener en cuenta los aspectos culturales para adaptar la aplicación al mercado de destino, pues, en definitiva, la localización consiste en la adaptación del producto a la cultura meta. Así pues, se deben adaptar elementos presentes en las aplicaciones como formatos de fecha, de hora, de número, la moneda, referencias culturales, etc.

3.4. Entrenamiento de motores de TAE

En la actualidad, existen distintos tipos de motores de traducción automática a los que se puede recurrir como recurso durante el proceso de traducción, ya sean neuronales, basados en reglas, estadísticos o híbridos. No obstante, se suele tratar de motores genéricos que no están especializados en un campo en concreto o en una tipología textual específica y, cuando se aplican al proceso de traducción de un tipo de texto específico, pueden dar peores resultados que un motor que se entrene específicamente con corpus pertenecientes a dicho campo o tipología textual. «In corpus-based approaches of

machine translation, the more specific the training corpus domain, the better the translation output will be» (Doğru, Martín-Mor, & Aguilar-Amat, 2018, pág. 14). De ahí la importancia de la existencia de herramientas que nos permitan entrenar motores de TAE según nuestras propias necesidades y el tipo de textos que vayamos a traducir.

Asimismo, muchos autores como Lumeras y Way (2017) o Koehn (2019) han señalado que los motores de TAE suelen dar mejores resultados cuando se ponen a prueba con textos similares a los corpus empleados para su entrenamiento. No obstante, aunque con un corpus de entrenamiento más específico, los resultados suelen ser mejores; por lo general, cuanto más específico sea el campo elegido para recopilar el corpus, la cantidad de recursos que podremos obtener será menor, y la situación se complica cuando se trata de lenguas minoritarias, pues son combinaciones lingüísticas de las que se dispone de pocos recursos paralelos con los que entrenar los motores (Doğru et al., 2018).

En primer lugar, debemos tener claro cómo funciona un motor de TAE. Al principio de su existencia, a finales de los 80 y principios de los 90, los motores de TAE se basaban en modelos de alineación palabra por palabra, por lo que no podían almacenar información sobre el contexto de la frase y daba lugar a ciertos problemas como elecciones léxicas erróneas (Sánchez-Martínez, 2012). Actualmente, estos motores estadísticos se basan en corpus bilingües alineados previamente para determinar de manera estadística que un grupo de palabras se corresponde con otro grupo de palabras en la lengua meta. Estos grupos de palabras también se conocen como n-gramas, es decir, una secuencia de n elementos (1, 2, 3, etc.) dentro de una secuencia de mayor tamaño (Koehn, 2019).

Existen diversas herramientas con las que cualquiera puede empezar a entrenar un motor de TAE con sus propios recursos lingüísticos. Entre las opciones que existen nos encontramos con sistemas de propietario como KantanMT o LetsMT, que además incluyen características adicionales como la estimación automática de la calidad; o herramientas de software libre como Moses o MTradumàtica (plataforma en línea que se basa en Moses y le proporciona una interfaz) (Martín-Mor, 2017).

Moses es un sistema de código abierto muy potente y utilizado, que ofrece numerosas funcionalidades de personalización de los motores, pero tiene la desventaja de que no dispone de interfaz, por lo que aquella persona que desee utilizarlo necesitará tener unos conocimientos avanzados en línea de comandos. Además, se necesita tener al menos

2 GB de RAM y unos 10 GB de espacio libre en disco en el equipo en el que se va a instalar (Koehn, 2019). De ahí el desarrollo de la plataforma MTradumàtica, la cual hace mucho más sencillo para el usuario la creación y entrenamiento de motores de TAE personalizados, pues ofrece una interfaz sencilla de usar, intuitiva y en línea, por lo que los requisitos del sistema ya no suponen una limitación y se puede usar en cualquier sistema operativo. La existencia de programas de código abierto como los que hemos mencionado «ha favorecido la adopción de sistemas de TA sin necesidad de invertir grandes cantidades de dinero en implementar técnicas y métodos de TA. [...]De este modo, instituciones y empresas se benefician de forma directa de los últimos avances en investigación, además de poder modificar los programas para ajustarlos a las necesidades concretas de su organización» (Sánchez-Martínez, 2012, pág. 153).

Una vez seleccionada la herramienta que se va a emplear para el entrenamiento del motor de TAE, se necesita disponer de dos corpus, uno bilingüe y otro monolingüe, y a partir de estos corpus se crea lo que se conoce como modelo de traducción y modelo de lengua respectivamente.

En realidad, no sería necesario disponer de dos corpus distintos (bilingüe y monolingüe) para crear un motor de traducción, pues se puede usar una parte del corpus bilingüe para entrenar el motor de lengua, sin embargo, el uso de recursos monolingües adicionales para entrenar este modelo suele ser favorable (Koehn, 2019). Pero para poder entrenar el motor con estos corpus, primero se deben llevar a cabo los siguientes procesos: segmentación, *truecasing* y limpieza (Koehn, 2019). La segmentación (*tokenization*) consiste en separar las palabras de la puntuación mediante espacios; durante la fase de *truecasing* las palabras se convierten a su forma más probable, es decir, si es más probable que aparezcan en mayúscula o en minúscula; y, durante la fase de limpieza, aquellas oraciones que son demasiado largas o que están mal alineadas se eliminan para evitar que provoquen problemas durante el entrenamiento. En el caso de Moses estos procesos se llevan a cabo manualmente ejecutando una serie de comandos, mientras que con MTradumàtica se realizan de forma automática.

Además, como los sistemas de TAE se basan por completo en corpus para su entrenamiento, es de vital importancia que dichos corpus tengan un gran tamaño y que sean de cierta calidad. Por este motivo, la elección y preparación del corpus paralelo condicionará la calidad del motor de TAE resultante (Doğru et al., 2018). Existen sitios

web como OPUS (Tiedemann, 2012), donde podemos acceder y descargar corpus paralelos de inglés a español, así como en otras combinaciones de idiomas, ya alineados y de gran extensión de forma rápida. Sin embargo, cuando lo que se necesita obtener es un corpus específico de un tipo de texto, como el de este estudio, la tarea de obtención de recursos lingüísticos para la creación del corpus se complica, pues requiere buscar los recursos manualmente, siempre teniendo en cuenta la importancia que un corpus de calidad tiene en el entrenamiento del motor y el posterior resultado de traducción. También podemos optar por potenciar la calidad de este resultado mediante la introducción de glosarios bilingües durante el proceso de entrenamiento del motor para aumentar la calidad de la terminología específica del campo elegido (Doğru et al., 2018).

Una vez se han preparado los corpus para el entrenamiento, es el momento de crear el modelo de lengua y el de traducción. El modelo de traducción contiene información sobre cómo deben traducirse grupos de palabras siguiendo probabilidades estadísticas, mientras que el modelo de lengua permite al motor determinar qué frase será más probable en la lengua meta basándose en textos reales de un corpus para asegurar la fluidez de la traducción (Martín-Mor, 2017). Este modelo de lengua se puede crear a partir del corpus bilingüe que se va a emplear para crear el modelo de traducción o a partir de nuevos corpus monolingües, la elección de uno u otro modelo de entrenamiento depende de la persona que entrena el motor y sus preferencias, no obstante, se pueden crear motores de distinta composición para comparar los resultados y ver cuál de ellos es más adecuado.

En cuanto a su composición, el modelo de traducción está formado por un modelo de lengua y una o más tablas probabilísticas que incluyen segmentos en lengua origen y meta, así como la probabilidad de que uno sea la traducción de otro. Este modelo también puede incluir tablas de reordenamiento, que describen los cambios que se producen en el orden de las palabras entre el original y su traducción. Una vez se han creado ambos modelos, se puede llevar a cabo una fase de optimización, durante la que se comparan diferentes modelos estadísticos para dar lugar a las mejores traducciones posibles mediante algoritmos (Koehn, 2019).

3.5. Posedición

Cada vez son más las empresas que implementan la traducción automática en sus procesos de traducción, ya sean las propias empresas que necesitan los servicios de traducción, como Adobe, eBay y TripAdvisor, o los propios proveedores de servicios de traducción (Aranberri, 2014). Así pues, durante los últimos años, la traducción automática seguida de posedición, «entendida como la edición de segmentos obtenidos mediante traducción automática» (Sánchez-Gijón, 2016, pág. 152), ha incrementado su uso entre los profesionales de la traducción en gran medida debido a diversas razones. Sánchez-Martínez (2012, pág. 150) opina que los motivos de este crecimiento han sido «la mejora en las técnicas de TA, la mayor disponibilidad de recursos tales como software y datos, el cambio en las expectativas de los usuarios en cuanto a lo que se puede esperar o no de un sistema de TA, y por último la mayor integración de sistemas TA en entornos de ayuda a la traducción». Tampoco debemos olvidar la existencia de diversas herramientas que nos permiten personalizar los motores de TA según nuestras propias necesidades, lo que también ha servido para favorecer este incremento.

Según un informe de TAUS (2017) donde plantean sus previsiones sobre la industria de la traducción en los próximos años, «post-editing is now the standard modus operandi for most professional translators, and it is going to stay that way, at least in the short-term». Por lo que podemos decir que, aunque siga habiendo muchos traductores reticentes al uso de traducción automática como herramienta de trabajo, esta se ha ido abriendo paso poco a poco dentro del sector, y no parece que vaya a cambiar esta situación en el futuro.

Asimismo, la inmediatez en la necesidad de los encargos de traducción es un factor que también ha influido en el incremento del uso de la posedición, pues cada vez se dispone de menos tiempo y, gracias a la TA, el proceso se acelera en gran medida (siempre y cuando la TA tenga cierto nivel de calidad). Debido a este incremento en el uso de posedición en el mercado de servicios lingüísticos, la Organización Internacional de Normalización (ISO) ha creado una norma específica para esta actividad, distinta a la norma que existe para los servicios de traducción, con el fin de establecer unos estándares mediante la norma ISO 18587:2017 (Translation services — Post-editing of machine translation output — Requirements). En la introducción de este documento se señalan las ventajas que supone la posedición desde el punto de vista del cliente:

Clients have come to realize that the use of such systems is a viable solution for translating projects that need to be completed within a very tight time frame and/or with a reduced budget. When an MT system is used, clients can have material translated that can otherwise not be translated; translation costs can be decreased and the launch of products on specific markets, as well as the flow of information, can be accelerated (ISO 18587, 2017).

Pero según esta norma, la posesición no solo ofrece beneficios al cliente, sino también a los proveedores de servicios lingüísticos, pues ayuda a mejorar la productividad, los tiempos de entrega y a mantener la competitividad en un entorno en el que los clientes demandan cada vez más traducción automática seguida de posesición (2017). Además, en caso de que se use un motor personalizado, gracias a los propios textos posesitados por profesionales a partir de la traducción automática, se puede seguir entrenando y mejorando dicho motor.

Cuando una empresa trata de decidir si implementar la traducción automática seguida de posesición o incluso decidir qué motor va a elegir entre varios disponibles, lo primero que tiene que plantearse es si este cambio le va a resultar beneficioso, y uno de los modelos que se emplean para medir la eficacia de la TA es la productividad de posesición, que consiste en «trata[r] de comparar la productividad de un traductor al realizar traducciones de manera tradicional o al poseitar una TA» (Aranberri, 2014, pág. 473).

Por otro lado, debemos tener en cuenta que la posesición puede aplicarse en distintos niveles según las necesidades del mercado al que va dirigido el texto traducido (Aranberri, 2014). Por un lado, estaría la posesición parcial (*light post-editing*), que consiste en realizar únicamente los cambios necesarios para que un texto sea comprensible, es decir, no haría falta corregir todos los errores, sino solo aquellos que impidan su comprensión, así que la traducción final puede contener errores que no interfieran con la correcta comprensión del mensaje. Por otra parte, encontramos la posesición completa (*full post-editing*), la cual tiene como objetivo la obtención final de una traducción con una calidad equiparable a la de un traductor humano, eliminando todo error presente en la TA, incluidos los errores estilísticos que disminuyen la fluidez de la traducción. La elección de un tipo u otro de posesición depende en definitiva de las preferencias del cliente, de la finalidad de la traducción y de la visibilidad y la función que vaya a tener el texto traducido. Por tanto, con respecto a los encargos de posesición,

podríamos decir que la calidad del producto final viene definida por las preferencias del cliente.

Por este motivo, según la guía de posesición de TAUS (2010), en lugar de hablar de dos niveles de posesición diferentes, podríamos hablar de dos niveles de calidad esperada, por un lado, una calidad suficientemente buena (*good enough quality*) y, por otro lado, una calidad humana o publicable (*human quality*). En el caso de la calidad suficientemente buena, el texto puede sonar como generado por un ordenador y con algunos pequeños errores, pero el mensaje es fiel al original. Para alcanzar una calidad humana, el texto también debe ser comprensible y fiel, pero, además, debe sonar bien estilísticamente.

Cabe destacar la importancia que tiene la creación de guías para la posesición por parte de los clientes con el fin de fijar unas pautas a seguir, las cuales facilitan la tarea en gran medida a los poseedores, pues informan sobre «las peculiaridades del sistema o modelo de MT que hay detrás, de cara a la elaboración de tipologías de errores, fórmulas correctivas y acciones reparadoras a acometer, etc., así como poner de manifiesto el grado o nivel de posesición que se solicita» (Yuste-Rodrigo, 2012).

Como señala Aranberri (2014), debido a la existencia de distintos niveles de calidad esperados, aquí entraría en juego la formación de los traductores en el ámbito de la posesición, pues llevar a cabo una posesición completa puede resultar bastante fácil para alguien que ha sido formado en traducción. Pero, cuando este mismo profesional se tiene que enfrentar a un encargo de posesición parcial, la tarea se complica, ya que ha sido formado para alcanzar la perfección en cualquier texto, algo que no sería necesario en estos niveles de posesición en los que no se necesita que el texto meta sea perfecto. De modo que, en definitiva, la tarea de posesición requiere de unas habilidades distintas a las que está habituado un traductor (Aranberri, 2014, pág. 473). Así pues, podemos decir que la posesición consiste en el proceso de mejora de la traducción generada por el motor con un mínimo de esfuerzo (Joscelyne & Brace, 2010). Esto quiere decir que todo aquello que no sea un error y sea solo un cambio preferencial, no se tiene que modificar, pues la finalidad de la posesición es realizar el menor número de cambios posibles de modo que se emplee el menor tiempo posible.

Pero actualmente no solo nos enfrentamos a este perfil de traductor formado en posesición que se dedica simplemente a editar los segmentos resultantes de la traducción

automática, sino que también se nos presenta el perfil de un traductor profesional que llega a dominar todos los procesos de un proyecto de traducción con traducción automática. La diferencia entre un perfil y otro radica principalmente en que el segundo cuenta con una formación tecnológica e instrumental que le permiten poder hacerse cargo de todo el proceso, desde la edición de los segmentos hasta la preparación de los materiales para la posesición, la preparación de los motores de traducción y la gestión de sistemas de traducción automática (Sánchez-Gijón, 2016).

3.6. La calidad de la traducción automática

El uso de traducción automática (TA) seguida de posesición ha supuesto sin duda un aumento en la productividad los traductores. No obstante, debemos tener en cuenta que la TA supone una ayuda siempre y cuando el resultado de esta tenga una cierta calidad y nos permita tardar menos en realizar un trabajo de lo que tardaríamos traduciéndolo desde cero; en caso contrario, no interesa su uso. De ahí la necesidad de determinar si la calidad de la traducción resultante de los motores de TA es la adecuada o no, ya sea para decantarse por el uso de un motor u otro, o bien durante el proceso de entrenamiento de motores de TAE, con el fin de mejorar los resultados de un motor o comparar la efectividad de varios motores creados a partir de corpus diferentes.

Como señala Babych (2014), a pesar de que, al principio, la evaluación de la TA comenzó como un subcampo dentro del desarrollo de sistemas de TA dirigido principalmente a medir las mejoras de los propios sistemas durante el proceso de desarrollo, con el tiempo se ha convertido en un campo independiente con un conjunto de objetivos más amplio y con más partes interesadas, que incluyen no solo a desarrolladores de sistemas de TA, sino también a traductores, equipos de traductores, gestores de proyectos de localización o usuarios finales, quienes tienen que decidir si los sistemas de TA pueden ayudarles a satisfacer sus necesidades específicas.

Existen diferentes métricas que nos permiten evaluar la calidad de los resultados de TA, pero podemos clasificarlas en dos grandes tipos de métodos empleados: métodos automáticos y métodos manuales. Los métodos manuales son aquellos en los que la evaluación de la calidad la realiza una persona, mientras que los métodos de medición automáticos no necesitan de la evaluación humana.

El principal objetivo de los métodos automáticos es calcular mediante algoritmos una puntuación numérica que represente el grado de calidad o de rendimiento de sistemas de TA específicos y, además, se espera que estas puntuaciones guarden correlación con las evaluaciones manuales realizadas por personas (Babych, 2014). Estos métodos suelen comparar el resultado de la traducción automática con el mismo texto traducido por un traductor humano a modo de referencia, ya sea usando una única referencia o varias para el mismo texto fuente; no obstante, también existen algunos métodos que no emplean una traducción de referencia para realizar la evaluación, conocidos como *Quality Estimation* (QE) (Han, 2018).

Una de las principales diferencias entre las métricas automáticas y manuales señalada por diversos autores como Denkowski y Lavie (2010) o Han (2018) es que la evaluación manual requiere más tiempo y recursos, además de que es más subjetiva, pues una persona que evalúa una traducción puede tener una opinión distinta a la de otra que realiza la misma tarea de evaluación de calidad, mientras que los métodos automáticos son más rápidos y requieren menos recursos.

A pesar de esto, se considera que las evaluaciones humanas de la TA son el estándar que las métricas automáticas deberían intentar alcanzar, pues las automáticas tienen una serie de limitaciones con respecto a las realizadas por personas (Han, 2018). Una de estas limitaciones es que no pueden tener en cuenta aspectos externos relevantes para la traducción como el hecho de que se vaya a usar la traducción en diferentes plataformas, la integración con otras herramientas TAO, consideraciones de privacidad del cliente, etc. (Babych, 2014). Además, algunos métodos automáticos también tienden a ignorar la relevancia que tienen ciertas palabras o conceptos dentro del texto, como conceptos clave que tienen más importancia que otros elementos del texto como pueden ser la puntuación o los determinantes (Han, 2018).

Algunos de los métodos automáticos más conocidos son el método BLEU (Bilingual Evaluation Understudy) y el método METEOR (Metric for Evaluation of Translation with Explicit ORdering). El método BLEU se basa en n-gramas, o grupos de palabras, para comparar la similitud entre una traducción que se pretende evaluar con una traducción de referencia realizada por una persona, y penaliza en caso de que la traducción sea demasiado larga o demasiado corta (Papineni et al., 2002). Por su parte, la métrica METEOR se creó para abordar las carencias de BLEU y se basa en la coincidencia palabra

por palabra entre la traducción y una o más referencias, pero no solo teniendo en cuenta palabras que son idénticas en forma, sino también sinónimos y palabras derivadas (Banerjee & Lavie, 2005).

A la hora de evaluar el resultado de la TA de forma manual, existen distintas aproximaciones, algunas de ellas las explicamos a continuación, pero no son las únicas. Por un lado, nos encontramos con evaluaciones de la calidad en las que los evaluadores dan una puntuación al resultado de la TA en cuanto a adecuación (*adequacy*) y fluidez (*fluency*). Por adecuación nos referimos a qué cantidad de significado o información original se ha incluido también en la traducción respondiendo a la pregunta «How much of the meaning expressed in the gold-standard translation is also expressed in the target translation?» (Linguistic Data Consortium, 2002); mientras que fluidez se refiere al grado en el que la traducción está bien formada, es decir, está bien formada gramaticalmente, sin errores de ortografía y usa un lenguaje natural (TAUS, 2013).

Otro método usado para evaluación manual es el de clasificación de motores (*engine ranking*), donde se muestran a los evaluadores humanos traducciones procedentes de distintos motores para que las clasifiquen de mejor a peor según un valor numérico; este tipo de evaluaciones sería adecuado para casos en los que se comparan sistemas de TA que dan resultados similares. No obstante, esta tarea puede resultar complicada para los evaluadores cuando aparecen frases muy largas, segmentos casi idénticos o errores difíciles de comparar entre sí (Denkowski & Lavie, 2010).

También se pueden realizar pruebas de productividad para medir la calidad de la traducción de un motor. En este tipo de pruebas, en lugar de evaluar directamente la calidad de la traducción, los evaluadores tienen que traducir y poseer segmentos comparables con el fin de calcular la diferencia de productividad entre la traducción y la posesión (de la Fuente, 2014).

Cabe destacar la existencia de la herramienta de evaluación de la calidad DQF (Dynamic Quality Framework) desarrollada en 2012 por TAUS en colaboración con una serie de empresas y organizaciones que incluyen a proveedores de servicios de traducción, instituciones académicas, clientes de los servicios de traducción y proveedores de tecnologías de la traducción (Görög, 2014). Esta herramienta permite evaluar los resultados de traducción, ya sean de TA o de traductores humanos, mediante una

plataforma en línea y, para realizar la evaluación, ofrece las opciones de comparar motores, evaluar su adecuación y fluidez, medir la productividad de posesición y puntuar los segmentos traducidos en base a una tipología de errores. Además, en DQF se considera que la calidad es dinámica, pues depende del contenido, la finalidad de dicho contenido y su audiencia, por ello, ofrece recomendaciones sobre qué método es más adecuado en cada caso.

Por último, no debemos olvidar que para poder evaluar la calidad de un motor de TA se necesita un texto sobre el que realizar pruebas, es decir, hay que elegir una muestra representativa con la que llevar a cabo el análisis. Como señala de la Fuente (2014), lo recomendable sería no realizar la selección de la muestra de manera aleatoria, sino de forma sistemática usando el conocimiento que se tiene sobre el contenido que se pretende traducir con dicho motor. A la hora de seleccionar dicha muestra, también se debe tener en cuenta que cada método de evaluación tiene un propósito y requisitos distintos a los de otros; por ejemplo, para los métodos automáticos se necesitan menos recursos, por lo que es más sencillo trabajar con muestras de mayor tamaño, al contrario que con los métodos manuales, para los que se necesita trabajar con muestras más pequeñas y seleccionadas cuidadosamente. Asimismo, este autor señala que, según su experiencia, la manera más efectiva de obtener resultados fiables sobre el rendimiento de la TA no es el uso de un único método de evaluación, sino la combinación de diferentes métodos.

4. Metodología

En este apartado nos centraremos en el proceso seguido para llevar a cabo el estudio y dar respuesta a los planteamientos expuestos previamente en el apartado 2. Así pues, se comparará la calidad de distintos motores de TAE entrenados a partir de diferentes corpus para una tipología textual concreta, las aplicaciones móviles, y se analizará cómo aborda cada uno de ellos la traducción de elementos concretos. Para ello, como ya hemos mencionado, se han creado tres motores de TAE mediante corpus de distinta composición para comparar el resultado de cada uno y ver cuál de ellos sería la opción más adecuada.

En primer lugar, en el apartado 4.1, expondremos la composición de los distintos corpus creados para llevar a cabo la investigación. En los apartados 4.2, 4.3 y 4.4 explicaremos el proceso que se ha seguido para la creación de dichos corpus mediante la

búsqueda de recursos para el entrenamiento de los motores y explicaremos las herramientas empleadas durante este proceso. A continuación, en el apartado 4.5, veremos paso a paso cómo se han creado los motores gracias a la plataforma MTradumàtica. Y, por último, en el apartado 4.6, presentaremos los archivos procesados por los motores para el posterior análisis de los resultados.

4.1. Motores de TAE creados

Para la realización de este estudio se han creado tres motores de TAE de distinta composición para determinar cuál de ellos da mejor resultado, pero en los tres casos se han empleado corpus creados a partir de textos de la tipología textual elegida, las aplicaciones móviles. Lo primero que hemos hecho ha sido establecer una posible clasificación de aplicaciones móviles en función de su contenido para poder clasificar posteriormente los recursos recopilados. Hemos llevado a cabo esta clasificación debido a que, como diversos autores han señalado (véase el apartado 3.4), un motor de TAE da mejores resultados cuando se entrena con textos similares a aquellos que se van a traducir, y cuanto más especializado sea el motor, mejores resultados dará. Puesto que las aplicaciones móviles existentes son muy heterogéneas, hemos creído oportuno clasificarlas según el contenido de cada una para procurar que el motor creado sea lo más específico posible para un tipo de texto. No obstante, también debemos tener en cuenta que, a pesar de que cuanto más especializado sea el motor, mejores resultados dará, si acotamos demasiado cada categoría nos resultaría difícil conseguir una cantidad adecuada de recursos fiables. De este modo, se han dividido las aplicaciones en cinco categorías según la temática de cada una: herramientas, servicios/estilo de vida, productividad y multimedia. Además, para garantizar que las aplicaciones se clasificaran sin variación de criterio, se ha definido cada una de estas categorías como podemos ver en la siguiente tabla:

Contenido	Explicación
Comunicación	Cualquier aplicación que sirva para ponerse en contacto con otras personas. Incluye aplicaciones de redes sociales, de mensajería, de llamadas, de correo electrónico, etc.
Herramientas	Cualquier aplicación que añada una funcionalidad o utilidad al teléfono. Por ejemplo, navegadores que te permiten realizar búsquedas en internet, gestores de archivos, antivirus, lanzadores etc.

Servicios/estilo de vida	Aplicaciones creadas por empresas u organizaciones que ofrecen un servicio o que facilitan algún aspecto de la vida diaria. En teoría, serían más generales, por ejemplo, aplicaciones de compras en línea, de viajes, de salud, de domótica, etc.
Productividad	Cualquier aplicación que ayude a la productividad personal del usuario o que tenga un fin educativo, como aplicaciones de edición de documentos, de lectura, calendarios, etc.
Multimedia	Aplicaciones que sirven para reproducir o editar contenido multimedia. Por ejemplo, aplicaciones de edición de fotografías o vídeos, servicios de música en <i>streaming</i> , visualización de contenidos en línea, reproductores, cámaras, etc.

Tabla 2. Tabla con la clasificación de las aplicaciones que integran el corpus

A pesar de que las clasificaciones existentes en los distintos sitios de descarga de aplicaciones suelen ser más exhaustivas, hemos optado por no crear demasiadas categorías, pues en tal caso se complicaría la tarea de crear un corpus de tamaño considerable para cada una de las clasificaciones, teniendo también en cuenta que no todas las aplicaciones que existen tienen por qué tener una traducción de calidad la cual podamos emplear para entrenar el motor. Además, no hemos incluido dentro de este estudio los videojuegos para dispositivos móviles, pues estos presentan diferencias con respecto al resto de aplicaciones y hemos considerado que se incluirían dentro de la localización de videojuegos dedicados a una plataforma concreta, como pueden ser PC, web o consolas.

En cuanto a los motores creados, el primero de ellos es específico de la tipología textual, integrado por el corpus bilingüe completo que incluye todas las aplicaciones móviles que hemos recopilado a lo largo del proceso de búsqueda de recursos. El segundo motor se centra en un tipo en concreto de aplicaciones; en este caso hemos seleccionado la categoría «comunicación» para el entrenamiento mediante un corpus formado por las aplicaciones de este tipo y repetido cuatro veces con el fin de aumentar su tamaño, para así ver si de esta manera se mejoran los resultados en la traducción de una aplicación de esta categoría. El último 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.

4.2. Búsqueda de recursos

Para entrenar los motores de TAE, lo primero que debemos hacer es recopilar un corpus bilingüe para poder entrenar el motor, el cual debe estar formado por textos originales con sus respectivas traducciones. Pero, además, este corpus debe tener un gran tamaño en cuanto a número de palabras para poder obtener unos resultados más fiables, pues, ya que este tipo de motores de traducción se basan por completo en los corpus para generar las traducciones, cuantos más recursos bilingües tenga, mayor será la probabilidad de que el resultado sea adecuado. La necesidad de encontrar una gran cantidad de recursos no comporta un gran problema en nuestro caso, pues hoy en día cualquier persona tiene a su alcance una gran cantidad de aplicaciones móviles cuyo idioma de partida suele ser el inglés. Además, el español es un idioma muy extendido en todo el mundo, por lo que la gran mayoría de aplicaciones cuyo idioma original es el inglés están disponibles traducidas al español. Por tanto, podemos afirmar que la cantidad de recursos a los que tenemos acceso para entrenar un motor con corpus de aplicaciones móviles es ideal en cuanto a volumen.

En este caso hemos optado por centrarnos en las aplicaciones de la plataforma Android por varios motivos. En primer lugar, es el sistema operativo más empleado a nivel mundial, cuenta con aproximadamente un 75 % de la cuota de mercado según StatCounter (2019). Además, la descarga de estas aplicaciones es muy sencilla. Los corpus paralelos de este estudio están compuestos por los archivos originales y traducidos de distintas aplicaciones móviles que están disponibles en la Play Store de Google y en otros portales de descarga, así como las cadenas de texto de la interfaz de usuario de las aplicaciones nativas que forman parte de Android. Este sistema operativo, además de ser el más usado en la actualidad en dispositivos móviles, también es de código abierto, lo que quiere decir que cualquier persona que así lo desee puede acceder a los archivos del proyecto ([Android Open Source Project](https://android.googlesource.com/)) y descargarlos¹.

No obstante, hay un aspecto que debemos tener en cuenta, y es que se debe procurar que los recursos que empleamos para el entrenamiento del motor tengan cierta calidad en la traducción. Esto nos plantea un problema en la búsqueda de recursos, pues hoy en día cualquier persona puede crear una aplicación con los conocimientos necesarios y una

¹ Archivos disponibles en el repositorio de Google Git al que se puede acceder a través del siguiente enlace: <https://android.googlesource.com/>.

inversión reducida, sin necesidad de que esta tenga una cierta calidad lingüística. En ocasiones, los desarrolladores de aplicaciones no disponen de los medios necesarios para traducir su aplicación contratando a profesionales de la localización nativos y que proporcionen un resultado de una calidad lingüística aceptable, por lo que optan por métodos menos costosos como pueden ser la traducción automática o recurrir a personas que no son traductores profesionales para realizar esa labor.

Por estos motivos, a la hora de descargar las aplicaciones y con el fin de reunir los recursos lingüísticos necesarios, se han establecido una serie de criterios que nos permitan tener una cierta seguridad de la calidad lingüística de los textos recopilados. Según estos criterios hemos ido seleccionando y descargando las aplicaciones que forman el corpus; dichos criterios son los siguientes:

- ◁ Aplicaciones desarrolladas por empresas de un cierto tamaño, como Google LLC o Microsoft, ya que disponen de los medios necesarios para invertir en la localización de sus productos. Además, estas empresas tienen experiencia en el sector y son conscientes de la importancia y las implicaciones de la calidad de sus traducciones, así como la repercusión que estas tienen en la satisfacción de los usuarios.
- ◁ Aplicaciones que cuenten con un determinado número de descargas; en este caso hemos establecido un límite aproximado de un millón de descargas.
- ◁ Aplicaciones más rentables según la Play Store.
- ◁ Aplicaciones gratuitas que cuentan con una versión de pago.

Estos criterios sirven para establecer unas bases a la hora de realizar la búsqueda de recursos y filtrar para intentar que los recursos utilizados en este estudio sean fiables, sin embargo, debemos tener en cuenta que nunca podemos estar seguros por completo de la calidad lingüística de una aplicación a menos que se analicen en profundidad los archivos con las cadenas de texto.

A la hora de descargar una aplicación móvil para Android, la forma más habitual es hacerlo directamente desde nuestro terminal móvil a través de la Play Store de Google, pero si lo que queremos es descargarla desde otro dispositivo como desde un ordenador, también existe la opción de descargarlas en formato APK. A nuestro alcance contamos con distintos sitios web desde donde es posible descargar infinidad de aplicaciones

[sitio web](#)⁴, así como los comandos que hay que emplear para llevar a cabo todas las acciones que nos ofrece.

Para poder descompilar cualquier aplicación APK mediante Apktool, primero hay que abrir la carpeta donde se encuentran los archivos que se quieren descompilar desde la línea de comandos. Una vez que nos encontramos dentro de ese directorio, el comando empleado para descompilar una aplicación (en este caso la de Amazon) es el siguiente:

```
$ apktool d amazon - shopping - 18- 4- 0- 100.apk -s
```

Para descompilar la aplicación debemos escribir `apktool d` y, a continuación, el nombre completo del archivo. Como podemos ver, a la hora de descompilar las aplicaciones, se ha optado por emplear la opción `-s` que ofrece la herramienta para no descompilar los archivos `.dex` (archivos ejecutables de las aplicaciones), pues no son necesarios para nuestro objetivo y de este modo se agiliza también el proceso. Cuando esta ha terminado de descompilarse, se nos crea una carpeta con el mismo nombre en el mismo directorio que incluye todos los archivos de la aplicación. La estructura de la carpeta que se crea es la siguiente:

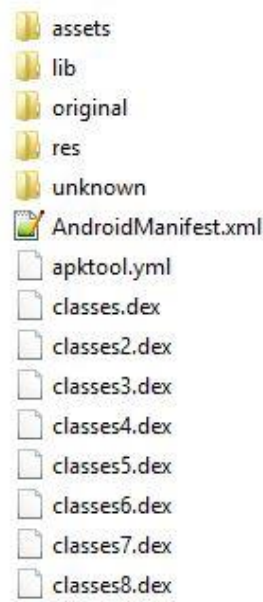


Figura 7. Ejemplo de la estructura de carpetas que se crea al descompilar el archivo APK de una aplicación

⁴ Wiśniewski, R., & Tumbleson, C. (3 de marzo de 2019). Apktool (v 2.4.0) [Software]. Recuperado el 8 de febrero de 2019, de <https://ibotpeaches.github.io/Apktool/>

En la carpeta *res* se encuentran todos los archivos con las cadenas de texto de dicha aplicación, así como imágenes y otros recursos. Dentro de esta carpeta hay a su vez diversas carpetas *values* con un sufijo que se corresponde al idioma y que incluyen los archivos de ese idioma. Esta estructura de carpetas varía en función de cada aplicación, mientras algunas incluyen una única carpeta para español, que corresponde con el español neutro (*values-es*), otras incluyen también carpetas con la configuración regional (*values-es-rES*), esto se debe a que el proceso de localización varía en función de los intereses de cada desarrollador.

Así pues, se han extraído todos los archivos relevantes para su posterior alineación y creación de los corpus. Además, las aplicaciones de las que se han extraído los archivos se han incluido en una tabla descriptiva (véase Anexo 1. Composición de corpus) para clasificar cada aplicación según su contenido en las categorías mencionada en el apartado anterior: herramientas, productividad, servicios/estilo de vida, comunicación y multimedia.

4.3. Creación de los corpus paralelos

Una vez clasificados todos los recursos, el siguiente paso es crear el corpus bilingüe. El objetivo es obtener diversas memorias de traducción (una para cada tipo de aplicación) en formato TMX, ya que este es compatible con la herramienta de entrenamiento MTradumàtica que emplearemos más adelante.

En un principio, se barajaron diversas opciones para llevar a cabo la alineación, como emplear el alineador de OmegaT o MemoQ, puesto que incluyen filtros para los archivos *strings.xml*, pero finalmente se descartó el uso de los alineadores integrados en herramientas de traducción asistida debido a sus limitaciones. El de OmegaT en concreto daba resultados considerablemente buenos, ya que tiene la opción de alinear los segmentos en función del identificador, sin embargo, te obliga a alinear siempre por pareja de archivos; así que, debido a la gran cantidad de archivos que habría que añadir manualmente, no es viable esta opción, pues supondría una gran carga de tiempo que se podría emplear en otras tareas como la obtención de más recursos.

Así pues, la herramienta seleccionada para llevar a cabo la alineación ha sido la de Okapi Rainbow, que también es de código abierto y totalmente gratuita. Okapi Rainbow

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 y basándose en los identificadores de cada uno, es decir, une aquellas cadenas que cuentan con el mismo identificador en inglés y en español (en caso de que una cadena en inglés no tenga una correspondiente en español, la omite).

Para comenzar la alineación simplemente hay que añadir todos los archivos originales a la lista 1 (*Input List 1*) y los archivos traducidos correspondientes a la lista 2 (*Input List 2*), y, a continuación, asociar todos al filtro de archivo que les corresponde haciendo clic en *Input > Edit Document Properties*. La herramienta reconoce directamente que se trata de un archivo XML, pero también ofrece filtros específicos para distintos tipos de XML; en este caso se ha elegido el filtro de Android: *okf_xml-AndroidStrings*.

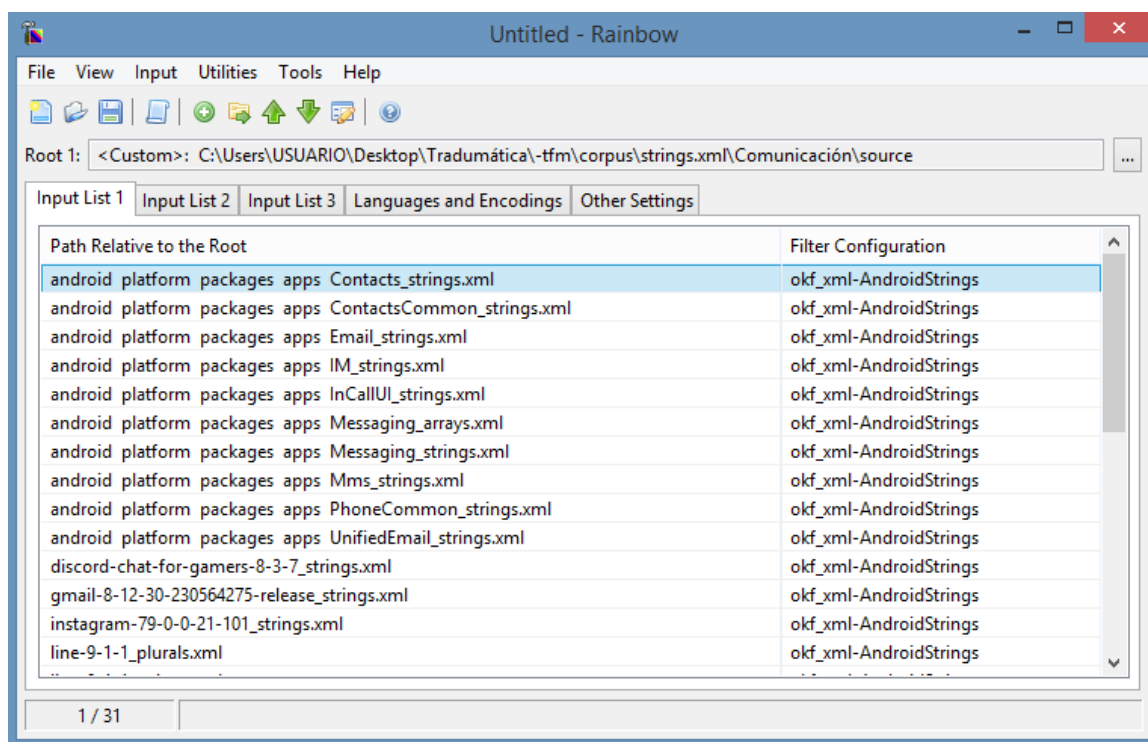


Figura 8. Interfaz de Okapi Rainbow con los archivos de las aplicaciones de «comunicación»

A continuación, es necesario elegir los idiomas de cada una de las listas y la codificación de los archivos en la pestaña *Languages and Encodings* y, una vez hecho esto, solo hay que hacer clic en la opción *ID-Based Alignment*, seleccionar el directorio donde queremos que se guarde la memoria y, tras unos segundos, se completa la

alineación y se crea el archivo TMX. En este caso, hemos creado cinco memorias de traducción pues cada una de ellas se corresponde con uno de los cinco tipos de aplicaciones que hemos establecido en la clasificación previa.






 Comunicacion-TMX.tmx	26/02/2019 0:59	TMX Translation Memory	1.742 KB
 Herramientas_TMX.tmx	01/03/2019 1:21	TMX Translation Memory	11.926 KB
 Multimedia-TMX.tmx	26/02/2019 0:57	TMX Translation Memory	6.288 KB
 Productividad-TMX.tmx	26/02/2019 0:57	TMX Translation Memory	8.319 KB
 Servicios_estilo-de-vida_TMX...	26/02/2019 0:56	TMX Translation Memory	9.110 KB

Figura 9. Memorias de traducción en formato TMX creadas tras la alineación

4.4. Limpieza de las memorias de traducción

Una vez creados los archivos TMX correspondientes a cada categoría de aplicaciones, lo que hemos hecho ha sido realizar la limpieza de los mismo y eliminar todos los elementos innecesarios que no interesan en el entrenamiento de los motores. Para realizar esta limpieza, hemos empleado la herramienta Okapi Olifant y el editor de texto Notepad ++.

Okapi Olifant es una herramienta que permite gestionar memorias de traducción TMX y, con ella, hemos separado aquellos segmentos que eran demasiado largos en dos o más manualmente, puesto que, como se señala en el manual de Moses (Koehn, 2019), los segmentos bilingües demasiado largos pueden dar problemas en el entrenamiento de los motores y así se agiliza el proceso. La existencia de segmentos tan largos es resultado de la alineación basada en identificadores de Okapi Rainbow, pues no tiene en cuenta reglas de segmentación a no ser que se añada un archivo SRX adicional. También se han tenido que eliminar segmentos vacíos y segmentos que no eran relevantes para nuestro propósito, por ejemplo, segmentos en otros idiomas

Además, los archivos de las aplicaciones nativas de Android, para marcar el texto que no se debe traducir, utilizan la etiqueta <xliff:g> con el fin de evitar que los traductores puedan modificarlo. Okapi Olifant te da la opción de eliminarlas de manera muy sencilla en Entries>Remove Codes... y, a continuación, seleccionando «Remove TMX tags, keep their content» para eliminar las etiquetas, pero no su contenido.

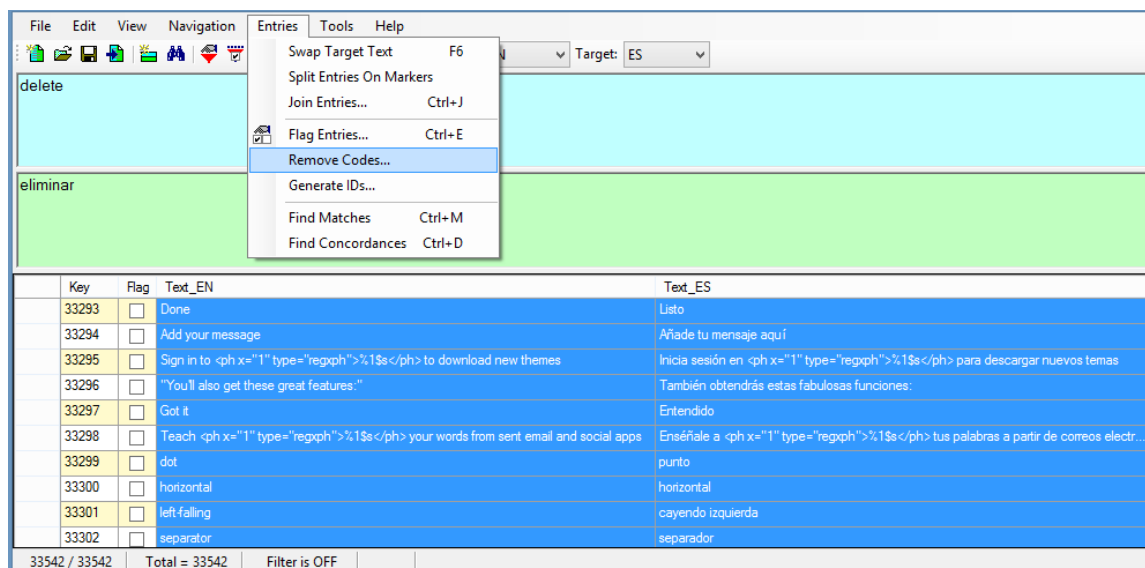


Figura 10. Vista del programa Okapi Olifant al abrir una memoria TMX

Una vez hecho esto, y tras percatarnos de todo aquello presente en los corpus y que es innecesario, hemos llevado a cabo una serie de búsquedas y sustituciones sobre los archivos TMX de las memorias, primero en Okapi Olifant y luego en Notepad ++, para eliminar los siguientes elementos: saltos de línea y espacios innecesarios; comillas dobles que encierran algunos segmentos; espacios al comienzo y al final de segmento; y las etiquetas HTML de formato.

Hemos eliminado las etiquetas HTML puesto que no son relevantes para el entrenamiento del motor, ya que, durante la fase de segmentación en MTradumàtica, todas las palabras se separan de los signos de puntuación y símbolos, y la presencia de tantos símbolos y palabras que no tienen relación con la frase que se pretende traducir afectaría negativamente al entrenamiento.

El proceso que hemos seguido ha sido el siguiente:

1. Hemos buscado en Okapi Olifant todos los saltos de línea marcando la casilla de expresiones regulares e introduciendo en la búsqueda **\n**, y hemos sustituido estos saltos de línea por un espacio.
2. Asimismo, los saltos de línea que usan la etiqueta HTML **
** (
) los hemos reemplazado por espacios.
3. Después hemos introducido en la caja de búsqueda dos espacios y los hemos sustituido por un único espacio, repitiendo esta operación las veces que sea necesaria hasta que ya no queden dobles espacios.

4. En muchas ocasiones, mientras un segmento estaba rodeado por comillas dobles ("..."), el segmento equivalente no. Esto se debe a que las comillas dobles sirven para escapar ciertos caracteres que tienen un uso especial o, por ejemplo, para conseguir que los espacios que se insertan al final o al principio del segmento o espacios adicionales no desaparezcan cuando se muestre la cadena en pantalla. Esta falta de uniformidad en la presencia de comillas hace que sea necesario eliminarlas y, para ello, se ha buscado en Notepad ++ `<seg>`" (comillas de inicio de un segmento) para sustituirlo por `<seg>` y, a continuación, `"</seg>` (comillas de cierre), para sustituirlo por `</seg>`.
5. Para buscar los espacios situados al principio de segmento basta con buscar en Notepad ++, con la función Expresión regular marcada, `<seg>\s` (donde `\s` puede corresponderse con un espacio, un carácter de nueva línea o una tabulación) y sustituir por `<seg>`. Esta operación se repite las veces que sea necesario hasta que ya no haya coincidencias.
6. Para el espacio de final de segmento, hay que buscar `\s</seg>` y sustituirlo por `</seg>`. Se repite las veces que sea necesario.
7. Por último, también hemos eliminado todas las etiquetas HTML presentes en los textos originales y traducidos. Para eliminarlas, hemos buscado `<.*?>` en Notepad ++ con el modo de búsqueda Normal marcado y los hemos eliminado todos (los de cierre y apertura) dejando el cuadro «Reemplazar con» vacío. Aquí, `<` sería la entidad de `<`, la expresión regular `.*?` busca cualquier carácter que se repita una o más veces, y `>` es el cierre de la etiqueta `>`. Al estar escritas con la entidad, solo eliminamos las etiquetas del contenido y evitamos eliminar también las etiquetas del archivo TMX.

Una vez llevado a cabo este proceso, hemos obtenido los corpus finales que se emplearán para el entrenamiento de los motores. En la siguiente tabla podemos ver el número de segmentos y palabras de cada uno de ellos.

Corpus	N.º de segmentos	N.º de palabras	
		inglés	español
Comunicación	53 776	239 331	267 265
Multimedia	24 819	107 055	122 233
Productividad	29 269	135 638	159 104
Servicios/estilo de vida	53 840	259 901	292 646
Herramientas	50 064	220 068	255 521
Total:	211 768	961 993	1 096 769

Tabla 3. Tabla de la composición de cada uno de los corpus tras el proceso de limpieza

4.5. Creación de los motores mediante MTradumàtica

Una vez hemos realizado la limpieza de todos los corpus que vamos a emplear en el entrenamiento de los motores, el siguiente paso es la creación de los motores con MTradumàtica⁵. Para comenzar a trabajar, lo primero que debemos hacer es acceder a la plataforma con una cuenta de Google, de tal manera que nuestros archivos y motores estarán asociados a nuestra cuenta.

La versión actual de MTradumàtica admite tanto corpus en formato de memorias de traducción TMX como archivos de texto plano, algo que antes no ocurría, pues solo se podían introducir los archivos en formato Moses, formato de texto plano segmentado línea a línea y donde cada idioma se incluye en un archivo diferente. Esta funcionalidad es muy útil, ya que el TMX es un formato mucho más usado dentro del mundo de la traducción y es más sencillo de crear, además, es el formato más frecuente de exportación de los proyectos de alineación en las distintas herramientas de alineación que existen. Por tanto, no tendremos la necesidad de realizar la conversión de los archivos para entrenar el motor.

Para entrenar un motor, el proceso que deberemos llevar a cabo será el siguiente:

1. Subida de los archivos: se suben todos los archivos con los que se van a entrenar los motores.

⁵ Se puede acceder a la plataforma desde el siguiente enlace: <https://mtradumatica.uab.cat/>.

2. Creación de los monotextos: se crean los corpus monolingües a partir de los archivos subidos en el paso anterior y que se emplearán para crear los modelos de lengua del paso siguiente.
3. Creación de los modelos de lengua (ML): se crean los modelos de lengua, es decir, modelos estadísticos creados con corpus monolingües que empleará el motor para mejorar la fluidez del resultado de traducción en el idioma meta, en este caso en español.
4. Creación de los bitextos: se crean los corpus bilingües a partir de los archivos subidos en el paso número 1 y que se emplearán para entrenar los motores de traducción.
5. Creación de los traductores: en este paso se crean los motores de TAE a partir de los modelos de lengua y los bitextos creados que se emplean para crear el modelo de traducción.

En primer lugar, las memorias de traducción en formato TMX con las que vamos a entrenar el motor hay que introducirlas en el apartado Ficheros y, una vez se han cargado, la plataforma separa los dos idiomas de forma automática como podemos ver en la siguiente captura de pantalla:

<input type="checkbox"/>	Nombre	Idioma	Líneas	Words (unique)	Caracteres
<input type="checkbox"/>	Comunicacion-TMX.tmx.es	es	53776	267265 (10305)	1670848
<input type="checkbox"/>	Comunicacion-TMX.tmx.en	en	53776	239331 (6803)	1430685

Figura 11. Separación de los archivos TMX en el apartado Ficheros de MTradumàtica

Esta pestaña también nos permite ver el número de líneas, de palabras, de palabras únicas (entre paréntesis) y caracteres de cada archivo que hemos subido. Aquí también se añaden el corpus monolingüe que emplearemos posteriormente para el entrenamiento del motor que incluye un modelo de lengua genérico, pero debemos tener en cuenta que el tamaño máximo de subida de cada archivo es de 1024 MiB, por lo que hemos tenido que separar algunos de los corpus monolingües descargados en varios archivos de texto, pues superaban este límite. Al tratarse de archivos tan grandes no podemos abrirlos con

cualquier editor de texto, por lo que necesitaremos alguno que permita abrir archivos de gran tamaño como EditPad para poder separarlos en varios.

A continuación, en el apartado Monotextos, podemos crear, a partir de los archivos subidos anteriormente, los corpus monolingües que se emplearán para generar los modelos de lengua. En este apartado únicamente es necesario crear los monotextos para el idioma de llegada, en este caso el español. Como podemos ver en la siguiente captura de pantalla, también tenemos información sobre el número de líneas de cada uno de los corpus monolingües:

<input type="checkbox"/>	Nombre	Idioma	Líneas
<input type="checkbox"/>	Corpus-específico-aplicaciones_ES	es	211768

Figura 12. Monotexto creado para el motor de traducción específico de la tipología textual

El siguiente paso consiste en entrenar los modelos de lengua mediante los corpus monolingües creados en el paso anterior. Aquí simplemente tenemos que poner nombre al modelo de lengua, seleccionar el monotexto que queremos que emplee en su entrenamiento y elegir el idioma. Entonces, una vez ha pasado el tiempo necesario, el modelo de lengua ya estará listo.

Seguidamente, en la pestaña Bitextos, debemos crear los corpus bilingües con los que entrenaremos más adelante los modelos de traducción. En este apartado debemos señalar la dirección de cada bitexto (inglés>español) y añadir los corpus subidos previamente en el apartado Ficheros.

Como último paso, a partir de los corpus bilingües y los modelos de lengua creados en pasos anteriores, entrenaremos los modelos de traducción en el apartado Traductores. Aquí es donde crearemos los tres motores de traducción que mencionábamos en el apartado 4.1 un motor genérico y especializado, uno especializado en la traducción de aplicaciones de cualquier tipo y otro especializado únicamente en la traducción de aplicaciones de la categoría comunicación.

<input type="checkbox"/>	Motor especializado apps comunicación	en-es	Corpus-especializado-comunicación_EN-ES	ML_específico-aplicaciones_ES
<input type="checkbox"/>	Motor especializado aplicaciones	en-es	Corpus-especializado-aplicaciones_EN-ES	ML_específico-aplicaciones_ES
<input type="checkbox"/>	Motor genérico+especializado aplicaciones	en-es	Corpus-especializado-aplicaciones_EN-ES	ML_genérico+específico_ES

Figura 13. Motores de TAE creados con MTradumàtica

A continuación, podemos ver las tablas de la composición de cada uno de los motores, con el número de segmentos de cada uno de ellos. En el caso de los corpus monolingües en español, todos ellos se han obtenido a través del sitio web de OPUS (Tiedemann, 2012).

Motor especializado en aplicaciones + genérico	
Corpus monolingüe: ES	Corpus bilingüe EN-ES
Corpus de todas las aplicaciones:	Corpus de todas las aplicaciones:
§ Comunicacion-TMX.tmx.ES	§ Comunicacion-TMX.tmx
§ Multimedia-TMX.tmx.ES	§ Multimedia-TMX.tmx
§ Productividad-TMX.tmx.ES	§ Productividad-TMX.tmx
§ Servicios_estilo-de-vida_TMX.tmx.ES	§ Servicios_estilo-de-vida_TMX.tmx
§ Herramientas_TMX.tmx.ES	§ Herramientas_TMX.tmx
Corpus general:	
§ DGT_es.txt . DGT es un conjunto de memorias de traducción de textos legislativos de la UE.	
§ EUROPARL_V7_es.txt . Corpus extraído de la web del Parlamento Europeo.	
§ Globalvoices.es.txt . Corpus procedente de noticias publicadas en el sitio web Global Voices.	
§ EUbookshop_es.txt , EUbookshop_es_2.txt y EUbookshop_es_3.txt . Corpus formado	

por textos de la biblioteca en línea de publicaciones de la UE.	
<p>Número de segmentos:</p> <p>14 534 589 (corpus monolingüe) y 211 768 (corpus bilingüe)</p>	

Tabla 4. Tabla descriptiva del motor especializado en aplicaciones + genérico

Motor especializado en aplicaciones	
Corpus monolingüe: ES	Corpus bilingüe EN-ES
Corpus de todas las aplicaciones:	Corpus de todas las aplicaciones:
§ Comunicacion-TMX.tmx.ES	§ Comunicacion-TMX.tmx
§ Multimedia-TMX.tmx.ES	§ Multimedia-TMX.tmx
§ Productividad-TMX.tmx.ES	§ Productividad-TMX.tmx
§ Servicios_estilo-de-vida_TMX.tmx.ES	§ Servicios_estilo-de-vida_TMX.tmx
§ Herramientas_TMX.tmx.ES	§ Herramientas_TMX.tmx
<p>Número de segmentos:</p> <p>211 768 (corpus bilingüe y corpus monolingüe)</p>	

Tabla 5. Tabla descriptiva del motor especializado en aplicaciones

Motor especializado en aplicaciones: comunicación	
Corpus monolingüe: ES	Corpus bilingüe EN-ES
Corpus de todas las aplicaciones:	Corpus de las aplicaciones de la categoría comunicación cuatro veces:
§ Comunicacion-TMX.tmx.ES	§ Comunicacion-TMX.tmx
§ Multimedia-TMX.tmx.ES	§ Comunicacion-TMX.tmx
§ Productividad-TMX.tmx.ES	§ Comunicacion-TMX.tmx
§ Servicios_estilo-de-vida_TMX.tmx.ES	§ Comunicacion-TMX.tmx
§ Herramientas_TMX.tmx.ES	§ Comunicacion-TMX.tmx
<p>Número de segmentos:</p> <p>211 768 (corpus monolingüe) y 215 104 (corpus bilingüe)</p>	

Tabla 6. Tabla descriptiva del motor especializado en aplicaciones de comunicación

4.6. Elección de la muestra para realizar el análisis

Para realizar la evaluación de la calidad se ha tenido que seleccionar una aplicación con el fin de utilizarla como muestra. Hemos optado por realizar una evaluación manual para tener una imagen más profunda del rendimiento de cada uno de los motores. Para

ello, hemos seleccionado una aplicación que fuera representativa y tuviera cierta similitud con la composición de los corpus que forman los motores de traducción y, de este modo, hemos seleccionado la aplicación Signal⁶, una aplicación de mensajería instantánea y llamadas de código abierto. Hemos elegido esta aplicación dado que a la hora de entrenar el motor específico de un tipo de aplicaciones habíamos elegido las aplicaciones de la categoría comunicación. Además, nuestro corpus incluye diversas aplicaciones similares de mensajería, por lo que, a la hora de seleccionarla, hemos tenido en cuenta también este aspecto.

Además, hemos optado por no realizar la evaluación de la TA de todas las cadenas de la aplicación, ya que supondría invertir una gran cantidad de tiempo en llevar a cabo dicha tarea. En su lugar, hemos seleccionado una muestra representativa de segmentos teniendo en cuenta una serie de factores que veremos a continuación.

Se han realizado dos tipos de análisis a la hora de comparar los resultados de los tres motores creados. Por un lado, se han extraído una serie de segmentos en relación a aspectos problemáticos de la localización de aplicaciones para Android que hemos observado tras analizar esta aplicación. Estos aspectos son la presencia de variables, de caracteres escapados con la contrabarra, de saltos de línea mediante la secuencia de escape `\n` y la traducción de segmentos de una cierta longitud. Los segmentos los hemos comparado entre sí y también con la traducción que ofrece el traductor de Deepl para dichos segmentos. De este modo, hemos clasificado el resultado de cada motor para los tres primeros problemas otorgando una puntuación del 1 al 4 en función de cómo aborda cada motor dichos aspectos con respecto al nivel de posesición necesario:

1. No aprovechable, requiere un mayor número de ediciones que traducirlo desde cero.
2. Poco aprovechable, requiere gran número de ediciones y puede hacer dudar al traductor.
3. Bastante aprovechable.
4. Totalmente aprovechable, no necesita ser editado.

En los tres primeros casos (variables, apóstrofes escapados y saltos de línea) no se han tenido en cuenta otros aspectos como la adecuación o la fluidez del resto de la frase, pues lo que pretendemos observar es cómo aborda cada uno de los motores estos aspectos

⁶ URL: <https://signal.org/es/>.

en concreto. En el caso de las oraciones largas sí que hemos puntuado tanto adecuación como fluidez siguiendo la misma escala (del 1 al 4) para cada una, pero, en este caso, la descripción de cada puntuación varía: 4. Totalmente fluido/adecuado, 3. Bastante fluido/adecuado, 2. Poco fluido/adecuado, 1. Nada fluido/adecuado.

- a) **Variables.** En cuanto a la presencia de variables, los principales problemas que se nos presentan para la traducción automática son la concordancia y la posición de la variable con respecto al resto de elementos de la frase, puesto que el motor de traducción no las reconoce como un solo elemento, sino como una secuencia de símbolos que, además, se separan el uno del otro durante la fase de segmentación.
- b) **Caracteres escapados.** Los problemas con caracteres escapados se presentan sobre todo en el caso de los apóstrofes en inglés, pues se emplean bastante a menudo. A pesar de que los apóstrofes en Android también se pueden escapar colocando la cadena de texto completa entre comillas, algunos desarrolladores optan por hacerlo colocando antes la contrabarra ('\'). Esta diferencia en el tratamiento de este carácter está también presente en el corpus empleado para el entrenamiento de los motores, pues cada aplicación tiene su manera de actuar al respecto. Por ejemplo, en la versión de la aplicación de Signal que hemos empleado para el análisis se ha optado por la opción de usar la contrabarra. Así pues, hemos decidido prestar atención a la manera en la que los motores entrenados resuelven la aparición de este elemento.
- c) **Salto de línea.** En el caso de los saltos de línea introducidos mediante la secuencia de escape `\n`, se nos presenta el problema de que la letra *n* se encuentra pegada a la siguiente palabra, por lo que, en teoría, los motores de traducción no la separarían de la palabra posterior durante la fase de segmentación, pues se considera todo como una sola palabra, mientras que la contrabarra, al ser un símbolo, sí que se separaría del resto.
- d) **Oraciones largas.** También hemos considerado relevante analizar la traducción de oraciones largas (de más de diez palabras) debido a que el corpus que hemos empleado para el entrenamiento de los motores está compuesto en gran medida de segmentos cortos de pocas palabras y que, en muchas ocasiones, no constituyen siquiera una frase completa; por este motivo, creemos que la presencia de oraciones de mayor tamaño afectará a la calidad de la traducción obtenida.

Por otro lado, una vez llevado a cabo este primer análisis, hemos realizado una comparación de los tres motores entrenados gracias a la plataforma DQF de TAUS para

determinar qué composición de corpus sería la ideal. Para ello hemos incluido los segmentos empleados en el análisis de aspectos problemáticos y hemos añadido algunos más para tener una imagen más global y representativa de la aplicación. Hemos tenido en cuenta que no hubiera segmentos repetidos ni muy parecidos entre sí, que los segmentos tuvieran longitud variable y que incluyeran elementos representativos de las aplicaciones móviles (ya sean vocabulario, estructuras sintácticas, variables, etc.).

Para obtener las traducciones de la muestra, hemos hecho uso de la herramienta OmegaT, ya que cuenta con un filtro de archivos específico para los archivos XML de aplicaciones Android y ofrece la posibilidad de añadir un motor de traducción automática de tipo Moses como el de MTradumàtica, siempre y cuando este disponga de una URL a la que conectarse. 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. También se planteó la posibilidad de emplear Memsource debido a la funcionalidad que ofrece para modificar el filtro de archivos XML, ya que permite establecer mediante expresiones regulares elementos como las variables para que se transformen en etiquetas, sin embargo, no es posible conectar esta plataforma a MTradumàtica, por lo que se descartó su uso.

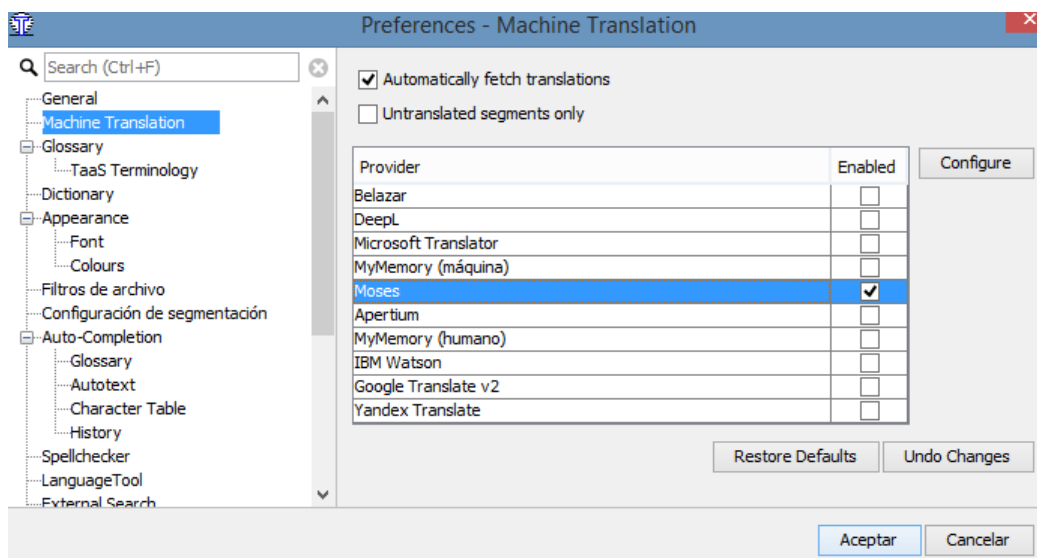


Figura 14. Configuración de MTradumàtica en OmegaT

5. Análisis de los resultados

En este apartado veremos el resultado de las dos partes del análisis que se ha llevado a cabo. En primer lugar, en el apartado 5.1, expondremos y explicaremos los resultados del análisis de los aspectos problemáticos mencionados en el apartado 4.6, comparando los tres motores creados con el de DeepL. Por su parte, en el apartado 5.2, se comentarán los resultados de la comparación de los tres motores realizada a través de la herramienta DQF. A modo de referencia, se han incluido en el Anexo 2. Segmentos extraídos de la aplicación procesada por los motores todos los segmentos de la aplicación de Signal originales y traducidos empleados en ambos análisis.

5.1. Aspectos problemáticos

Para realizar el análisis, se han incluido todos los segmentos seleccionados en una tabla, se han clasificado según el problema y se les ha ido otorgando una puntuación con un baremo del 1 al 4, siendo 4 el que ofrece un mejor resultado y 1 un peor resultado de traducción (1. No aprovechable, 2. Poco aprovechable, 3. Bastante aprovechable, 4. Totalmente aprovechable). Asimismo, se han analizado las posibles causas de los fallos que se han observado para poder sacar una serie de conclusiones que veremos a continuación.

a) Variables

Hemos seleccionado un total de doce segmentos que incluían variables para comprobar cómo aborda su traducción cada motor. En la siguiente tabla podemos ver el número de segmentos de cada motor según la puntuación obtenida:

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 7. Puntuación de los segmentos seleccionados para el análisis de las variables

De estos resultados podemos deducir que el motor que ha proporcionado unos mejores resultados en el tratamiento de las variables ha sido el motor especializado en aplicaciones, en el que diez de los doce segmentos analizados se considera que están bien traducidos y se pueden aprovechar la traducción en gran medida. El motor que peores resultados ha dado ha sido el de DeepL, con nueve segmentos calificados de no aprovechables, mientras que los dos restantes han obtenido también unos resultados bastante aceptables.

Los tres motores especializados que hemos entrenado, por lo general, han dado buenos resultados con respecto a concordancia, colocación adecuada dentro de la oración y un uso adecuado de las preposiciones. Por ejemplo, los tres motores han traducido «Saving %1\$d attachments» como «Guardando % 1 \$ d archivos adjuntos», mientras que DeepL lo ha traducido por «Ahorro %1\$d de archivos adjuntos». Esto se puede deber a que los motores se han entrenado con corpus que también incluían numerosas variables, así que, a la hora de reordenar la oración, tienen referencias de elementos similares dentro de oraciones y son capaces de realizar elecciones léxicas más adecuadas. El único problema repetitivo que presentan estos tres motores es que siempre devuelven la variable con todos sus 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. Esto ralentizaría en cierta medida el proceso de posesición, por ello, a todos los segmentos que han dado muy buenos resultados se les ha otorgado una puntuación de 3 (bastante aprovechable) en lugar de 4 (totalmente aprovechable). No obstante, este proceso de eliminación de espacios se podría automatizar en un caso real de posesición con el fin de agilizar el proceso, por ejemplo, mediante el uso de búsquedas y sustituciones una vez se ha finalizado la traducción.

Por su parte, el motor de DeepL sí que devuelve las variables sin espacios, pero tiende a realizar elecciones léxicas erróneas o colocar las variables en lugares de la oración erróneos que pueden provocar un cambio de significado (p. ej., ha traducido «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 y ha cambiado el sentido original de la frase.

b) Apóstrofo

En el caso de los apóstrofes, cabe mencionar que en un primer momento observamos los resultados de las traducciones obtenidas de OmegaT y nos percatamos de que no había conseguido traducir ninguna palabra que llevara apóstrofo, es decir, aparecían exactamente igual que en el original. Sin embargo, al traducir los mismos segmentos en los motores dentro de la plataforma de MTradumàtica, sí que conseguía traducir dichas palabras. Esto se puede deber a que, dentro del propio editor de esta herramienta, al procesar un archivo con el filtro de archivos XML Android, el traductor solo ve en pantalla los apóstrofes (sin la contrabarra), pero una vez se ha terminado de traducir todo, en caso de que se haya incluido algún apóstrofo en el texto traducido, este aparecerá en archivos traducido como \ ' sin necesidad de haberlo escrito como tal. Por tanto, hemos realizado el análisis de los segmentos a partir de las traducciones obtenidas directamente de la plataforma en línea de MTradumàtica, no las de OmegaT, ya que se trata de un fallo provocado por el procesamiento del archivo en la herramienta TAO, y no por el propio proceso de entrenamiento de los motores.

Se han seleccionado un total de cuatro segmentos para analizar y los resultados arrojados han sido que el motor con mejores resultados es el de DeepL, cuyas traducciones de los cuatros segmentos han abordado perfectamente la presencia de los apóstrofes 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 8. Puntuación de los segmentos seleccionados para el análisis del apóstrofo escapado

Los tres motores especializados han obtenido exactamente las mismas puntuaciones, pues todos han abordado los segmentos de la misma manera. 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), lo motores entrenados no han sabido identificarlo como tal debido a que la contrabarra lo separa de la palabra a la que debería ir unido, así pues, los motores los han identificado como

elementos separados sin relación ninguna. Por su parte, el motor de DeepL sí que ha sabido identificarlo y, en consecuencia, lo ha traducido de manera perfecta.

c) Saltos de línea

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 9. Puntuación 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, separado de la palabra que le sigue. Por ejemplo, en el caso del segmento en el que aparece «\n\nContinue?», la traducción que nos proporcionan estos tres motores es « \ n \ ncontinue?». Por otro lado, DeepL tiene la desventaja de que, en el caso del ejemplo anterior, ha eliminado por completo la secuencia «\n\nContinue?», no aparecen ni los saltos de línea ni la traducción de la palabra. En el resto de segmentos, se le ha otorgado una puntuación de 3 (bastante aprovechable), pues ha conseguido traducir de manera correcta las palabras que van pegadas a la letra *n*, pero los saltos de línea (\n) o bien han desaparecido por completo, o solo se ha conservado la contrabarra en la traducción resultante.

d) Oraciones largas

Para comprobar cómo aborda cada traductor las oraciones de más de 10 palabras, hemos seleccionado un total de cinco oraciones y las hemos clasificado siguiendo la misma escala del 1 al 4, pero aplicándola a la fluidez y la adecuación de cada una. A continuación, podemos ver las tablas con las puntuaciones de cada motor para ambos parámetros:

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 10. Puntuación 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 11. Puntuación 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, un motor de traducción general en cuyo entrenamiento se emplean textos que contienen oraciones de mayor tamaño y con mayor continuidad y conexión entre unas y otras. Cabe destacar el hecho de que, en el caso del motor especializado con el modelo de lengua genérico, se ha conseguido aumentar la fluidez y no se ha valorado ningún segmento como nada fluido, pero esto no ha supuesto ningún cambio significativo en la adecuación obtenida con respecto al motor especializado en aplicaciones. Por su parte, el motor especializado en aplicaciones de comunicación ha sido el que peores resultados ha obtenido en ambos parámetros.

Bien es cierto que la cantidad de segmentos analizados en este apartado no ha sido muy elevada, pero nos permite dar una imagen general sobre cada uno de los motores y ver qué errores se nos pueden presentar, para así poder buscar soluciones y optimizarlos en el futuro.

5.2. 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*) a través de la plataforma en línea DQF de TAUS. Esta prueba se ha llevado a cabo con el fin de determinar qué composición de corpus sería la más adecuada en caso de que se quisiera seguir entrenando un motor especializado en la traducción de aplicaciones móviles a partir de los motores de los que ya disponemos. En esta prueba, se ha clasificado cada motor con un baremo del 1 al 3 (siendo 1 el mejor resultado y 3 el peor). En caso de que uno de los resultados fuera igual de bueno o igual de malo que otro, se ha otorgado la misma puntuación.

A la hora de calcular las puntuaciones de cada motor para decidir cuál ha sido el preferido, esta herramienta otorga una puntuación de 3 cada vez que un motor ha recibido una clasificación de 1, una puntuación de 2 cuando se lo clasifica con un 2, y una puntuación de 1 cuando se le otorga un 3. De este modo se suman todos los puntos para dar las puntuaciones globales de cada uno de los motores, y se calcula el porcentaje que representa cada una de ellas. Así pues, la clasificación global que hemos obtenido ha sido la siguiente:

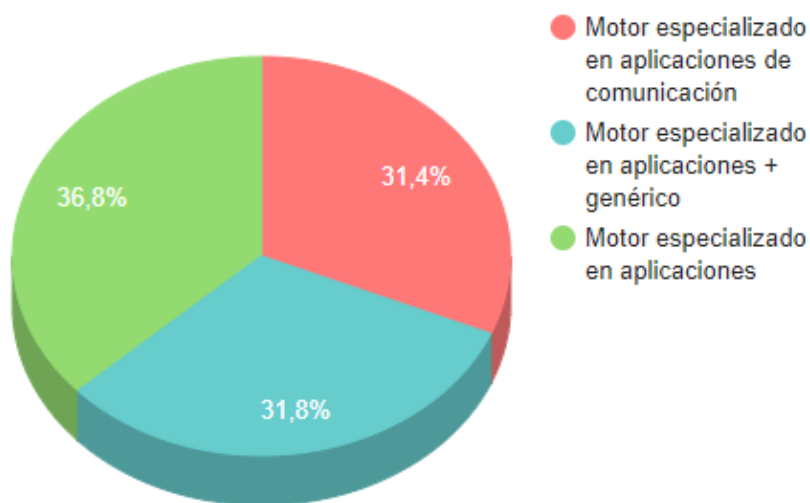


Figura 15. Gráfica de la clasificación de los motores entrenados

Como podemos ver en la gráfica, el motor especializado en aplicaciones (aquel que incluye el corpus de todos los tipos de aplicaciones) ha sido el que mejores resultados ha obtenido con un 36,8 %. En cuanto a los otros dos motores, ambos tienen un porcentaje

similar (31,4 % y 31,8 %). Cabe mencionar que no existen diferencias demasiado notorias entre los tres debido a que los tres motores comparten en cierta medida los corpus que los forman, pero son las diferencias en la composición de cada uno lo que nos hace decantarnos por la elección de uno u otro.

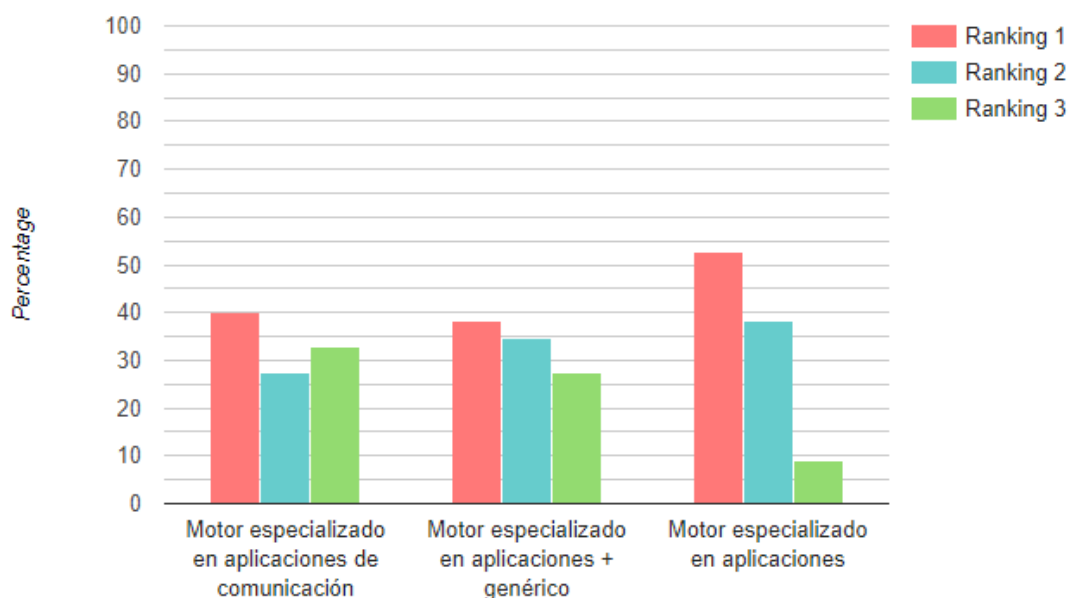


Figura 16 . Gráfica con las valoraciones otorgadas a cada motor

En el diagrama de barras anterior podemos observar cómo el motor especializado en aplicaciones ha sido el que ha recibido una clasificación de 1 en más ocasiones (en torno a un 52 % de las valoraciones que ha recibido) y, al mismo tiempo, ha sido el que ha obtenido una clasificación de 3 en menos ocasiones (tan solo un 9 % de las valoraciones). Por tanto, podemos afirmar que los resultados de este motor especializado en aplicaciones han sido bastante favorables y que, por tanto, sería la opción ideal para continuar con el entrenamiento de un motor especializado destinado a la traducción de aplicaciones móviles. Asimismo, estos resultados desmienten nuestra hipótesis inicial de que la utilización de un corpus que incluyera aplicaciones de cualquier tipo perjudicaría a la traducción de una aplicación perteneciente a una categoría concreta.

Por otro lado, a pesar de que los otros dos motores (especializado + genérico y especializado en 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 así conseguir 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 la ha empeorado, de este modo, lo que pensábamos que sería una mejora, ha resultado ser un detrimento.

Por su parte, como hemos mencionado, el motor especializado en aplicaciones de comunicación ha obtenido unos peores resultados que el especializado en todo tipo de aplicaciones. Esto puede deberse al hecho de que el corpus empleado para su entrenamiento, al ser más especializado, tiene un menor tamaño y, a pesar de que se ha intentado suplir esta falta de volumen en cierta manera con la introducción del mismo corpus cuatro veces, no ha sido suficiente para conseguir unos mejores resultados al traducir una aplicación de la misma categoría.

6. Conclusiones

A lo largo de este trabajo hemos comentado las especificidades que lleva asociada la localización de aplicaciones móviles y su relevancia dentro del mercado de la traducción, así como las ventajas que supondría la implementación de la traducción automática en el proceso de localización. Asimismo, nos hemos dado cuenta de la complejidad que conlleva el entrenamiento de motores de traducción personalizados según las necesidades específicas de cada uno, ya sean empresas o traductores autónomos que pretenden ampliar su abanico de recursos para emplear durante el ejercicio de su trabajo.

Por todo esto, es de gran ayuda la existencia de herramientas de entrenamiento de motores de traducción automática 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 en gran medida la tarea de entrenamiento otorgando una interfaz intuitiva y una serie de funcionalidades añadidas para ayudarnos en el entrenamiento de los motores. Por ejemplo, MTradumàtica dispone de la función Inspeccionar, donde nos permite ver cómo funciona el modelo de lengua, el modelo de traducción, el diccionario bilingüe probabilístico y los pasos que realiza el motor a la hora de traducir un fragmento de texto (desde la segmentación del texto de entrada, hasta el texto de salida). Por otro lado, sería interesante que MTradumàtica también ofreciera la opción de seleccionar qué

traducción es la más adecuada para cada grupo de palabras (n-grama) de entre aquellas que tienen una mayor probabilidad. También sería conveniente que se añadiera una funcionalidad que permita dividir archivos de más de 1024 MiB de forma automática al subirlos en el apartado Ficheros, para que no haya que hacerlo de forma manual usando un editor de texto externo para archivos de gran tamaño.

La integración del flujo de trabajo en OmegaT nos ha permitido también darnos cuenta de algunas carencias de esta herramienta. Por un lado, el hecho de que el filtro de XML de Android oculte los apóstrofes escapados es positivo, ya que permite ver el texto tal y como se vería en pantalla, y en caso de que el traductor escriba un apóstrofo, este ya aparecería en el archivo traducido con el carácter de escape. Sin embargo, el hecho de que al conectar el motor de traducción automática nos haya devuelto las palabras que contienen estos caracteres sin traducir es un aspecto que se podría mejorar para evitar que esto ocurra, pues puede provocar que se traduzca mal toda la frase y que no se pueda aprovechar la traducción automática del segmento traducido. Por otro lado, otro punto a mejorar es el hecho de que no disponga de una opción de pretraducción integrada, ya que, en caso de que se quisiera realizar una evaluación de traducción automática que incluyera un gran número de segmentos, habría que ir introduciendo la traducción uno a uno, tarea que llevaría mucho tiempo.

Asimismo, sería interesante incluir, ya sea en el propio motor de traducción o en la herramienta TAO, la opción de introducir elementos como las variables como una sola unidad, y no la suma de símbolos, números y letras separados. Esta opción ya existe en Memsource, donde se permite introducir una expresión regular al filtro de archivos XML para localizar los elementos que queramos (siempre que podamos definirlos con una expresión regular), y estos elementos se sustituyen por una etiqueta del programa. No obstante, MTradumàtica no se puede conectar con Memsource, por lo que también sería interesante que la plataforma ampliara las herramientas TAO en las que se puede integrar. En el caso de los saltos de línea (\n) también sería beneficiosa la posibilidad de indicarle a la herramienta de entrenamiento de motores de traducción que se trata de un único elemento que debe ir separado de la palabra que va pegada a continuación, para así evitar que dicha palabra se quede sin traducir.

Por último, los resultados obtenidos durante la fase de análisis nos han permitido comprobar que se pueden obtener unos resultados bastante aceptables mediante el

entrenamiento de un motor de TAE con corpus formados por aplicaciones móviles disponibles al alcance de cualquiera. A pesar de que se trata de una tarea que requiere de tiempo y de cierto conocimiento técnico, hay una gran cantidad de recursos disponibles. Además, hemos podido observar que la inclusión de un corpus monolingüe genérico no nos ayuda a ofrecer unos mejores resultados, ya que las características textuales de las aplicaciones son diferentes a las de otro tipo de textos, así pues, la presencia de estos textos en el corpus de entrenamiento del modelo de lengua, solo ha servido para empeorar los resultados.

7. Bibliografía

- Android Open Source Project. (s.f.). *Developer Guides*. Recuperado el 17 de marzo de 2019, de <https://developer.android.com/guide>
- APKPure. (2019). Recuperado el 15 de abril de 2019, de APKPure: <https://apkpure.com/es/>
- Apple Inc. (2019). *Apple Developer Documentation*. Recuperado el 18 de marzo de 2019, de <https://developer.apple.com/documentation/>
- Apple Inc. (2019). *Build Apps for the World*. Recuperado el 13 de mayo de 2019, de <https://developer.apple.com/internationalization/>
- Apple Inc. (2019). *Xcode Help*. Recuperado el 18 de marzo de 2019, de <https://help.apple.com/xcode/mac/current/#/devc8c2a6be1>
- Aragón-Lumeras, M., & Way, A. (2017). On the Complementarity between Human Translators and Machine. *Hermes Journal of Language and Communication in Business* (56), 21-42. Recuperado el 13 de abril de 2019, de <https://tidsskrift.dk/her/article/view/97200>
- Aranberri, N. (2014). Posedición, productividad y calidad. *Revista Tradumàtica* (12), 471-477. Recuperado el 13 de abril de 2019, de <https://ddd.uab.cat/record/130152/>
- Babych, B. (2014). Automated MT evaluation metrics and their limitations. *Revista Tradumàtica* (12), 464-470. Recuperado el 21 de mayo de 2019, de <https://revistes.uab.cat/tradumatica/article/view/n12-babych>
- Banerjee, S., & Lavie, A. (2005). METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments. *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization* (págs. 65–72). Ann Arbor, Michigan: Association for Computational Linguistics. Recuperado el 23 de mayo de 2019, de <https://www.aclweb.org/anthology/W05-0909>

- de la Fuente, R. (2014). Sampling for machine translation evaluation. *Revista Tradumàtica* (12), 476-483. Recuperado el 19 de mayo de 2019, de <https://ddd.uab.cat/record/130149>
- Denkowski, M., & Lavie, A. (2010). Choosing the Right Evaluation for Machine Translation: an Examination of Annotator and Automatic Metric Performance on Human Judgment Tasks. Recuperado el 21 de mayo de 2019, de <https://www.cs.cmu.edu/~alavie/papers/AMTA-10-Denkowski.pdf>
- Doğru, G., Martín-Mor, A., & Aguilar-Amat, A. (2018). Parallel Corpora Preparation for Machine Translation of Low-Resource Languages: Turkish to English Cardiology Corpora. En M. Melero, M. Krallinger, & A. González-Aguirre (Ed.), *Proceedings of the LREC 2018 Workshop "MultilingualBIO: Multilingual Biomedical Text Processing"*, (págs. 12-15). Miyazaki, Japón. Recuperado el 13 de abril de 2019, de <https://multilingualbio.bsc.es/>
- Esselink, B. (2000). *A Practical Guide to Localization*. Ámsterdam: John Benjamins Publishing.
- Fundación Telefónica. (2019). *Sociedad Digital en España 2018*. Barcelona: Penguin Random House Grupo Editorial. Recuperado el 20 de abril de 2019, de www.fundaciontelefonica.com/publicaciones
- Google LLC. (s.f.). *Git repositories on android*. Recuperado el 20 de marzo de 2019, de Google Git: <https://android.googlesource.com/>
- Görög, A. (2014). Quantifying and benchmarking quality: the TAUS Dynamic Quality Framework. *Revista Tradumàtica* (12). Recuperado el 23 de mayo de 2019, de <https://ddd.uab.cat/record/130151>
- Han, L. (2018). Machine Translation Evaluation Resources and Methods: A Survey. Recuperado el 21 de mayo de 2019, de <https://arxiv.org/abs/1605.04515v8>
- ISO, International Organization for Standardization. (2017). ISO 18587:2017: Translation services — Post-editing of machine translation output — Requirements. Recuperado el 19 de mayo de 2019, de <https://www.iso.org/obp/ui/#iso:std:iso:18587:ed-1:v1:en>

- Joscelyne, A., & Brace, C. (2010). *Postediting in Practice: a TAUS report*. De Rijp: TAUS. Recuperado el 18 de mayo de 2019, de <https://www.taus.net/think-tank/reports/postedit-reports/postediting-in-practice>
- Koehn, P. (2019). *Moses User Manual and Code Guide*. Recuperado el 13 de abril de 2019, de <http://www.statmt.org/moses/manual/manual.pdf>
- Li, C. (2 de enero de 2018). *2017 Localization Trends & Predictions for 2018*. Recuperado el 12 de mayo de 2019, de OneSky Blog: <http://www.oneskyapp.com/blog/localization-trends-2018/>
- Linguistic Data Consortium. (2002). *Linguistic Data Annotation Specification: Assessment of Fluency and Adequacy in Arabic-English and Chinese-English Translations*. Recuperado el 23 de mayo de 2019, de <https://catalog.ldc.upenn.edu/docs/LDC2003T17/TransAssess02.pdf>
- Martín-Mor, A. (2017). MTradumàtica: Statistical Machine Translation Customisation for Translators. *Skase journal of translation and interpretation*, 11 (1), 25-40.
- Muñoz-Sánchez, P., & Lopez-Sánchez, R. (2016). The ins and outs of the video game localization process for mobile devices. *Revista Tradumàtica* (14), 16-35. Recuperado el 3 de mayo de 2019, de <https://revistes.uab.cat/tradumatica/article/view/n14-munoz-lopez>
- Papineni, K., Roukos, S., Ward, T., & Zhu, W.-J. (2002). BLEU: A Method for Automatic Evaluation of Machine Translation. *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics* (págs. 311-318). Philadelphia, Pennsylvania: Association for Computational Linguistics. Recuperado el 23 de mayo de 2019, de <https://www.aclweb.org/anthology/P02-1040>
- Roturier, J. (2015). *Localizing Apps*. Londres y Nueva York: Routledge.
- Sánchez Ramos, M. d. (2017). Retos docentes en la didáctica de traducción: diseño y evaluación de la enseñanza de la localización de aplicaciones móviles. *Revista Digital de Investigación en Docencia Universitaria*, 11 (2), 118-135. Recuperado el 10 de marzo de 2019, de <http://dx.doi.org/10.19083/ridu.11.566>
- Sánchez-Gijón, P. (2016). La posesición: hacia una definición competencial del perfil y una descripción multidimensional del fenómeno. *Sendebat*(27), 151-16.

- Recuperado el 10 de mayo de 2019, de <http://revistaseug.ugr.es/index.php/sendeban/article/view/4016>
- Sánchez-Martínez, F. (2012). Motivos del creciente uso de traducción automática seguida de posesición. *Revista Tradumàtica* (10), 150-156. Recuperado el 12 de abril de 2019, de <https://ddd.uab.cat/record/105634?ln=ca>
- Santiago, R., Trbaldo, S., Kamijo, M., & Fernández, Á. (2015). *Mobile learning: nuevas realidades en las aulas*. Barcelona: Editorial Océano.
- Schäler, R. (2011). Localization. En M. Baker, & S. Gabriela (Edits.), *Encyclopedia of Translation Studies* (Segunda ed., págs. 157-161). Abingdon y Nueva York: Routledge.
- 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* (14), 5-15. Recuperado el 17 de marzo de 2019, de https://ddd.uab.cat/pub/tradumatica/tradumatica_a2016n14/tradumatica_a2016n14p5.pdf
- StatCounter. (2019). *Mobile Operating System Market Share Worldwide*. Recuperado el 12 de mayo de 2019, de <http://gs.statcounter.com/os-market-share/mobile/worldwide>
- Tabacchi, A. (4 de junio de 2018). *The Evolution of Project Management: Waterfall to Agile to Continuous Delivery*. Recuperado el 12 de mayo de 2019, de Memsource blog: <https://www.memsource.com/blog/2018/06/04/the-evolution-of-project-management-from-waterfall-to-agile-to-continuous-delivery/>
- TAUS. (2010). *MT Post-editing Guidelines*. Recuperado el 14 de abril de 2019, de <https://www.taus.net/academy/best-practices/postedit-best-practices/machine-translation-post-editing-guidelines>
- TAUS. (2013). *TAUS Best Practice Guidelines: Quality Evaluation using Adequacy and or Fluency Approaches*. Recuperado el 24 de mayo de 2019, de <https://taus.net/academy/best-practices/evaluate-best-practices/adequacy-fluency-guidelines>

TAUS. (2017). *The Translation Industry in 2022*. Amsterdam.

Tiedemann, J. (2012). Parallel Data, Tools and Interfaces in OPUS. *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012)*. Recuperado el 10 de abril de 2019, de <http://opus.nlpl.eu/>

Uptodown. (2019). Recuperado el 15 de abril de 2019, de Uptodown: <https://www.uptodown.com/android>

Wiśniewski, R., & Tumbleson, C. (3 de marzo de 2019). Apktool (v2.4.0) [Software]. Recuperado el 8 de febrero de 2019, de <https://ibotpeaches.github.io/Apktool/>

Yuste-Rodrigo, E. (2012). La posesición en el flujo de producción de contenido multilingüe: tendencias, actantes e implicaciones tecnológicas. *Revista Tradumàtica* (10), 157-165. Recuperado el 10 de mayo de 2019, de <https://revistes.uab.cat/tradumatica/article/view/n10-yuste-rodrigo>

8. Anexos

Anexo 1. Composición de corpus

Tabla descriptiva de las aplicaciones empleadas para el entrenamiento de los motores

App	Versión	Descripción	URL de descarga	Desarrollador	Versión de pago	Descargas	Contenido
Android app Settings	Android 9.0.0_r38	Ajustes	https://android.googlesource.com/platform/packages/apps/AccountsAndSyncSettings/+master/res/	Open Handset Alliance / Google LLC	N/A	N/A	Herramientas
Android app AccountsAndSyncSettings	Android 2.2.3_r2	Sincronización de cuentas	https://android.googlesource.com/platform/packages/apps/AccountsAndSyncSettings/+master/res/	Open Handset Alliance / Google LLC	N/A	N/A	Herramientas
Android app AlarmClock	Android 2.2.3_r2.1	Alarma	https://android.googlesource.com/platform/packages/apps/AlarmClock/+master/res/	Open Handset Alliance / Google LLC	N/A	N/A	Servicios / estilo de vida
Android app Bluetooth	Android 9.0.0_r38	Bluetooth	https://android.googlesource.com/platform/packages/apps/Bluetooth/+master/res/	Open Handset Alliance / Google LLC	N/A	N/A	Herramientas
Android app Browser	Android 6.0.1_r81	Buscador web	https://android.googlesource.com/platform/packages/apps/Browser/+master/res/	Open Handset Alliance / Google LLC	N/A	N/A	Herramientas
Android app Calendar	Android 9.0.0_r38	Calendario	https://android.googlesource.com/platform/packages/apps/Calendar/+master/res/	Open Handset Alliance / Google LLC	N/A	N/A	Productividad
Android app Camera	Android 6.0.1_r81	Cámara	https://android.googlesource.com/platform/packages/apps/Camera/+master/res/	Open Handset Alliance / Google LLC	N/A	N/A	Multimedia

Android app Camera2	Android 9.0.0_r38	Cámara	https://android.googlesource.com/platform/packages/apps/Camera2/+master/res/	Open Handset Alliance / Google LLC	N/A	N/A	Multimedia
Android app CertInstaller	Android 9.0.0_r38	Instalación de certificados	https://android.googlesource.com/platform/packages/apps/CertInstaller/+master/res/	Open Handset Alliance / Google LLC	N/A	N/A	Herramientas
Android app Contacts	Android 9.0.0_r38	Contactos	https://android.googlesource.com/platform/packages/apps/Contacts/+master/res/	Open Handset Alliance / Google LLC	N/A	N/A	Comunicación
Android app ContactsCommon	Android 7.0.0_r36	Contactos	https://android.googlesource.com/platform/packages/apps/ContactsCommon/+master/res/	Open Handset Alliance / Google LLC	N/A	N/A	Comunicación
Android app DeskClock	Android 9.0.0_r38	Alarma	https://android.googlesource.com/platform/packages/apps/DeskClock/+master/res/	Open Handset Alliance / Google LLC	N/A	N/A	Servicios /estilo de vida
Android app DocumentsUI	Android 9.0.0_r38	Gestión de archivos	https://android.googlesource.com/platform/packages/apps/DocumentsUI/+master/res/	Open Handset Alliance / Google LLC	N/A	N/A	Herramientas
Android app Email	Android 9.0.0_r38	Correo electrónico	https://android.googlesource.com/platform/packages/apps/Email/+master/res/	Open Handset Alliance / Google LLC	N/A	N/A	Comunicación
Android app EmergencyInfo	Android 9.0.0_r38	Información de emergencia	https://android.googlesource.com/platform/packages/apps/EmergencyInfo/+master/res/	Open Handset Alliance / Google LLC	N/A	N/A	Servicios /estilo de vida
Android app ExactCalculator	Android 9.0.0_r38	Calculadora	https://android.googlesource.com/platform/packages/apps/ExactCalculator/+master/res/	Open Handset Alliance / Google LLC	N/A	N/A	Herramientas
Android app FMRadio	Android 6.0.1_r81	Radio	https://android.googlesource.com/platform/packages/apps/FMRadio/+master/res/	Open Handset Alliance / Google LLC	N/A	N/A	Multimedia
Android app Gallery	Android 9.0.0_r38	Galería multimedia	https://android.googlesource.com/platform/packages/apps/Gallery/+master/res/	Open Handset Alliance / Google LLC	N/A	N/A	Multimedia
Android app Gallery2	Android 9.0.0_r38	Galería multimedia	https://android.googlesource.com/platform/packages/apps/Gallery2/+master/res/	Open Handset Alliance / Google LLC	N/A	N/A	Multimedia

Android app Gallery3D	Android 2.2.3_r2	Galería multimedia	https://android.googlesource.com/platform/packages/apps/Gallery3D/+master/res/	Open Handset Alliance / Google LLC	N/A	N/A	Multimedia
Android app IM	Android 2.2.3_r2.1	Mensajes de texto	https://android.googlesource.com/platform/packages/apps/IM/+master/res/	Open Handset Alliance / Google LLC	N/A	N/A	Comunicación
Android app InCallUI	Android 6.0.1_r81	Llamadas	https://android.googlesource.com/platform/packages/apps/InCallUI/+master/res/	Open Handset Alliance / Google LLC	N/A	N/A	Comunicación
Android app Launcher	Android 2.1_r2.1p2	Lanzador	https://android.googlesource.com/platform/packages/apps/Launcher/+master/res/	Open Handset Alliance / Google LLC	N/A	N/A	Herramientas
Android app Launcher2	Android 9.0.0_r38	Lanzador	https://android.googlesource.com/platform/packages/apps/Launcher2/+master/res/	Open Handset Alliance / Google LLC	N/A	N/A	Herramientas
Android app Launcher3	Android 9.0.0_r38	Lanzador	https://android.googlesource.com/platform/packages/apps/Launcher3/+master/res/	Open Handset Alliance / Google LLC	N/A	N/A	Herramientas
Android app LegacyCamera	Android 9.0.0_r38	Cámara	https://android.googlesource.com/platform/packages/apps/LegacyCamera/	Open Handset Alliance / Google LLC	N/A	N/A	Multimedia
Android app Messaging	Android 9.0.0_r38	Mensajes de texto	https://android.googlesource.com/platform/packages/apps/Messaging/+master/res/	Open Handset Alliance / Google LLC	N/A	N/A	Comunicación
Android app Mms	Android 5.1.1_r38	Mensajes multimedia	https://android.googlesource.com/platform/packages/apps/Mms/+master/res/	Open Handset Alliance / Google LLC	N/A	N/A	Comunicación
Android app Music	Android 9.0.0_r38	Reproducción de música	https://android.googlesource.com/platform/packages/apps/Music/+master/res/	Open Handset Alliance / Google LLC	N/A	N/A	Multimedia
Android app Nfc	Android 9.0.0_r38	Conexión NFC	https://android.googlesource.com/platform/packages/apps/Nfc/+master/res/	Open Handset Alliance / Google LLC	N/A	N/A	Herramientas
Android app PackageInstaller	Android 9.0.0_r38	Instalación de aplicaciones	https://android.googlesource.com/platform/packages/apps/PackageInstaller/+master/res/	Open Handset Alliance / Google LLC	N/A	N/A	Herramientas
Android app PhoneCommon	Android 9.0.0_r38	Llamadas	https://android.googlesource.com/platform/packages/apps/PhoneCommon/+master/res/	Open Handset Alliance / Google LLC	N/A	N/A	Comunicación
Android app SoundRecorder	Android 9.0.0_r38	Grabación de sonido	https://android.googlesource.com/platform/packages/apps/SoundRecorder/+master/res/	Open Handset Alliance / Google LLC	N/A	N/A	Multimedia

Android app StorageManager	Android 9.0.0_r38	Gestión de archivos	https://android.googlesource.com/platform/packages/apps/StorageManager/+master/res/	Open Handset Alliance / Google LLC	N/A	N/A	Herramientas
Android app TV	Android 9.0.0_r38	Control de televisión	https://android.googlesource.com/platform/packages/apps/TV/+master/res/	Open Handset Alliance / Google LLC	N/A	N/A	Multimedia
Android app UnifiedEmail	Android 9.0.0_r38	Gestión de cuentas de correo electrónico	https://android.googlesource.com/platform/packages/apps/UnifiedEmail/+master/res/	Open Handset Alliance / Google LLC	N/A	N/A	Comunicación
Android app VideoEditor	Android 4.4.4_r2.0.1	Editor de vídeos	https://android.googlesource.com/platform/packages/apps/VideoEditor/+master/res/	Open Handset Alliance / Google LLC	N/A	N/A	Multimedia
Youtube	14.05.53	Reproducción y compartición de vídeos	https://youtube.uptodown.com/android	Google LLC	No	Más de 5000 M	Multimedia
WhatsApp Messenger	2.19.31	Servicio de mensajería y voz IP	https://whatsapp-messenger.uptodown.com/android	WhatsApp Inc.	No	Más de 1000 M	Comunicación
Telegram	5.3.0	Servicio de mensajería y voz IP	https://telegram.uptodown.com/android	Telegram LLC	No	Más de 100 M	Comunicación
Snapseed	2.19.0.201907 232	Edición de fotos	https://snapseed.uptodown.com/android	Google LLC	No	Más de 50 M	Multimedia
Calendario Menstrual - Fertilidad y ovulación	1.653.193	Calendario menstrual	https://period-calendar.uptodown.com/android	Simple Design Ltd.	No	Más de 100 M	Servicios /estilo de vida
Nova Launcher	5.5.4	Lanzador	https://nova-launcher.uptodown.com/android	TeslaCoil Software	Sí	Más de 50 M	Herramientas
Instagram	79.0.0.21.101	Red social	https://instagram.uptodown.com/android	Instagram	No	Más de 1000 M	Comunicación

Google Fotos	4.8.0.2299929 26	Gestor de fotos	https://google-photos.uptodown.com/android	Google LLC	No	Más de 1000 M	Multimedia
Google Docs	1.19.032.03.35	Creación y edición de documentos	https://google-docs.uptodown.com/android	Google LLC	No	Más de 500M	Productividad
Google Calendar	6.0.20-231367491-release	Calendario/agenda	https://google-calendar.uptodown.com/android	Google LLC	No	Más de 500M	Productividad
Google App	9.5.3.21.arm	Buscador web	https://google-quick-search-box.uptodown.com/android	Google LLC	No	Más de 1000 M	Herramientas
Gmail	8.12.30.23056 4275.release	Servicio de correo electrónico	https://gmail.uptodown.com/android	Google LLC	No	Más de 1000 M	Comunicación
Firefox	64.0.2	Navegador web	https://firefox.uptodown.com/android	Mozilla Corporation	No	Más de 100 M	Herramientas
Aliexpress	7.1.0	Compras en línea	https://aliexpress.uptodown.com/android	Alibaba Mobile	No	Más de 100 M	Servicios /estilo de vida
Avg Antivirus Free	6.16.4	Antivirus	https://antivirus-security-free.uptodown.com/android	AVG Mobile	Sí	Más de 100 M	Herramientas
Clean Master	7.0.1	Herramienta de limpieza	https://clean-master-cleaner.uptodown.com/android	Cheetah Mobile	No	Más de 1000 M	Herramientas
Discord-Chat for gamers	8.3.7	Servicio de mensajería para gamers	https://discord-chat-for-gamers.uptodown.com/android	Discord Inc.	No	Más de 50 M	Comunicación

Amazon shopping	18.4.0.100	Compras	https://amazon-shopping.uptodown.com/android	Amazon Mobile LLC	No	Más de 100 M	Servicios /estilo de vida
TripAdvisor	29.4.1	Viajes	https://apkpure.com/es/tripadvisor-hotels-flights-restaurants-attractions/com.tripadvisor.tripadvisor	TripAdvisor	No	Más de 100 M	Servicios /estilo de vida
Kindle	8.15.0.68	Lectura	https://apkpure.com/es/amazon-kindle/com.amazon.kindle	Amazon Mobile LLC	No	Más de 100 M	Productividad
Spotify	8.4.94.817	Servicio de música en streaming	https://apkpure.com/es/spotify-premium-music/com.spotify.music	Spotify Ltd.	No	Más de 100 M	Multimedia
SwiftKey Keyboard	7.2.3.24	Teclado	https://apkpure.com/es/swiftkey-keyboard/com.touchtype.swiftkey	SwiftKey	No	Más de 100 M	Herramientas
Trello	5.5.0.11966	Aplicación de notas colaborativa	https://apkpure.com/es/trello/com.trello	Trello, Inc.	No	5 M	Productividad
Airdroid	4.2.1.12	Acceso remoto y transferencia de archivos	https://airdroid.uptodown.com/android	SAND STUDIO	No	Más de 10 M	Herramientas
Shareit	4.6.88_ww	Transferencia de archivos	https://shareit-connect-and-transfer.uptodown.com/android	SHAREit Technologies Co. Ltd	No	Más de 500 M	Herramientas
Booking.com	17.0	Viajes	https://booking-com.uptodown.com/android	Booking.com	No	Más de 100 M	Servicios /estilo de vida
Netflix	6.26.0 build 1331685	Películas y series en streaming	https://netflix.uptodown.com/android	Netflix, Inc.	Aplicación gratuita, servicio de pago	Más de 500 M	Multimedia
Kinemaster	4.8.13.12545.GP	Edición de vídeos	https://kinemaster.uptodown.com/android	NexStreaming Corp	No	Más de 50 M	Multimedia
Wish	4.26.5	Compras en línea	https://wish.uptodown.com/android	Wish Inc.	No	Más de 100 M	Servicios /estilo de vida

Google Drive	2.19.072.05.34	Almacenamiento en la nube	https://google-drive.uptodown.com/android	Google LLC	No	Más de 1000 M	Productividad
Google Sheets	1.19.052.01.35	Creación y edición de hojas de cálculo	https://google-sheets.uptodown.com/android	Google LLC	No	Más de 100 M	Productividad
WPS Office	11.4.4	Creación y edición de documentos	https://apkpure.com/es/wps-office-word-docs-pdf-note-slide-sheet/cn.wps.moffice_eng	Kingsoft Office Software Corporation Limited	No	Más de 100 M	Productividad
TikTok	9.9.0	Red social de vídeos cortos	https://apkpure.com/es/musically/com.zhiliaoapp.musically	musical.ly	No	Más de 100 M	Comunicación
Fitbit	2.87	Deporte y vida sana	https://fitbit.uptodown.com/android	Fitbit, Inc.	No	Más de 10 M	Servicios /estilo de vida
Mx Player	1.10.44	Reproductor de vídeo	https://reproductor-mx.uptodown.com/android	J2 Interactive	Sí	Más de 500 M	Multimedia
CamScanner - PDF creator	5.9.2.20190221	Escaneo y conversión de documentos a PDF	https://camscanner-pdf-creator.uptodown.com/android	IntSig Information Co.,Ltd.	Sí	Más de 100 M	Productividad
Avast Cleanup	4.12.0	Herramienta de limpieza	https://avast-grimefighter.uptodown.com/android	Avast Software	No	Más de 10 M	Herramientas
Avast Battery Saber	2.8.1	Ahorro de batería	https://avast-battery-saver.uptodown.com/android	Avast Software	No	Más de 10 M	Herramientas
Adobe Illustrator Draw	3.5.1	Creación de diseños vectoriales	https://adobe-illustrator-draw.uptodown.com/android	Adobe	No	Más de 5 M	Multimedia
File Commander	5.2.30001	Gestión de archivos	https://file-commander.uptodown.com/android	MobiSystems	No	Más de 100 M	Herramientas

Microsoft Outlook	3.0.14	Servicio de correo electrónico	https://apkpure.com/es/microsoft-outlook/com.microsoft.office.outlook	Microsoft Corporation	No	Más de 100 M	Comunicación
Microsoft Powerpoint	16.0.11328.20080	Creación y edición de presentaciones	https://microsoft-powerpoint.uptodown.com/android	Microsoft Corporation	No	Más de 500 M	Productividad
Evernote	8.8.1	Notas	https://evernote.uptodown.com/android	Evernote Corporation	No	Más de 100 M	Productividad
Eventbrite	6.4.0	Gestión de eventos	https://eventbrite.uptodown.com/android	Eventbrite	No	Más de 10 M	Servicios /estilo de vida
Deezer	6.0.6.79	Servicio de música en streaming	https://deezer.uptodown.com/android	Deezer Mobile	No	Más de 100 M	Multimedia
Duolingo	4.3.1	Aprendizaje de idiomas	https://duolingo.uptodown.com/android	Duolingo	No	Más de 100 M	Productividad
Opera Browser	50.2.2426.136249	Navegador web	https://opera-browser.uptodown.com/android	Opera	No	Más de 100 M	Herramientas
Pinterest	7.5.0	Red social para compartir imágenes	https://pinterest.uptodown.com/android	Pinterest	No	Más de 100 M	Comunicación
Adobe Photoshop Express	5.9.571	Edición de imágenes	https://apkpure.com/es/adobe-photoshop-express-photo-editor-collage-maker/com.adobe.psmobile	Adobe	No	Más de 100 M	Multimedia
Flipboard	4.2.8	Noticias	https://flipboard.uptodown.com/android	Flipboard	No	Más de 500 M	Servicios /estilo de vida
Snapchat	10.51.0.0	Red social de mensajería multimedia	https://snapchat.uptodown.com/android	Snap Inc	No	Más de 500 M	Comunicación
FindNow	0.6.20	Ubicaciones en tiempo real	https://findnow.uptodown.com/android	Ratech	No	Más de 10 M	Servicios /estilo de vida

Android Auto	4.0.590433-release	Asistente de viaje	https://android-auto.uptodown.com/android	Google LLC	No	Más de 10 M	Servicios /estilo de vida
Vlc for Android	3.0.13	Reproductor multimedia	https://vlc.uptodown.com/android	Videolabs	No	Más de 100 M	Multimedia
Dropbox	130.2.6	Almacenamiento en la nube	https://dropbox.uptodown.com/android	Dopbox, Inc.	No	Más de 500 M	Productividad
Gboard	7.8.8.224901760	Teclado	https://google-keyboard.uptodown.com/android	Google LLC	No	Más de 1000 M	Herramientas
Solid Explorer File Manager	2.6.1	Gestión de archivos	https://solid-explorer-file-manager.uptodown.com/android	NeatBytes	Sí	Más de 1 M	Herramientas
Avast Mobile Antivirus	6.16.0	Antivirus	https://avast-mobile-security.uptodown.com/android	Avast Software	No	Más de 100 M	Herramientas
LastPass Password Manager	4.8.3632	Gestión de contraseñas	https://lastpass-password.uptodown.com/android	LogMeIn, Inc.	No	Más de 5 M	Herramientas
Twitter	7.83.0-release.33	Red social	https://twitter.uptodown.com/android	Twitter, Inc.	No	Más de 500 M	Comunicación
Uber Eats	118.310.001	Comida a domicilio	https://ubereats.uptodown.com/android	Uber Technologies, Inc.	No	Más de 50 M	Servicios /estilo de vida
eBay	5.28.1.1	Compras	https://app-oficial-de-ebay.uptodown.com/android	eBay Mobile	No	Más de 100 M	Servicios /estilo de vida
LinkedIn	4.1.236	Red social profesional	https://linkedin.uptodown.com/android/old	LinkedIn	No	Más de 100 M	Comunicación

Ikea Store	2.9.3	Compras	https://ikea-store.uptodown.com/android	IKEA	No	Más de 10 M	Servicios /estilo de vida
Tinder	10.4.2	Aplicación de citas	https://tinder.uptodown.com/android	Tinder	No	Más de 100 M	Comunicación
Sketch	8.4.A.3.2	Dibujo y edición de fotos	https://apkpure.com/es/sketch-draw-paint/com.sonymobile.sketch	Sony Mobile Communications	No	Más de 100 M	Multimedia
Adobe Acrobat Reader	19.0.0.8512	Lector de PDF	https://apkpure.com/es/adobe-acrobat-reader/com.adobe.reader	Adobe	No	Más de 100 M	Productividad
Google Family Link	1.33.0.1.232598558	Control parental	https://google-family-link.uptodown.com/android	Google LLC	No	Más de 5 M	Herramientas
Youtube Music	3.03.55	Servicio de música en streaming	https://youtube-music.uptodown.com/android	Google LLC	No	Más de 100 M	Multimedia
TV Ttime	7.5.0.19021901	Seguimiento de series de TV	https://apkpure.com/es/tv-time-1-show-tracker/com.tozelabs.tvshowtime	Toze Labs	No	Más de 5 M	Servicios /estilo de vida
Alarmy	4.7.1	Alarma	https://apkpure.com/es/alarm-clock-loud-smart-puzzle-wake-up-alarms/droom.sleepIfUCan	Alarmy	Sí	Más de 10 M	Servicios /estilo de vida
21 Buttons	4.7.3	Red social de moda	https://21-buttons.uptodown.com/android	21 Buttons	No	Más de 1 M	Servicios /estilo de vida
Samsung Smart Switch Mobile	3.5.03.7	Transferencia de datos entre dispositivos	https://apkpure.com/es/samsung-smart-switch-mobile/com.sec.android.easyMover	Samsung Electronics Co., Ltd.	No	Más de 100 M	Herramientas
Samsung Smart View	2.1.0.107	Transferencia de contenido multimedia a la TV	https://apkpure.com/es/samsung-smart-view/com.samsung.smartviewad	Samsung Electronics Co., Ltd.	No	Más de 10 M	Herramientas
Uber	425.010.001	Aplicación de transporte para particulares	https://uber.uptodown.com/android	Uber Technologies, Inc.	No	Más de 100 M	Servicios /estilo de vida

Huawei Health	9.0.4.332	Salud y deporte	https://apkpure.com/es/huawei-health/com.huawei.health	Huawei Internet Service	No	Más de 100 M	Servicios /estilo de vida
Calls Blacklist	3.2.36	Bloqueo de llamadas y SMS	https://calls-blacklist.uptodown.com/android	Vlad Lee	Sí	Más de 10 M	Comunicación
DU Recorder	1.7.9.7	Grabadora de pantalla	https://du-recorder.uptodown.com/android	DU Recorder Team	No	Más de 100 M	Multimedia
Microsoft Edge	42.0.22.3333	Navegador web	https://microsoft-edge.uptodown.com/android	Microsoft Corporation	No	Más de 5 M	Herramientas
Shazam	9.20.0.190215	Reconocimiento de canciones	https://shazam.uptodown.com/android	Apple, Inc.	No	Más de 100 M	Multimedia
Amazon Prime Video	3.0.242.14741	Series y películas en streaming	https://amazon-video.uptodown.com/android	Amazon Mobile LLC	Aplicación gratuita, servicio de pago	Más de 100 M	Multimedia
Deliveroo	2.59.1	Comida a domicilio	https://deliveroo.uptodown.com/android	Deliveroo	No	Más de 5 M	Servicios /estilo de vida
SmartThings	1.7.27-25	Control de dispositivos (domótica)	https://apkpure.com/es/smarthings/com.samsung.android.oneconnect	Samsung Electronics Co., Ltd.	No	Más de 100 M	Servicios /estilo de vida
Mi Remote	5.7.2	Mando para aparatos eléctricos	https://apkpure.com/es/mi-remote-controller-for-tv-stb-ac-and-more/com.duokan.phone.remotecontroller	Xiaomi, Inc,	No	Más de 10 M	Herramientas
Google Duo	47.1.234325686.DR47_RC14	Videollamadas	https://apkpure.com/es/google-duo-high-quality-video-calls/com.google.android.apps.tachyon	Google LLC	No	Más de 1000 M	Comunicación
File Manager - free and easily	V1-190128	Gestión de archivos	https://apkpure.com/es/file-manager-free-and-easily/com.mi.android.globalFileexplorer	Xiaomi, Inc,	No	Más de 100 M	Herramientas

Google Allo	27.0.326_rc03 (armeabi-v7a_xhdpi)	Servicio de mensajería	https://google-allo.uptodown.com/android	Google LLC	No	Más de 10 M	Comunicación
Google Home	2.9.40.16	Control de dispositivos (domótica)	https://apkpure.com/es/google-home/com.google.android.apps.chrome.cast.app	Google LLC	No	Más de 50 M	Servicios /estilo de vida
McAfee-Mobile Security	5.2.0.152	Antivirus	https://apkpure.com/es/mobile-security-antivirus-wi-fi-vpn-anti-theft/com.wsandroid.suite	McAfee LLC	No	Más de 10 M	Herramientas
Viber Messenger	10.1.0.1	Servicio de mensajería y llamadas	https://apkpure.com/es/viber-messenger/com.viber.voip	Viber Media SARL	No	Más de 500 M	Comunicación

Anexo 2. Segmentos extraídos de la aplicación procesada por los motores

Cadenas extraídas de la aplicación de Signal, versión 4.40.3 (extraída el 18/5/19 de <https://github.com/signalapp/Signal-Android/tree/master/res>).

Variables					1. No aprovechable 2. Poco aprovechable 3. Aprovechable en gran medida 4. Aprovechable al completo			
Segmento original	Motor especializado en aplicaciones de comunicación	Motor especializado en aplicaciones	Motor especializado en aplicaciones + genérico	DeepL				
This will permanently delete all %1\$d selected messages.	Esta acción ¿Quieres eliminar permanentemente todos los % 1 \$ d mensajes seleccionados.	Esto eliminará permanentemente todos los % 1 \$ d mensajes seleccionados.	Esto eliminará permanentemente todos los % 1 \$ d mensajes seleccionados.	Esto eliminará permanentemente todos los mensajes seleccionados de %1\$d.	3	3	3	1
%s reset the secure session.	% S ha cambiado la sesión seguro.	% S ha cambiado la sesión seguro.	% S ha cambiado la sesión de seguros.	restablece la sesión segura.	3	3	3	1
Last active %s	Última conexión % s	Último estado activo % s	Último estado activo % 1 \$ s	Últimos %s activos	3	3	2	1
Linked %s	Vinculado a % s	Vinculado a % s	Vinculado a % s	% S vinculado	3	3	3	1
Called %s	Llamada % s	Llamar a % s	Llamada % s	Llamado %s	1	1	1	1
%d others	% D más	% D más	% D más	de otros	3	3	3	1
Moved %d conversations to inbox	Se ha movido % d conversaciones a la bandeja de entrada	Se ha movido % d conversaciones a la bandeja de entrada	Mover a bandeja de entrada de % d conversaciones	Mover conversaciones %d a la bandeja de entrada	2	2	1	1
You are now %d steps away from submitting a debug log.	Ahora eres % 1 \$ d pasos de la vista de enviando un registro de depuración.	Ahora estás % d pasos de enviando un registro de depuración.	Te quedan % d pasos de enviar un registro de depuración.	Ahora está a %d pasos de enviar un registro de depuración.	2	3	3	4
Message to %s	Mensaje para % s	Mensaje para % s	Mensaje para % s	Mensaje a %s	3	3	3	2
You marked your safety number with %s verified	Has indicado tu seguridad número con % s verificado	Se ha marcado el número de seguridad con % s verificado	Marcaste tu número de seguridad con % s verificado	Marcó su número de seguridad con %s verificados	2	3	3	2

Block %s?	Bloquear a % s?	Bloquear a % s?	Bloquear % s?	¿Bloque %s?	3	3	2	1
Saving %1\$d attachments	Guardando % 1 \$ d archivos adjuntos	Guardando % 1 \$ d archivos adjuntos	Guardando % 1 \$ d archivos adjuntos	Ahorro %1\$d de archivos adjuntos	3	3	3	1
Apóstrofo escapado (\')								
You have selected a contact that doesn't support Signal groups, so this group will be MMS.	Has seleccionado un contacto que no admite señal grupos, por lo que este grupo se MMS.	Ha seleccionado un contacto que no admite Signal grupos , por lo que este grupo se MMS.	Ha seleccionado un contacto que no admite Signal grupos, para que este grupo se MMS.	Has seleccionado un contacto que no soporta grupos de señales, así que este grupo será MMS.	4	4	4	4
Would you like to import your existing text messages into Signal's encrypted database?	¿Quieres importar el los mensajes de texto en señal's cifrados base de datos?	¿Quieres importar el los mensajes de texto es la señal's cifrados base de datos?	¿Quieres importar tus mensajes de texto a Signal's encriptados base de datos?	¿Desea importar sus mensajes de texto existentes a la base de datos cifrada de Signal?	1	1	1	4
It looks like you don't have any apps to share to.	Parece que no tienes ninguna aplicación para compartir.	Parece que no tienes ninguna aplicación para compartir.	Parece que no tiene aplicaciones para compartir.	Parece que no tienes ninguna aplicación para compartir.	4	4	4	4
Please be patient, we'll notify you when the import is complete.	Por favor, ten paciencia, te lo te notifiemos cuando la importación.	Por favor, ten paciencia, te respondo notificarte cuando se haya completado la importación.	Por favor, ten paciencia, le respondo te notificaré cuando la importación está completo.	Por favor, tenga paciencia, le notificaremos cuando la importación esté completa.	2	2	2	4
Salto de línea (\n)								
Saving this media to storage will allow any other apps on your device to access it.\n\nContinue?	Guardar este contenido multimedia al almacenamiento, se permite ningún otras aplicaciones en tu dispositivo para acceder a él. \n \n continue?	Guardar este contenido multimedia en el almacenamiento, permite que otras aplicaciones en tu dispositivo para acceder a él. \n \n continue?	Guardar este contenido multimedia para el almacenamiento, de otras aplicaciones en tu dispositivo para acceder a ella. \n \n continue?	Guardar este medio en el almacenamiento permitirá que cualquier otra aplicación de tu dispositivo pueda acceder a él.	1	1	1	1

Take privacy with you.\nBe yourself in every message.	Tomar la privacidad contigo. \nbe a ti mismo en cada mensaje.	Tomar privacidad contigo. \nbe en cada mensaje.	Tomar la privacidad. \nbe en cada mensaje.	Llévate la privacidad contigo. Sé tú mismo en cada mensaje.	1	1	1	3
\n\nRegistration transmits some contact information to the server.	\n \nregistration transmits algunos datos de contacto con el servidor.	\n \nregistration transmits algunos datos de contacto con el servidor.	\n \nregistration transmits algunos datos de contacto con el servidor.	\nEl registro transmite información de contacto al servidor.	1	1	1	3
\n\nSend unsecured message?	\n \nsend unsecured mensaje?	\n \nsend desasegurar mensaje?	\n \nsend desasegurar mensaje?	\n¿Enviar un mensaje no seguro?	1	1	1	3
Segmentos largos					(A: adecuación, F: fluidez)			
Registration of this phone number will be possible without your Registration Lock PIN after 7 days have passed since this phone number was last active on Signal.	El registro de este número de teléfono estará posibles sin tu registro Lock PIN después de 7 días ha entregado desde este número de teléfono ha sido última actividad en la señal	Registro de este número de teléfono se posibles sin el bloqueo de registro PIN después de 7 días haya pasado desde este número de teléfono ha sido última actividad en Signal.	El registro de este número será posible sin el PIN de bloqueo del registro tras 7 días han pasado desde el número de teléfono ha sido la actividad de Signal.	El registro de este número de teléfono será posible sin su PIN de bloqueo de registro después de que hayan pasado 7 días desde que este número de teléfono estuvo activo por última vez en Signal.	A: 1 F: 1	A: 3 F: 2	A: 3 F: 3	A: 4 F: 4
The scanned QR code is not a correctly formatted safety number verification code.	Al escanear el código QR no es un correctamente formatted seguridad el código de verificación.	Escanear el código QR no es un número de de seguridad formateada correctamente el código de verificación.	Escanear el código QR no es un número de seguridad correctamente formatear tu código de verificación.	El código QR escaneado no es un código de verificación del número de seguridad formateado correctamente.	A: 1 F: 1	A: 2 F: 1	A: 2 F: 2	A: 4 F: 4
Signal requires MMS settings to deliver media and group messages through your wireless carrier.	Señal requiere MMS ajustes para obtener contenido multimedia y mensajes de grupo mediante su operador inalámbrica.	Señal requiere MMS ajustes para transferir contenido multimedia y mensajes de grupo mediante su operador de servicios.	La señal requiere MMS para transferir archivos multimedia y mensajes de grupo a través de la de operador.	La señal requiere la configuración de MMS para enviar mensajes multimedia y de grupo a través de su proveedor de servicios inalámbricos.	A: 2 F: 2	A: 3 F: 4	A: 3 F: 2	A: 3 F: 4
Friends and contacts who already know how to contact you by phone will be able to	Tus amigos y contactos que ya conoces cómo en contacto contigo por	Tus amigos y contactos que ya conoces cómo en contacto contigo por teléfono podrá fácilmente	Amigos y contactos que ya sabemos cómo contactar por teléfono	Los amigos y contactos que ya saben cómo ponerse en contacto con usted por teléfono podrán	A: 2 F: 1	A: 2 F: 1	A: 2 F: 3	A: 4 F: 4

easily get in touch by Signal.	teléfono podrán fácilmente contacta por señal.	ponte en contacto con Signal.	podrá fácilmente contactar por Signal.	ponerse en contacto fácilmente con Signal.				
Signal makes it easy to communicate by using your existing phone number and address book.	Señal, es fácil comunicarse utilizando tu número de teléfono existente y libreta de direcciones.	Señal te facilita la comunicación al usar tu número de teléfono y tu libreta de direcciones.	La señal te facilita la comunicación mediante tu número de teléfono y tu libreta de direcciones.	Signal facilita la comunicación utilizando su número de teléfono y su libreta de direcciones.	A: 3 F: 4	A: 3 F: 4	A: 3 F: 4	A: 4 F: 4

Otros segmentos			
Segmento original	Motor especializado en aplicaciones de comunicación	Motor especializado en aplicaciones	Motor especializado en aplicaciones + genérico
Error while retrieving full resolution GIF	Error al recuperar el resolución completa GIF	Error al recuperar resolución completa GIF	Error al recuperar toda resolución GIF
Attachment Thumbnail	Miniatura de archivo adjunto	Miniatura de archivo adjunto	Imagen en miniatura de archivo
Send failed, tap for details	Error de envío, pulsa para más detalles	No se ha podido enviar el mensaje, pulsa para ver detalles	No se ha podido enviar el mensaje, pulsa para ver los detalles
Your safety number with %1\$s has changed.	Tu número de seguridad con % 1 \$ s ha cambiado.	Tu número de seguridad con % 1 \$ s ha cambiado.	El número de seguridad con % 1 \$ s ha cambiado.
This message will not be encrypted because the recipient is no longer a Signal user.	Este mensaje se no estar cifrado porque el destinatario ya no es una señal de usuario.	Este mensaje no se cifrarán porque el destinatario ya no es un Signal del usuario.	Este mensaje, no se cifrará porque el destinatario ya no es una señal al usuario.
Sorry, there was an error setting your attachment.	Lo sentimos, hubo un error al configurar el archivo adjunto.	Lo sentimos, se ha producido un error al configurar el archivo adjunto.	Lo sentimos, se ha producido un error al configurar su adjunto.
Signal now supports secure video calling.	Señal ahora asiste proteger las videollamadas.	Señal de forma segura ahora admite videollamadas.	La señal segura ahora admite videollamadas.
You can disable or enable this feature anytime in your Signal settings (Privacy & Send link previews).	Puedes desactivar o habilitar esta función en cualquier momento en los ajustes de	Puedes desactivar o habilitar esta función en cualquier momento en los ajustes de	Puedes desactivar o habilitar esta función cuando quieras en los ajustes de privacidad de

	privacidad de señal (> enviar las vistas previas).	privacidad Signal (> enviar las vistas previas)	Signal (> enviar las vistas previas).
An item was removed because it exceeded the size limit	Un elemento se eliminó porque superado el límite de tamaño de	1 elemento se eliminó porque ha superado el límite de tamaño	Un artículo fue retirado porque se ha superado el límite de tamaño
Sorry, you have too many devices linked already, try removing some	Lo sentimos, tienes demasiados dispositivos vinculados ya, prueba a quitar algunos	Lo sentimos, tienes demasiados dispositivos vinculados ya, intenta eliminar algunos	Lo sentimos, tienes demasiados dispositivos vinculados ya, intenta eliminar algunos
Messages sent and received in this conversation will disappear %s after they have been seen.	Mensajes enviados y recibidos en esta conversación se desaparezcan % s después de que se haya visto.	Los mensajes enviados y recibidos en esta conversación, desaparecerá de % s después de que se haya visto.	Los mensajes enviados y recibidos en esta conversación, desaparecerá % s después de que se han visto.
The number you specified (%s) is invalid.	El número que has especificado (% s) no es válido.	El número que ha especificado (% s) no es válido.	El número especificado (% s) es válido.
You will receive a verification code. Carrier rates may apply.	Recibirás un código de verificación. Es posible que se apliquen las tarifas del operador.	Recibirás un código de verificación. Es posible que se apliquen las tarifas del operador.	Recibirás un código de verificación. Operadora puede aplicarte sus precios.
Signal needs the Camera permission in order to scan a QR code, but it has been permanently denied.	Señal necesita el permiso de cámara para escanear un código QR, pero se ha denegado permanentemente.	Señal necesita el permiso de cámara para escanear un código QR , pero se ha denegado permanentemente.	La señal necesita el permiso de cámara para escanear un código QR, pero se ha denegado permanentemente el.
No web browser installed!	No hay un navegador Web!	No hay navegador instalado!	No hay navegador instalado.
Now you can share a profile photo and name with friends on Signal	Ahora puedes compartir una foto de perfil y nombre con tus amigos en señal	Ahora puedes compartir una foto de perfil y nombre con tus amigos en Signal	Ahora puedes compartir una foto de perfil y el nombre con amigos en Signal
Tap and hold to record a voice message, release to send	Toca y mantén pulsado para grabar un mensaje de voz, suelta para enviar	Toca y mantén pulsado para grabar un mensaje de voz, soltar para enviar	Toca y mantén pulsado para grabar un mensaje de voz, soltar para enviar
TextSecure and RedPhone are now one app: Signal. Tap to explore.	TextSecure y RedPhone ahora es una aplicación: Señal. Toca para explorar.	TextSecure y RedPhone ahora es una aplicación : Señal. Toca para explorar .	TextSecure y RedPhone son ahora una aplicación: La señal. Toca para explorar.

Permanent Signal communication failure!	Permanente señal comunicación error!	Error de comunicación Signal permanente!	Signal permanente comunicación fracasar!
Working in the background...	Trabajando en segundo plano...	Ejecutándose en segundo plano...	Trabajando en el plano...
Messages will no longer send successfully.	Mensajes dejará de enviar correctamente.	Mensajes dejará de enviar correctamente.	Mensajes no se podrán enviar correctamente.
Please upgrade this device to receive future Signal updates.	Por favor, actualiza este dispositivo para recibir futuro señal actualizaciones.	Por favor, actualiza tu cuenta de este dispositivo para recibir futuro Signal actualizaciones.	Por favor actualice este dispositivo para recibir actualizaciones futuras Signal.
Backups will be saved to external storage and encrypted with the passphrase below.	Las copias de seguridad se guardarán en el almacenamiento externo y cifrados con la frase de acceso a continuación.	Copias de seguridad se guardará en el almacenamiento externo y cifrados con la frase de paso siguiente.	Copias de seguridad se guardan para el almacenamiento externo y cifrado con la frase a continuación.
You must enter your Registration Lock PIN	Debes introducir tu PIN registro de acceso	Debes introducir tu código PIN de registro de bloqueo	Debes introducir tu código PIN de bloqueo de registro
Unlock to view pending messages	Desbloquea para ver pendiente de mensajes	Desbloquea para ver pendiente de mensajes	Desbloquea para ver los mensajes pendientes
Toggle attachment drawer	Alternar el archivo adjunto de navegación	Alternar panel de archivo adjunto	Alternar panel adjunto
There is no app available to handle this link on your device.	No hay ninguna aplicación disponible para procesar este enlace en tu dispositivo.	No hay ninguna aplicación disponible para este enlace en tu dispositivo.	No hay ninguna aplicación disponible para emplear este enlace en tu dispositivo.
This device does not appear to support dial actions.	Este dispositivo no aparece en el soporte técnico marcar acciones.	Este dispositivo no aparece en el servicio de asistencia de marcado acciones.	Este dispositivo no parece compatible con las acciones de marcado.
Unsupported media type	Tipo de contenido multimedia no compatible	No admitimos este tipo de medio	Este tipo de dispositivo
System database import is complete.	Importación de la base de datos del sistema.	Se ha completado la importación de la base de datos del sistema.	El sistema es la base de datos de importación