



Universidad de Valladolid

OFICIAL POSTGRADUATE MASTER
**TRADUCCIÓN
PROFESIONAL
E INSTITUCIONAL**

FACULTAD DE TRADUCCIÓN E INTERPRETACIÓN

Máster en Traducción Profesional e Institucional

TRABAJO FIN DE MÁSTER

La traducción automática para dispositivos
móviles: Siri y Google Translator

Presentado por Cristina García Aguilar

Tutelado por Susana Álvarez Álvarez

Soria, 2018

ÍNDICE DE CONTENIDOS

RESUMEN.....	6
<i>ABSTRACT</i>	7
INTRODUCCIÓN.....	8
JUSTIFICACIÓN DEL TRABAJO.....	8
OBJETIVOS Y COMPETENCIAS	10
METODOLOGÍA Y PLAN DE TRABAJO	11
ESTRUCTURA DEL TRABAJO	12
CAPÍTULO 1. TRADUCCIÓN AUTOMÁTICA: HISTORIA Y SISTEMAS DE TRADUCCIÓN	15
1.1 EL PROCESAMIENTO DE LENGUAJES NATURALES Y LA INTELIGENCIA ARTIFICIAL.....	15
1.2 ANTECEDENTES Y SITUACIÓN ACTUAL.....	16
1.3 SISTEMAS DE TRADUCCIÓN Y LIMITACIONES	18
1.3.1 SISTEMAS BASADOS EN REGLAS	18
1.3.2 SISTEMAS BASADOS EN CORPUS.....	19
1.3.3 SISTEMAS DE TRADUCCIÓN NEURONAL	20
1.3.3.1 COMPUTACIÓN Y REDES NEURONALES	21
1.3.3.2 LA TRADUCCIÓN BASADA EN REDES NEURONALES	22
1.4 LIMITACIONES DE LA TRADUCCIÓN AUTOMÁTICA	24
CAPÍTULO 2. SIRI: ASISTENTE PERSONAL DE APPLE.....	26
2.1 AGENTES CONVERSACIONALES Y ASISTENTES VIRTUALES PERSONALES	26
2.2 SIRI: NACIMIENTO Y EVOLUCIÓN.....	27
2.2.1 EL MULTILINGÜISMO DE SIRI.....	28
2.2.2 EL TRADUCTOR INTEGRADO EN SIRI	29
CAPÍTULO 3. GOOGLE TRANSLATOR: EL TRADUCTOR DE ANDROID.....	31
3.1 GOOGLE: NACIMIENTO Y EVOLUCIÓN	31
3.1.1 GOOGLE TRANSLATOR PARA DISPOSITIVOS MÓVILES	31
3.2 GOOGLE TRANSLATOR	32
3.2.1 SISTEMA DE TRADUCCIÓN DE GOOGLE TRANSLATOR	33
3.2.1.1 GOOGLE NEURAL MACHINE TRANSLATION (GNMT).....	34

3.2.2	EL MULTILINGÜISMO DE GOOGLE TRANSLATOR.....	35
CAPÍTULO 4.	EL CONCEPTO DE ERROR EN LOS ESTUDIOS DE TRADUCCIÓN	38
4.1	MÉTODO DE CLASIFICACIÓN DE ERRORES DE NORD (1995).....	38
4.2	MÉTODO TRADICIONAL DE CLASIFICACIÓN DE ERRORES (1998).....	39
4.3	MÉTODO DE CLASIFICACIÓN DE ERRORES DE HURTADO ALBIR (1999/2001)	40
4.4	MÉTODO DE CLASIFICACIÓN DE ERRORES DE DIÉGUEZ MORALES (2001).....	42
4.5	PROPUESTA DE CLASIFICACIÓN DE ERRORES DE TRADUCCIÓN AUTOMÁTICA	42
4.5.1	ERRORES LÉXICO-SEMÁNTICOS	43
4.5.2	ERRORES ORTOTIPOGRÁFICOS	43
4.5.3	ERRORES GRAMATICALES	44
4.5.4	ERRORES SINTÁCTICOS.....	44
4.5.5	ERRORES RETÓRICOS	44
4.5.6	ERRORES PRAGMÁTICOS.....	45
4.5.7	ERRORES CULTURALES	45
CAPÍTULO 5.	METODOLOGÍA DE LA INVESTIGACIÓN	46
CAPÍTULO 6.	RESULTADOS DEL ANÁLISIS DE LA PRODUCCIÓN DE SIRI	54
6.1	CRITERIO 1: INGLÉS BRITÁNICO E INGLÉS ESTADOUNIDENSE	54
6.2	CRITERIO 2: NÚMERO DE CARACTERES	59
6.3	CRITERIO 3: GENERAL O ESPECIALIZADO.....	63
6.4	CRITERIO 4: TEXTOS DE PRODUCCIÓN PROPIA.....	67
6.5	CONCLUSIONES EXTRAÍDAS DEL ANÁLISIS	69
CAPÍTULO 7.	RESULTADOS DEL ANÁLISIS DE LA PRODUCCIÓN DE GOOGLE TRANSLATOR.....	73
7.1	CRITERIO 1: INGLÉS BRITÁNICO E INGLÉS ESTADOUNIDENSE	73
7.2	CRITERIO 2: NÚMERO DE CARACTERES.....	75
7.3	CRITERIO 3: GENERAL O ESPECIALIZADO.....	81
7.4	CRITERIO 4: TEXTOS DE PRODUCCIÓN PROPIA.....	85
7.5	CONCLUSIONES EXTRAÍDAS DEL ANÁLISIS	86
CAPÍTULO 8.	COMPARATIVA DE LAS PRODUCCIONES DE SIRI Y GOOGLE TRANSLATOR.....	89
	CONCLUSIONES	92

REFERENCIAS BIBLIOGRÁFICAS 96
REFERENCIAS BIBLIOGRÁFICAS DEL CORPUS 102

ÍNDICE DE TABLAS

Tabla 1. Comparativa de los sistemas de TA	24
Tabla 2. Datos de los textos del corpus.....	53
Tabla 3. Diferencia del número de caracteres entre el T.O. y el T.M en Siri.....	61
Tabla 4. Medias de error (I).....	63
Tabla 5. Medias de error (II).....	64
Tabla 6. Diferencia del número de caracteres entre el T.O. y el T.M. en Google Translator	79

ÍNDICE DE ILUSTRACIONES

Ilustración 1. Porcentaje poblacional de usuarios de dispositivos móviles a escala mundial	10
Ilustración 2. Representación del funcionamiento de un sistema de traducción neuronal	21
Ilustración 3. Funcionamiento de un sistema de TA neuronal.....	23
Ilustración 4. Funciones de Siri (I)	28
Ilustración 5. Funciones de Siri (II)	28
Ilustraciones 6, 7 y 8. Traductor de Siri (I, II y III)	30
Ilustraciones 9, 10 y 11. Google Translator para Android (I, II y III)	33
Ilustración 12. Respuesta cuando no reconoce el idioma	55
Ilustración 13. Respuesta cuando no procesa el texto.....	55
Ilustración 14. Sugerencias del traductor.....	72
Ilustración 15. Comunidad de Google Translator.....	90

ÍNDICE DE FIGURAS

Figura 1. Esquema explicativo de la división de los fragmentos de textos escogidos para el estudio	47
Figura 2. Porcentaje de textos no traducidos en inglés británico	54
Figura 3. Porcentaje de textos no traducidos en inglés estadounidense	54
Figura 4. Porcentaje de error en los textos en inglés británico	58
Figura 5. Porcentaje de error en los textos en inglés estadounidense	58

Figura 6. Porcentaje de textos no traducidos por número de caracteres en inglés británico ...	59
Figura 7. Porcentaje de textos no traducidos por número de caracteres en inglés estadounidense	59
Figura 8. Porcentaje de textos no traducidos generales y especializados en inglés estadounidense	63
Figura 9. Porcentaje de textos no traducidos generales y especializados en inglés británico ..	63
Figura 10. Porcentaje de error para generales y especializados.....	64
Figura 11. Porcentajes de error en textos generales	66
Figura 12. Porcentajes de error en textos especializados.....	67
Figura 13. Porcentajes de error por categoría de error	69
Figura 14. Porcentajes de error en los textos en inglés británico	75
Figura 15. Porcentajes de error en los textos en inglés estadounidense	75
Figura 16. Porcentajes de inadecuaciones (I).....	80
Figura 17. Porcentajes de inadecuaciones (II).....	81
Figura 18. Porcentajes de error por categoría en los textos generales británicos	83
Figura 19. Porcentajes de error por categoría en los textos generales estadounidenses	83
Figura 20. Porcentajes de error por categoría en los textos especializados británicos.....	84
Figura 21. Porcentajes de error por categoría en los textos especializados estadounidenses..	84
Figura 22. Porcentajes de error en los textos de producción propia.....	86

RESUMEN

La traducción y la tecnología son dos ámbitos de trabajo que, desde sus inicios, han ido evolucionando de la mano, gracias, especialmente, a la inteligencia artificial que es capaz de simular los pensamientos, conocimientos y acciones de la mente humana. La traducción automática, no se doblega ante las adversidades y continúa mejorando día a día, haciendo que las barreras que separan la traducción automática de la traducción humana sean cada vez más débiles, aunque todavía nos encontramos con profesionales que rechazan los grandes avances y la ayuda que nos proporcionan.

Este Trabajo de Fin de Máster tiene como finalidad contribuir a los estudios que se realizan para mejorar esta interacción traductor humano-traductor automático, en la que, además, la Inteligencia Artificial hace acto de presencia. Para ello, se ha estudiado, por una parte, cómo funcionan de forma general los traductores automáticos, cómo surgieron, con que limitaciones cuentan, etc. Estos conocimientos se han llevado a la práctica gracias al estudio de los motores de traducción desarrollados por las empresas Apple y Google Inc., Siri y Google Translator, respectivamente, en sus versiones para dispositivos móviles.

La investigación recorre los distintos sistemas y disciplinas de traducción englobadas dentro del concepto de traducción automática, centrándose especialmente, en la más novedosa del mercado: la neuronal. A través del análisis de un corpus monolingüe y las traducciones propuestas por cada uno de los traductores automáticos presentes en el estudio, se han pautado las fortalezas y debilidades de cada uno de ellos y se presenta una comparativa que recoge los puntos de mayor relevancia.

Palabras clave: traducción automática, dispositivos móviles, Inteligencia Artificial, Siri, Google Translator.

ABSTRACT

Translation and technology are two areas of work that, since their inception, have evolved hand in hand, thanks, especially, to artificial intelligence, which is capable of simulating the thoughts, knowledge and actions of the human mind. This means the barriers that separate automatic translation from human translation are weakening.

This Final Degree Project aims to contribute to the studies that are carried out to improve this human-translator automatic interaction, in which, Artificial Intelligence makes an appearance. For this, it has been studied, on the one hand, how automatic translators work in general, how this idea came about, what limitations it has, etc. This knowledge has been put into practice thanks to the study of translation engines developed by Apple companies and Google Inc., Siri and Google Translate, respectively, in their versions for mobile devices.

The research covers the different systems and disciplines of translation encompassed within the concept of "Automatic Translation", focusing especially on the most novel on the market: the neuronal. Through the analysis of a monolingual corpus and the translations proposed by each of the automatic translators present in the study, the strengths and weaknesses of each one of them have been outlined and a comparative is presented that collects the most relevant points.

***Keywords:** automatic translation, mobile devices, Artificial Intelligence, Siri, Google Translate.*

INTRODUCCIÓN

La búsqueda de la innovación en el campo de las comunicaciones entre diversas culturas ha sido el impulso que necesitaban los estudios sobre traducción automática (TA). Desde su nacimiento a finales de los años cincuenta, la TA ha conseguido automatizar numerosos procesos que forman parte de la labor traductora, intentando buscar la perfección, y, aunque este objetivo sigue estando fuera del alcance de los investigadores, no supone un impedimento que aúne las mismas cualidades que tienen los mejores traductores profesionales.

Hutchins y Somers (1992) identificaron una serie de factores que mueven a los investigadores a indagar sobre las posibilidades de automatizar el proceso de traducción:

- Factores prácticos: las investigaciones en el campo de la traducción automática no se mueven por una necesidad científica, sino práctica, ya que se ha constatado que, cada día, se realizan traducciones de carácter técnico equivalentes a 20 millones de palabras en todo el mundo, cifra que, además, crece de forma exponencial (Abaitua, 1997).

- Factores sociopolíticos: la creciente demanda de traducciones en los mercados internacionales, la gran cantidad de textos de diferente índole y el mayor número de pares de lenguas suponen tanto elevados costes de tiempo y de dinero que pueden reducirse gracias a la TA. Abaitua (1997) recoge en su informe que, desde hace 30 años, solo en Estados Unidos, se editan alrededor de las 9 000 publicaciones periódicas al año y que organismos políticos internacionales como los europeos gastan entorno al 50 % de sus presupuestos en traducciones e interpretaciones.

- Factores comerciales y tecnológicos: el incremento de los intercambios comerciales a nivel mundial, así como las nuevas tecnologías que generan documentos simultáneos en diversas lenguas, también han contribuido a que el número de traducciones se eleve.

En este contexto de referencia, surge nuestro Trabajo Fin de Máster (TFM) que tiene como base la traducción automática aplicada a los dispositivos móviles. Como veremos, los dos motores de TA que se han escogido para su análisis —Siri y Google Translator— difieren en numerosos aspectos, lo que hace que su forma de trabajar sea completamente distinta y digna de nuestra atención.

Justificación del trabajo

Académicamente, este Trabajo Fin de Máster tiene una vinculación directa el mismo, puesto que no hemos tenido ocasión de trabajarlo, por lo que este TFM supone una gran oportunidad para profundizar en el tema y tener una perspectiva más amplia que complete mi formación.

Sin embargo, el principal motivo que me ha llevado a elegir este tema es mi propia experiencia personal como estudiante de máster. Como futura magíster con gran interés en el campo de las tecnologías aplicadas a la traducción y, en concreto, de la traducción automática, así como de los enormes avances que está presentando la industria de la Inteligencia Artificial (IA), me pareció interesante, no solo a nivel personal, sino también a nivel profesional, seguir investigando dentro de esta materia, debido a que creo firmemente que el futuro de la traducción será más que una colaboración entre los humanos y las máquinas. Además, su lento desarrollo hace de él un tema candente y novedoso, del que apenas se está comenzando a investigar, a excepción de un limitado número de trabajos académicos. Como futura profesional del mundo de la traducción, soy consciente de que es imposible plantarse ante las nuevas disciplinas de traducción que se apoyan en la tecnología, como la traducción asistida por ordenador o la traducción automática, pues considero que, gracias a ella, los traductores son capaces de ampliar sus horizontes.

Su redirección hacia los dispositivos móviles en concreto se debe a la gran invasión que estos han protagonizado desde hace una década escasamente, convirtiéndose, así, en proveedores de cultura y creación de estrechos vínculos entre personas que hablan distintos idiomas. Según el último informe publicado por la empresa sueca de redes y telecomunicaciones Ericsson (2018), desarrolladora, también, del primer prototipo de *smartphone* que salió al mercado, el número de usuarios de estos dispositivos crece de manera exponencial, en torno a un 6 % interanual. En concreto, en este primer trimestre del año 2018, se ha contabilizado la cantidad de 7 900 millones de dispositivos, 98 millones más que el año anterior, a nivel mundial, que está muy por encima del número de personas que viven en el mundo actualmente —entorno a los 7 650 millones (ONU, 2017). Estas cifras, las cuales se pueden poner de relieve el crecimiento de las comunicaciones interculturales y, por lo tanto, la necesidad de traductores que permitan a las personas comunicarse entre ellas. La traducción automática nace de esta necesidad de derribar las fronteras lingüísticas que separan los países y, de ahí, que las empresas dediquen parte de sus inversiones en investigación en el desarrollo de sistemas eficaces y de calidad dentro del sector de la traducción.

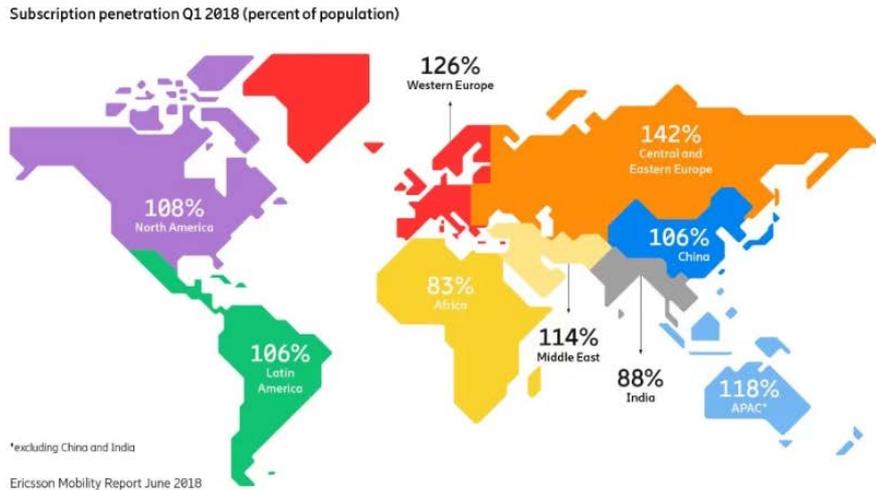


Ilustración 1. Porcentaje poblacional de usuarios de dispositivos móviles a escala mundial (Ericsson, 2018).

Objetivos y competencias

Este trabajo surgió como una oportunidad para estudiar algo que hasta el momento se desconoce. Como usuaria de productos Apple y como traductora, descubrir que en la nueva actualización de Siri, el asistente personal que se incluye en todos sus productos, se había incluido una función de traducción despertó mi interés por saber si era competencia para el conocido Google Translator que tantísimas personas descargan en sus dispositivos. Por ello, el objetivo principal de este estudio es analizar tanto las traducciones de Siri como las de Google Translator con el fin de saber si contienen muchos errores o no, para lo que se han escogido textos generales de temática diferente y de diferentes textos especializados, y compararlas para poder dar una valoración apropiada a cada uno de ellos.

A raíz de este objetivo principal, surgieron los objetivos secundarios. Cada uno de ellos guarda relación con los diferentes capítulos de este trabajo:

- Comprender la forma de trabajo de los motores de traducción automática.
- Estudiar dos traductores automáticos conocidos, en este caso, Siri y Google Translator.
- Localizar distintos tipos de inadecuaciones en las traducciones y clasificarlas.
- Establecer una metodología de análisis y extraer las debilidades y las fortalezas de cada motor de traducción.

Desde una perspectiva más global, el presente TFM me ha permitido desarrollar las siguientes competencias. Haciendo referencia a las competencias generales recogidas en la Memoria presentada desde el centro, se puede destacar que, gracias a este trabajo, hemos podido mejorar su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios, o multidisciplinares, relacionados con la traducción profesional (G2), que

es capaz de comunicar sus conclusiones —y los conocimientos y razones últimas que las sustentan a públicos especializados y no especializados de un modo claro y sin ambigüedades (G4) y que ha mejorado su habilidad de aprendizaje, de forma que podrá continuar sus estudios de forma autónoma (G5).

Haciendo referencia a las competencias específicas, basadas en las directrices de la red EMT (*European Master's in Translation*), cabe destacar hemos trabajado para dominar las técnicas y estrategias de relectura y revisión de una traducción (E6), hemos comprendido las estructuras gramaticales, léxicas e idiomáticas, así como las convenciones gráficas y ortotipográficas de las lenguas implicadas en la traducción (E8), hemos sido capaces de analizar textos y discursos generales y especializados en las lenguas de trabajo, identificando los rasgos lingüísticos y de contenido relevantes para la traducción (E15), hemos gestionado con eficacia y valorado críticamente las fuentes y los recursos de información y documentación necesarios para el ejercicio de la traducción general/especializada en las lenguas de trabajo, incluyendo la consulta a expertos (E17), hemos sido capaz de utilizar las herramientas informáticas básicas como instrumento específico de ayuda a la traducción en las diferentes fases del proceso traductológico — procesamiento de textos, corrección ortográfica y gramatical, revisión y control de cambios, formateado textual etc.— (E18), hemos desarrollado un método de trabajo organizado y optimizado gracias al empleo de herramientas informáticas (E19), hemos manejado las últimas tecnologías documentales aplicadas a la traducción (E20), hemos aprendido a desarrollar el conocimiento temático en los diferentes campos de especialidad —dominio de conceptos y contenidos básicos, terminología, fraseología, etc.— (E25) y, por último y quizás más importante por ser el argumento principal de este TFM, hemos estudiado las posibilidades y las limitaciones de la traducción automática (TA).

Metodología y plan de trabajo

Para cumplir los objetivos que se han presentado en el apartado anterior, se ha aplicado la siguiente metodología y plan de trabajo.

Se comenzó por la parte teórica, para la cual se llevó a cabo una revisión bibliográfica relacionada con los conceptos de traducción automática, inteligencia artificial y asistentes virtuales, principalmente. La amplia búsqueda documental recoge desde documentos que se pueden encontrar en fuentes y repositorios académicos (Dialnet, Google Académico), páginas web especializadas, periódicos, revistas y documentos impresos extraídos de la biblioteca de la Universidad de Valladolid a través de su catálogo digital, Almena.

Tras la elaboración de los cuatro capítulos teóricos, explicamos de forma breve algunos de los aspectos que han influido en la parte de resultados del análisis, puesto que la explicación más detallada se puede encontrar en el Capítulo 5 de este TFM. Inicialmente, se sopesaron varias

opciones sobre las que focalizar este estudio, se planteó la posibilidad de que se contemplara un estudio multilingüe del inglés hacia el español y el francés y, también, la posibilidad de establecer una comparación entre varios motores de TA. Al final, nos decantamos por una única combinación de idiomas entre Siri y Google Translator dado que había muchos aspectos que analizar. Una vez creado el corpus de textos monolingüe, se propuso una clasificación de errores propia a partir de varias fuentes de referencia, lo que nos ha permitido adaptar el análisis a los aspectos que requerían más atención. Este análisis se ha realizado a partir de la recopilación de errores, en varias tablas de Microsoft Excel, las cuales se pueden encontrar en los Anexos III y V. Estos dos últimos pasos mencionados, han permitido extraer numerosos resultados y presentar reflexiones que mejoren el trabajo de ambos motores de TA.

Dos puntos que también queremos destacar en este apartado, es que, dado que se trabaja con términos en inglés y español, se diferenciarán del cuerpo de texto con cursiva y con comillas angulares, respectivamente, —para las citas se han utilizado las comillas altas, tal y como indican las normas APA— y que los términos «error» e «inadecuación» en este TFM se han usado como sinónimos para evitar la repetición, pese a que existen matices en sus significados que los diferencian.

Estructura del trabajo

Tal y como se ha expuesto anteriormente, el objetivo principal de este trabajo es analizar las traducciones propuestas por Siri en su función de motor de traducción automática y Google Translator para poder hacer una comparativa entre ambos. Para abordar este objetivo, este TFM se ha dividido en tres partes: un marco teórico, un marco práctico y, finalmente, en una última parte, se presentan las conclusiones obtenidas de esta investigación.

Más en detalle, el marco teórico está formado por cuatro capítulos: en el Capítulo 1, titulado «Traducción automática: historia y sistemas de traducción», se realiza un repaso de la historia de la TA, desde sus orígenes basados en la inteligencia artificial como de sus inicios y su evolución hasta el día de hoy, así como de los numerosos sistemas que existen para trabajar con ella y de las limitaciones que presenta trabajar con este tipo de traductores no humanos. Este capítulo se presenta en cuatro apartados principales. En el segundo capítulo, titulado «Siri: asistente personal de Apple», se recoge toda la información que concierne uno de los objetos de nuestro estudio, Siri, el asistente virtual de Apple: su creación, sus funciones, la gran cantidad de idiomas que habla y su relación con el mundo de la traducción. Para poder abarcar toda esta información, se ha subdividido en dos apartados; el primero de ellos versa sobre los orígenes de estos asistentes y el segundo, centrado en Siri y el traductor. El Capítulo 3 abarca todos los aspectos teóricos relacionados con el otro objeto de nuestra investigación, Google Translator. En este capítulo, la información acerca del traductor es más detallada debido a que, al contrario que el asistente y

como explicaremos más adelante, es un traductor colaborativo y abierto. El Capítulo 4, titulado «El concepto de error en los estudios de traducción», es el encargado de cerrar la parte teórica; en él, se ha llevado a cabo un estudio de los tipos de errores que podemos encontrarnos en un texto meta, teniendo en cuenta varias de las muchísimas clasificaciones que existen y en presentar la que se ha creado propiamente para esta investigación.

El Capítulo 5, titulado «Metodología de la investigación», explica por qué se han seleccionado los textos que forman el corpus, qué metodología se ha seguido para el análisis y otros aspectos que han intervenido en el proceso de estudio.

En el Capítulo 6, titulado «Resultados del análisis de la producción de Siri», muestra los datos, acompañados de gráficos y tablas que los recojan de una manera más visual y clara. Se ha subdividido en cuatro apartados que se corresponden con los diferentes grupos generales de estudio en los que está dividido el corpus. El Capítulo 7 tiene la misma función que el anterior, pero en este caso se centra en la producción de Google Translator. Como último capítulo de la parte práctica, el octavo una comparación de los resultados obtenidos de los análisis, en el que se ponen en común los puntos fuertes y débiles del traductor de Siri y Google Translator.

Para terminar, se exponen las conclusiones extraídas de todo el trabajo realizado, en relación con los objetivos mencionados, con lo aprendido a lo largo del desarrollo de esta investigación, y la compilación de todas las fuentes documentales sobre las que se fundamenta el trabajo.

Además, el trabajo se complementa con cinco anexos que contienen la siguiente información:

- Anexo I: recoge los fragmentos de textos que forman el corpus monolingüe (EN), divididos por temáticas —económicos, literarios, científicos, técnicos y jurídicos—, número de caracteres —inferior o superior a 500— y grado de especialización —general, especializado— o de producción propia. Además, cada uno de los fragmentos está acompañado por sus referencias bibliográficas.
- Anexo II: recoge las traducciones proporcionadas por Siri para cada uno de los textos del corpus.
- Anexo III: a diferencia de los dos anteriores, este anexo es un archivo Excel que contiene los datos numéricos extraídos de cada uno de los textos en materia de errores, en relación con la producción del traductor de Siri. Está formado por tres hojas, la primera para los textos británicos, la segunda para los estadounidenses y la tercera para los textos de producción propia.
- Anexo IV: recoge las traducciones proporcionadas por Google Translator para cada uno de los textos del corpus.

- Anexo V: al igual que el anexo número tres, es un archivo Excel que contiene los datos numéricos extraídos de cada uno de los textos en materia de errores, en relación con la producción del Google Translator. Está formado por tres hojas, la primera para los textos británicos, la segunda para los estadounidenses y la tercera para los textos de producción propia.

CAPÍTULO 1. TRADUCCIÓN AUTOMÁTICA: HISTORIA Y SISTEMAS DE TRADUCCIÓN

La traducción automática forma parte de la lingüística computacional, área que investiga el uso de *software* para traducir texto de un lenguaje natural a otro. Su forma de trabajo se basa en una sustitución simple de las palabras por sus equivalentes en otro idioma para las traducciones más sencillas o a través de corpus lingüísticos para traducciones más complejas.

Cada día la traducción automática está más presente en la vida de los profesionales del sector, pero desde su nacimiento ha dado lugar a importantes implicaciones académicas —entre las que se encuentran las conceptuales, terminológicas, etc.—, implicaciones industriales y políticas, creando un entorno de competitividad, confidencialidad y secretismo (Cerdá, 1995: 2).

1.1 El procesamiento de lenguajes naturales y la inteligencia artificial

Las personas nos comunicamos gracias a distintos tipos de lenguajes. Los lenguajes naturales son aquellos que han evolucionado y continúan haciéndolo a lo largo del tiempo, desarrollando nuevas reglas gramaticales o creando nuevos vocablos para describir nuevos inventos, con el fin de mejorar las comunicaciones humanas. Este tipo de lenguajes nos proporcionan una gran capacidad para expresarnos incluso en las situaciones más complejas gracias a su «poder expresivo» y a su sintaxis fácilmente modelable. Los lenguajes formales son aquellos que se corresponden con los procesos de las matemáticas y la lógica, los cuales, al contrario que los lenguajes naturales, tienen una gramática preestablecida cuya sintaxis es clara y precisa y el «componente semántico es mínimo». Por último, gracias al lenguaje de programación los humanos hemos conseguido comunicarnos con las máquinas, dando lugar a la inteligencia artificial. Pese a que el lenguaje de programación, teóricamente, es un lenguaje formal, se hace la distinción diciendo que se trata de un lenguaje cuyos elementos (los componentes léxicos) están organizados a través de constructores (las reglas gramaticales) que dan lugar a programas que el propio computador es capaz de procesar y hacer funcionar en muchos sistemas a la vez (Cortez Vásquez, 2009: 46-47).

Junto a los lenguajes de programación nació la inteligencia artificial, rama que se engloba dentro de las ciencias computacionales. Recibe este nombre ya que estudia aquellas máquinas capaces de realizar las mismas actividades que una persona humana, es decir, simulan los procesos mentales de analizar, aprender, razonar de forma lógica o resolver problemas, entre otros. Es capaz de analizar datos a gran escala, así como identificar patrones y tendencias de forma rápida y precisa, gracias a lo cual puede formular predicciones de forma automática (López Takeyas, 2015).

Una de las principales actividades que desarrolla la inteligencia artificial es el procesamiento de los lenguajes naturales usando las herramientas computacionales que le brindan los lenguajes de programación; a este proceso se le denomina procesamiento computacional del lenguaje natural (PLN). El PLN está dividido en diferentes niveles —fonológico, morfológico, sintáctico, semántico y pragmático— gracias a los cuales la computadora es capaz de interpretar y analizar las oraciones que el usuario ha introducido (Cortez Vásquez, 2009: 46-47).

Gracias al PLN, la inteligencia artificial se está abriendo un marcado camino en el mundo de la traducción, creando un lenguaje artificial propio que le permite realizar traducciones de una calidad que mejora con el tiempo y hace vislumbrar que el futuro no es de los traductores humanos, sino de una colaboración entre estos y las máquinas. Los numerosos proyectos que existen a día de hoy sobre traducción automática tienen como finalidad conseguir que la inteligencia artificial imite las funciones neuronales del cerebro humano, es decir, que analice las oraciones en su totalidad, teniendo en cuenta el contexto (Intertext, 2017).

1.2 Antecedentes y situación actual

Los primeros paradigmas de TA surgieron en el siglo XVII cuando Descartes y Leibniz hablaban de crear diccionarios basados en códigos numéricos o en el siglo XX, en la década de los años 30, cuando se presentaron algunas máquinas capaces de traducir textos palabra por palabra (Alonso Martín, 2003: 94-99).

Sin embargo, lo que conocemos hoy en día por traducción automática surgió a finales de los años cincuenta y principios de los sesenta junto con la creación de los primeros ordenadores. Los primeros informáticos se propusieron un nuevo reto: introducir programas que tradujeran pares de lenguas basados en entradas léxicas, ya que, antes, se creía que las mejores traducciones surgían a partir de una enorme compilación de diccionarios almacenados dentro del ordenador (*ibidem*).

Se dedicaron una enorme cantidad de inversiones para poder financiar los proyectos de investigación en el campo de la TA tras un prototipo que se presentó en Estados Unidos en 1954, en el que se recogían 250 entradas léxicas y 6 reglas gramaticales para traducir un texto del ruso al inglés. No obstante, a medida que se profundizaba en el campo, los expertos comprobaron que su idea de una traducción basada en diccionarios no era viable, ya que no proporcionaba traducciones de calidad. Los primeros problemas a los que se tuvieron que enfrentar los desarrolladores están directamente relacionados con la sintaxis y las ambigüedades (*ibidem*).

A partir de ese momento, en los años ochenta, la estrategia que se utilizó fue la transferencia en sistemas de TA como KBMT de la Universidad de Carnegie Mellon, en Estados Unidos. En los años noventa, los programas de traducción automática salieron de los laboratorios

de investigación para dar el gran salto a las tiendas informáticas y estar al alcance de las personas; salieron al mercado programas para MS-DOS/Windows como Globalink, PC-Translator y Trascend, los cuales basaban su producto final en la traducción directa, por lo que los textos meta que producían tenían poca potencia lingüística y los resultados eran pobres (*ibidem*).

En el siglo XXI, tras haber alcanzado un punto en el que las traducciones automáticas han mejorado considerablemente, la ingente cantidad de traducciones que se solicitan cada día ha obligado a los profesionales a ver la TA como un aliado y no como un enemigo, a adaptarse a los nuevos medios digitales. Los ordenadores e Internet se han vuelto un recurso básico para la traducción, facilitando enormemente el trabajo de los traductores.

Las líneas de investigación actuales en el campo de la TA se centran en:

- adaptar los sistemas actuales hacia ámbitos con datos de entrenamiento limitados que incluyen técnicas de aprendizaje en línea y selección de frases;
- traducir entre pares de lenguas minoritarias;
- desarrollar sistemas interactivos multimodales;
- construir sistemas de TA que partan de corpus no paralelos;
- desarrollar sistemas de autoaprendizaje;
- usar los llamados «modelos continuos»;
- unir distintas tecnologías;
- incorporar información lingüística;
- traducir el habla;
- y, aplicar traductores basados en redes neuronales recurrentes a la descripción de imágenes y vídeos (Casacuberta Nolla y Vidal Ruiz, 2016: 86-94).

Los proyectos de TA han sido numerosos y cada vez avanzan más para crear un sistema de traducción automática avanzado, desarrollando modelos lingüísticos, bases de datos léxicas y una tecnología lingüística básica, la cual se conseguirá con la creación de recursos como gramáticas, diccionarios, recopilaciones de terminología y corpus textuales. Además, también deben existir normas formuladas y "directrices para la codificación e intercambio de datos lingüísticos, una investigación y experimentación prenormativa, que preste especial atención a los diccionarios y al corpus de texto" (Beltrán Gandullo, 2001: 143-147).

En algunos proyectos de estos proyectos se investiga un nuevo concepto de TA, especialmente dirigido a las empresas, que incorpora servicios personalizables de calidad que incluyen herramientas como personalización del léxico, terminología, expresiones y entrenamiento de la máquina, así como metodologías y herramientas de postedición humanas, la *traducción automática asegurizada*, haciendo que los recursos que mencionábamos anteriormente sean muy especializados. Las empresas de traducción se abren al mercado introduciendo programas de TA

personalizados para otras empresas que necesitan este tipo de servicios, ofreciéndoles una «poderosa herramienta» que da accesibilidad a todos los trabajadores de forma segura a contenidos en idiomas que desconocen, evitando que esos documentos se pasen por otros traductores automáticos que no garantizan la seguridad y la privacidad de los mismos (Linguaserve, 2018).

1.3 Sistemas de traducción y limitaciones

Tanto la traducción humana como la automática tienen sus propios desafíos. Por ejemplo, dos traductores individuales no pueden producir traducciones idénticas del mismo texto en el mismo par de idiomas, y es posible que se requieran varias rondas de revisiones para lograr la satisfacción del cliente. Pero el mayor desafío reside en cómo se pueden producir traducciones de calidad aptas para ser publicadas mediante la traducción automática (Systran, 2018).

Tradicionalmente, los distintos sistemas de traducción automática se clasificaban en función de la «potencia lingüística» de cada uno de ellos, entendiendo por potencia lingüística la cantidad de información léxica, morfológica, sintáctica, semántica, fonológica y pragmática que utiliza el sistema a la hora de traducir. La empresa Systran (2018) presenta los sistemas de traducción automática que se utilizan actualmente, la TA basada en reglas y la TA basada en estadísticas. No obstante, la investigación dentro de este campo ha conseguido modernizar estas técnicas hasta crear novedosos motores de traducción automática que trabajan con redes neuronales.

1.3.1 Sistemas basados en reglas

Inicialmente los traductores automáticos producían los textos meta en base a una serie de reglas, es decir, estaban dotados de un carácter deductivo que permitía al motor encontrar equivalencias semánticas gracias al conocimiento lingüístico que se le había introducido por medio de esas reglas. Este tipo de TA "se basa en incontables reglas lingüísticas integradas y en millones de diccionarios bilingües para cada par de idiomas", aunque los nuevos softwares ya incluyen analizadores sintácticos que permiten crear una «representación transitoria» que permite generar un texto en otra lengua. Las reglas que componen este tipo de sistemas son muy complejas, pero ofrecen la posibilidad de alimentar los diccionarios en los que se apoyan aumentando la calidad de las traducciones a largo plazo (Systran, 2018). Existen diferentes métodos de producir un texto meta según estas directrices:

- Los sistemas basados en la traducción directa tienen su base en amplios léxicos monolingües y bilingües que contenían todas las entradas lingüísticas posibles de cada

lema, es decir, que de cada verbo se introducían todas las personas en todos los tiempos, condicional en todos los modos, etc. y, después, se aplicaban reglas de flexión, concordancia y reordenamiento, gracias a las cuales se pretendía resolver los problemas principales que presentara la traducción. Es el sistema más simple; en la mayoría de los casos no incluían módulos que analizaran ni morfológicamente ni sintácticamente las oraciones del texto, por lo que la calidad de la traducción era muy limitada y sus resultados eran parcialmente utilizables, ya que no se garantizaba la correcta traducción de, entre otros, oraciones ambiguas o idiomáticas.

- Los sistemas basados en transferencia, que fueron los primeros a los que se les aplicaron técnicas de lingüística formal, llevan a cabo un proceso de traducción dividido en tres fases denominadas como fase de análisis —el programa hace un análisis a nivel morfológico, sintáctico, semántico, estructural y contextual—, fase de transferencia —el programa elige la traducción más adecuada de cada palabra teniendo en cuenta los aspectos anteriormente mencionados— y fase de generación —el programa realiza las tareas propias de adecuación de la lengua de destino del texto meta, es decir, cambia el orden de la oración, introduce o elimina material léxico, genera las formas flexivas adecuadas de las palabras en la frase de salida, combina y contrae elementos léxicos, etcétera—. Las traducciones obtenidas de estos motores de traducción eran de mayor calidad, por lo que experimentaron una etapa de auge; sin embargo, seguían presentando errores a nivel sintáctico, por lo que se concluyó que este tipo de sistema tenía mayor eficacia si se utilizaba con lenguas cercanas que tuvieran estructuras similares (Hutchins y Somers, 1992: 128).
- Los sistemas basados en una interlingua desarrollaban su propia lengua entre cada par de lenguas, es decir, descomponen el texto original para relacionarlo con las oraciones de la lengua intermedia creada específicamente para ese par de lenguas y a partir de esa «primera traducción» se vuelve a traducir hacia la lengua meta (Mendoza García, 2017: 83-84).

No obstante, esta clasificación de la que nos habla Alonso Martín (2003) no es absoluta puesto que ninguno de los tres tiene unas líneas definidas que le separen del resto, sino que, en ocasiones, pueden entremezclarse.

1.3.2 Sistemas basados en corpus

Los sistemas que emanan del análisis de corpus monolingües, bilingües o multilingües también son denominados sistemas de TA estadística. Este tipo de traducciones ha sido el más desarrollado y el que más cambios ha experimentado desde finales de los años 80.

Los sistemas basados en estadísticas parten de la premisa de que al traducir una oración de un idioma origen 'e' a un idioma meta 'f', todas las oraciones del idioma meta 'f' son potencialmente la traducción correcta de la oración en cuestión 'e1' en el idioma origen. Sin embargo, algunas oraciones tienen una probabilidad más alta de ser la traducción correcta. Para poder encontrar esta probabilidad el sistema utiliza textos bilingües o corpus alineados y mediciones de probabilidad (Mendoza García, 2017: 85-87).

Estos sistemas requieren de un entrenamiento anterior para el que se utilizan corpus paralelos –textos que ya han sido traducidos se alinean con sus originales, es decir, se relacionan las oraciones del texto original con sus correspondientes traducciones–, que han sido segmentados en oraciones, normalmente con ayuda de los signos de puntuación o del propio formato del texto. En estas combinaciones, no todas las posibilidades tienen el mismo coeficiente de probabilidad; no obstante, gracias a este método se pueden obtener diccionarios bilingües basados en el cálculo de probabilidades, algo conocido como *modelo de traducción palabra por palabra*, aunque se dice que, en muchos casos, este tipo de sistemas no son los más funcionales puesto que la gran mayoría de documentos no se encuentran en distintos idiomas, además de que necesitan al menos dos millones de palabras para que un área de estudio específica o general alcance el nivel mínimo exigido de calidad (Koehn, 2010).

Sin embargo, Microsoft (2017) afirma que *la traducción automática estadística cuenta con numerosas limitaciones como por ejemplo que sólo traduce palabras en el contexto de unas pocas palabras antes y después de la palabra traducida. Esto quiere decir que cuando trabaja con frases pequeñas, funciona bastante bien. Para los más largos, la calidad de la traducción puede variar de muy bueno a, en algunos casos, la frontera absurda. Es casi siempre posible ver que ha sido generado por la máquina.*

Al igual que Microsoft los investigadores en este campo de la traducción también han llegado a la conclusión de que *las posibilidades de mejorar el rendimiento de estos sistemas eran limitadas*, es decir, que algunos de los errores se podían corregir con conocimientos lingüísticos básicos, pero insuficientes para alcanzar la idea de un *fully automatic high quality translation* o FAHQT (Tertoolen, 2010).

1.3.3 Sistemas de traducción neuronal

La traducción automática neuronal o NMT (por sus siglas en inglés, *neural machine translation*) se acerca a la TA basada en corpus, pero los resultados que obtenemos de ella son mejores que los que se han obtenido con los sistemas basados en la estadística (Mendoza García, 2017: 98-99) dado que estos traductores han desarrollado la capacidad de «pensar y razonar» por

si mismos. Sin embargo, algo característico de las NMT es que tanto las palabras como las frases se representan con vectores y números, gracias a lo cual se han podido aplicar nuevas técnicas de aprendizaje automático.

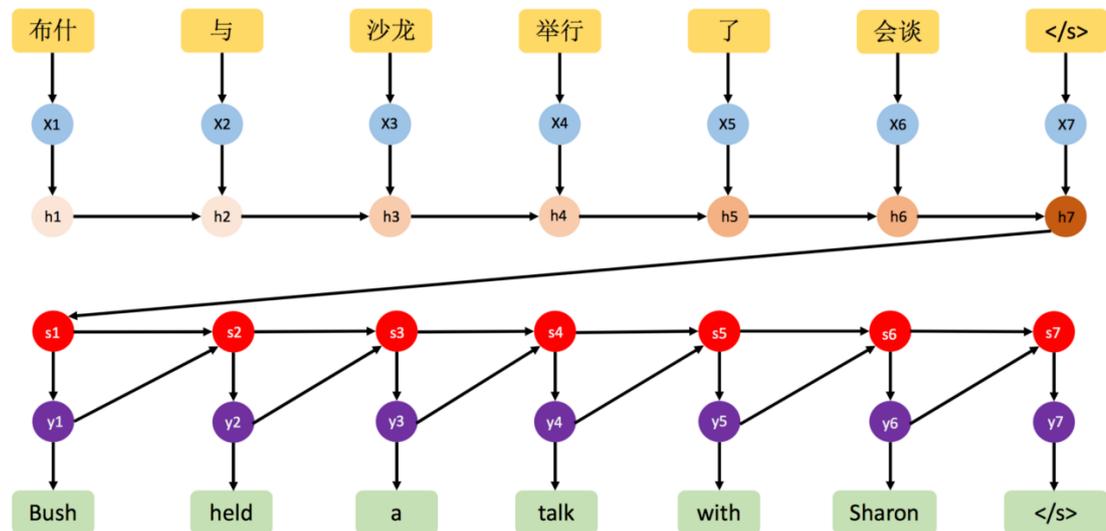


Ilustración 2. Representación del funcionamiento de un sistema de traducción neuronal (Sutskever et al., 2014).

1.3.3.1 Computación y redes neuronales

Las neuronas artificiales, también denominadas unidades básicas de la red, son elementos formales conceptualmente inspirados en las propias neuronas biológicas, creando así, «cerebros» para las máquinas organizados de la misma manera que el cerebro humano. Estas configuraciones son las Redes de Neuronas, las cuales, entre sí, reciben información del resto de módulos o del propio interno, la integran, la computan y emiten una única salida que se transmite idéntica a las neuronas posteriores (Wasserman, 1989).

Las redes neuronales de las que están compuestas estos sistemas de traducción son unidades de procesamiento de la información simples, es decir, que están formadas por dos capas, una por la que entra dicha información y otra por la que sale. Estas neuronas, también conocidas como perceptrones, están densamente conectadas entre sí y su función es crear un producto, que en este caso, sería la traducción (Bishop, 2006). Gracias a estas redes es posible calcular de manera sencilla para el usuario las respuestas lógicas posibles (en caso de que existan) para una entrada desconocida, en relación con un conjunto de datos de entrada y de salida de una determinada función, ya sea lineal o no. La red neuronal más conocida y la que se utiliza para desarrollar sistemas de TA es más compleja y se conoce como perceptrón multicapa, ya que está compuesta de numerosas capas de neuronas, de forma que la salida de una de ellas es la entrada de la siguiente y así se cree un sistema totalmente interconectado (Casacuberta Nolla y Peris Abril, 2017). Según Elman y Jordan (1990), existe otro tipo de redes conocidas como redes neuronales

recurrentes, que son aquellas que se realimentan con sus propias salidas, lo que las hace potentes modelos de secuencias temporales.

Uno de los principales objetivos de los sistemas autónomos es emular la habilidad que posee el hombre para interactuar con el ambiente y aprender de dichas interacciones. Es así como necesita de una estructura flexible, capaz de desempeñarse en ambientes de operación dinámicos sujetos a diversas incertezas y perturbaciones. Dentro de las aproximaciones existentes en la teoría de Inteligencia Artificial, surgen Redes de Neuronas Artificiales como elementos capaces de proveer de dicha estructura flexible, mediante la integración con diversos sistemas de aprendizaje (Gestal Pose, 2009).

Estos sistemas, como mencionábamos anteriormente, tienen la capacidad de aprender de forma automática y lo hacen de dos formas: sintetizando reglas, ya que son "capaces de clasificar casos no vistos en un conjunto finito de categorías discretas", y los que trabajan con funciones (Escobar, 2009). Aun así, este aprendizaje se estimula a partir de un corpus de entrenamiento mediante la ampliación de un algoritmo popularmente conocido como *algoritmo de retro-propagación del error* (Casacuberta Nolla y Peris Abril, 2017).

1.3.3.2 La traducción basada en redes neuronales

Microsoft (2017) plantea que "durante muchos años (desde los años sesenta), los informáticos intentaron construir sistemas de traducción automática basados en reglas gramaticales y estructuras de lenguajes individuales. En la mayoría de los casos los resultados fueron, digamos, menos que grandiosos". De estos avances, surgió un nuevo concepto aplicado a la TA: el aprendizaje automático. Estas nuevas tecnologías, denominadas como *Deep Learning* o *Deep Neural Networks*, tratan de imitar el funcionamiento del cerebro humano, aunque únicamente lo consiguen de forma parcial.

El entrenamiento que se solía utilizar para que estas máquinas aprendiesen a traducir se basaba en ofrecerles una gran cantidad de datos y tenerlas bajo supervisión durante el proceso; se les introducían diccionarios y textos paralelos que contuvieran las mismas frases en distintos pares de idiomas y su labor era adivinar la traducción adecuada para cada oración. Durante sus primeros pasos, cuando la máquina se equivocaba, recibía las respuestas correctas de los ingenieros que la estaban desarrollando con el fin de que aprendieran de sus propios errores y, así, se ajustase el proceso. Este tipo de métodos eran muy eficaces cuando los pares de lenguas escogidos tenían un gran número de hablantes, puesto que se podían encontrar muchísimos documentos que enriquecen el conocimiento de la máquina. Sin embargo, cuando el traductor automático quería abordar idiomas raros o poco comunes de los que no se pueden extraer suficientes textos paralelos, la calidad de la traducción final, se reducía (*ibidem*).

Este «impedimento» ha sido la causa por la que se han desarrollado los nuevos sistemas de aprendizaje automático que permiten a la máquina aprender por si sola. Estos métodos permiten a las redes neuronales agrupar palabras en ambos idiomas de forma que crean sus propios diccionarios, glosarios específicos y bases de datos. "Una vez han creado su propio diccionario, llega la hora de traducir frases". Se utilizan dos métodos denominados «traducción inversa» —se realiza una traducción aproximada de la frase de un idioma a otro y la frase meta se vuelve a traducir al idioma original; en caso de la frase retraducida no sea idéntica a la inicial, las redes neuronales se ajustan por si mismas para aumentar la efectividad en su siguiente tarea— y «eliminación de ruido», también conocido como *Denoising*. En traducción, se denomina «ruido» a todos aquellos elementos de la oración que se reorganizan o se eliminan palabras durante el proceso antes de volver a traducirlo al idioma original. "Combinando estos dos métodos las redes neuronales aprenden las estructuras de los idiomas y cómo traducirlo correctamente" (Yúbal, 2017).

El funcionamiento de estos sistemas está dividido en dos etapas. La primera es la encargada de «modelar» la palabra que necesita ser traducida, basando este modelaje en el contexto de dicha palabra (y sus posibles traducciones) dentro de la oración completa. De esta forma, entorno a esta palabra, se crea una «red de nervios». Esta red es la que será traducida según el contexto de la oración, en la lengua meta (Microsoft, 2018). La siguiente imagen explica de manera visual cómo se crea esta red y como se desarrolla el proceso de traducción a posteriori.

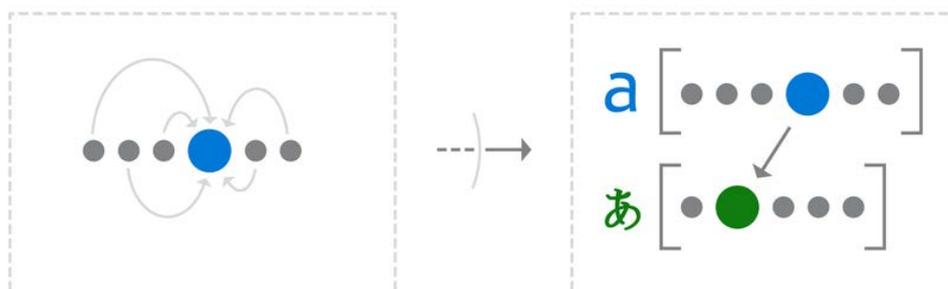


Ilustración 3. Funcionamiento de un sistema de TA neuronal (Microsoft, 2018).

Con el fin de exponer de una forma más clara los conceptos recogidos en estos apartados, hemos elaborado la siguiente tabla, en la que se recogen las principales características de cada uno de los sistemas de traducción automática con el fin de mostrar de forma más visual las relaciones que existen entre todos ellos y que se han comentado en los apartados interiores:

	Sistema basado en reglas	Sistema basado en estadística	Sistema basado en redes neuronales
1	Entrenamiento previo a través de diccionarios y glosarios	Entrenamiento previo a través de corpus paralelos	Entrenamiento previo con diccionarios y corpus (autoaprendizaje)
2	Mejores traducciones con lenguas cercanas		No importa el par de lenguas que se introduzca
3	Traducción literal, palabra por palabra o basada en una interlingua	Traducción en contexto	
4	Análisis de la estructura y cambios sintácticos		

Tabla 1. Comparativa de los sistemas de TA. Elaboración propia.

1.4 Limitaciones de la traducción automática

Como vemos, independientemente de que usemos un sistema u otro de traducción automática, todos ellos tienen unas limitaciones para las que no se ha encontrado la solución aún. Uno de los principales problemas a los que se enfrentan estos tipos de sistemas de traducción es que la complejidad del lenguaje natural hace que no siempre se construya de forma lógica. Un ejemplo de ello es la polisemia con la que están dotadas las palabras, dependiendo del contexto en el que se utilicen, su posición dentro de la oración, de las variantes regionales dentro de un mismo idioma, de la información que se quiera transmitir, de la intención del mensaje, etc. (Hernández Mercedes, 2002).

Alonso Martín (2003: 98-100) propone un claro ejemplo de la ambigüedad del lenguaje humano. Las frases *ayer estuve en un banco leyendo el periódico* y *ayer estuve en un banco para hablar con el director* presentan dos significados altamente alejados de la palabra banco; en la primera oración, banco es un elemento físico en el que podemos sentarnos, mientras que, en la segunda oración, banco es una entidad bancaria. Si tuviéramos que hacer una traducción de la palabra banco en base a esta oración, por ejemplo, al francés (*banc-banque*, respectivamente) o

al inglés (*bench-bank*, respectivamente) nos damos cuenta de lo importante que es una buena interpretación de una oración en base al contexto. A estas pequeñas diferencias polisémicas se les denomina como «conocimiento del mundo» y ellas son las líneas que dividen la traducción humana y la traducción automática. El conocimiento del mundo es tan extenso que condensarlo todo para introducirlo en un programa informático es casi imposible.

Además, los sistemas de TA no tienen la capacidad de reaccionar ante entradas inesperadas —en el texto original pueden aparecer estructuras o términos novedosos o errores, dando lugar a oraciones gramaticalmente incorrectas— o de procesar la idiomática de algunas lenguas —con las traducciones automáticas se corre el riesgo de crear una «nueva lengua» basada en la traslación o traducción literal del contenido del texto (Krauer, 2003).

Por otro lado, Arnold (2003: 121) habla de cuatro limitaciones que los ordenadores no pueden superar:

- *Realizar tareas no específicas*: el lenguaje natural no tiene unas reglas lo suficientemente precisas como para que su procesamiento sea 100% fiable y proporcione ideas precisas.
- *Aprender cosas nuevas*: el mundo y la lengua cambian de forma constante por lo que los traductores tienen que estar al día de todas las novedades en su ámbito de trabajo para no cometer errores, lo que requiere creatividad para crear nuevas reglas y no seguir las que ya están establecidas.
- *Realizar razonamientos de sentido común*: los ordenadores son incapaces de razonar por sí mismos conocimientos generales como que el agua es húmeda o el fuego caliente.
- *Enfrentarse a problemas con un gran número de soluciones posibles*: los traductores automáticos, en ocasiones, se enfrentan a infinitas posibilidades para resolver un problema de traducción, como, por ejemplo, usando sinónimos de las palabras que forman la oración o cambiando el orden sintáctico de la frase.

No obstante, como podemos ver, el estudio de estas limitaciones se hizo hace ya unos años, por lo tanto, en el estudio de la parte práctica se observará si estas limitaciones se han solucionado gracias a las mejoras en IA.

CAPÍTULO 2. SIRI: ASISTENTE PERSONAL DE APPLE

2.1 Agentes conversacionales y asistentes virtuales personales

Los agentes conversacionales inteligentes son seres virtuales que surgieron hace unos años en las páginas web, con aspecto humano, que tienen la capacidad de responder las preguntas que se les planteen, "algunas veces de forma correcta y otras no tanto". Orientan y dan respuesta a las inquietudes de los usuarios sobre un tema específico. Inicialmente eran conocidos como *chatbots* —*chat*, término de origen inglés que significa hablar, charlar, chatear; *bot*, diminutivo de robot, programa informático que imita el comportamiento de una persona—, un término acuñado por Michael Mauldin en el año 1994. Con el paso del tiempo, se han ido perfeccionando, dotándoles de conocimiento y de las capacidades necesarias para aprender, gracias a lo cual consiguen transmitir a la persona con la que están interactuando la sensación de estar hablando con una persona física y no con una máquina. Sin embargo, las respuestas más fiables son aquellas que contestan a preguntas o frases que son fáciles de comprender y coherentes, de las cuales reconocen una serie de palabras que les permiten dar respuestas que ya se les han introducido de ante mano.

Años después, aparecieron los asistentes virtuales personales cuya diferencia con los agentes conversacionales es que utilizan simuladores de voz, lo que les da un mayor realismo a la hora de interactuar. Encontramos este tipo de asistentes en los dispositivos móviles, es decir, *smartphones*, *tablets* y *smartwatches*. Están considerados como la segunda generación de *chatbots* y sus funciones van más allá de funciones específicas dentro de una web, entre ellas, facilitar al usuario del dispositivo sus tareas diarias. "Algunas de las funciones mínimas que se piden a estos asistentes virtuales son las siguientes":

- Que esté pensado y diseñado desde el punto de vista del usuario, al que tienen que asistir.
- Que identifique correctamente las necesidades y sepa atenderlas.
- Que se pueda acceder de manera ágil a cada función.
- Que sus funciones sean manejables.
- "Que su pronunciación sea entendible". No obstante, tanto la entonación como el tono de voz de estos asistentes aun dista mucho de asemejarse a la humana.

El salto de los *chatbots* alojados en páginas webs a asistentes virtuales personales en dispositivos móviles, ha desatado también una feroz batalla entre estos gigantes por ver quién es el primero que es capaz de crear al asistente virtual perfecto y con ello intentar ganar una cuota indeterminada del mercado, lo que supondrá miles o millones de posibles consumidores de sus productos (Jiménez Martín y Sánchez Allende, 2015: 21).

2.2 Siri: nacimiento y evolución

Siri fue el primer asistente virtual que salió al mercado. Surgió después de varias décadas de investigación en la empresa SRI, especializada en inteligencia artificial, hace más de 70 años, en 1946. Este tipo de tecnología fue desarrollada siguiendo los pasos del Asistente Cognitivo de la empresa que tenía funciones como aprender y organizar (CALO, por sus siglas en inglés: *Cognitive Assistant that Learns and Organizes*). Se desarrolló dentro del proyecto de inteligencia artificial más conocido en la historia de Estados Unidos, PAL (*Personalized Assistant that Learns*), mientras trabajaban en conjunto con el Instituto Suizo de Tecnología EPFL. Su principal creador es Adam Cheyer, cofundador de la empresa que dio a conocer este asistente virtual, Siri Inc, y pionero en inteligencia artificial. Siri Inc., antes de ser comprada por Apple, era una organización sin ánimo de lucro de la Universidad de Stanford, que tenía como objetivo llevar a cabo iniciativas que cambiaran el mundo. Cheyer trabajaba en ella como Ingeniero de Computación, junto con el Dag Kittlaus, CEO noruego-estadounidense, Tom Gruber, genio de la web semántica, y Norman Winarsky (SRI Internacional, 2007).

La *startup*, que nació en 2007, buscaba un nombre fácil de recordar, que no fuera largo, cómodo de pronunciar en cualquier lengua, un nombre humano que no fuera muy común con el objetivo de poder obtener el nombre de dominio por poco dinero. Siri se proclamó como mejor opción entre todos los miembros del equipo y cada uno de ellos explicó por qué le gustaba esa opción: Kittlaus dijo que le gustaba por su significado en noruego, «mujer hermosa que te llevará hasta la victoria». Para Cheyer, fue el significado en suajili, «secreto». Por otro lado, también es la palabra Iris escrita a la inversa, un sistema de software que este último ayudó a construir como parte del proyecto CALO. A otros miembros también les gustó el parecido con el nombre de su empresa desarrolladora SRI y su significado en cingalés, «belleza» (Cheyner, 2012).

Siri fue adquirida por Apple el 28 de abril de 2010 por la cantidad de 200 millones de dólares y sus cuatro creadores fueron contratados tras la venta de la aplicación. Inicialmente, solo era capaz de trabajar con el inglés, por lo que no se presentó integrado dentro del sistema operativo de iOS tal y como lo conocemos hoy en día, sino que se comercializó como una aplicación independiente. Sin embargo, pronto se vio con claridad que la intención de Apple era trasladarlo rápidamente a otros idiomas, haciendo de este sistema uno de los productos más codiciados por los usuarios. Esta actualización se presentó en octubre de 2011, integrada en el sistema operativo del iPhone 4s, debutando como una «pequeña revolución».

Este asistente pasó a interactuar en inglés con acentos de Estados Unidos, Australia y Reino Unido, en francés y en alemán; aunque no todas las funcionalidades que se podían encontrar en la versión en inglés, estaban en las versiones desarrolladas para los otros idiomas. Con la

versión actualizada de iOS 6 (2012), entraban a formar parte del amplio elenco nuevos idiomas, entre ellos, el español (Archanco, 2016).

El reconocimiento de voz, la funcionalidad y nivel de asistencia han sido mejorados significativamente desde que Siri salió al mercado. Entre sus funciones más actualizadas se encuentran mostrar la previsión del tiempo, el calendario, escribir consultas, configurar alarmas, acceder a las aplicaciones, realizar llamadas y videollamadas, acceder a la cámara y al GPS, enviar mensajes o e-mails, reproducir música y reconocer canciones, etc., como podemos ver en las imágenes; aunque la que más nos interesa por ser objeto de nuestro estudio, es la nueva función de traducción que integra este asistente, de la cual hablaremos más adelante (Santiago Pérez, 2014).

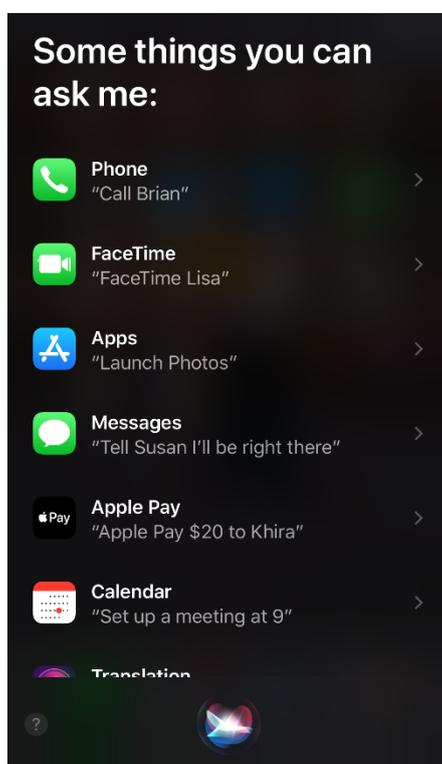


Ilustración 4. Funciones de Siri (I). Elaboración propia.

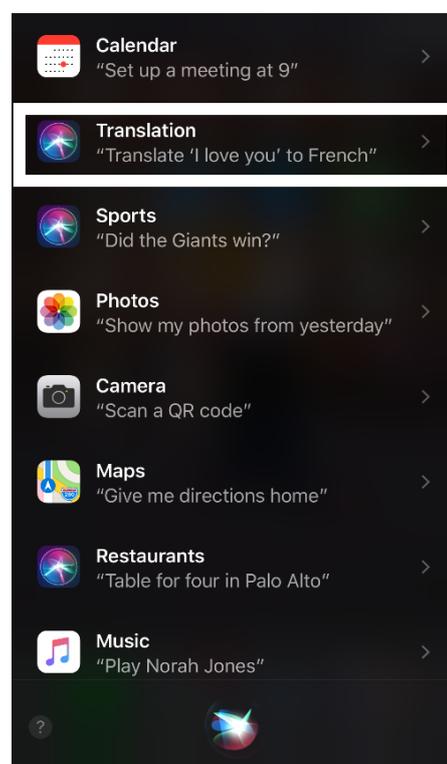


Ilustración 5. Funciones de Siri (II). Elaboración propia.

2.2.1 El multilingüismo de Siri

Este sistema destaca por la gran cantidad de idiomas en los que es capaz de procesar información, un total de 20, con sus respectivos dialectos y variedades:

- | | | |
|-----------|-----------|------------|
| - inglés | - alemán | - chino |
| - ruso | - japonés | - italiano |
| - francés | - coreano | - español |

- | | | |
|--------------|---------|----------|
| - neerlandés | - danés | - finés |
| - tailandés | - turco | - malayo |
| - portugués | - sueco | - hebreo |
| - noruego | - árabe | |

Este asistente aprende a hablar un idioma gracias a la colaboración de personas reales que leen palabras y párrafos de diferentes áreas geográficas. Estas lecturas se graban, se transcriben por humanos, con el objetivo de verificar cómo se pronuncian las palabras más comunes de un idioma, y se ofrecen a un programa dotado de inteligencia artificial que sea capaz de aprender de ello. De esta forma, tras procesar la información que se le ha dado, comienza a predecir cómo se pronunciarían el resto de las palabras que no se encuentran en su base de datos, formando oraciones y cadenas de palabras que mejoren su vocalización (Rus, 2017).

2.2.2 El traductor integrado en Siri

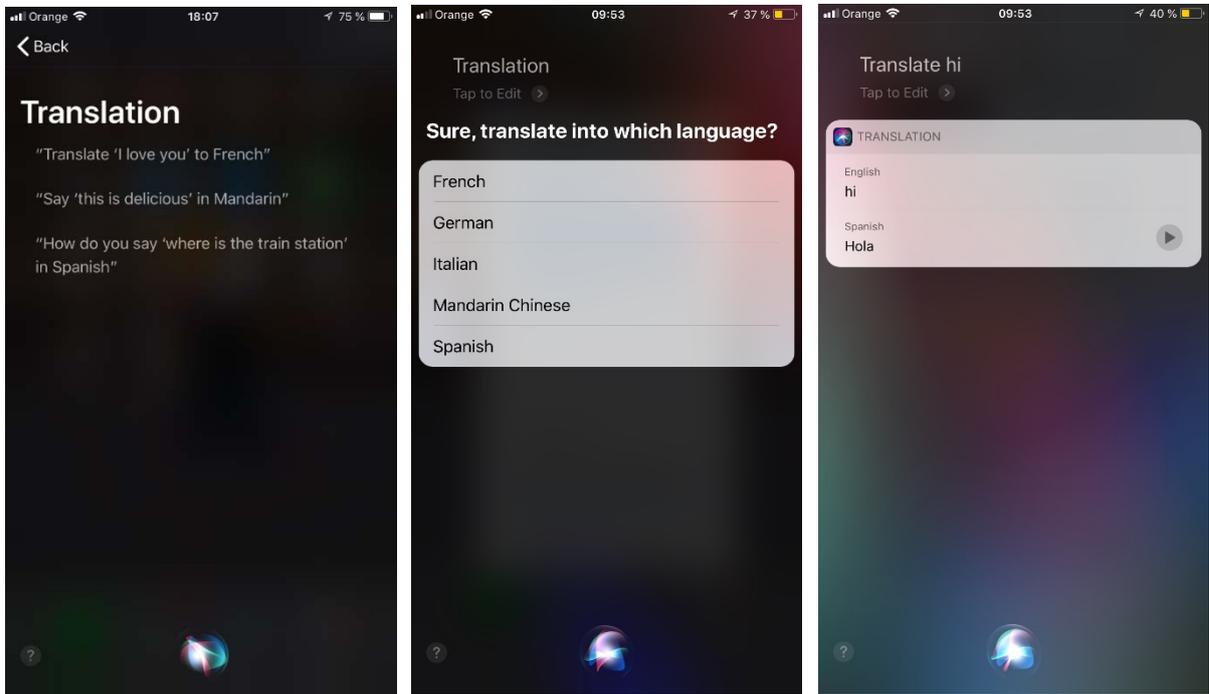
En sus primeras versiones, podíamos ampliar la funcionalidad de Siri gracias a un ajuste llamado Lingual, creado por el equipo de CodeThemed, permitiendo la traducción del inglés a más de 30 idiomas. Esta aplicación exclusiva para Apple se podía instalar utilizando la API de Microsoft *Bing Translate* (Souppouris, 2012).

Sin embargo, el pasado año 2017, con la actualización a iOS 11, se introdujo por primera vez en este asistente la capacidad de traducir sin necesidad de utilizar aplicaciones externas como Google Translator o iTranslate. Aunque cuenta con un pequeño inconveniente, para poder utilizar esta nueva función de Siri, debe estar configurado en su versión inglesa de Estados Unidos. Sin embargo, es capaz de realizar traducciones de forma instantánea del inglés de frases y palabras hacia el mandarín, el francés, el alemán, el italiano y el español (Martínez, 2017). No obstante, se espera que en próximas actualizaciones se integren nuevos idiomas para que se convierta en un traductor universal (González, 2017).

Esta nueva función, presentada en la conferencia para desarrolladores WWDC (*Apple Worldwide Developers Conference*) ante más de 5 000 asistentes, le permite convertirse en el nuevo traductor portátil disponible para los 375 millones de usuarios de productos Apple (Chicharro, 2017).

Para poder traducir con este motor de traducción, primero el dispositivo tiene que estar configurado en *Inglés (Estados Unidos)*. A continuación, hay que ejecutar el asistente, ya sea mediante el reconocimiento de voz que integra —diciendo *Hey Siri*— o mediante una pulsación mantenida del botón *Home*. Para que inicie la función de traductor, bastaría con decir la siguiente frase *How do I say (texto que queremos traducir) in (lengua de destino)*, aunque también tenemos

la opción de introducir de manera manual a través del teclado palabras como *translate* o *translation*, seguido del texto que queremos traducir. Unos instantes después, Siri responde con su traducción. Además, te permite escuchar cómo se pronuncia el texto meta (González, 2017). Las siguientes imágenes muestran el proceso que acabamos de describir:



Ilustraciones 6, 7 y 8. Traductor de Siri (I, II y III). Elaboración propia.

CAPÍTULO 3. GOOGLE TRANSLATOR: EL TRADUCTOR DE ANDROID

3.1 Google: nacimiento y evolución

Google nació en 1995 en el seno de la Universidad de Stanford como un proyecto llamado *Backrub* entre un estudiante de esa universidad, Sergey Brin, y un posible estudiante a quien tenía que enseñarle la institución, Larry Page. Su motivación era crear un motor de búsqueda que usara todos los vínculos para determinar la importancia de las páginas individuales en la *World Wide Web* (WWW). Le cambiaron el nombre poco tiempo después tras inspirarse en una expresión matemática que representaba lo que querían conseguir con su proyecto: el número 1 seguido de 100 ceros es el gúgol y ellos tenían como misión organizar la información del mundo para que fuera accesible y útil para todos (Google, 2018).

"En los años siguientes, Google acaparó la atención no solo de la comunidad académica, sino también de los inversores de Silicon Valley. En agosto de 1998, el cofundador de Sun, Andy Bechtolsheim, les dio a Larry y Sergey un cheque por USD 100,000". Con esta inversión, Google Inc. salió al mercado y se abrieron las primeras oficinas en los suburbios de Menlo Park (California). La persona que les arrendó la propiedad fue Susan Wojcicki —actual directora ejecutiva de YouTube, adquirido por Google Inc. en 2006 a cambio de 1650 millones de dólares. Poco después, la empresa comenzó a expandirse rápidamente, se contrataron ingenieros y se creó un equipo de ventas. El local se quedó pequeño por lo que trasladaron las oficinas actuales en Mountain View (California) (*ibidem*).

Actualmente, su pasión sigue siendo proporcionar mejores respuestas a través de la tecnología y, para ello, gracias a la ayuda de sus más de 60 000 empleados en más de 50 países, la empresa Google es la creadora de cientos de productos que usan miles de millones de personas por todo el mundo, desde Android, pasando por el motor de búsqueda, hasta llegar al traductor (*ibidem*).

3.1.1 Google Translator para dispositivos móviles

Android es un sistema operativo para dispositivos móviles, basado en Linux —núcleo de sistema operativo libre, gratuito y multiplataforma—, que permite programar aplicaciones adaptadas a distintas interfaces que accedan a las distintas funciones de un móvil, como el GPS, las llamadas, la agenda, etc. en lenguaje de programación Java. Esta sencillez que lo caracteriza permite que existan una cantidad inmensa de aplicaciones disponibles que mejoran la experiencia del usuario y su código abierto permite que se puedan incluir en los dispositivos sin coste alguno, consiguiendo una alta demanda entre los fabricantes y desarrolladores de dispositivos móviles. De

la misma manera, cualquier usuario tiene la posibilidad de descargarlo para su posterior inspección, la cual puede incluir o no cambios (Nieto González, 2011).

Android nació en el año 2003, en Palo Alto (California), gracias a un grupo de ingenieros, pero no tuvo repercusión en el mundo de las tecnologías hasta el año 2005, momento en el que Google decidió comprar la empresa Android Inc. En noviembre de 2007, a través de la *Open Handset Alliance* —alianza comercial que aúna a 84 compañías que trabajan para desarrollar estándares abiertos para dispositivos móviles— se presentó la primera versión de Android con el fin de que las empresas comenzaran a desarrollar aplicaciones que integrar en el primer móvil, el «HTC Dream» en 2008 (*ibidem*).

Desde que Google adquirió este proyecto, las aplicaciones propias de la compañía vienen preinstaladas en la versión de fábrica del dispositivo, lo que permite acceder a todos los productos de la empresa de forma rápida y sencilla —"Google Mobile Services incorpora las apps y API más populares de Google a tu dispositivo Android de manera gratuita"— (Android, 2018).

3.2 Google Translator

Desde que entró en nuestras vidas en los años 90, Google Translator es el traductor más usado del mundo. Aunque empezó siendo uno de los mejores recursos para salir del paso con traducciones simples, hoy puede presumir de ser uno de los más potentes gracias a las continuas mejoras de traducción y, cómo no, al desarrollo de su aplicación móvil, una de las preferidas por los usuarios Android (Signewords, 2017).

El mundo de la traducción se revolucionó gracias a uno de los productos más exitosos de la empresa: el Traductor de Google o *Google Translator*. Su sistema de traducción automática supuso que todas aquellas personas que necesitaban traducciones simples y rápidas, especialmente de palabras, dejaran de lado los diccionarios bilingües. Por aquellos años, esta *machine translation* solo era capaz de traducir oraciones formadas por cuatro palabras como máximo y, en numerosas ocasiones, el resultado era ininteligible (Signewords, 2017).

En 2006, su año de lanzamiento, su sistema de traducción estaba basado en reglas; sin embargo, un año más tarde, en octubre de 2007, pasó a ser un sistema de traducción automática estadística durante nueve años de su vida. En ese momento, soportaba 58 idiomas. Actualmente, desde 2016, Google le incorporó el sistema de traducción automática neuronal con el fin de perfeccionar la calidad de sus traducciones a través de las consultas de sus usuarios, es decir, del análisis de la composición de las oraciones (IIEMD, 2016).

Este nuevo sistema, permite traducir frases completas, en lugar de hacerlo por partes. Gracias a ello, las traducciones serán más acertadas y con un resultado más humano, lo cual se ve repercutido en una frase más fácil de leer y de entender una vez ha sido traducida. Según Google, estos idiomas son suficientes como para cubrir más de un 35% de las solicitudes de traducción de los usuarios alrededor de todo el mundo (IEMD, 2016).

Otra de las novedades que se incluyó en 2016 fue la limitación de la extensión de los textos a 5 000 caracteres en su versión gratuita (Mazin, 2016).



Ilustraciones 9, 10 y 11. Google Translator para Android (I, II y III). Elaboración propia.

La *API Translation* —*Application Programming Interface*— son los servicios de pago que Google restringió para ofrecer traducciones de mayor calidad aún a usuarios que quisieran integrarlos en sus páginas web y otras aplicaciones. "Entre bastidores, la *API Translation* aprende de los análisis de registros y los ejemplos de traducción humana. Esto permite mejorar los pares de idiomas disponibles y añadir otros nuevos sin coste adicional". Se adapta al volumen de trabajo del cliente, aplicando una tarifa por carácter, más concretamente, por millones de caracteres, siendo el precio 20 dólares estadounidenses por cada millón (Google Cloud, 2018).

3.2.1 Sistema de traducción de Google Translator

Como ya hemos anticipado, el motor de traducción de Google trabajaba en su gran mayoría con estadísticas. Su desarrollador, el alemán Franz Josef Och, fue crítico con los sistemas de algoritmos basados en reglas; por eso duró tan poco el sistema de reglas con el que funcionaba

inicialmente. Según este investigador, la mejor forma de crear un sistema de traducción automática estadística con pares de idiomas nuevos es crear corpus paralelos de entre 150 y 200 millones de palabras y dos corpus monolingües en cada uno de los idiomas con más de mil millones de palabras. De esta forma, los datos estadísticos permitían traducir cualquier texto introducido (Och, 2005). Google se nutrió de los documentos de las Naciones Unidas y el Parlamento Europeo para adquirir esta gran cantidad de datos lingüísticos (Tanner, 2007). No obstante, sigue teniendo problemas de comprensión (Adams, 2010).

En su funcionamiento interno el Google Translator no traduce directamente de un idioma a otro ($L1 \rightarrow L2$). En cambio, a menudo traduce primero del idioma fuente al inglés y luego del inglés a la lengua de destino ($L1 \rightarrow EN \rightarrow L2$) (Boitet, Blanchon, Seligman y Bellynck, 2010). Cuando Google Translator genera una traducción, previamente, ha buscado patrones en cientos de millones de documentos que ya han sido traducidos con anterioridad por traductores humanos con el fin de proporcionar la traducción más adecuada (Google, 2010).

3.2.1.1 Google Neural Machine Translation (GNMT)

Esta tecnología se presentó en septiembre de 2016 y tiene su origen en el concepto de *Deep Learning* que estudiábamos antes (Schuster, Jonhson y Thorat, 2016). Nace del proyecto *Google Brain*, desarrollado por el equipo formado por los investigadores Jeff Dean, Greg Corrado y Angrew Ng (Dean y Ng, 2012), y su misión es aumentar la precisión y la fluidez a la hora de traducir y mejorar la calidad de las traducciones basándose en millones de ejemplos (Turovsky, 2016). Esta forma de trabajo se ha probado en más de 100 idiomas: se habilitó por primera vez para ocho idiomas: desde y hacia inglés y francés, alemán, español, portugués, chino, japonés, coreano y turco en 2016 (Turovsky, 2016). En marzo de 2017, se habilitaron tres idiomas adicionales: ruso, hindi y vietnamita junto con tailandés para el cual se agregó soporte más adelante (Turovsky, 2017). El soporte para el hebreo y el árabe también se agregó con la ayuda de *Google Translate Community* en el mismo mes (Novet, 2017). A mediados de abril de 2017, los Países Bajos anunciaron su apoyo a los idiomas holandés y otros idiomas europeos relacionados con el inglés (Finge, 2017). Se agregó más apoyo para nueve idiomas indios: hindi, bengalí, marathi, gujarati, punjabi, tamil, telugu, malayalam y kannada a finales de abril de 2017 (Turovsky, 2017). No obstante, todos estos idiomas, en la actualidad, solo admiten la traducción al inglés y viceversa.

Además, la herramienta GNMT usa una enorme red neuronal artificial que amplía el número de contextos posibles, reorganiza la información y la adapta con el fin de acercarse lo máximo posible al lenguaje natural humano basándose en la gramática. Además, utiliza una interlingua que conecta todos los pares de idiomas, es decir, busca puntos en común entre todos ellos. Este tipo de traducción se denomina *Zero-Shot translation*. Este enfoque permite que se

pueda realizar una traducción entre pares de lenguas para la que el traductor no ha sido entrenado con anterioridad, recurriendo a la inteligencia artificial para que la máquina genere, por si sola, conexiones entre conceptos y palabras (Johnson et al., 2016).

3.2.2 El multilingüismo de Google Translator

Google Translator tiene la capacidad de traducir 103 idiomas diferentes:

- | | | |
|------------------------|-------------------|----------------|
| - Afrikáans | - Estonio | - Kurdo |
| - Albanés | - Euskera | - Lao |
| - Alemán | - Filipino | - Latín |
| - Amhárico | - Finés | - Letón |
| - Árabe | - Francés | - Lituano |
| - Armenio | - Frisón | - Luxemburgués |
| - Azerí | - Gaélico escocés | - Maltés |
| - Bengalí | - Galés | - Macedonio |
| - Bielorruso | - Gallego | - Malayo |
| - Birmano | - Georgiano | - Malayalam |
| - Bosnio | - Griego | - Malgache |
| - Búlgaro | - Guyaratí | - Maorí |
| - Camboyano | - Haitiano | - Marathi |
| - Canarés | - Hmong | - Mongol |
| - Catalán | - Hausa | - Neerlandés |
| - Cebuano | - Hawaiano | - Nepalí |
| - Chichewa | - Hebreo | - Noruego |
| - Cingalés | - Hindi | - Panyabí |
| - Chino (simplificado) | - Húngaro | - Pastún |
| - Chino (tradicional) | - Igbo | - Persa |
| - Coreano | - Indonesio | - Polaco |
| - Corso | - Inglés | - Portugués |
| - Checo | - Irlandés | - Rumano |
| - Croata | - Islandés | - Ruso |
| - Danés | - Italiano | - Somalí |
| - Eslovaco | - Japonés | - Sesotho |
| - Esloveno | - Javanés | - Sندانés |
| - Español | - Kazajo | - Suajili |
| - Esperanto | - Kirguís | - Sueco |

- | | | |
|-------------|--------------|----------|
| - Serbio | - TAMIL | - Xhosa |
| - Samoano | - T lugu | - Yidis |
| - Shona | - Ucraniano | - Yoruba |
| - Tailand s | - Urdu | - Zul  |
| - Tayiko | - Uzbeko | |
| - Turco | - Vietnamita | |

Sin embargo, no todas las funciones del traductor est n disponibles para todos los idiomas.

Nos permite comunicarnos:

- a trav s del teclado –se puede usar en los 103 idiomas,
- de una conversaci n biling e –se puede usar en 32 idiomas,
- de im genes con texto –se puede usar en 38 idiomas,
- de la c mara para ver traducciones instant neas –se puede usar en 37 idiomas,
- de dibujos de letras o caracteres que se pueden hacer con el dedo –se puede usar en 93 idiomas,
- o de traducciones sin necesidad de estar conectado a una red de Internet –se puede usar en 59 idiomas.

Adem s, actualmente incorpora una funci n denominada «Vocabulario», que destaca y guarda traducciones en cualquier idioma para consultarlas posteriormente y, ahora, se acopla a otras aplicaciones, como WhatsApp, entre otras, para poder copiar el texto y generar su traducci n. Esta funci n se llama «Toca para Traducir» (Google Traductor, 2018).

En su versi n para Android, Google Translator est  disponible en el Play Store como una aplicaci n de descarga gratuita en dispositivos con la versi n del sistema operativo 2.1 y superiores. Se encuentra en el cuarto puesto dentro del *ranking* de aplicaciones m s populares con m s de 500 millones de descargas. Para poder activar algunas de las funciones que mencion bamos anteriormente, la aplicaci n requiere de permisos por parte del usuario para poder acceder al micr fono, a la c mara, a los SMS, al almacenamiento externo para almacenar mensajes sin conexi n y a cuentas y credenciales para iniciar sesi n y sincronizar varios dispositivos (Play Store de Android, 2018).

La traducci n en Google tambi n se desarrolla gracias a una plataforma, la *Google Translate Community*, en la que voluntarios participan sin  nimo de lucro para mejorar las traducciones de su lengua o para introducir una lengua con la que a n no trabaja, como, por ejemplo, el w lof. Dependiendo de las necesidades de cada idioma, los participantes contribuyen de diferentes formas:

- Traducción: traducción de palabras y frases a su idioma.
- Calificación: evaluando la calidad de la traducción.
- Validación: tarea simple de calificación de traducción de sí/no.
- Comparación: elegir una mejor traducción o ejemplo (Google Translate Community, 2018).

En resumen, con el paso de los años, las mejores empresas en tecnología se encuentran a la carrera por desarrollar los mejores motores de traducción automática. Esta necesidad no viene fomentada por una índole científica, sino por una necesidad práctica: la globalización ha unido naciones y ha expandido culturas, por lo que el número de traducciones aumenta día tras día de la mano de la comunicación y la información. Google Translator es uno de los muchos proyectos de esta índole que existen y, por lo que hemos estudiado, es uno de los más desarrollados también. Su herramienta de GNMT es una de las tecnologías más avanzadas en el sector de la traducción y el trabajo colaborativo que realiza con la inteligencia artificial es capaz de establecer conexiones neuronales que procesen múltiples idiomas. Siri es un motor de traducción completamente diferente. Su sistema de traducción está poco desarrollado debido a que su función principal es ser el asistente de los usuarios de productos Apple.

CAPÍTULO 4. EL CONCEPTO DE ERROR EN LOS ESTUDIOS DE TRADUCCIÓN

"De modo genérico, se puede definir el error de traducción, (aquél que se detecta en un texto, en tanto que traducción) como una ruptura de las reglas de coherencia de un TT, sean éstas de tipo gramatical, de combinabilidad léxica, congruencia semántica o de conformidad al conocimiento del mundo y de la experiencia acumulada" (Cruces Colado, 2001).

Nord (1995) recoge en uno de sus artículos que los errores en traducción se suelen describir como una "desviación de un sistema de normas o reglas". Sin embargo, partimos de la premisa de que en traducción no existen normas estrictas para cada uno de los problemas de traducción, por lo que tampoco es fácil crear un método infalible que recoja todos los tipos de errores, ya que "una determinada expresión puede considerarse «correcta» en el sentido de que es funcional o «incorrecta», según las situaciones."

Desde la experiencia personal puedo decir que errores se pueden clasificar en dos subcategorías: aquellos que se presentan durante el proceso traductor y los que surgen después de finalizar este proceso. No obstante, como el principal objetivo de este estudio es valorar la analizar las traducciones, nos fijaremos principalmente en aquellos errores que aparecen en el texto meta, es decir, en el producto final. A continuación, se presentan diferentes clasificaciones de errores que se tomarán como punto de partida para la clasificación que hemos elaborado para el análisis de nuestro corpus.

4.1 Método de clasificación de errores de Nord (1995)

Christiane Nord (1995), cuyo método de clasificación también sigue las ideas funcionalistas, está basado en una serie de exigencias intratextuales que condicionan de forma directa aquellos errores que ella tendrá en cuenta a la hora de establecer sus criterios y que son modificables dependiendo del texto que se esté trabajando. Las exigencias de carácter general por las que se rige son:

- "Para conseguir las funciones deseadas, la formulación del texto debe corresponder a las normas y convenciones estilísticas generales y genéricas de la cultura meta, así como a un registro más bien formal.
- En la producción del texto hay que tener en cuenta la distancia cultural: los conocimientos y las expectativas de los destinatarios de la cultura meta pueden ser diferentes de los de la cultura original".
- El texto debe integrarse en el formato dado y debe ser coherente con las figuras que pueda incluir.

- En caso de conflicto funcional, las informaciones sobre específicas deben tener prioridad sobre las informaciones más generales.

En base a eso, todas las «desviaciones» dentro del texto meta que no se correspondan con estas exigencias, serán consideradas errores de traducción:

- Errores pragmáticos: este grupo engloba todos aquellos que entorpecen de forma directa la funcionalidad del texto, ya que no siguen las reglas pragmáticas del encargo.
- Errores culturales: este grupo engloba todos aquellos errores específicos que perjudican a la traducción de forma indirecta puesto que no cumplen las normas y convenciones estilísticas generales de la cultura meta.
- Errores lingüísticos: este grupo engloba las faltas gramaticales, de léxico, de ortografía y puntuación, etc. (en Hurtado Albir, 2001: 94-99)

4.2 Método tradicional de clasificación de errores (1998)

Espiñeira Caderno y Caneda Fuentes (1998), en su trabajo *El error y su corrección*, presentan una clasificación de los errores basándose en los resultados que han obtenido de los textos producidos por sus alumnos en las clases de Español Lengua Extranjera. En él, tipifican los errores de la siguiente manera:

1. Primeramente, los dividen basándose en criterios de carácter lingüístico, etiológico y pedagógico.
2. Dentro de cada uno de estos grupos, encontramos subdivisiones más específicas entre las que se encuentran:
 - a. Errores lingüísticos:
 - Errores de adición: entre los que destacan el uso de preposiciones a/de en frases de infinitivo, el uso redundante del artículo indeterminado o el dequeísmo.
 - Errores de omisión: entre los que destacan la omisión de conectores, la omisión de los artículos obligatorios que preceden al sustantivo y al relativo, concordancias de género y número, etc.
 - Errores de selección falsa: entre los que destacan aquellos se producen principalmente por la confusión de elementos de una misma categoría como relativos, pronombres y posesivos, adverbios, posesivos y reflexivos pleonásticos, así como confusiones.

- Errores de colocación falsa: son aquellos que afectan al orden de los elementos en los diferentes tipos de frases.
- b. Errores etiológicos: "hacen referencia los mecanismos desarrollados por los alumnos para la producción durante el proceso de aprendizaje".
 - i. Errores interlinguales: como la transferencia.
 - ii. Errores intralinguales: como los errores de diccionario, las españolizaciones o palabras inventadas, los falsos amigos o la utilización de palabras en desuso.
 - iii. Errores de simplificación.
- c. Errores pedagógicos: entre los que encontramos los errores inducidos (para evitar la explicación de excepciones), transitorios (dependiendo del estadio de aprendizaje de los alumnos), permanentes, comunicativos o que entorpecen la comunicación.

La importancia de esta clasificación para este TFM reside en que los textos para los que se ha redactado. Esta clasificación no recoge errores de traducción, si no de producción y, como veremos más adelante, lo que se van a analizar son las producciones de ambos traductores, ya que, aunque sean traducciones, podremos encontrar algún tipo de los errores anteriormente mencionados.

4.3 Método de clasificación de errores de Hurtado Albir (1999/2001)

Hurtado Albir (2001) sigue las tendencias funcionalistas del error a la hora de desarrollar su clasificación de errores para afirmar que:

[...] la gravedad del error no depende en abstracto de su naturaleza (falso sentido, ortografía, etc.), sino que sólo puede analizarse desde una perspectiva textual, contextual y funcional que considere el elemento en cuestión en relación con el conjunto del texto, con el contexto en el que se efectúa la traducción (época, destinatario, etc.), con la finalidad de la traducción y el método elegido, con el tipo y la modalidad de traducción de que se trate, y, sobre todo en el ámbito didáctico, con la dirección de la traducción (directa o inversa). Esta concepción funcionalista desautoriza cualquier análisis descontextualizado del error. De este modo, incluso un, en principio, falso sentido puede ser una equivalencia adecuada en un contexto traductor determinado (Hurtado Albir, 2001: 303).

Tomando como base esta cita, la autora propone una clasificación de errores en la que destaca tres categorías que se aplican a las traducciones humanas escritas:

- Inadecuaciones que afectan a la comprensión del texto original: este apartado recoge omisiones, falsos sentidos, sin sentidos, no mismo sentido, adiciones, supresiones, referencias extralingüísticas mal solucionadas e inadecuaciones de variación lingüística.
- Inadecuaciones que afectan a la expresión en la lengua de llegada: ortografía y puntuación, gramática, léxico, aspectos textuales (coherencia, progresión, temática, elementos de referencia, conectores) y redacción (formulación defectuosa o poco clara, falta de riqueza expresiva, pleonasmos, etc.).
- Inadecuaciones pragmáticas, por no ser acordes con la finalidad de la traducción (en relación con el tipo de encargo, el destinatario a que va dirigida), el método elegido, el género textual y sus convenciones, etc. (Hurtado Albir, 2001: 303).

Además, completa estas explicaciones dando ejemplos de literalismos —calcos de puntuación, calcos léxicos, calcos morfológicos, calcos sintácticos, calco de orden de palabras, calco de frases, calco de conectores y calco de elementos de referencia— e inadecuaciones que podemos encontrar en las traducciones —en la comprensión (contrasentidos, falsos sentidos, sinsentidos, adición o supresión de información), en la reformulación (errores de lengua relacionados con el léxico o la gramática o formulaciones defectuosas, poca riqueza en las expresiones, etc.), al enfrentarse a problemas lingüísticos, extralingüísticos, textuales y pragmáticos. También recoge un listado de otras inadecuaciones que se deben tener en cuenta a la hora de evaluar una traducción:

(a) Que afectan a la comprensión del texto original:

- no mismo sentido
- referencia cultural mal solucionada
- inadecuación de variación lingüística
 - o inadecuación de tono
 - o inadecuación de estilo
 - o inadecuación de dialecto (social, geográfico, temporal)
 - o inadecuación de idiolecto

(b) Que afectan a la expresión en la lengua de llegada:

- ortografía y puntuación
- gramática
- léxico
- textual
 - o de coherencia
 - o de progresión temática —repeticiones innecesarias, lagunas informativas

- o de referencia —mal uso de pronombres o redes léxicas, entre otros
- o de conectores
- de redacción (Hurtado Albir, 1999: 114-120).

4.4 Método de clasificación de errores de Diéguez Morales (2001)

Esta autora hace una clasificación de los errores encontrados en las traducciones desde el punto de vista del aprendizaje. Su método nace de sus ganas por introducir y familiarizar a los alumnos de traducción para que trabajen con la TA y la posesión y porque, además, este método "ofrece a los alumnos la posibilidad de optimizar el proceso de aprendizaje de la traducción humana al detectar, describir y evaluar los aciertos y errores de la TA".

Sus criterios se rigen por patrones de carácter lingüístico y pragmático:

- Por error léxico entenderemos que se trata de error cuando el término se mantiene en inglés y cambia el sentido de la frase, a excepción de aquellas que se mantienen en su versión original.
- Por error sintáctico entenderemos que "si bien todos los términos y palabras están correctamente traducidos, la oración no respeta el orden sintáctico normal del español (traducción literal que no afecta el sentido de la oración)".
- Por error léxico-sintáctico entenderemos que no se utiliza la acepción correcta de un término, dando lugar a una oración sin sentido.
- Por error léxico-semántico entenderemos que "la acepción que ha entregado el programa de TA es incorrecta y por ello altera el sentido de la oración".
- Por sintáctico-semántico entenderemos que si el orden sintáctico se aleja del orden español y altera el sentido.
- Por error semántico- sintáctico-léxico entenderemos que "en este caso, el error se produce por la mala elección de una o más acepciones de términos, a lo cual se suma una sintaxis incorrecta en español, lo cual redundará en un problema grave de sentido" (Diéguez Morales, 2001).

4.5 Propuesta de clasificación de errores de traducción automática

Tras analizar las principales propuestas de clasificación de errores de traducción que encontramos en la literatura, observamos que no responden a las necesidades de nuestro objeto de estudio. Por lo tanto, hemos considerado pertinente diseñar nuestra propia propuesta de clasificación de errores a partir de los enfoques y modelos desarrollados por los autores anteriormente destacados.

4.5.1 Errores léxico-semánticos

En esta categoría se tendrá en cuenta la lingüística, en particular, la lexicografía y la semántica. Desconocemos que tipo de bancos terminológicos utiliza cada uno de los traductores automáticos, por lo que pueden surgir errores que abarcan desde alternancias terminológicas, neologismos, lagunas semánticas, sinonimia y antonimia contextuales, contigüidad semántica y redes léxicas. Suelen producirse cuando no se tienen en cuenta las diferencias en los vocabularios de cada lengua.

Los aspectos que se tendrán en cuenta a la hora de valorar si una traducción contiene errores de este tipo son:

- Falsos sentidos,
- mantenimiento del término original
- siglas
- lenguaje especializado y lenguaje coloquial
- neologismos,
- y topónimos.

4.5.2 Errores ortotipográficos

Martínez de Sousa (2014: 399) define la ortotipografía o escritura *tipográfica* como "*la parte de la ortografía técnica que se ocupa del estudio y la aplicación de las reglas de escritura tipográfica; es decir, el conjunto de reglas por que se rige la confección de un impreso mediante recursos tipográficos*". En sí esto no son errores de traducción propiamente dichos si no que más bien suelen ocurrir debido a un descuido o al desconocimiento de las reglas ortográficas y de puntuación de la lengua de recepción, problemas que no deberían ocurrir al enfrentarnos a un traductor automático, pero igualmente, es un ámbito interesante para nuestro estudio.

Los aspectos que se tendrán en cuenta a la hora de valorar si una traducción contiene errores de este tipo son:

- Mayúsculas y minúsculas
- cursivas
- puntuación
- comillas
- signos de interrogación y exclamación
- términos mal escritos
- y tildes.

4.5.3 Errores gramaticales

"Por gramática entendemos la organización de las palabras dentro de una oración, y sus reglas y principios" (Barnés, 2013). Por lo tanto, cuando hablamos de errores gramaticales dentro del ámbito de la traducción, nos estaremos refiriendo todas las cuestiones relacionadas con los verbos, los pronombres, los adjetivos, los sustantivos, etc.

Los aspectos que se tendrán en cuenta a la hora de valorar si una traducción contiene errores de este tipo son:

- Tiempos verbales
- aspecto verbal
- concordancia género-número
- pronombres y determinantes
- formas *-ing* del inglés
- preposiciones
- adverbios en *-mente*
- y gerundios y participios.

4.5.4 Errores sintácticos

Los errores sintácticos se producen cuando el traductor es demasiado fiel a la sintaxis de la lengua origen. Da lugar a paralelismos sintácticos, uso indebido de la voz pasiva en español, inversiones del orden natural del discurso, repeticiones de una misma palabra, omisión del sujeto, etc.

Los aspectos que se tendrán en cuenta a la hora de valorar si una traducción contiene errores de este tipo son:

- Voz pasiva y voz activa
- sujeto elíptico
- colocación de los adjetivos con respecto al núcleo del sintagma (sustantivo)
- oraciones subordinadas
- y distribución de los componentes de la oración.

4.5.5 Errores retóricos

Los errores retóricos, como su propio nombre indica, tienen relación con las figuras retóricas, tales como metáforas, comparaciones, metonimias, personificaciones, hipérbolos, etc., las cuales se encuentran con frecuencia en textos periodísticos.

Los aspectos que se tendrán en cuenta a la hora de valorar si una traducción contiene errores de este tipo serán juzgados por la forma en la que se hayan traducido, si en español se mantiene la figura retórica, si no se traduce o si se hace una traducción del sentido sin usar el lenguaje retórico.

Por otro lado, también se incluirán como error retórico aquellas traducciones en las que se haya incorporado alguna figura retórica que en texto original no aparece.

4.5.6 Errores pragmáticos

La pragmática es la parte de la lingüística que estudia el lenguaje desde el punto de vista de las circunstancias de la comunicación y de los usuarios. Cuando se comenten errores de tipo pragmático son aquellos que se comenten en la producción y en la recepción de la información entre los usuarios, mayormente debido a un desconocimiento de las normas de la lengua meta o de los valores socioculturales. Teniendo esto en cuenta, los errores pragmáticos solo pueden cometerse por fuerza de los factores extralingüísticos (Centro Virtual Cervantes, 2018).

Los aspectos que se tendrán en cuenta a la hora de valorar si una traducción contiene errores de este tipo son:

- Uso de «tú» y «usted»
- juegos de palabras
- refranes
- locuciones
- frases idiomáticas
- y el humor y la ironía.

4.5.7 Errores culturales

La lejanía entre culturas no solo se advierte en las diferencias lingüísticas existentes, sino que se amplía a las costumbres y tradiciones, entre otras cosas.

Dada la amplitud de errores que se pueden recoger dentro de esta categoría, simplemente mencionaremos algunos de los más destacados:

- Nombres propios de personas famosas,
- ciudades y países,
- fiestas,
- etc.

CAPÍTULO 5. METODOLOGÍA DE LA INVESTIGACIÓN

Este estudio se realizará sobre la base de una selección de 26 textos completos agrupados de la siguiente manera:

Diez de los textos seleccionados están extraídos de publicaciones británicas mientras que la otra mitad han sido sacados de publicaciones estadounidenses. La siguiente división se ha realizado entre textos generales y textos especializados. Los textos generales son de publicaciones de periódicos y revistas, mientras que los textos especializados los hemos extraído de manuales técnicos, novelas, estados financieros, artículos científicos y leyes. Tras realizar esta división, de cada uno de los textos completos seleccionados, se han escogido fragmentos de los mismos con una longitud inferior o superior a los 500 caracteres, respectivamente. Los últimos seis textos se dividen en cuatro producciones propias escritas por la estudiante de nivel básico según el Marco Común Europeo de Referencia y dos producciones de voluntarios con conocimientos de inglés bajos.

La razón de haber hecho una selección tan amplia, la cual se presenta de forma más visual en el esquema de más abajo, es que nos permitirá hacer un análisis exhaustivo, con datos fiables y lo más concretos posible, es decir, tener el número suficiente de ejemplos con los que presentar todos los resultados. Además, el hecho de desconocer cómo trabaja el traductor de Siri también nos obligaba, en parte, a hacer una amplia selección de textos que nos permitiera determinar su sistema de trabajo. Haber escogido textos en dos variantes diferentes del inglés como son el británico y el estadounidense se debe a que, como bien hemos mencionado anteriormente, el traductor solo funciona con la versión de iOS de EE. UU., por lo que buscamos comprobar si el hecho de trabajar con diferentes variedades diatópicas supone un problema para Siri o si, en su defecto, para Google Translator, que tampoco hace esta diferenciación. Por otro lado, trabajar con textos generales y especializados se debe a que analizaremos las traducciones con el propósito de averiguar si sus sistemas de traducción están tan desarrollados como para afrontar cualquier tipo de traducción o si, por el contrario, necesitarán limar ciertas asperezas. Por último, la decisión de trabajar simplemente con fragmentos de cada texto es porque se desconoce el número de caracteres que admite el motor de traducción de Siri —recordamos que Google Translator procesa hasta 5 000 caracteres en su versión gratuita— y si las traducciones contienen un mayor o menor número de errores según los criterios establecidos.

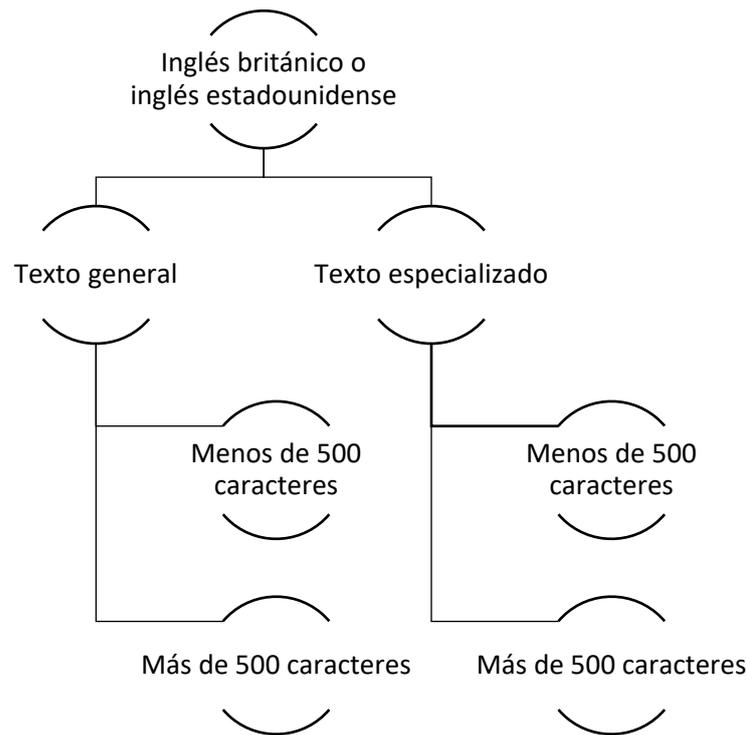


Figura 1. Esquema explicativo de la división de los fragmentos de textos escogidos para el estudio. Elaboración propia.

Tal y como se puede ver en la imagen previa, los criterios de búsqueda de textos para el corpus estaban organizados de manera jerárquica. La división en textos británicos y estadounidenses es la más importante por su amplitud, es la categoría más general de todas ellas. Después, los textos se han seleccionado en base a su grado de especialización, dividiéndolos en generales y especializados dentro de cada variedad lingüística. Por último, cada texto completo se ha fragmentado y se han seleccionado fragmentos del texto original, dividiéndolos, así, en textos que superan o no el límite de 500 caracteres.

El artículo *As the global economy picks up, inflation is oddly quiescent* de *The Economist* del cual se han extraído los fragmentos para el estudio de los textos generales económicos en inglés británico se escogió por su alto contenido en cifras, mientras que el artículo *Trump Comments Slam Puerto Rico Debt* extraído del periódico estadounidense *Wall Street Journal* ofrece dos fragmentos cargados de términos especializadas como *bankruptcy*, *investors* o *hard-hit and cash-strapped island*. Los fragmentos extraídos para el estudio de un texto especializado económico, tanto en inglés de Gran Bretaña como de Estados Unidos, han sido extraídos de los estados financieros de dos empresas de estos países, Vodafone y Apple, respectivamente. Se han elegido por ser dos textos integrados dentro del marco de la traducción especializada que nos han enseñado a reconocer en varias de nuestras asignaturas del máster.

Los fragmentos de carácter general dentro del ámbito literario se han extraído del servicio de publicaciones escritas de la *BBC* y de una revista norteamericana de poesía llamada *Ploughshares at Emerson College*. Estos dos textos son muy importantes dado que en sus líneas encontramos numerosas referencias culturales de ambos países como *Sotheran's*, *Gabriel Wells*, *Louise Glück* o el *BAP*. Por el contrario, trabajar con fragmentos de textos literarios especializados en las dos variedades es de especial interés debido a que, como bien es sabido, el lenguaje literario es, quizás uno de los más complejos de traducir, debido a que se utilizan tiempos verbales o expresiones menos comunes, figuras retóricas, etc. En este caso, se escogieron dos novelas, *The woman who stole my life* (de autora británica: Marian Keyes) y *Game of thrones: A clash of kings* (de autor estadounidense: George R. R. Martin).

Con respecto a los dos textos escogidos en representación de los textos científicos de carácter general, pertenecen a dos periódicos muy conocidos en Gran Bretaña y Estados Unidos, *The Guardian* y *The New York Times*, respectivamente. Se han seleccionado por tratar un tema similar, que guarda relación con la Prehistoria, es decir que se introducen términos de origen latino, entre otros. En el grupo opuesto, los fragmentos especializados pertenecen a artículos científicos del campo de la medicina, más concretamente, relacionados con las enfermedades cardiovasculares y oculares, respectivamente. Se han escogido por su alto contenido en vocabulario con un altísimo grado de especialización.

Los dos textos generales técnicos versan sobre empresas pioneras en el sector de la tecnología: Apple y Windows. El texto sobre Apple, que habla sobre productos de la marca, se ha escogido precisamente por la relación que guarda con uno de los sujetos de nuestro estudio y se ha extraído de *The Guardian*. En *The New York Times*, encontrábamos un artículo que habla sobre dispositivos móviles, en particular sobre *tablets*. Por el contrario, los textos especializados eran completamente distintos entre sí. El que se ha extraído de una publicación británica guardaba relación con la tecnología mientras que el estadounidense se ha extraído de un manual de instrucciones de controladores digitales de señal.

Por último, los textos jurídicos generales, una vez más extraídos de *The Guardian* y *The New York Times*, se escogieron por ser actuales y guardar relación con los distintos partidos políticos estadounidenses, así como para tratar el tema político también dentro de este análisis y por los nombres propios de personas que no son conocidas en España, con el fin de estudiar su comportamiento de adaptación a la cultura meta. Los textos jurídicos especializados, que se centran en la traducción de dos leyes de gran importancia como son la Constitución de los Estados Unidos y la ley antidrogas de Gran Bretaña, se escogieron por los aspectos culturales que contienen este tipo de textos.

Tras la selección de los textos, se han ido introduciendo en el traductor con el fin de estudiar las respuestas que nos proporciona. En base a la clasificación de errores presentada en el capítulo anterior, se ha hecho un exhaustivo análisis de las inadecuaciones encontradas, el cual se ha realizado en tres etapas:

1. Lectura de la traducción.
2. Detección de los errores.
3. Clasificación de los errores.

Se ha seguido esta estrategia con el fin de, primeramente, tener una visión global de la traducción, en la que poder detectar inadecuaciones en la lengua de llegada y denotar si existe una falta de naturalidad o no. Después, se han detectado los errores a nivel de producción, entre los que entran los errores recogidos en la clasificación y, además, la coherencia y la cohesión. Y, por último, se han clasificado esos errores en base a las categorías anteriormente mencionadas. Algunos de estos errores podrían encontrarse dentro de varias.

Tabla recapitulativa

TEXTOS DEL CORPUS		
Tipo de texto	Número de caracteres	Fuente
Texto general económico de menos de 500 caracteres británico	499	The Economist
Texto general económico de más de 500 caracteres británico	752	
Texto especializado económico de menos de 500 caracteres británico	321	Vodafone Group
Texto especializado económico de más de 500 caracteres británico	602	

Texto general literario de menos de 500 caracteres británico	457	BBC
Texto general literario de más de 500 caracteres británico	790	
Texto especializado literario de menos de 500 caracteres británico	373	The Woman Who Stole My Life: Keyes, M.
Texto especializado literario de más de 500 caracteres británico	1111	
Texto general científico de menos de 500 caracteres británico	207	The Guardian
Texto general científico de más de 500 caracteres británico	695	
Texto especializado científico de menos de 500 caracteres británico	273	British Medical Journal
Texto especializado científico de más de 500 caracteres británico	728	
Texto general técnico de menos de 500 caracteres británico	167	The Guardian
Texto general técnico de más de 500 caracteres británico	764	

Texto especializado técnico de menos de 500 caracteres británico	397	IBASE Technology Inc
Texto especializado técnico de más de 500 caracteres británico	559	
Texto general jurídico de menos de 500 caracteres británico	489	The Guardian
Texto general jurídico de más de 500 caracteres británico	743	
Texto especializado jurídico de menos de 500 caracteres británico	419	Drugs Act 2005
Texto especializado jurídico de más de 500 caracteres británico	731	
Texto general económico de menos de 500 caracteres estadounidense	380	Wall Street Journal
Texto general económico de más de 500 caracteres estadounidense	649	
Texto especializado económico de menos de 500 caracteres estadounidense	408	Apple Inc.
Texto especializado económico de más de 500 caracteres estadounidense	592	

Texto general literario de menos de 500 caracteres estadounidense	498	Ploughshares at Emerson College
Texto general literario de más de 500 caracteres estadounidense	679	
Texto especializado literario de menos de 500 caracteres estadounidense	376	Game of thrones: A clash of kings, R.R. Martin, G.
Texto especializado literario de más de 500 caracteres estadounidense	734	
Texto general científico de menos de 500 caracteres estadounidense	325	The New York Times
Texto general científico de más de 500 caracteres estadounidense	557	
Texto especializado científico de menos de 500 caracteres estadounidense	194	American Journal of Ophthalmology
Texto especializado científico de más de 500 caracteres estadounidense	616	
Texto general técnico de menos de 500 caracteres americano	418	The New York Times

Texto general técnico de más de 500 caracteres estadounidense	603	
Texto especializado técnico de menos de 500 caracteres estadounidense	374	Texas Instruments
Texto especializado técnico de más de 500 caracteres estadounidense	796	
Texto general jurídico de menos de 500 caracteres estadounidense	220	The New York Times
Texto general jurídico de más de 500 caracteres estadounidense	743	
Texto especializado jurídico de menos de 500 caracteres estadounidense	329	The Constitution of the United States of America. U.S.
Texto especializado jurídico de más de 500 caracteres estadounidense	619	

Tabla 2. Datos de los textos del corpus. Elaboración propia.

Una vez establecida la metodología de la investigación, pasamos a analizar los resultados de nuestros análisis. Los hemos clasificado en dos capítulos diferentes, con el fin de hacer una diferenciación entre los obtenidos con el sistema de iOS y con el de Android y, así, poder establecer una comparativa a todos los niveles, una vez detectados las fortalezas y debilidades de cada uno de ellos.

CAPÍTULO 6. RESULTADOS DEL ANÁLISIS DE LA PRODUCCIÓN DE SIRI

El análisis que se ha llevado a cabo para poder estudiar las inadecuaciones de traducción que comete este asistente tiene su base en la clasificación de errores desarrollada de forma personalizada anteriormente para este proyecto. Se presenta en cuatro apartados, correspondientes con cada una de las categorías generales en las que se centra este estudio: las diferencias entre variedades, diferencias por número de caracteres, diferencias por tipo de texto y, por último, los errores encontrados en los textos básicos. Dentro de cada apartado, se presentan los porcentajes de error en forma de gráficos, mientras que los datos numéricos a los que se hace referencia se pueden encontrar en el Anexo III de este trabajo. Es importante dejar constancia de que la versión de iOS con la que se ha trabajado es la 11.3.1, ya que se han realizado actualizaciones en el *software* posteriores a la fecha en la que se comenzó el análisis.

6.1 Criterio 1: inglés británico e inglés estadounidense

La primera clasificación de errores se recoge en materia de variedad. Iniciamos el análisis comentando aquellos textos que el traductor no fue capaz de procesar. Hemos constatado que hay un porcentaje mayor de textos en inglés estadounidense que no ha podido traducir, un total de 12, frente a los 7 en inglés británico. Las siguientes figuras recogen de manera gráfica estos datos y las ilustraciones son un ejemplo de las respuestas que da cuando ocurren estas dos situaciones:

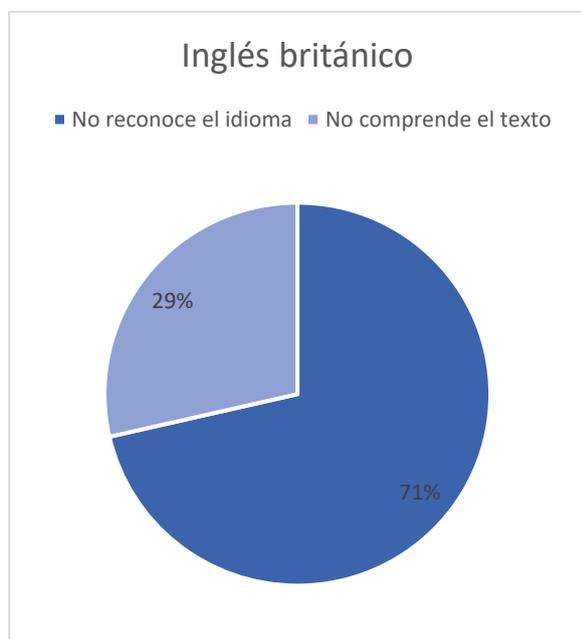


Figura 2. Porcentaje de textos no traducidos en inglés británico. Elaboración propia.

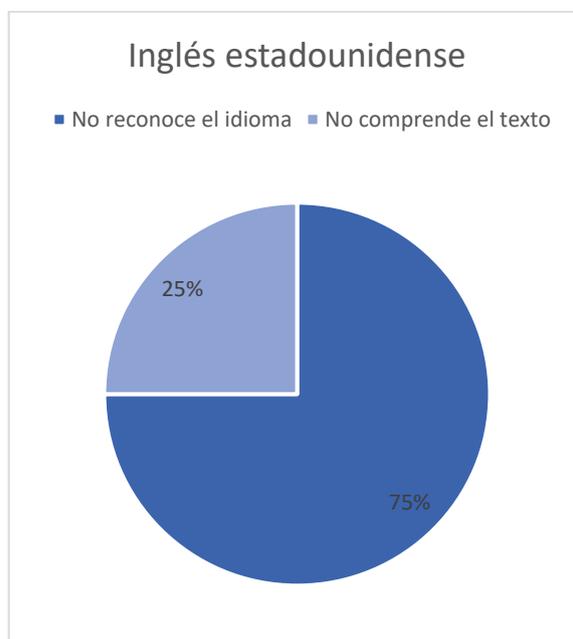


Figura 3. Porcentaje de textos no traducidos en inglés estadounidense. Elaboración propia.

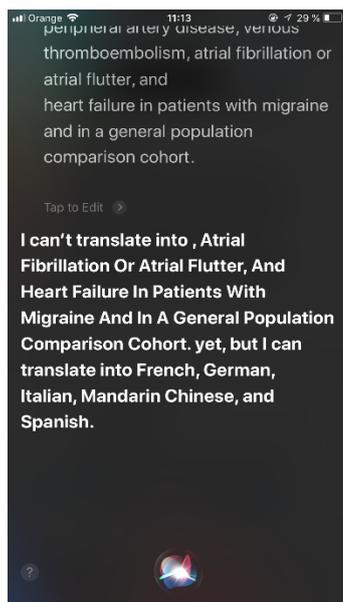


Ilustración 12. Respuesta cuando no reconoce el idioma.

Elaboración propia.

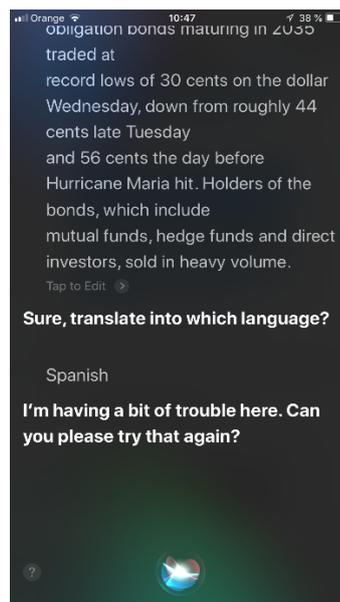


Ilustración 13. Respuesta cuando no procesa el texto.

Elaboración propia.

Este dato es curioso puesto que como mencionábamos anteriormente, para que el traductor funcione, el dispositivo debe estar configurado en su versión *Inglés (Estados Unidos)* y, sin embargo, reconoce un mayor número de textos de Gran Bretaña.

Enfocando ahora nuestro análisis en la traducción de los 21 textos restantes, podemos decir que, en ambos casos, el porcentaje más alto de error se encuentra en la categoría léxico-semántica. Se han encontrado falsos sentidos como el verbo *go off*, el cual se ha traducido como «se apagó» y en este contexto significa lo contrario, «sonar» —texto especializado literario de menos de 500 caracteres en inglés británico— o el sustantivo *lawsuit*, el cual se ha traducido de distintas formas dentro del mismo contexto, «ley» y «demanda» —textos generales técnicos de menos y más de 500 caracteres en inglés británico—. Además, se han registrado casos en los que de una misma palabra se ha mantenido su versión original en inglés, como *Neanderthal* o casos en los que, dentro de una misma frase, se dan tres soluciones distintas a un mismo sintagma, *Peripheral Frame*: «marco periférica», en el cual falla la concordancia de género entre el sustantivo y el adjetivo, «marco periférico», la traducción correcta, y «periphérica», una solución en la que solo se presenta el adjetivo, en femenino, que mezcla la palabra inglesa con su equivalente en español —texto especializado técnico de más de 500 caracteres en inglés estadounidense. En cifras, las inadecuaciones léxico-semánticas son más abundantes en los textos en inglés estadounidense, sin embargo, la diferencia es muy pequeña entre ambos, 49 y 41, respectivamente.

Por otro lado, se ha localizado un error ortotipográfico que se repite con gran frecuencia en todos los textos analizados, poner una coma delante de la «y» en oraciones en las que actúa como

nexo en enumeraciones, un claro ejemplo de error en la lengua meta influenciado por la ortotipografía inglesa:

- En el fragmento de texto especializado técnico en inglés estadounidense con menos de 500 caracteres encontramos la siguiente enumeración: *32 líneas de datos, 32 líneas de información, y tres cambios de transporte.*
- En el fragmento de texto especializado científico en inglés británico con menos de 500 caracteres encontramos la siguiente enumeración: *el infarto, el golpe de argentina, el rombralismo de la artesanía, y el corazón de la justicia.*

En cifras, las inadecuaciones ortotipográficas son más abundantes en los textos en inglés británico, sin embargo, la diferencia es notable entre ambos, 40 y 25, respectivamente.

A nivel gramatical, los errores más comunes que se comparten en ambas variedades son la traslación inadecuada de los números, como, por ejemplo, en el texto general económico de menos de 500 caracteres en inglés británico, en el cual aparece la cifra original de *1,4%* y en el texto meta, la cifra de *4%*. Revisando la traducción, podemos comprobar como este error procede de una mala comprensión del texto, ya que en el texto meta aparece de la nada la siguiente oración: *en la zona del euro es un, 4%*. Podemos pensar que el *1* ha sido traducido pero que, al haber seguido la estrategia de ponerlo en letra (*un*) haya perdido el sentido de porcentaje que tiene en la versión original del fragmento. En cifras, las inadecuaciones gramaticales son más abundantes en los textos en inglés británico; de hecho, se han registrado 46 y 26, respectivamente, lo que supone casi el doble de errores de una variedad a la otra.

Los errores sintácticos, en su gran mayoría, vienen dados por el conjunto de errores encontrados dentro de las tres categorías anteriores. Las inadecuaciones a nivel ortotipográfico, en cuanto a diferencias a la hora de mantener la división original de los párrafos, han provocado que las ideas de diferentes oraciones queden entremezcladas o que las frases queden inacabadas, especialmente al final de cada traducción, ya que no ha producido textos en los que encontremos una frase coherente y bien cohesionada, que acabe en punto final. Además, otro tipo de error sintáctico que se repite son las omisiones de verbos o la introducción de otro que no aparece en el original, lo cual, en ambos casos, cambia totalmente el sentido de la frase. El ejemplo más significativo de este último tipo de error mencionado se encuentra en el texto especializado jurídico con más de 500 caracteres en inglés británico. Como esclarecíamos en la parte teórica de este trabajo, para que Siri sepa que la función que queremos utilizar de todas las posibilidades que ofrece es el traductor, es necesario poner delante del texto que queremos traducir las palabras *translate* o *translation*; en el caso de este texto, siendo la única vez en la que este traductor automático ha cometido este tipo de error de entre todos los fragmentos analizados, ha introducido el verbo *translate* como parte de la oración original, haciendo que la ley cambie totalmente el tema

principal que trata: *traducir una valoración inicial es una cita con una persona cualificada*. Como vemos, el contexto inicial de esta ley, *Drugs Act 2005*, pasa de tratar de una evaluación o valoración (v.o.: *an initial assessment*) a la traducción de la misma. En cifras, las inadecuaciones sintácticas son mucho más abundantes en los textos en inglés estadounidense y, al igual que con la categoría anterior, supone casi el doble de errores de una variedad a la otra, siendo 32 las inadecuaciones registradas en los textos estadounidenses frente a 18 en los británicos.

Las figuras retóricas que aparecían en los textos originales, como el refrán *be as disciplined as a monk*, que se encuentra en el texto especializado literario de menos de 500 caracteres en inglés británico, tampoco se han traducido por un equivalente o se ha hecho una traducción literal como *ser tan disciplinado como un monje*, una opción totalmente válida que, aunque le quite ese tono especial que le quería dar la autora a su obra. Por el contrario, pese a no ser una figura retórica como tal, dentro de este apartado también se ha admitido un tipo de error que se repite a lo largo de todos los textos meta: la repetición sin causa determinada aparente de palabras, sintagmas u otro tipo de elementos que forman la oración. Esta inadecuación se comprende mejor con ejemplos:

- (b) *A lo mejor, a Nueva York, a los de Sotheran, a los perfumes, los perfumes, los perfumes, a Inglaterra, a Inglaterra, a Inglaterra, donde le compraron el puesto de economía de Sotheby's* (texto general literario de menos de 500 caracteres en inglés británico)
- (c) *La ley sigue en el sentido de los elementos, los medios de comunicación, y una petición con más de 21.000 firmas, y una petición con más de 21.000 firmas, que se ha hecho el nuevo modelo, y luego, en el nuevo* (texto general técnico de más de 500 caracteres en inglés británico)

Este error se repite con mayor frecuencia en los textos en inglés británico, sin influir el número de textos traducidos en cada una de las variedades, ya que esta inadecuación únicamente se ha encontrado en dos ocasiones en los textos en inglés estadounidense frente a las 12 que se han registrado en los textos extraídos de publicaciones de Gran Bretaña.

Por último, dado que errores de carácter pragmático no se han encontrado, haremos referencia a los problemas culturales que presentaban los textos originales, los cuales se encontraban principalmente dentro de la selección de textos estadounidenses. Los más destacados eran los relativos a las Librerías Sotheran's —texto general literario de menos de 500 caracteres en inglés británico—, problema que ha solventado sin errores, quizás porque, al no existir equivalente en español, se mantiene la palabra original, y al político Timmermans —texto general jurídico de más de 500 caracteres en inglés jurídico. Esta referencia cultural llama especialmente la atención ya que ha sido eliminada del texto meta sin causa aparente, lo que cambia totalmente el sentido de la frase, dado que actuaba como sujeto explícito en la oración. Otro claro ejemplo de error cultural, curioso cuanto menos puesto que se trata de un texto en parte

escogido por la relación que establece entre la empresa Apple y el objeto de estudio de este trabajo, es la mala traducción de los productos de la misma. En el fragmento de texto general técnico de menos de 500 caracteres en inglés británico encontrábamos *MacBook* y *MacBook Pro*, productos cuyo nombre se mantiene en español, y, sin embargo, en la traducción de Siri aparece una nueva versión de estos ordenadores llamada «Macbook Pro Delta». En cifras, entre los 13 textos en inglés británico tan solo se ha registrado un único error cultural, mientras que en los ocho textos estadounidenses se han encontrado cinco errores. Estos datos se pueden ver reflejados en las figuras 4 y 5 de este TFM.

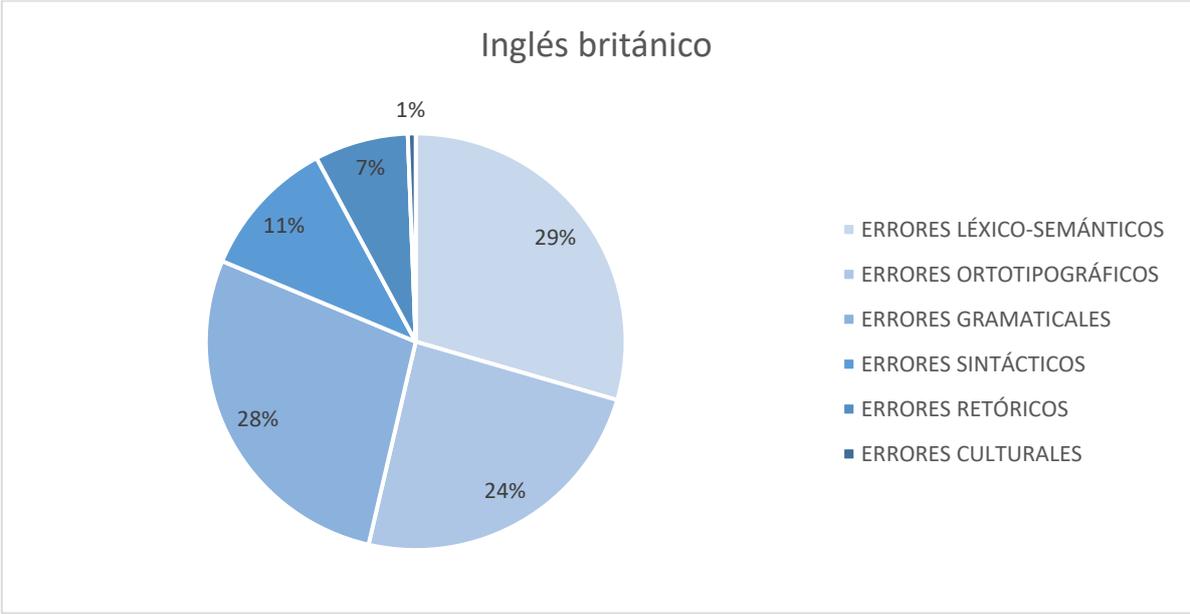


Figura 4. Porcentaje de error en los textos en inglés británico. Elaboración propia.

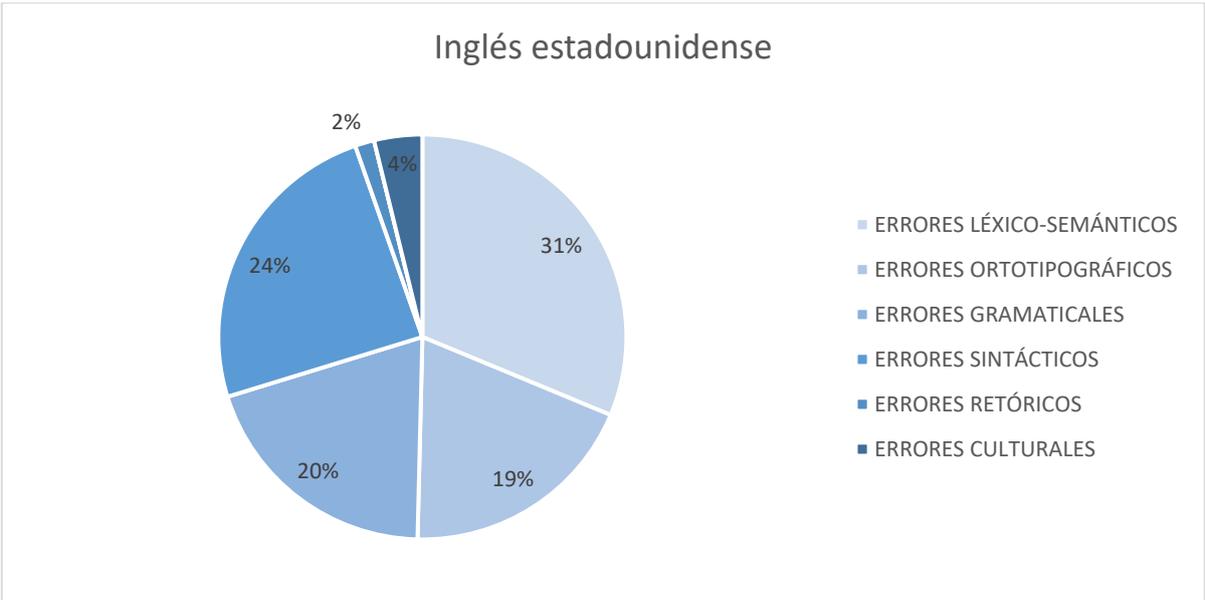


Figura 5. Porcentaje de error en los textos en inglés estadounidense. Elaboración propia.

6.2 Criterio 2: número de caracteres

Al igual que en el apartado anterior, comenzaremos nuestro análisis con aquellos textos de los cuales no hemos obtenido una traducción. En este caso, tanto en inglés británico como en inglés estadounidense, los textos que contienen más inadecuaciones son los que superan los 500 caracteres, aunque también ha habido casos en los que los textos con una media de 300 caracteres aproximadamente. En las siguientes figuras, podemos encontrar los porcentajes de textos no traducidos en función del número de caracteres.

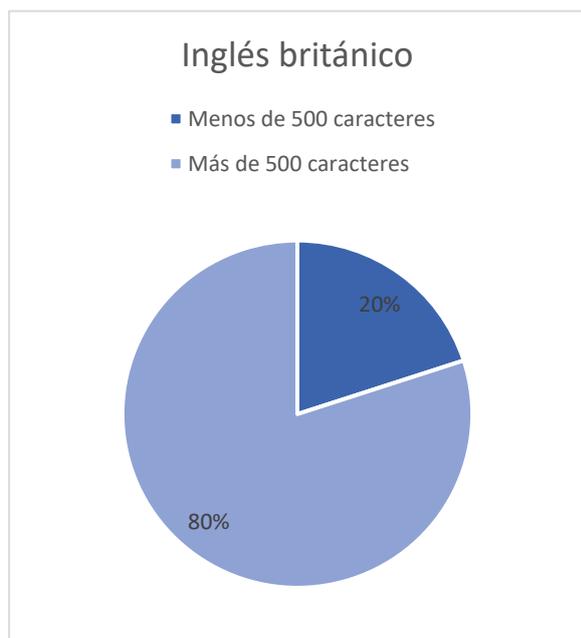


Figura 6. Porcentaje de textos no traducidos por número de caracteres en inglés británico. Elaboración propia.

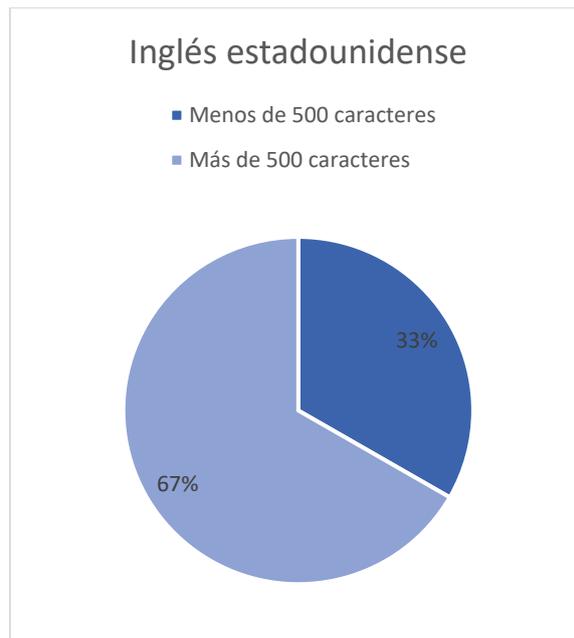


Figura 7. Porcentaje de textos no traducidos por número de caracteres en inglés estadounidense. Elaboración propia.

Uno de los datos más curiosos que hemos contrastado dentro de esta categoría es que el traductor no tiene un número limitado de caracteres que traduce sin problema, como por ejemplo su competidor Google Translator, limitado a 5 000. De hecho, no se ha conseguido establecer un patrón de actuación, puesto que los textos que ha traducido y los que no, como se puede ver en el Anexo I, varían entre los 100 y los 1 200 caracteres.

Por otro lado, si dirigimos el análisis hacia los textos cuya traducción se ha obtenido, hemos visto cómo, independientemente de su longitud, los textos meta no superaban los 250 caracteres. Los textos más largos producidos por Siri tienen una longitud de 238 caracteres y, curiosamente, uno de ellos pertenece al grupo seleccionado de publicaciones de Gran Bretaña —el texto especializado jurídico de menos de 500 caracteres, en concreto, 419—, mientras que el otro es estadounidense —el texto especializado técnico de menos de 500 caracteres, en concreto, 374.

En ambos casos, supera la mitad de los caracteres totales del texto. Sin embargo, esta tendencia no ha sido la más común en comparación con el resto de los textos meta obtenidos. La siguiente tabla, recoge los datos que comentaremos más adelante en relación a este fenómeno:

	Caracteres del texto original (T.O.)	Caracteres del texto meta (T.M.)
Texto general económico de menos de 500 caracteres británico	499	190
Texto general literario de menos de 500 caracteres británico	457	184
Texto especializado literario de menos de 500 caracteres británico	373	174
Texto general científico de menos de 500 caracteres británico	207	121
Texto especializado científico de menos de 500 caracteres británico	273	170
Texto especializado científico de más de 500 caracteres británico	728	158
Texto general técnico de menos de 500 caracteres británico	165	174
Texto general técnico de más de 500 caracteres británico	764	208
Texto especializado técnico de menos de 500 caracteres británico	397	208
Texto general jurídico de menos de 500 caracteres británico	489	230
Texto general jurídico de más de 500 caracteres británico	743	229

Texto especializado jurídico de menos de 500 caracteres británico	419	238
Texto especializado jurídico de más de 500 caracteres británico	731	228
Texto general económico de menos de 500 caracteres estadounidense	380	189
Texto general literario de menos de 500 caracteres estadounidense	498	223
Texto general literario de más de 500 caracteres estadounidense	679	220
Texto especializado literario de menos de 500 caracteres estadounidense	376	176
Texto general científico de menos de 500 caracteres estadounidense	325	179
Texto general técnico de menos de 500 caracteres americano	418	207
Texto especializado técnico de menos de 500 caracteres americano	374	238
Texto especializado técnico de más de 500 caracteres americano	796	99

Tabla 3. Diferencia del número de caracteres entre el T.O. y el T.M en Siri. Elaboración propia.

Inicialmente, como comentábamos anteriormente, esta selección de textos se hizo tras comprobar que el traductor no indicaba el número límite de caracteres que tenía la capacidad de procesar. Se puede reflexionar en base a esto para decir que no se ha encontrado un patrón de división de textos traducibles y no traducibles a nivel de caracteres ya que, los textos que no pudo procesar tenían una longitud de 559, 603, 649, 734 y 1 111, todos ellos pertenecientes al grupo de más de 500 caracteres, pero con la problemática de que textos que se encuentran dentro del mismo rango de caracteres y han sido traducidos.

De los 21 fragmentos de textos seleccionados que se han traducido, se ha hecho una estimación aproximada basada en el número de caracteres obtenidos en los textos meta. Tan solo uno de ellos tiene una longitud mayor que su original: es el caso del texto técnico general de menos de 500 caracteres en inglés británico, el cual cuenta con 165 caracteres en el T.O. y 174 en el T.M. Esta tendencia debería haber sido la general puesto que los T.M. hacia el español desde inglés suelen tener una longitud mayor. No obstante, en siete de los restantes, se ha comprobado que la traducción cuenta con un número superior a la mitad de los T.O. Se trata de los casos de los textos en inglés británico científico general y especializado y jurídico especializado de menos de 500 caracteres y en inglés estadounidense, científico general y técnico general y especializado de menos de 500 caracteres.

Por otro lado, se ha comprobado que la media de error es más alta en los textos de menos de 500 caracteres, tanto para los de inglés británico como los de inglés estadounidense, mientras que, a pesar de que una gran parte no se han traducido, la media para los que superan los 500 caracteres es más baja con respecto a sus opuestos.

	Inglés británico		Inglés estadounidense	
	Menos de 500 caracteres	Más de 500 caracteres	Menos de 500 caracteres	Más de 500 caracteres
Texto general económico	15		18	
Texto especializado económico				
Texto general literario	N.C.		16	14
Texto especializado literario	13		16	
Texto general científico	5		9	
Texto especializado científico	12	10		
Texto general técnico	9	13	17	
Texto especializado técnico	24		25	16
Texto general jurídico	16	17		

Texto especializado jurídico	16	12		
TOTAL DE ERRORES	110	52	101	30
Media de error por texto	13,75	8,6	16,8	15

Tabla 4. Medias de error (I). Elaboración propia.

Como se muestra en la tabla, la diferencia de medias dentro del grupo de textos de Gran Bretaña es mayor —los textos con menos de 500 caracteres tienen una media de error aproximada de 14 errores, mientras que los que los superan, tienen una media de error aproxima de nueve errores; supone una diferencia cinco inadecuaciones por texto—, mientras que entre los de Estados Unidos tan solo encontramos una diferencia de dos errores de media.

6.3 Criterio 3: general o especializado

Por último, comenzaremos este apartado hablando, una vez más, por los textos no procesados. Los datos obtenidos en este apartado son, quizás, los más equilibrados de entre las tres categorías, puesto que hay un número muy similar de textos generales y especializados que no ha reconocido el traductor. No obstante, son los de carácter especializado los que encontramos en mayor número, tal y como muestran estas figuras:

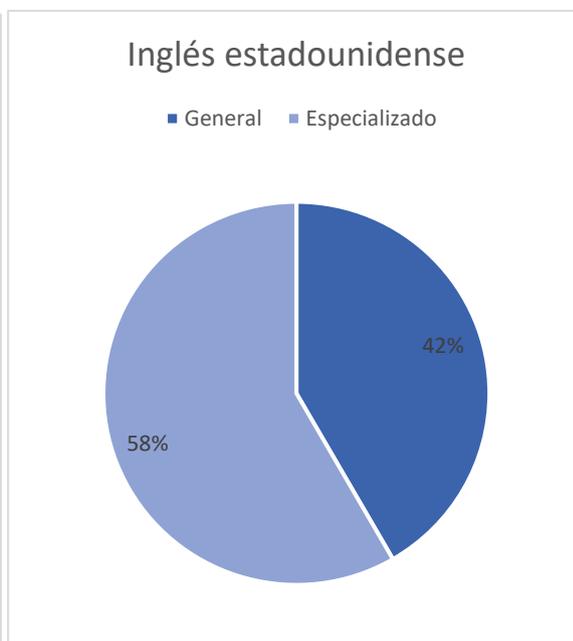
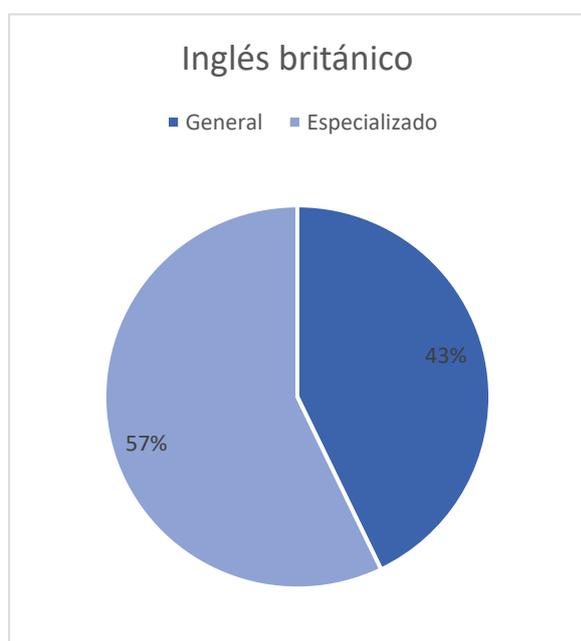


Figura 8. Porcentaje de textos no traducidos generales y especializados en inglés británico. Elaboración propia. Figura 9. Porcentaje de textos no traducidos generales y especializados en inglés estadounidense. Elaboración propia.

Para poder establecer la siguiente comparativa de errores entre los textos generales y específicos, se ha optado por aunar los dos fragmentos que conforman cada grupo, es decir, dejar de lado la división de textos por número de caracteres. De esta forma, podemos decir que, en los textos generales, la tasa de error es menor con respecto a la de los especializados, como se muestra en la Figura 10. Los datos demuestran que existe una diferencia aproximada de cinco inadecuaciones con respecto al uno del otro – 149 errores para 11 textos generales y 144 errores para ocho textos especializados; en particular, los extraídos de publicaciones británicas contienen menos inadecuaciones que los extraídos de publicaciones americanas –los grupos de traducción económica, científica y técnica de los cinco iniciales dentro de la categoría de generales contienen menos errores. Podemos ver estos datos reflejados en la siguiente tabla:

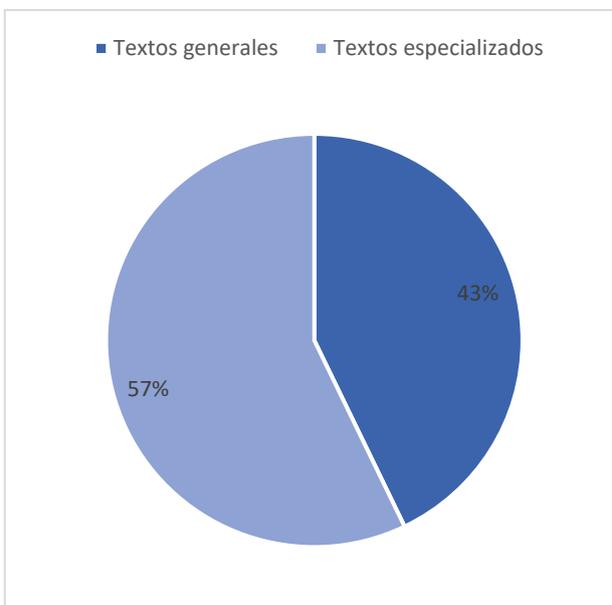


Figura 10. Porcentaje de error para generales y especializados. Elaboración propia.

	General (nº de inadecuaciones medio por texto)		Especializado (nº de inadecuaciones medio por texto)	
	Inglés británico	Inglés estadounidense	Inglés británico	Inglés estadounidense
Textos económicos	15	18		
Textos literarios	N.C. ¹	15	13	16
Textos científicos	5	9	22	
Textos técnicos	11	17	24	21
Textos jurídicos	17		14	

Tabla 5. Medias de error (II). Elaboración propia.

¹ N.C. o «No Contabilizado»: debido a que este fragmento de texto era inadecuado en su totalidad, sus errores no se han incluido en la clasificación de errores.

El porcentaje más alto de error lo sustentan los errores léxico-semánticos en general y, como era de esperar, el vocabulario especializado ha sido el que peor se ha traducido, por ejemplo, en el texto especializado científico en inglés británico de menos de 500 caracteres, el cual recordamos que había sido elegido por su alto contenido en vocabulario especializado, encontramos palabras como *myocardial infarction*, *stroke*, *peripheral artery disease*, *venous thromboembolism*, *atrial fibrillation*, *atrial flutter*, y *heart failure*. No se sabe de dónde Siri ha extraído estas traducciones: «risos de miocardio», «infarto», «golpe de argentina», «rombralismo de la artesanía», «troteador de artesanía» y «corazón de la justicia». Son traducciones que no solo no son equivalentes al original ni son términos científicos relacionados con el tema de la medicina, sino que, además, se ha investigado la existencia de estas expresiones y no hay resultados salvo para miocardio e infarto.

En la misma línea, los errores en materia de gramática se encuentran mayormente en los fragmentos especializados y esto se debe especialmente a que o bien los verbos no están expresados en el tiempo que corresponde, o los sintagmas nominales formados por sustantivo + adjetivo no concuerdan o por la omisión de palabras importantes dentro de la oración que cambian totalmente el sentido original:

- Ejemplo de tiempo verbal incorrecto encontrado en el texto especializado jurídico con menos de 500 caracteres en inglés británico: *has been arrested* se ha traducido por «se ha detenido», una pasiva refleja, cuando en realidad se debería mantener una pasiva normal, «ha sido detenida».
- Ejemplo de mala concordancia en el sintagma nominal encontrada en el texto especializado científico con más de 500 caracteres en inglés británico: en el T.O. se presentaba la secuencia *neurological, gastrointestinal, and autonomic symptoms*, que se ha traducido como «legislación legislativa de neurológicas, gastrointestinal y autónomo autónomo». Como vemos, ya desde un primer momento la secuencia del T.M. es incorrecta dado que síntomas no ha traducido correctamente, por lo que al final, todos los adjetivos que depende de ese núcleo quedan supeditados a otras palabras. Por consiguiente, la concordancia que se debe establecer entre el núcleo del sintagma y sus complementos es nula. Lo único que no se podría considerar como error es que ha intentado que el sustantivo se coloque delante de los adjetivos, orden preestablecido en las oraciones en español, pese a que la traducción no sea la adecuada.
- Ejemplo de omisión encontrada en el texto especializado científico con más de 500 caracteres en inglés británico: *this asynchronous interface consists of 20 address lines* es la oración original y «en las últimas zonas de los agentes» es la traducción. Como se observa, hay graves omisiones, entre ellas, la del sujeto y la del verbo.

Por otro lado, si fijamos la atención en los resultados obtenidos en la categoría de errores sintácticos, comprobamos que la diferencia entre la media es de 12 puntos entre los textos generales y los especializados, la más alta de todas ellas. En este caso, son los textos especializados los que contienen menos errores de este tipo, siendo un total de 19 errores encontrados en las traducciones, frente a los 31 de los textos generales. En las categorías de errores retóricos y culturales, los textos especializados también contienen un menor número de traducciones incorrectas, ya que, salvo en textos literarios como novelas o poemarios, las referencias a la cultura y las figuras retóricas no aparecen en leyes, manuales de instrucciones o artículos científicos, entre otros. Los porcentajes que se ven en las siguientes figuras, se corresponden con las cifras extraídas del Anexo III en relación a esta categoría.

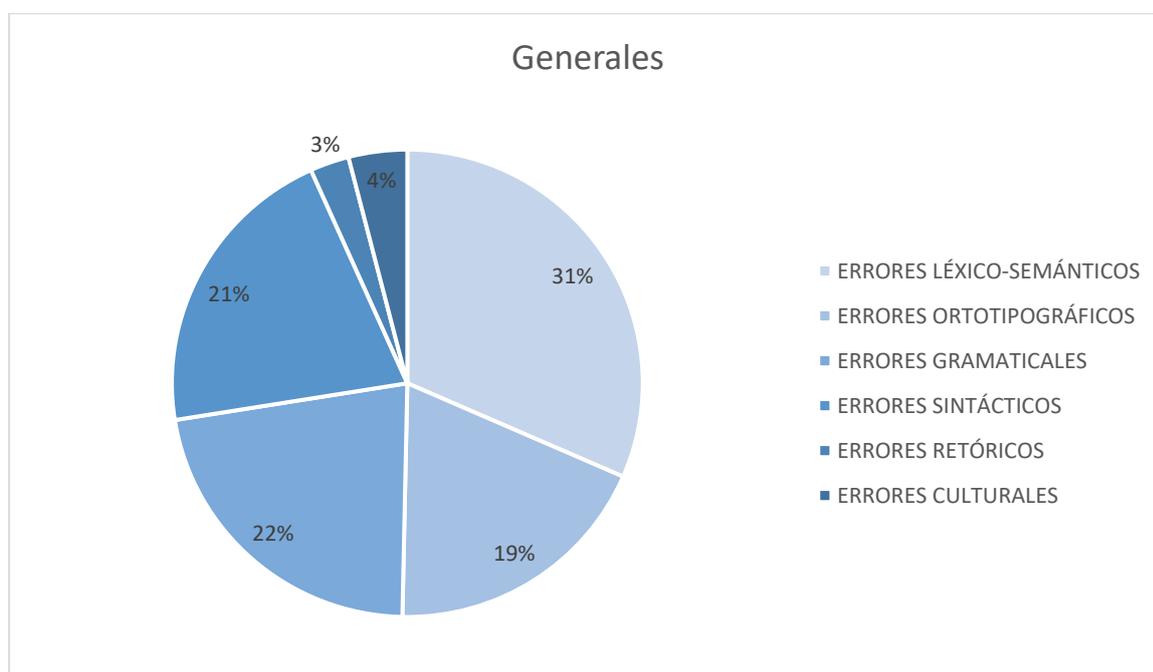


Figura 11. Porcentajes de error en textos generales. Elaboración propia.

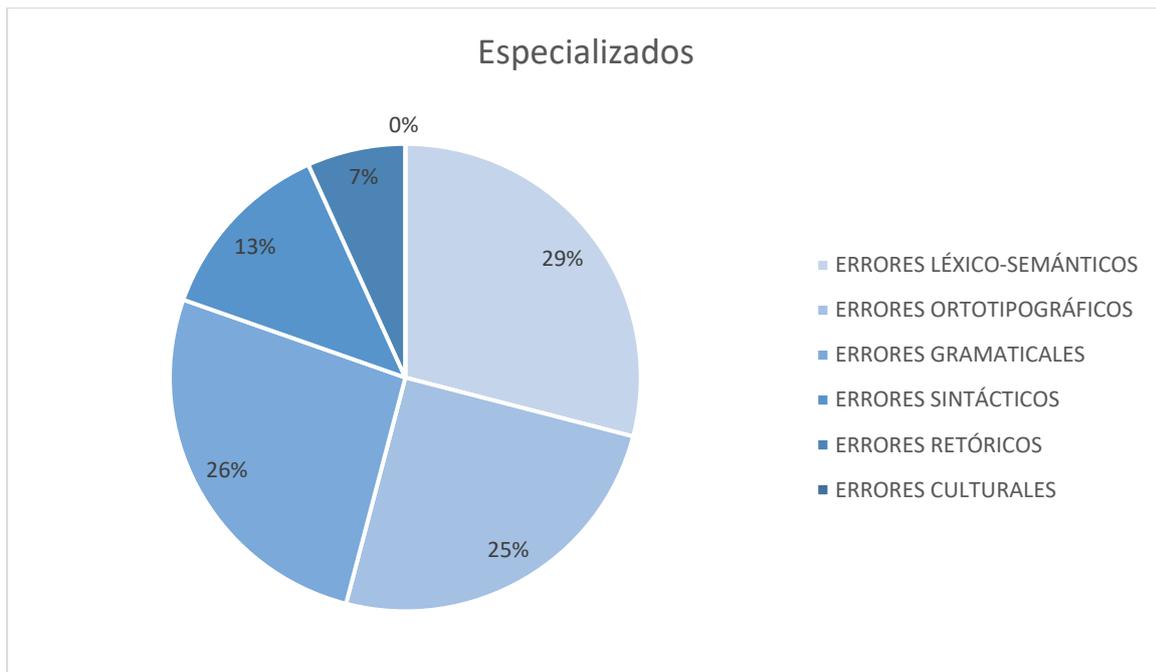


Figura 12. Porcentajes de error en textos especializados. Elaboración propia.

6.4 Criterio 4: textos de producción propia

Como se puede ver en el Anexo I, los textos escogidos para esta parte del estudio distan mucho de los que se habían escogido para los apartados anteriores, puesto que la finalidad era comprobar si Siri trabaja a través de un sistema de traducción basado en redes neuronales. De los seis textos utilizados, tres no se han traducido ya que no reconocía el idioma, lo que supone un 50 % del corpus de textos con estas características.

Para mayor sorpresa, es capaz de hacer una traducción bastante aproximada que un usuario con conocimientos en español pueda entender a la persona angloparlante cuando quieran comunicarse, ya que, desviando la atención a los textos que contenían errores en el T.O. Principalmente, las inadecuaciones que se han encontrado son de carácter léxico-semántico, al igual que en el resto de los textos, no obstante, esta cifra ha aumentado de manera considerativa dado que en el texto cinco, la enumeración de deportes no se ha traducido correctamente. Por lo tanto, en sí, los textos no tienen tantos errores de incomprensión semántica si no por no ser capaz de presentar resultados acertados en materia de deportes.

Por otro lado, una traducción cuanto menos acertada pero que choca a la vista del lector español es la que se ha presentado para *little sister* en el texto número seis. Es correcto decir «hermanita», puesto que es el diminutivo de hermana y como bien se especifica en el T.O., la hermana es menor a la persona que está redactando el párrafo. Sin embargo, a nivel contextual

suenan extrañas esta traducción puesto que justo a continuación especifica que la edad de la hermana es 20 años. «Hermanita», con el sufijo *-ita*, tiene en español esa connotación de no una persona pequeña, sino de un niño pequeño, un bebé, y, además, es más bien típico del lenguaje de los niños que de un adulto que quiera utilizar este traductor.

Con respecto al último de los textos meta de este apartado, es de destacar la diferencia entre la traducción de un topónimo español y uno estadounidense. Como bien mencionábamos anteriormente en el apartado de metodología, estas dos ubicaciones no se habían escogido de forma aleatoria, sino que se quería comprobar si reaccionaba de la misma forma cuando se encontraba con ciudades de diferentes países, aunque en este caso Miami se trate de una playa española, lleva por nombre el de una gran ciudad californiana. Efectivamente, tiene dificultades en reconocer ciudades que no sean conocidas a nivel internacional como ha pasado con Tarragona, ya que la ha traducido como «Tarrachan».

Se ha visto cómo los errores ortotipográficos que se encontraban en los textos cinco y seis no han sido un impedimento a la hora de realizar la traducción, es decir, los casos en los que se hacía omisión de la coma que une el sujeto y el verbo contraído o, en caso contrario, se ponía sin necesidad de ella —*I've born* en el texto cinco y *I'am Iris* en el texto seis, entre otros—, se han traducido correctamente, obteniendo los resultados de «nací» y «soy Iris» para los ejemplos mencionados. No obstante, no ha aplicado las redes neuronales con las que suponemos que trabaja para evitar errores como intercambio de las formas del verbo no personales —en el texto cinco, el autor escribió *I've born and grow up in Zaragoza* y cometió una inadecuación gramatical, mantener la forma en infinitivo del verbo *to grow* cuando en realidad debería estar en participio pasado como *born*, siendo *grown* la forma correcta en este caso. La traducción de Siri es: «nací y crecer en Zaragoza»; en ella, no queda reflejado que el verbo debería estar en pasado dado que la conjunción «y» está estableciendo una correlación que pone a los dos verbos al mismo nivel temporal, al igual que en inglés, motivo por el cual se omite la repetición del sujeto *I* y el verbo auxiliar *have*.

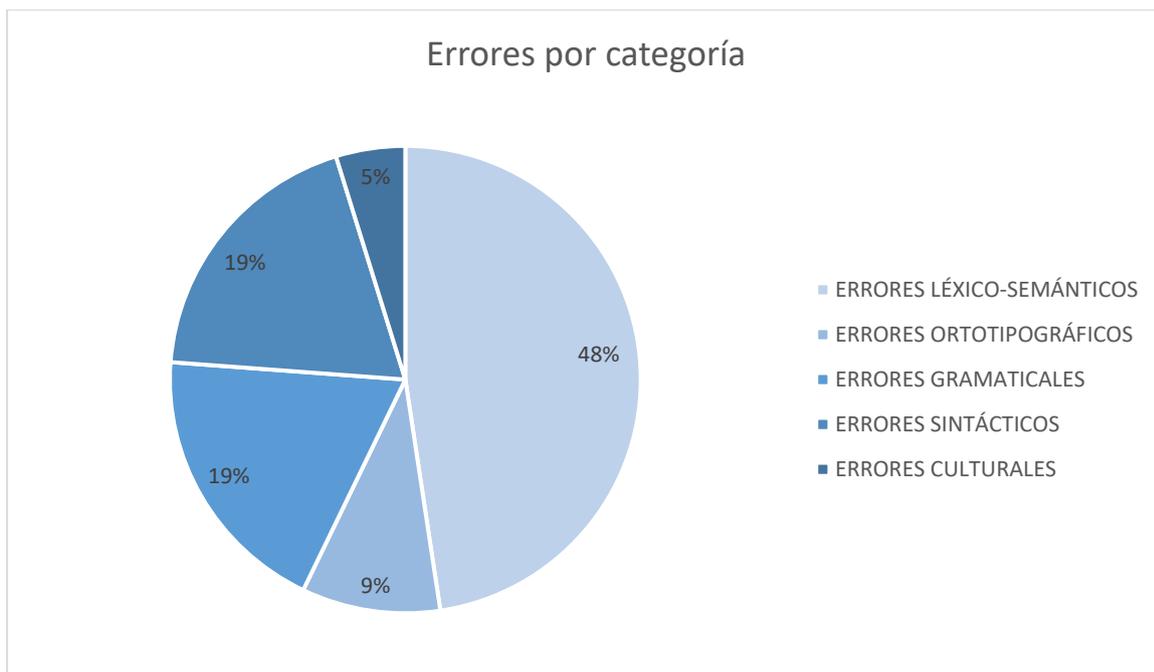


Figura 13. Porcentajes de error por categoría de error. Elaboración propia.

En este grupo del corpus, los caracteres no han sido un factor determinante a la hora de haber conseguido una mejor o peor traducción, con un mayor o menor número de errores, ya que los seis textos se enmarcaban entre los 220 y los 340. Dos de los textos que no se han traducido se corresponden con los de mayor número de caracteres —336 para el texto dos y 320 para el texto uno— y el tercero tenía una longitud de 258, habiendo dos textos que se han traducido con más caracteres.

Por último, cabe mencionar que estos seis textos no tenían ningún grado de especialización, por lo que, no se han cometido tantos errores a nivel gramatical o sintáctico. Únicamente se podrían considerar como error cultural el hecho de que solo reconoce algunas ciudades, puesto que, como se ha mencionado antes, Zaragoza es una ciudad que se ha reconocido a la perfección, mientras que de Tarragona se ha obtenido algo completamente alejado del original.

6.5 Conclusiones extraídas del análisis

El traductor del asistente virtual se puede considerar como una nueva gran ventaja que ofrece Apple al mundo, sin embargo, después de haber realizado el análisis del apartado anterior, he llegado a la conclusión de que aún le falta mucho para ser competitivo dentro del mercado de la traducción y salir de su estereotipo de asistente virtual.

Inicialmente, creía que este sistema trabajaría con un sistema de redes neuronales, dado que, como asistente virtual, la inteligencia artificial que le ha dado la vida le permite pensar y reaccionar por sí mismo. Sin embargo, en la práctica, no he podido comprobar esta hipótesis ya

que en algunas de las traducciones obtenidas se apreciaba como la base era una traducción literal mal interpretada, haciendo que pierda la coherencia, la cohesión y el sentido con respecto del original. He concluido que puede ser un sistema que no solo opere con un sistema de traducción basado en redes neuronales, sino que se trate de una fusión con los sistemas basados en corpus, ya que, tiene la capacidad de buscar en Internet cualquier cosa que desconozca y, hoy en día, existen grandes corpus paralelos que se pueden encontrar en la web, como los que proporciona la Unión Europea.

En relación a aquellos textos que no ha traducido, creo que el error se debe a que, durante el procesamiento, pone en mayúscula cada palabra. Esto podría producir en el traductor automático una confusión a nivel sintáctico y gramatical y por ello, no reconoce algunos topónimos u otras palabras que exigen mayúscula inicial. Sin embargo, no he podido llegar a una conclusión de por qué confunde partes de oraciones con idiomas ya que como asistente se puede configurar en prácticamente todas las lenguas de trabajo.

Siri no es capaz de mantener el tema principal de los textos, siguiendo una linealidad. Hace mezclas de las diferentes partes de los fragmentos, introduciendo información de unas y otras oraciones sin tener en cuenta los signos de puntuación que las dividen. Tampoco sabe determinar, en el caso de las enumeraciones de familias léxicas, como encontrábamos en el campo de la medicina (vocabulario especializado) y en el campo de los deportes (vocabulario general), de que tema está hablando, es decir, aún no piensa por sí mismo de que habla el texto original, procesarlo correctamente. Realiza omisiones de palabras que, quizás, no sabe traducir o que no encuentra en sus bases de datos con una frecuencia mayor de lo esperado, especialmente, porque en muchos de los casos se trata de palabras bastante comunes como el verbo *to work*, entre otros.

He registrado numerosos errores que por no reconocer el contexto en el que se desarrolla el fragmento de texto seleccionado. Un ejemplo de ello es la traducción literal de ciertos sintagmas que tienen una colocación específica: *the euro zone* se ha traducido como «zona del euro» cuando en realidad se trata de un término específico que se suele conocer como «zona euro» o «Eurozona».

Continuando en esta misma línea, no he podido extraer un patrón de traducción que explique por qué unos textos de la misma rama de conocimiento se traducían y otros no. En la selección de textos, se determinó que se escogerían del mismo texto, dos fragmentos, uno para el grupo con una longitud inferior a 500 caracteres y otro para la superior. Por motivos que se desconocen, se han dado casos en los que el traductor automático presentaba un T.M. de uno de los fragmentos, que solía coincidir con el que tenía menos de 500 caracteres.

Tampoco he encontrado un patrón que explique por qué algunas oraciones, sintagmas o palabras se repiten. Una de las teorías que he sacado es que se producen las repeticiones cuando desconoce lo que dice el original.

El único dato concluyente que hemos podido sacar de esta reflexión, es que el único tipo de textos que no ha podido traducir, en ninguno de las dos variedades e independientemente de su número de caracteres, es el texto especializado en materia de economía. Ni siquiera siendo un texto redactado por la propia empresa que lo desarrolla, ya que, para este análisis, se habían escogido dos memorias anuales de dos grandes multinacionales entre las que se encontraba la de Apple.

La mala traslación de las cifras, que son un elemento muy frecuente que se encuentra en los textos económicos, ha sido uno de errores que he encontrado y que considero de mayor gravedad. En numerosas de las traducciones, he visto como números sencillos, que no necesitan una traducción en sí, se expresaban de manera incorrecta en el T.M. A veces, también introducía elementos culturales en el texto meta, como, por ejemplo, el más curioso: «una demanda de acción de gracias», cuando en T.O. dice *class action lawsuit*. Acción de gracias es una festividad navideña que celebran en América, que no tiene nada que ver con el tema del que hablaba el fragmento técnico general.

También he comprobado que, pese a que comete inadecuaciones a nivel ortotipográfico, no se ha visto influenciado por el inglés en todos los aspectos que se contemplaban en la clasificación de errores. Destaca que no tiene problemas a la hora de colocar los signos de apertura de las preguntas y exclamaciones que únicamente se ponen en español. Asimismo, no se han registrado errores pragmáticos; de hecho, en el análisis no se ha encontrado ninguno en los textos, algo que sorprende, dada la tendencia de los extranjeros a usar las fórmulas de cortesía con el tratamiento de «usted», dotando a la traducción de un registro más formal que en español no se utiliza.

En relación a las limitaciones de la TA, a las que hacíamos referencia en el apartado teórico de este trabajo, puedo decir que no se ha detectado un uso incorrecto de sinónimos con connotaciones específicas, pero sí que se han detectado problemas con la polisemia de las palabras. Por ejemplo, en la preposición *for*, la cual en español puede significar «para», «por», «porque», «durante», etc. Este «conocimiento del mundo» es una de las grandes barreras que siempre separarán la traducción de las personas de la de las máquinas y, como hemos comprobado, Siri no es la excepción que pueda superar esa limitación.

En cuanto a la ambigüedad de la que nos hablaba Alonso Martín (2003: 98-100), puedo decir que en sí no se ha encontrado ningún error de este tipo que se haya contemplado en el análisis del apartado anterior, pero, concretamente en el texto literario especializado de menos de 500 caracteres y en inglés británico, he marcado que los pronombres deberían estar en femenino puesto que el personaje de la novela es una protagonista femenina que habla en primera persona. No se ha registrado como error de ambigüedad porque no se ha producido dentro de un contexto

en el que se pudiera conocer el sexo del autor, pero sí que es verdad que los pronombres posesivos en inglés y español, sin un contexto previo, no permiten identificar el género de la persona que está hablando.

Los T.M. que propone necesitan, por el momento, un enorme trabajo de posesición en el caso de que queramos utilizar este traductor, incluso cuando lo que se quiere traducir es un texto muy básico de presentación o párrafos con muy pocos caracteres. No obstante, después de la versión del sistema operativo con la que se ha trabajado (iOS 11.3.1, con fecha del 24 de abril de 2018) se introdujo una nueva actualización, iOS 11.4, el día 29 de mayo de 2018. Por curiosidad como traductora, introduje algunos de los fragmentos del corpus para comprobar si, poco a poco, van corrigiendo los errores que se han detectado en este trabajo. Y así es, alguno de los errores se había corregido, pero, aun así, la calidad de las traducciones sigue siendo extremadamente baja. Únicamente, quiero recalcar, que cuando te dispones a introducir un texto, hace sugerencias de palabras sueltas que, debido a la curiosidad traductora una vez más, se ha interactuado con él y se ha comprobado que las traducciones son bastante acertadas sin un contexto. Por eso, podemos decir que, quizás, lo que más desarrollado está actualmente no es tanto su función de traducción, si no de diccionario multilingüe.

Sin embargo, es de agradecer que la traducción se haga hacia el español de España, ya que el español latino ejerce una gran influencia en el país norteamericano y existen numerosas diferencias entre ambos.

Por último, hace un par de años, Alphatrad (2016) escribía en un artículo sobre traducción automática en su página web sobre las ventajas y desventajas que presentaba traducir con traductores automáticos, diciendo que las principales desventajas eran la falta de fluidez, la incapacidad de leer entre líneas o de diferencias entre palabras y composiciones gramaticales que pueden tener significados diferentes, producir frases extremadamente largas y carentes de sentido y que requieren de una gran posesición. Estas desventajas son algunos de los errores más graves que comete el traductor de Siri y para los cuales se proponen soluciones en el siguiente apartado.

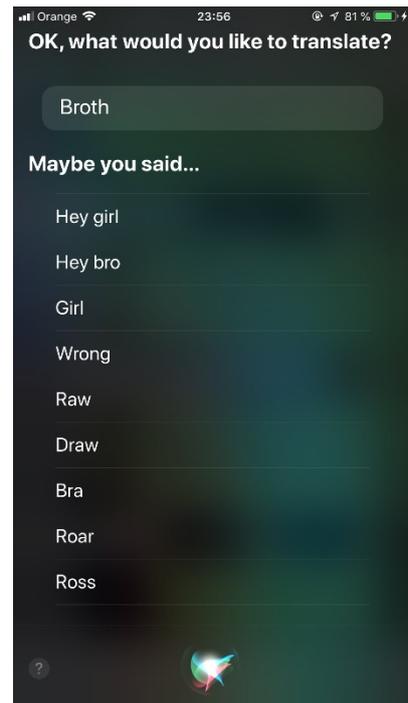


Ilustración 14. Sugerencias del traductor. Elaboración propia.

CAPÍTULO 7. RESULTADOS DEL ANÁLISIS DE LA PRODUCCIÓN DE GOOGLE TRANSLATOR

Como remarcábamos en el Capítulo 3, Google Translator es un motor de traducción que está mucho más entrenado que el asistente virtual de Apple, dados sus 10 años de andadura. Esto implica que no hemos registrado textos sin traducción, por lo tanto, los resultados se han extraído en base a los 40 textos del corpus.

7.1 Criterio 1: inglés británico e inglés estadounidense

Las inadecuaciones más llamativas son las ortotipográficas, especialmente en los textos de inglés británico, donde hemos registrado 61 errores en total. Para el corpus de inglés estadounidense se han registrado el mismo número de errores léxico-semánticos y ortotipográficos, un total de 43. Las inadecuaciones léxico-semánticas en los textos del corpus británico destacan menos, habiendo registrado un total de 35, un número inferior al número de textos que lo conforman. Algunos ejemplos que acompañan a estos datos son:

- Inadecuaciones ortotipográficas: aunque en este caso las más llamativas son las relacionadas con las cifras, también se han encontrado casos en los que se coloca una coma delante de la «y» en enumeraciones o, especialmente en los textos jurídicos, la falta de algunas mayúsculas como en «Estados miembros» o «la Comisión».
- Inadecuaciones léxico-semánticas: la mayoría de ellas se ven causadas por traducciones literales de algunos verbos, sobre todo, es decir, que el verbo existe en español, pero, por el contexto, no termina de adecuarse dentro de la oración. Un claro ejemplo de este fenómeno se encuentra en el texto general literario especializado de menos de 500 caracteres en inglés británico: la frase original decía *every day I will 'rise' early, 'ablute' in cold water and be as disciplined as a monk*. El verbo *ablute* es un verbo que literalmente significa « ducharse » y, sin embargo, suponiendo que, por desconocimiento de este verbo, Google Translator nos ha presentado la opción de «ablutaré en agua fría», verbo que no tiene ningún significado en español.

Entre los textos en inglés británico y estadounidense, se han registrado una enorme diferencia a nivel gramatical, 17 y 30 errores, respectivamente, casi el doble de inadecuaciones el uno con respecto al otro. Esto principalmente se debe a que el texto especializado literario en inglés estadounidense tenía prácticamente toda la traducción de los verbos mal, en el sentido de que los conjugó en un tiempo distinto en el que estaba redactado el original: el *past perfect* (por ejemplo: *they had called her*) se ha traducido por «la llamaron», que, más allá de que el verbo más

descriptivo sea «apodarla», ha perdido ese sentido de pasado del pasado que le otorga el pretérito pluscuamperfecto en español: «la habían llamado/apodado».

Por el contrario, son los textos extraídos de publicaciones de Gran Bretaña las que destacan por sus inadecuaciones sintácticas por encima de las de Estados Unidos. Esto se debe a que los británicos usan con una frecuencia superior la pasiva, la cual, en la mayoría de los casos, Google Translator traduce literalmente, quitando naturalidad a la redacción en español —«fue comprado por Gabriel Wells» o «fue devuelto a Inglaterra». Además, también se han encontrado frases que, debido a una mala traducción literal, cuesta entenderlas, como por ejemplo la oración «Pero en su estado actual, más feliz de actividad vigorosa,...».

Las inadecuaciones tanto pragmáticas como retóricas son muy escasas, tan solo se han encontrado en cuatro ocasiones a lo largo de toda la traducción del corpus. Tres de esos errores se encuentran en textos especializados (económico, jurídico y literario), que se corresponden con las inadecuaciones retóricas, mientras que tan solo una de ellas es pragmática. También hemos encontrado solo tres inadecuaciones en los textos estadounidenses de carácter cultural; por el contrario, en los textos británicos se han registrado un total de 16:

- En los textos económicos se ha traducido *Euro zone* de dos formas distintas: «zona euro», la más común, pero también como «zona del euro». El sintagma con artículo no es totalmente incorrecto, aunque se usa en un menor número de ocasiones. Lo ideal sería que los resultados fueran uniformes.
- En el texto general literario se ha traducido *swan song* como «canción del cisne», una expresión que en España no tiene el mismo sentido que en Gran Bretaña.
- En numerosas ocasiones se han encontrado traducciones hacia el español latino. Por ejemplo, «tienes que verte sonriente» o «roedores de cejas gruesas».

Las siguientes figuras recogen los datos explicados anteriormente en forma de porcentaje, siendo el distintivo la variedad de la lengua.

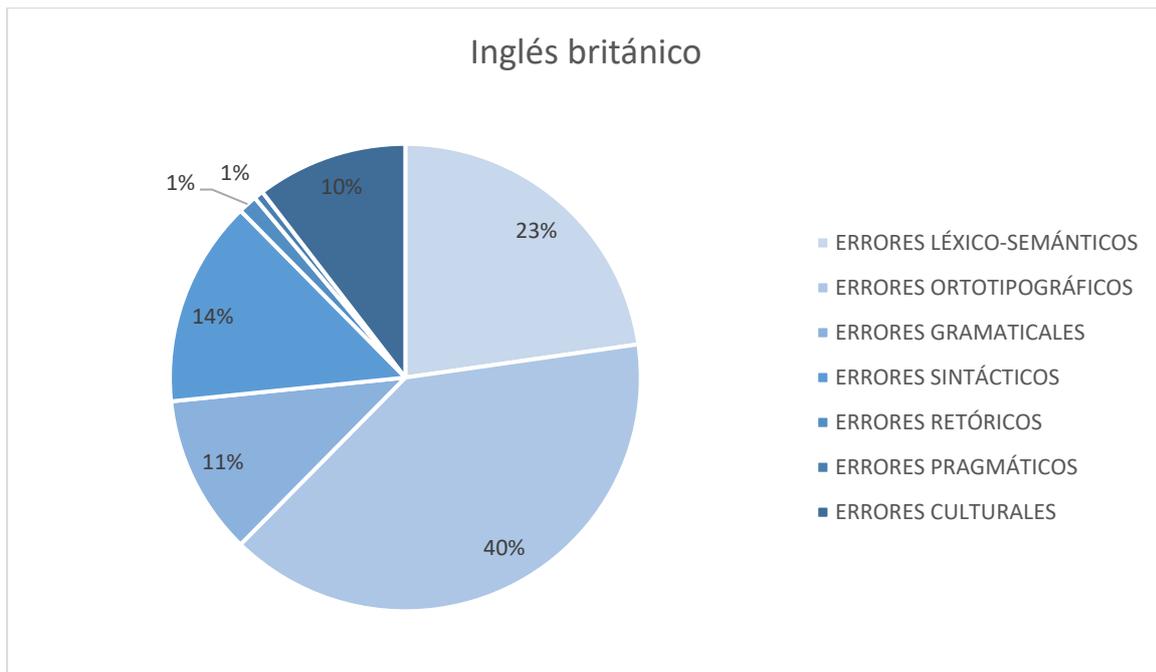


Figura 14. Porcentajes de error en los textos en inglés británico. Elaboración propia.

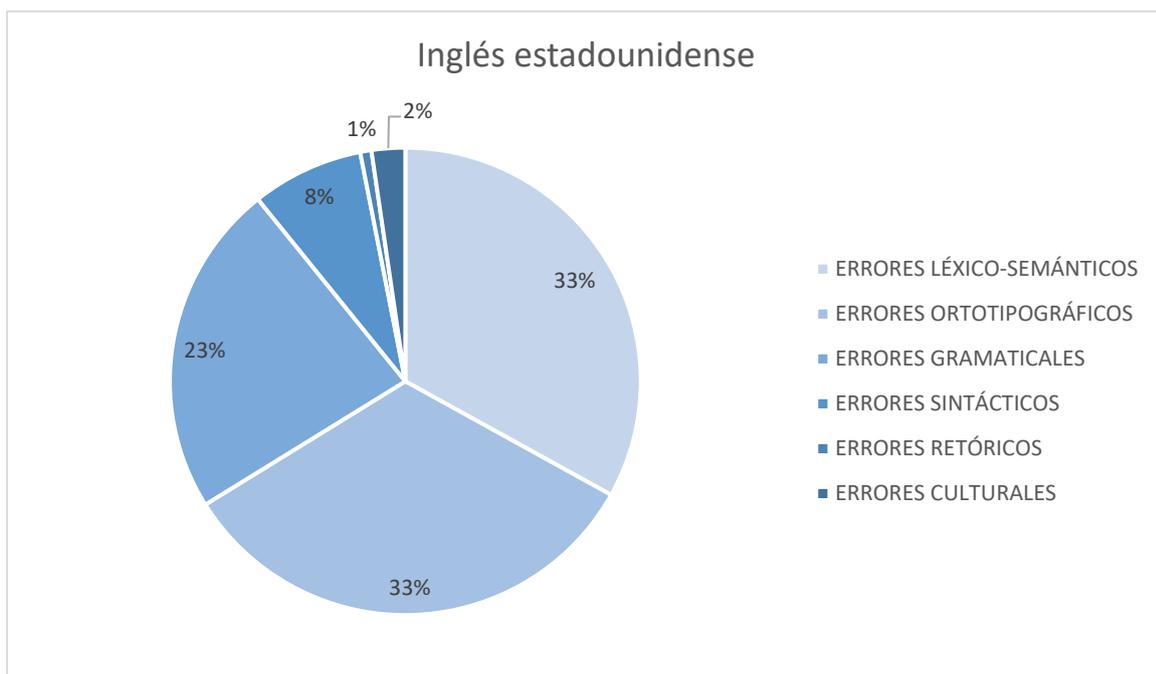


Figura 15. Porcentajes de error en los textos en inglés estadounidense. Elaboración propia.

7.2 Criterio 2: número de caracteres

Dado que son 5 000 los caracteres que reconoce Google Translator, no nos hemos encontrado con textos no traducidos por superar el número máximo. Además, los textos se ajustan

a las directrices de sintaxis y estilo, es decir, que los textos en español son más largos que el original. En la siguiente tabla, se puede ver la correspondencia de caracteres:

	Caracteres del texto original (T.O.)	Caracteres del texto meta (T.M.)
Texto general económico de menos de 500 caracteres británico	499	594
Texto general económico de más de 500 caracteres británico	752	853
Texto especializado económico de menos de 500 caracteres británico	321	459
Texto especializado económico de más de 500 caracteres británico	602	681
Texto general literario de menos de 500 caracteres británico	457	529
Texto general literario de más de 500 caracteres británico	790	838
Texto especializado literario de menos de 500 caracteres británico	373	439
Texto especializado literario de más de 500 caracteres británico	1111	1168
Texto general científico de menos de 500 caracteres británico	207	252
Texto general científico de más de 500 caracteres británico	695	805
Texto especializado científico de menos de 500 caracteres británico	273	310

Texto especializado científico de más de 500 caracteres británico	728	809
Texto general técnico de menos de 500 caracteres británico	167	177
Texto general técnico de más de 500 caracteres británico	764	896
Texto especializado técnico de menos de 500 caracteres británico	397	485
Texto especializado técnico de más de 500 caracteres británico	559	626
Texto general jurídico de menos de 500 caracteres británico	489	520
Texto general jurídico de más de 500 caracteres británico	743	756
Texto especializado jurídico de menos de 500 caracteres británico	419	432
Texto especializado jurídico de más de 500 caracteres británico	731	741
Texto general económico de menos de 500 caracteres estadounidense	380	444
Texto general económico de más de 500 caracteres estadounidense	649	732
Texto especializado económico de menos de 500 caracteres estadounidense	408	505
Texto especializado económico de más de 500 caracteres estadounidense	592	631

Texto general literario de menos de 500 caracteres estadounidense	498	530
Texto general literario de más de 500 caracteres estadounidense	679	777
Texto especializado literario de menos de 500 caracteres estadounidense	376	449
Texto especializado literario de más de 500 caracteres estadounidense	734	743
Texto general científico de menos de 500 caracteres estadounidense	325	410
Texto general científico de más de 500 caracteres estadounidense	557	616
Texto especializado científico de menos de 500 caracteres estadounidense	194	223
Texto especializado científico de más de 500 caracteres estadounidense	616	730
Texto general técnico de menos de 500 caracteres americano	418	520
Texto general técnico de más de 500 caracteres estadounidense	603	711
Texto especializado técnico de menos de 500 caracteres estadounidense	374	433
Texto especializado técnico de más de 500 caracteres estadounidense	796	948
Texto general jurídico de menos de 500 caracteres estadounidense	220	388

Texto general jurídico de más de 500 caracteres estadounidense	743	775
Texto especializado jurídico de menos de 500 caracteres estadounidense	329	348
Texto especializado jurídico de más de 500 caracteres estadounidense	619	662

Tabla 6. Diferencia del número de caracteres entre el T.O. y el T.M. en Google Translator. Elaboración propia.

Gracias a esta tabla podemos decir que las traducciones de Google Translator cumplen el criterio de ser más largas que los textos originales. La diferencia media se encuentra entre los 50 y los 100 caracteres. El texto especializado literario de más de 500 caracteres estadounidense, junto con su texto original, son los que se encuentran más próximos en cuanto a longitud, con 9 caracteres de diferencia. En el extremo con mayor diferencia de caracteres se encuentran el texto general jurídico de menos de 500 caracteres estadounidense y su original, con 168. Como se puede comprobar, el inglés británico lo traduce, en base a los caracteres, de una forma más literal y en el estadounidense se explaya más. Los textos extraídos de publicaciones en Gran Bretaña, en total, contienen una diferencia de 1 093 caracteres mientras que las de Estados Unidos contienen más del doble, 2 715.

Dentro de los grupos de menos de 500 caracteres, tanto británico como estadounidense, hemos encontrado que el texto científico especializado es, en ambos casos, el que menos errores contiene, siendo 0 y 2 las inadecuaciones encontradas. También comparten estos puestos con el texto técnico especializado en inglés británico y el texto jurídico especializado en inglés estadounidense. Para los de más de 500 caracteres, en el caso de los británicos, el texto con menos número de errores es el científico especializado también —4 inadecuaciones—, mientras que dentro de los estadounidenses se disputan el primer puesto entre el fragmento técnico especializado y el jurídico general, con una inadecuación cada uno de ellos.

Por el contrario, los textos con más número de errores son el económico general británico con 10 y el texto especializado literario estadounidense con 16 en aquellos que tienen menos de 500 caracteres y los textos literario especializado y científico general británicos con 18 inadecuaciones cada uno y el texto literario especializado estadounidense —26 errores— para los que conforman el corpus de fragmentos de más de 500 caracteres. Como podemos ver, los textos literarios especializados son los que se traducen con mayor dificultad, mientras que los científicos son los más sencillos. Esto se debe quizás a que los textos literarios, normalmente, no son parte

de grandes corpus multilingües con los que Google trabaja para sus traducciones, mientras que los textos científicos abundan por todo Internet.

Es por esto por lo que es necesario prestar especial atención al texto literario especializado de publicación estadounidense. La mayoría de las inadecuaciones se corresponden con los errores gramaticales en primer lugar y los léxico-semánticos, en segundo. Lo más chocante es el cambio de todos los tiempos verbales, como mencionábamos en el apartado anterior, podríamos pensar que este cambio se ha debe a que se busca que la traducción sea similar en extensión al original, puesto que, precisamente, este texto es el que menos caracteres de diferencia tiene con su T.O.

Además, hemos observado que entre los textos británicos de menos y más de 500 caracteres que forman el corpus, la diferencia a mayores es mínima, tan solo de 53 caracteres, siendo los que tienen un mayor número de caracteres los que más acumulan, mientras que, en los textos americanos, con diferencia, los que tienen un menor de caracteres son los que reúnen una mayor cantidad, siendo 848 la diferencia con respecto a los que tienen más de 500 caracteres. Sin embargo, son los textos de carácter general los que se alejan más del T.O. en materia de caracteres. Esta afirmación se complementa con las siguientes figuras:

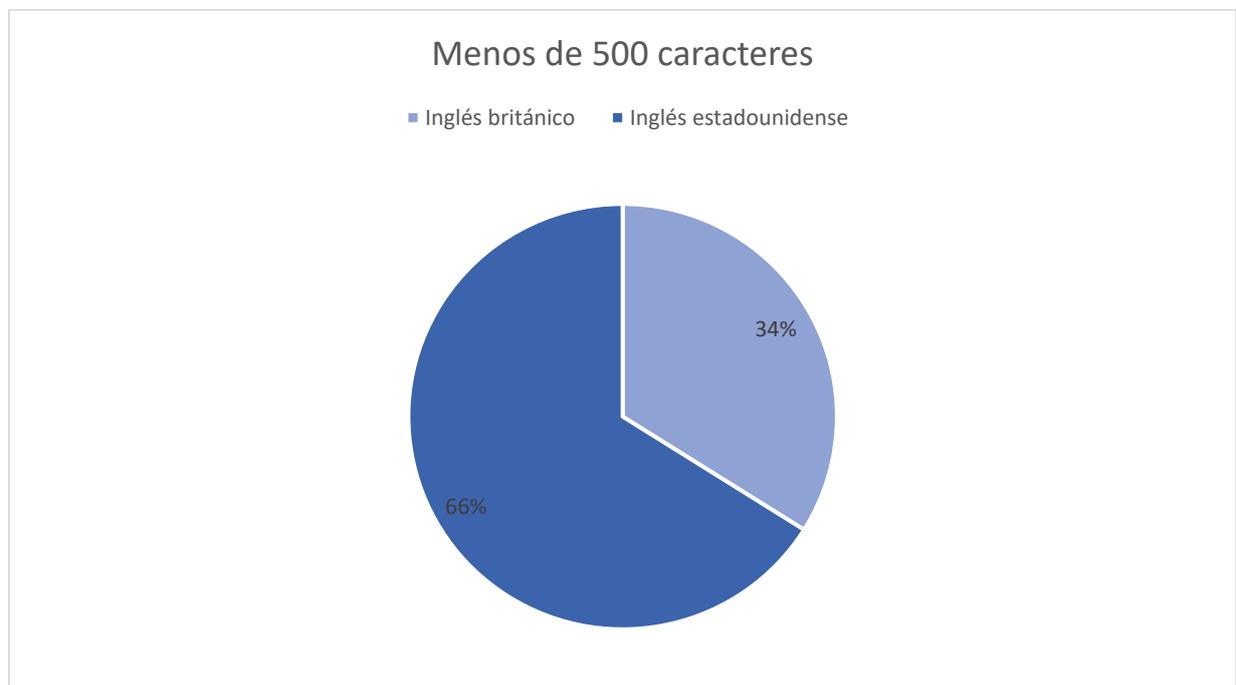


Figura 16. Porcentajes de inadecuaciones (I). Elaboración propia.

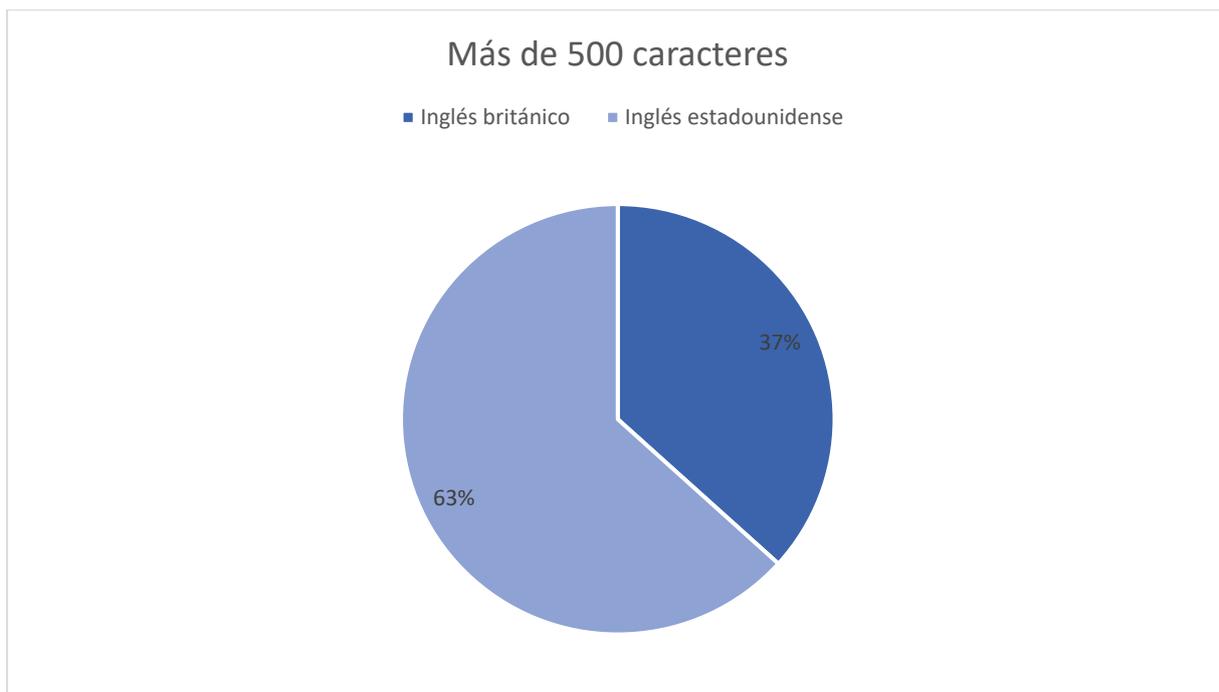


Figura 17. Porcentajes de inadecuaciones (II). Elaboración propia.

7.3 Criterio 3: general o especializado

En este caso, al igual que hicimos en el apartado de Siri, para poder establecer la siguiente comparativa de errores entre los textos generales y específicos, se ha optado por aunar los dos fragmentos que conforman cada grupo. Es por esto que, cuando hablemos de texto en este apartado, se referirá al conjunto de los fragmentos de menos y más de 500 caracteres.

Por lo general, son los textos especializados los que suelen presentar un mayor número de dificultades que solventar y en este caso, esta tendencia se ha demostrado. El texto en cuya traducción hemos encontrado más inadecuaciones es en el especializado estadounidense literario, un total de 42, en el que se engloban errores de todos los tipos en mayor o menor medida salvo los pragmáticos. Le siguen los textos literario especializado británico, económico general británico y literario general estadounidense, con 23 inadecuaciones cada uno. No obstante, también son los textos especializados los que contienen el menor número de errores en materia científica para los británicos y técnica para los estadounidenses. Estos datos podrían corresponderse con dos ideas que se suelen tener de estos países. Estados Unidos siempre ha sido uno de los países pioneros a la hora de innovación en nuevas tecnologías lo que hace de él un gran proveedor de información multilingüe de numerosos tipos de textos técnicos, mientras que Gran Bretaña siempre ha destacado por sus investigaciones en el campo de la medicina.

Los textos generales en este caso se encuentran más o menos equilibrados y el número de inadecuaciones oscila entre los cinco y los 25 como máximo, mientras que los especializados

tienen una media de error mucho más amplia que va desde los cuatro errores y los 42 que mencionábamos anteriormente.

En los textos británicos, la diferencia en la media de error entre los textos especializados y los generales es mínima. A nivel ortotipográfico, los más abundantes, se han encontrado 31 en el conjunto de todos los textos, sin embargo, no dista de los especializados, puesto que solo hemos registrado una inadecuación menos, 30. Los errores léxico-semánticos también son más frecuentes en los textos generales —17 inadecuaciones frente a las 15 de los textos especializados, aunque, en la contabilización, los gramaticales son los más próximos junto con los ortotipográficos, siendo una inadecuación la que los separa. Por el contrario, los errores sintácticos son los más llamativos puesto que en los textos de carácter general de este grupo, la reformulación tiende a veces hacia el español de Latinoamérica, lo que le resta calidad. El registro de inadecuaciones en los textos generales en sintaxis recoge 17, mientras que en los especializados solamente hemos encontrado cinco de ellos.

En los textos estadounidenses, las diferencias tampoco son muy grandes, pero si son más amplias que las anteriores. Por ejemplo, las inadecuaciones léxicas registradas en ambos casos son de 15 para los generales y 28 para los especializados. Las ortotipográficas se asemejan, igual que en el caso anterior, —21 y 22, respectivamente— y las gramaticales se diferencia por ocho, una cifra baja relativamente. Los errores sintácticos están muy próximos también, siendo tres la diferencia entre ambos. Como podemos comprobar, en el caso de los textos británicos, son los generales los que acumulan un mayor número de errores, mientras que en los americanos son justo los especializados los que necesitan una mayor revisión final que solucione los problemas encontrados en ellos.

Las siguientes cuatro figuras representan los porcentajes de los datos extraídos de los textos generales y especializados de ambas variedades lingüísticas.

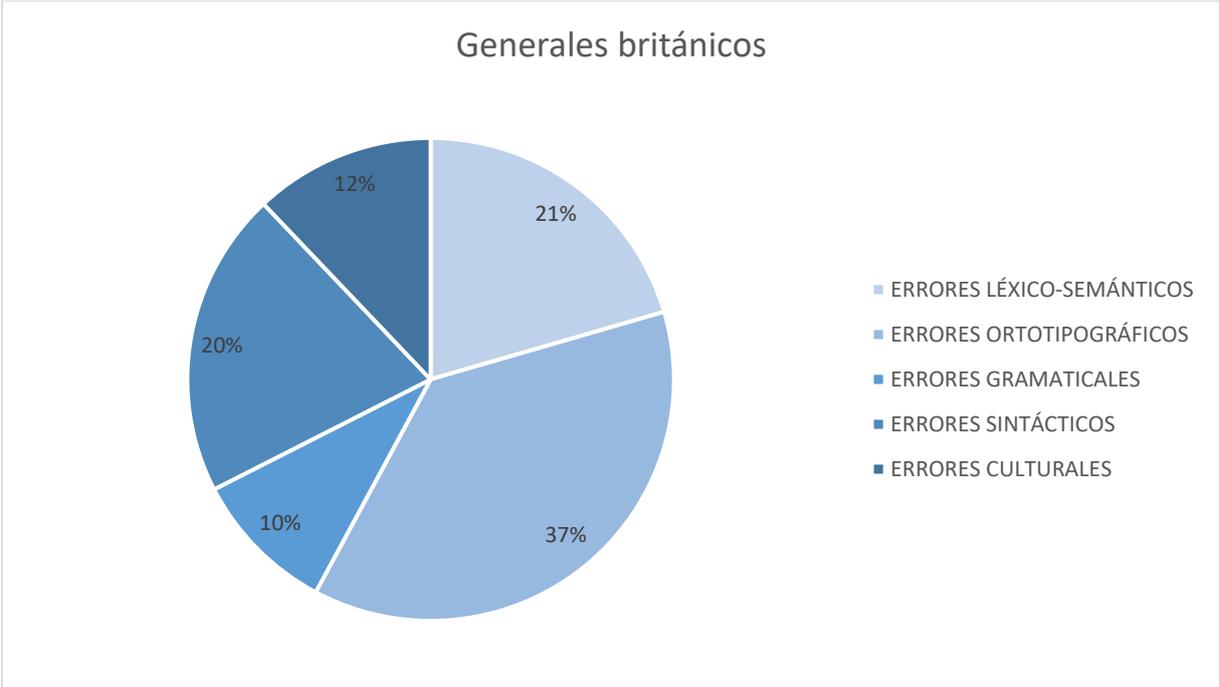


Figura 18. Porcentajes de error por categoría en los textos generales británicos. Elaboración propia.

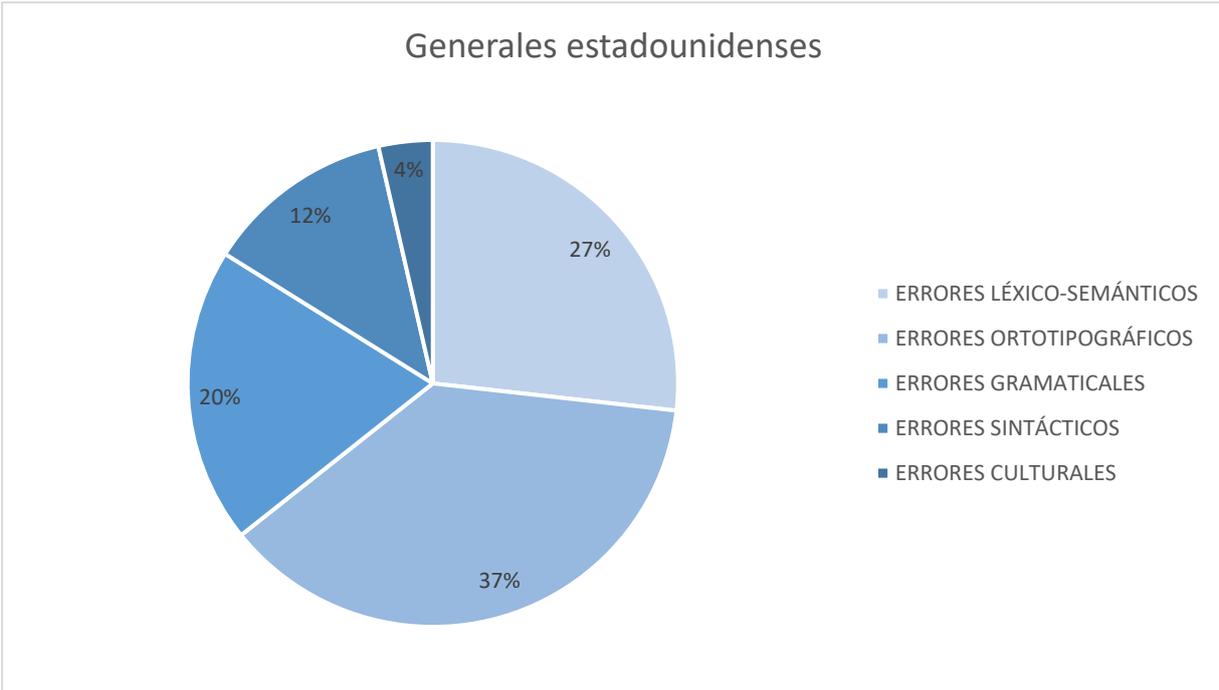


Figura 19. Porcentajes de error por categoría en los textos generales estadounidenses. Elaboración propia.

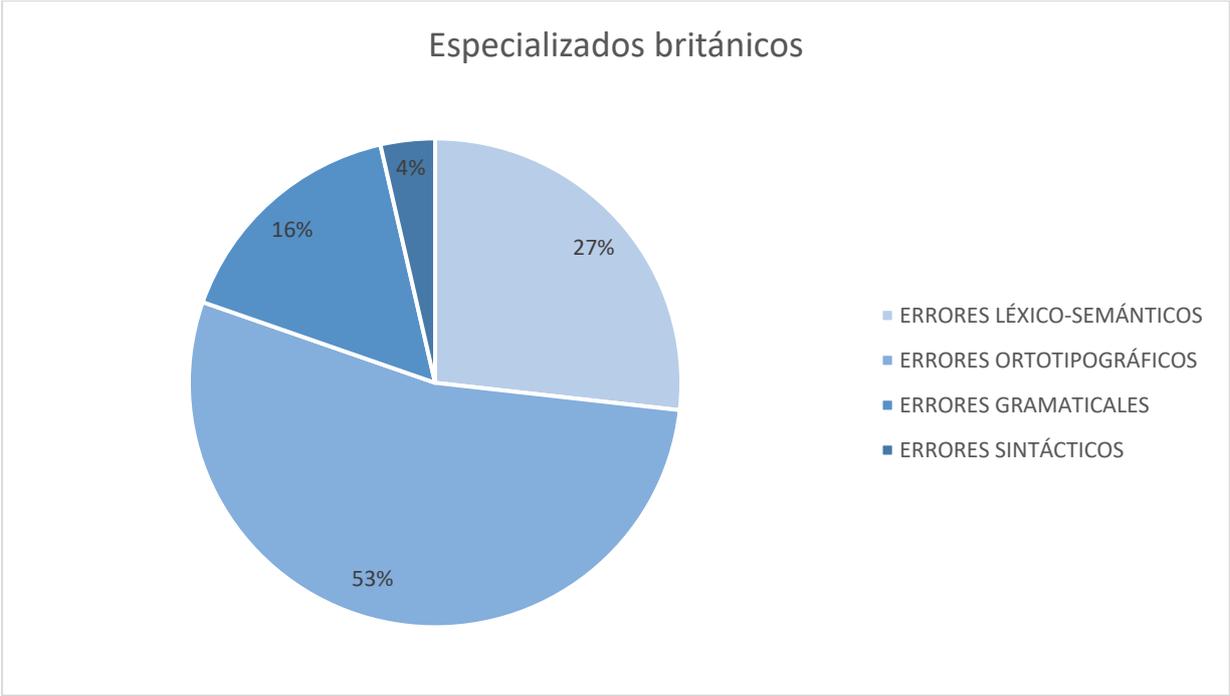


Figura 20. Porcentajes de error por categoría en los textos especializados británicos. Elaboración propia.

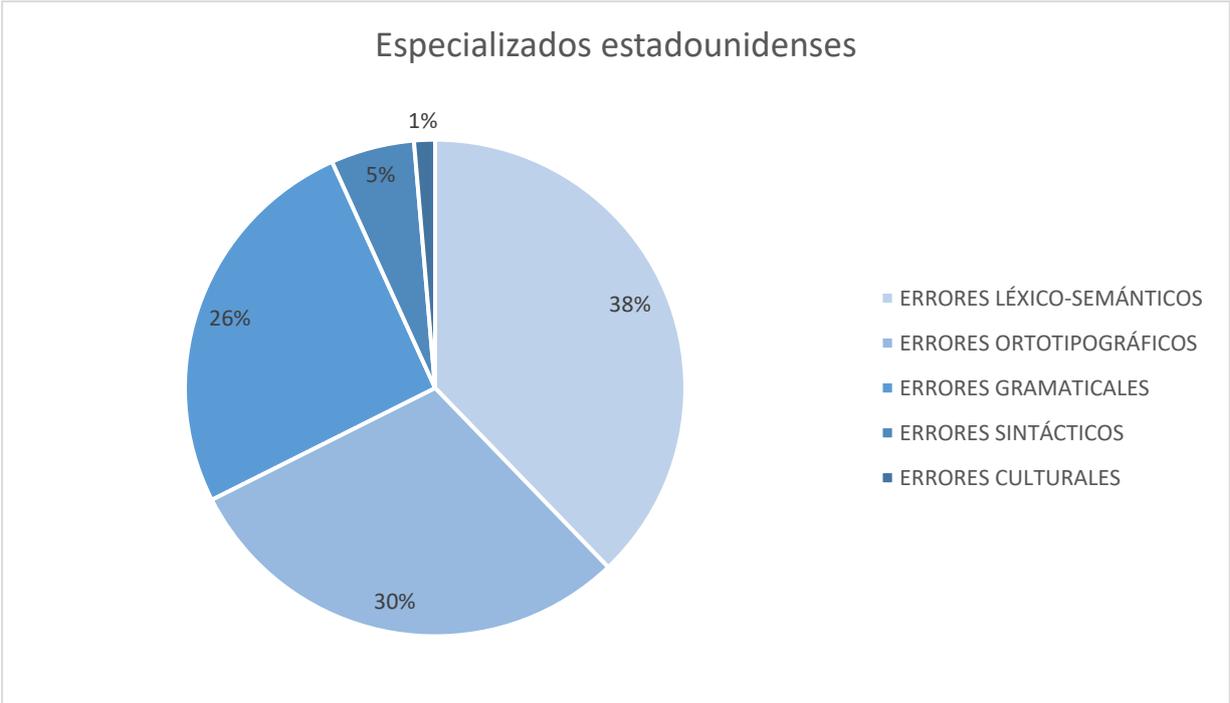


Figura 21. Porcentajes de error por categoría en los textos especializados estadounidenses. Elaboración propia.

7.4 Criterio 4: textos de producción propia

En este caso, las redes neuronales que forman Google Translator le han permitido afrontar los textos de producción propia casi al 100 % correcta. Estas son las siguientes inadecuaciones que hemos registrados para este grupo de textos:

- En el texto número uno, se ha marcado una única oración que podría confundir al receptor, «Me gustaría llegar a ser un gerente cuando sea grande». De ella, se han extraído cuatro inadecuaciones distintas: dos léxico-semánticas —en este caso el sentido de manager es el de representante y no tanto el de gerente y la expresión «ser grande» pertenece a un registro de lenguaje coloquial e infantil que no se amolda a este texto—, una inadecuación gramatical —en base a la expresión «ser un gerente», puesto que no llevaría el artículo delante del sustantivo— y una inadecuación sintáctica —por la falta de naturalidad en esta oración.
- En el texto número dos, hemos concluido que a nivel estructural no tiene ningún error, aunque presenta algunas inadecuaciones a nivel léxico-semántico y gramatical muy leves. Lo más destacable de este texto es que contenía un error a nivel cultural, geográfico más concretamente: en el texto original pone que Bali se encuentra en la India, la situación incorrecta, y, en la oración siguiente, se puso la correcta, que es Indonesia. El traductor no ha sido capaz de salvar esta dificultad.
- En el texto cuatro, hemos encontrado dos inadecuaciones bastante llamativas:
 - (a) por más que lo introdujimos en el traductor seguía dando el mismo resultado. El punto entre las oraciones «nos gusta ir a la playa» y «mi playa favorita se llama Miami» desapareció del mapa y no tienen separación. Por lo único que podemos intuir que son dos oraciones separadas es que la segunda comienza por mayúscula y que el T.O. está dividido de esta manera.
 - (b) Debido a que las oraciones «mi playa favorita se llama Miami», «es muy largo, con arena blanca y palmeras» y «es muy bonito» están separadas por puntos, el traductor no ha reconocido que los adjetivos *long* y *beautiful*, que en inglés son invariables, se refieren a la playa Miami, por lo que la traducción en masculino no es del todo correcta. Se entiende el significado puesto que suponemos que se habrá asociado a una palabra más general como «paisaje».
- Los textos cinco y seis, especiales por los errores que contienen, se han traducido de una forma sencilla que permite comprender el texto en su totalidad, pero les falta naturalidad. En el cinco, la oración que le quita más calidad al T.M. es «he jugado muchos deportes». Primeramente, porque el verbo «jugar» no se suele colocarse con deportes, sino que, más bien, se utilizan «practicar» o «hacer». Sin embargo, podríamos

mantener el verbo «jugar», pero, entonces, necesitaríamos la preposición «a» detrás del mismo. En el texto seis, el que más errores de sintaxis y vocabulario tenía, se ha traducido bastante literal, lo que supone que, por ejemplo, «trabajo de enseñanza en la escuela secundaria», al oído de un nativo español se entienda, pero suene extraño. Se sobreentiende que el trabajo del que habla es profesor y que es de instituto, con los estudiantes de secundaria.

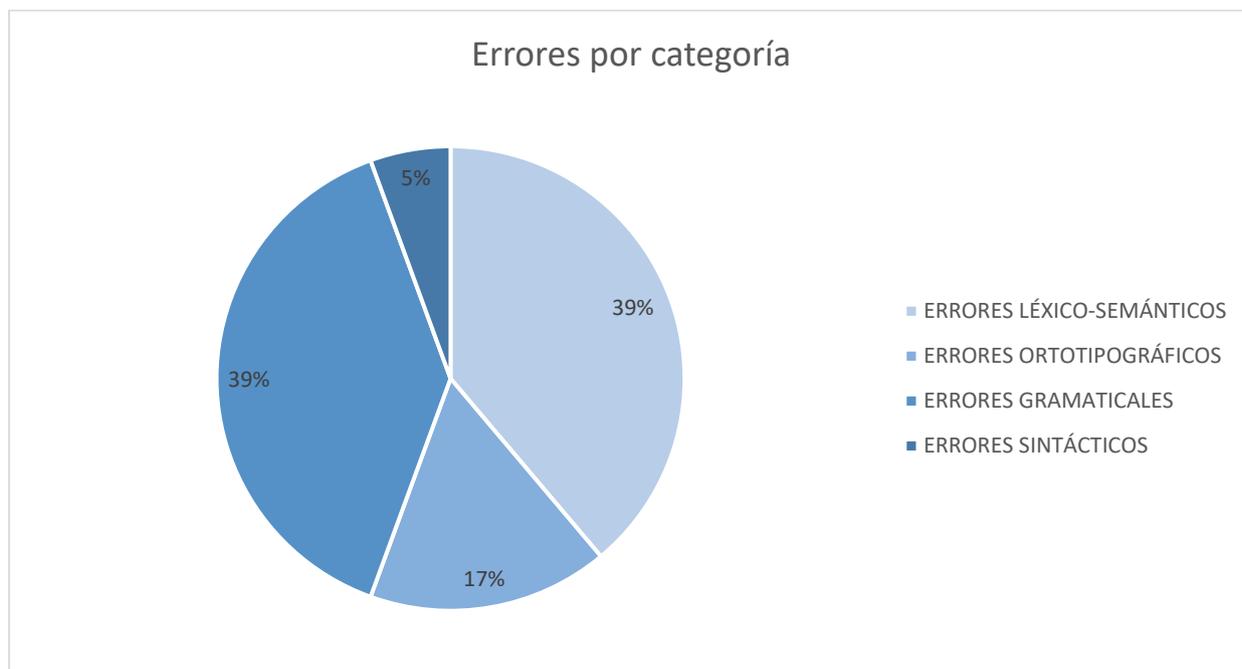


Figura 22. Porcentajes de error en los textos de producción propia. Elaboración propia.

7.5 Conclusiones extraídas del análisis

Google Translator es uno de los motores de traducción más avanzados gratuitos que existen en el mercado. Gracias a los voluntarios que ayudan con sus aportaciones, las inadecuaciones a nivel léxico-semántico o gramatical son las menos importantes. Sin embargo, a los errores ortotipográficos hay que prestarles mucha atención. Este problema creo que se debe principalmente a que la gente que participa en el desarrollo desconoce las normas de estilo, especialmente, las que están relacionadas con las cifras, las cuales reúnen la mayoría de los errores de este tipo:

- el símbolo del tanto por ciento se separa con un espacio de la cifra, normalmente indivisible,
- los símbolos de las divisas se colocan detrás de la cifra, al contrario que en inglés,
- los decimales no se separan con puntos, sino con comas,
- o que los miles se separan con un espacio y no con un punto.

Otro de los defectos importantes que hemos concluido de este análisis es que no hace distinción entre español de España y el de Latinoamérica. Varias de las expresiones que sonarían extrañas a un castellano son propias del habla de América Latina, por lo que, aunque el texto no pierde el sentido, la cercanía al lenguaje natural como la que ofrece un traductor humano no es fiable al 100 % para una «máquina». No obstante, también tiene sus puntos fuertes, que como hemos visto es capaz de abordar y proporcionar una traducción bastante cercana de varios tipos de textos, con diferentes temáticas, es decir, es capaz de cubrir de forma aceptable cualquier tema existente.

Sin embargo, se han registrado casos en los que las redes neuronales o los corpus con los que trabaja no consiguen que sea capaz de enlazar unas ideas con otras, especialmente cuando se trata de temas culturales. Un ejemplo de ello son los errores que se cometen con ciudades y países, que no se corresponden el uno con el otro, o con expresiones que el autor del texto inventa como, la que encontrábamos en el texto especializado literario británico de menos de 500 caracteres *Sleep Coaxing Routine*. De este tipo de expresiones que no se usan con frecuencia, incluso podríamos decir que es de las pocas ocasiones en las que lo veremos —en Internet únicamente aparece como resultado el T.O. original del que se ha extraído el segmento—, no se han obtenido traducciones fiables o, en su defecto, no se han traducido.

Uno de los puntos fuertes de este traductor es que, pese a que pueda cometer alguna inadecuación, los textos están bien cohesionados, no hemos encontrado ningún falso sentido que se haya cometido por una mala comprensión del sentido del texto o una mala traducción de los conectores. Esto implica que la coherencia tampoco supone un problema para el este motor de traducción.

Hemos constatado que los textos médicos son los que mejor consigue traducir, probablemente por los grandes corpus que lo integran, por su constante autoaprendizaje y por su colaboración con todos los recursos que podemos encontrar en Internet. No obstante, esta colaboración no siempre sirve de ayuda debido a que no es capaz de establecer la diferencia entre traducciones fiables y menos fiables, especialmente cuando se trata de separar el español de España con el español latino. Una traducción se considera correcta cuando se mantiene en la variedad lingüística que corresponde y, como explicábamos en los apartados anteriores, introducir otras variedades como en este caso, produce al lector extranjero una sensación de incoherencia o falta de naturalidad.

Otra de las características destacables de este traductor automático es que mantiene el formato del T.O., es decir, que, si se introducen varios párrafos separados o enumeraciones o cualquier otro tipo de configuración, en la traducción se mantienen. Cuando hablamos de formato, no se incluyen las oraciones, puesto que, al reformular lo principal es mantener el sentido y,

dependiendo de la lengua, la separación por puntos seguidos de éstas puede variar, en el sentido de que en inglés no utilizan tanto las subordinadas como en español.

En relación a las limitaciones de la TA, a las que hacíamos referencia en el apartado teórico de este trabajo, puedo decir que, al igual en las traducciones de Siri, no se ha detectado un uso incorrecto de sinónimos con connotaciones específicas, pero sí que se han detectado problemas con la polisemia de las palabras con algunos adjetivos, como en el texto general científico británico de más de 500 caracteres, en el que *thuggish* se ha traducido como «matón» y en el contexto en el que se presenta, hablando de los Neandertales, podemos afirmar que los estudios históricos no representan a estos antepasados como matones si no como «brutos». Además, es curioso ver como en ciertas ocasiones, la misma palabra se presenta con distintas tipografías dentro de un mismo texto:

El laboratorio ya ha insertado genes de Neanderthal para el desarrollo craneofacial en ratones (no se anticipan roedores de cejas gruesas) y genes de percepción del dolor neandertal en huevos de ranas, lo que podría indicar si tenían un umbral de dolor diferente para los humanos. Ahora el laboratorio está volviendo su atención al cerebro (Anexo IV).

En cuanto a la ambigüedad de la que nos hablaba Alonso Martín (2003: 98-100), podemos decir que en el texto literario especializado de más de 500 caracteres y en inglés británico, Google Translator tampoco ha tenido en cuenta las señales que indicaban que la persona que está escribiendo el texto en primera persona es una mujer.

CAPÍTULO 8. COMPARATIVA DE LAS PRODUCCIONES DE SIRI Y GOOGLE TRANSLATOR

Los traductores de Siri y de Google, desde el comienzo del estudio, ya presentaban una serie de diferencias a nivel de desarrollo que nos hacían intuir que los resultados iban a ser completamente opuestos. Cuando estudiábamos en la parte teórica que el motor de traducción de Apple era una función recién integrada, previmos que sus producciones no iban a ser de tanta calidad como las de Google, motor que lleva más de 10 años en constante evolución. También influía mucho el hecho de que el de Google es un traductor automático caracterizado por su apertura, es decir, por aceptar colaboración de todos los voluntarios alrededor del mundo que quieran prestar sus conocimientos para alcanzar la perfección en TA, mientras que los códigos del lenguaje del asistente virtual son exclusivos para sus desarrolladores.

Esta diferencia ha influido en la traducción de los textos, puesto que como veíamos, el traductor de Apple no era capaz de reconocer 19 de ellos, al contrario que el de Google que ha dado resultados para todos los que formaban el corpus monolingüe. Además, la calidad de los textos de Google Translator es bastante más alta. Los errores léxico-semánticos son más abundantes en los textos de Siri dado que no trabaja todavía con grandes corpus lingüísticos si no que funciona más bien como una especie de diccionario bilingüe. Por otro lado, la capacidad de reformulación a nivel sintáctico es superior, sus frases son más complejas, incluyen todo tipo de subordinadas y/o coordinadas, entre otros elementos.

Además, los textos producidos por Google Translator, como mencionábamos anteriormente, son más coherentes y están mejor cohesionados que los textos producidos por Siri, dado que la mayoría de ellos no seguían el orden original de la oración ni de la información y, aparte, no estaban acabados, sino que solo obtuvimos una parte del texto original, lo que no nos ha permitido hacer un estudio más profundo en este aspecto. Tampoco se han encontrado gran cantidad de conectores que unan unas frases con otras, ni tampoco signos ortográficos que las separen. De hecho, veíamos como Siri mezclaba unas frases con otras cambiando el sentido del texto, dando lugar a falsos sentidos dentro de la traducción.

Podría decirse que el traductor de Apple aún está en pañales, es decir, que puede servir como apoyo en un momento muy puntual en el que la situación en la que se necesite la traducción no sea de un alto grado de especialización, mientras que Google Translator, como también tiene aplicación para móviles e incluso te ofrece la posibilidad de trabajar sin conexión habiéndote descargado previamente los datos que necesita, es un recurso que te proporciona traducciones de calidad aceptable en una décima de segundo.

Si enfocamos ahora la comparativa hacia los datos extraídos de ambos análisis, vemos como, claramente y como se podía prever, Siri tiene una cantidad mayor de inadecuaciones en sus traducciones. Para un total de 13 textos británicos se han registrado 126 y 131 para los ocho estadounidenses poniendo en conjunto todos los tipos de errores que se contemplaban en la clasificación del Capítulo 4, mientras que, para los textos producidos por Google, sabiendo que de todos ellos se ha obtenido un texto final, se han registrado un total de 154 para los textos británicos y 130 para los estadounidenses. La media de error más alta la tienen los textos estadounidenses traducidos por Siri, siendo de 16 errores por texto y la más baja es también para los estadounidenses pero traducidos por Google, siendo de entre 6 y 7.

La inadecuación ortotipográfica que encontrábamos con mayor frecuencia en la producción de Siri —la coma delante de la «y»— también es un error que se repite con frecuencia en la de Google, por eso es necesario insistir en la importancia de conocer bien las normas de ortografía y tipografía de los idiomas antes de dar por válida una de las opciones que se muestran en el panel de la Comunidad de traductores de este último, cuya interfaz podemos ver en la siguiente imagen.

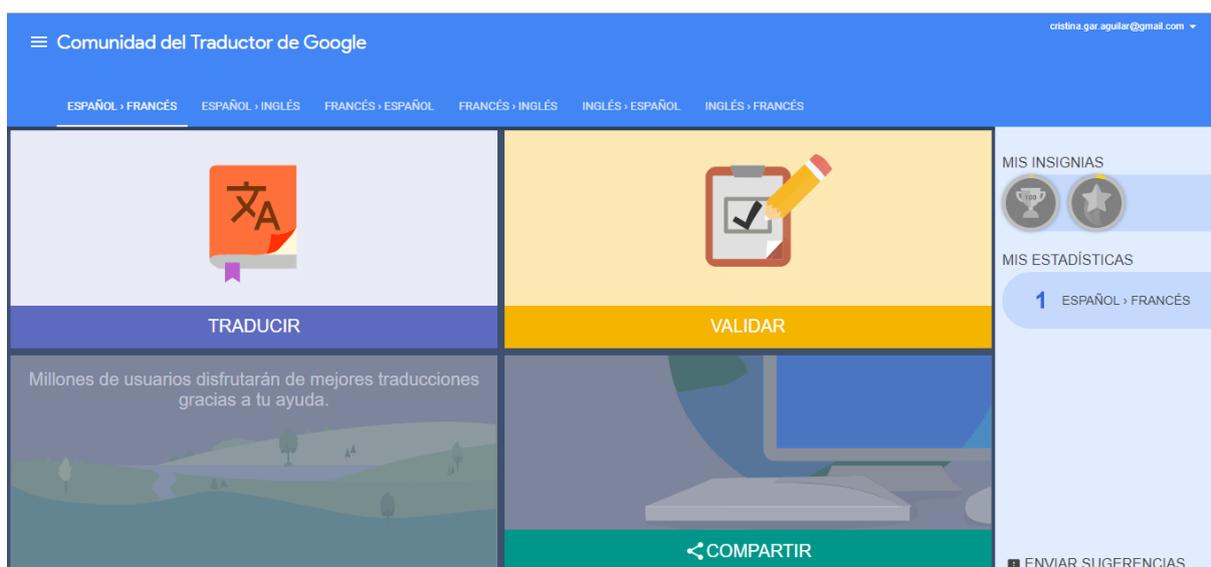


Ilustración 15. Comunidad de Google Translator. Google Translate Community (2018).

A nivel léxico-semántico, sintáctico y gramatical, se ha comprobado que los textos traducidos por Google las reformulaciones son menos literales, pero, a la vez, esto hace que aumente el peligro de cometer falsos sentidos o cambiar la finalidad del texto o su momento dentro del espacio temporal, un hecho que no ocurre con el traductor de Siri.

Los errores de carácter pragmático son prácticamente inexistentes en todas las traducciones que hemos analizado, ninguno de los dos motores de traducción tiene dificultad a la hora de cambiar la persona a la que se dirige dependiendo de la lengua a la que se traduzca, en este caso, entre «tú» y «usted».

Las inadecuaciones culturales están presentes en todas las traducciones de textos que se escogieron por este mismo motivo. Es complicado para ellos adecuar ciertos aspectos de la cultura británica o estadounidense a la española, puesto que no tienen conocimientos de adaptación a la lengua meta, si no que se basan en otras traducciones que pueden o no estar bien. Esta conclusión se puede ver con claridad en los textos jurídicos especializados, puesto que las leyes de cada país contienen puntos culturales que en España no existen, como por ejemplo la clasificación de drogas que hacen en el sistema legislativo británico.

CONCLUSIONES

Como muy bien expresaba Tertoolen (2010: 27) en su estudio, *la complejidad de la TA está vinculada con la ingeniosidad de las lenguas naturales. Una lengua natural no sólo es muy precisa y especializada, sino también arbitraria y ofrece una infinita cantidad de construcciones. Cuando vinculamos la complejidad de las lenguas naturales con las limitaciones de los ordenadores, nos enfrentamos con muchos problemas a la hora de automatizar el proceso de traducción. Uno de los problemas más complejos es la ambigüedad. Incluso para un traductor humano la ambigüedad puede producir problemas de interpretación. No obstante, el traductor automático se enfrenta con muchas más ambigüedades que el traductor humano. Contrariamente a los traductores humanos, el traductor automático no puede deducir la interpretación adecuada del contexto y tampoco de su conocimiento del mundo real.*

Coincido totalmente en esta reflexión que hace y resume muy bien la conclusión principal a la que he llegado durante la realización de este proyecto. Los traductores automáticos son una gran ventaja que nos ofrece la tecnología y que deberíamos aprovechar ya que nos da la oportunidad de aumentar el rendimiento de trabajo. Cada vez son más fiables gracias a las importantes inversiones que se están haciendo dentro de este campo y, con este trabajo, hemos puesto un pequeño grano de arena para que sigan evolucionando. Pese a que actualmente parece improbable que un motor de traducción proporcione una traducción tan acertada como la que podría nacer de las manos de un traductor humano, es un hecho que se van acercando mucho y son un recurso que no debería quedar tan renegado y ser tan rechazado por los profesionales del sector.

A través de este TFM, se ha ofrecido una breve exposición de la situación actual de la traducción automática dentro del sector a nivel internacional. Hacer un trabajo relacionado con esta temática, te hace abrir los ojos hacia un nuevo futuro en el que la traducción sea una colaboración entre humanos y máquinas. Personalmente, soy una de esas personas que no es nada reticente a trabajar con este tipo de sistemas artificiales, por lo que me pareció muy interesante acercarme más a este mundo, el cual, dentro del aula, no se explora en profundidad. Gracias a él, he podido ver cómo trabaja una *machine translation* y valorar cuales son los aspectos a los que más atención debemos prestar antes de entregar la traducción final al cliente.

No obstante, esta actividad debe estar regulada de alguna forma dentro del sector, puesto que lo que no se puede permitir es que se utilicen estas herramientas como una forma de eludir el trabajo, es decir, entregar traducciones pasadas por estos motores de traducción sin una revisión posterior con fines económicos. Por eso, incido en que desde las universidades se debe

comenzar a formar a los estudiantes armonizando el trabajo humano con el trabajo automático y asentando las bases de la posesión de textos para, así, evitar el fraude dentro de esta profesión y seguir defendiendo que una persona que sabe idiomas no es un traductor.

Como explicábamos en el Capítulo 8, al hacer la comparativa general de todos los aspectos, concluimos que los traductores automáticos son cada vez más útiles para facilitar la vida a los profesionales del sector, pero no se debe pensar que ellos acabarán haciéndolo todo debido a que como estudiábamos, los motores automáticos de traducción no son capaces de entender o traducir correctamente ciertos aspectos de la cultura del país, que se encuentran en constante cambio y de los cuales, en muchas ocasiones no existe una traducción oficial, por lo que los adaptan de manera literal, cuando en la otra cultura puede ser un término desconocido o, directamente, el mismo desconoce el equivalente.

Las inadecuaciones léxico-semánticas, junto con las gramaticales y las sintácticas son una batalla en la que ambos traductores automáticos siguen peleando. Estas dificultades son complicadas de salvar puesto que las lenguas naturales son producto de una evolución que no tiene una fecha concreta en la que se detengan. El procesamiento del lenguaje natural no puede seguir su ritmo de crecimiento puesto que entrenar un motor de traducción requiere de tiempo y grandes cantidades de textos multilingües que contengan las novedades en distintos idiomas, ya sean adaptaciones, traducciones o mantenimiento de la palabra en su lengua original.

Sin embargo, es importante decir que, gracias a ellos, un número muy personas que no saben lenguas tienen la posibilidad de ir a cualquier parte del mundo y pueden comunicarse con otras personas.

Por otro lado, una vez analizados los resultados obtenidos en nuestro análisis, en este capítulo final, también nos gustaría retomar los objetivos que se planteaban en las primeras líneas de este TFM. Inicialmente, dejábamos constancia de que el objetivo principal era analizar las inadecuaciones que se encontraban en la traducción realizada por Siri y Google, gracias a este análisis hemos podido comprobar el funcionamiento de dos motores de traducción totalmente opuestos y ver de qué forma trabajan cada uno de ellos, haciendo posible su comparación y la presentación de una calificación de ambos y las líneas futuras que deberían seguir investigándose.

La forma de trabajo de los motores de traducción de Google y Apple son completamente diferentes; Siri trabaja principalmente con diccionarios, proporcionando traducciones literales y de baja calidad, mientras que Google Translator trabaja a través de redes neuronales que le permiten reformular las oraciones y ofrecer mejores producciones. Por otro lado, su estudio nos ha permitido determinar que las limitaciones que recogíamos en la parte teórica ya no se cumplen en su totalidad, ya que con el paso de los años, obstáculos como las ambigüedades o la polisemia.

Líneas futuras

Como decíamos al principio de este trabajo, la función que permite a Siri traducir del inglés a otras cinco lenguas es algo muy novedoso, que salió al mercado a finales del año 2017. Por eso, considerando que es un primer prototipo, podemos decir que aún le faltan muchas cosas que pulir para que se pueda poner a la altura de otros traductores automáticos, empezando por el hecho de que solo pueden acceder a él personas que tengan configurado su dispositivo en inglés estadounidense. Además, no es fácil acceder a él, ya que para hacer cada una de las traducciones hay que estar entrando y saliendo de Siri, preguntándole para que te abra esa opción. Quizás sería de mayor utilidad si se incluyera en una aplicación a parte dentro del sistema operativo o que fuera descargable como Google Translator.

Por otro lado, una de las grandes carencias que se puede ver a simple vista en las traducciones de Siri es léxica. Para solventar este problema, invertir en la creación de bases de datos y glosarios más amplios puede ser una opción muy beneficiosa para aumentar la confianza en este traductor automático. Otra posibilidad es recurrir al entrenamiento de Siri a través de una enorme compilación de corpus paralelos, ya que hoy en día, Internet es una gran fuente de información que podría mejorar la calidad de las traducciones. En relación esto último y dado que Apple es una de las empresas más importantes a nivel mundial, que suponemos que trabajará con traductores profesionales pese a que no exista un departamento específico que desarrolle esta actividad, contar con la opinión experta de este sector sería, también, una forma de mejorar el resultado de los T.M. La nueva tecnología de Google está consiguiendo grandes avances dentro de este sector y sería de gran utilidad que se perfeccionara en estos años.

A nivel de formato, sería necesario introducir una actualización que reduzca o limite, en su defecto, el número de caracteres que puede procesar, puesto que se ha comprobado que la calidad de las traducciones baja cuantos más caracteres trata de analizar, e incluso, también sería necesario que se intentase mantener el formato original de los textos, como, por ejemplo, las enumeraciones separadas en diferentes párrafos, con listas de viñetas o numeradas, al igual que lo hace Google Translator.

Por último, me gustaría hacer una reflexión personal que no es tanto una mejora, si no una sugerencia. Como comentaba al final del apartado de «Resultados» dentro del capítulo de análisis de los textos de Siri, las traducciones más fiables se recogen cuando utilizas sus funciones como diccionario multilingüe y no como traductor automático, de momento. Por ello, para evitar confundir a los usuarios, quizás debería llamarse diccionario y no traductor. En cuanto al Traductor de Google, es bastante práctico cuando estás realizando traducciones generales, pero necesitan mejorar en ciertos aspectos como la ortotipografía y la mezcla de distintas variedades lingüísticas que dan poca naturalidad a los textos.

Se ha quedado en el tintero continuar investigando como trabaja el traductor integrado en Siri, por ejemplo, si tiene presente una interlingua, como el de Google, o si hace las traducciones directamente de una lengua a otra. Esta rama de la investigación no se ha podido concretar puesto que solo hemos trabajado con un par de lenguas en una única dirección debido a las limitaciones existentes.

Por el contrario, las líneas de investigación de Google se deberían centrar en el desarrollo de la inteligencia artificial, en particular, para la introducción de métodos a través de los cuales consigan reconocer las figuras literarias. Este problema es uno de los más preocupantes a la hora de hacer traducciones con motores automáticos puesto que no son capaces de entrever el significado que puede tener una metáfora, una ironía o un sarcasmo entre otros. En el momento en que este reto se cumpla, los traductores automáticos serán una verdadera competencia dentro de este sector.

REFERENCIAS BIBLIOGRÁFICAS

Abaitua, J. (1997). *La traducción automática: presente y futuro*. En línea. Consultado el 27 de marzo de 2018. Recuperado de <http://www.foreignword.com/es/Technology/art/Abaitua/>

Adams, T. (2010, 19 de diciembre). Can Google break the computer language barrier?. *The Guardian*. En línea. Consultado el 25 de junio de 2018. Recuperado de <https://www.theguardian.com/technology/2010/dec/19/google-translate-computers-languages>

Alonso Martín, J.A. (2003). "La traducción automática". En M.A. Martí Antonín (Ed.), *Tecnologías del lenguaje* (pp. 94-129). Barcelona, España: Editorial UOC.

Alphatrad (2016, 18 de mayo). Traducción automática – lo que puede y no puede hacer. *Alphatrad*. En línea. Consultado el 25 de junio de 2018. Recuperado de <https://www.alphatrad.es/noticias/traductor-automatico-versus-humano>

Archanco, E. (2016). La evolución de Siri: es hora de abrir el asistente de Apple. *Applesfera*. Consultado el 30 de marzo de 2018. Recuperado de <https://www.applesfera.com/apple-1/la-evolucion-de-siri-es-hora-de-abrir-el-asistente-de-apple>

Arnold, D. (2003). "Why translation is difficult for computers". En H. Somers (Ed.), *Computers and Translation: A translator's guide* (pp. 119-142). Ámsterdam, Países Bajos: John Benjamins.

Barnés, H. G. (2013, 4 de junio). 7 errores gramaticales (muy comunes) que debemos evitar. *El Confidencial*. En línea. Consultado el 9 de julio de 2018. Recuperado de https://www.elconfidencial.com/alma-corazon-vida/2013-06-04/7-errores-gramaticales-muy-comunes-que-debemos-evitar_196010/

Beltrán Gandullo, M. (2001). "Traducción automática versus traducción humana del lenguaje especializado: alemán-español". En C. Valero Garcés y I. de la Cruz Cabanillas (Eds.), *Traducción y nuevas tecnologías: herramientas auxiliares del traductor* (pp. 143-151). Madrid, España: Universidad de Alcalá, Servicio de Publicaciones.

Boitet, C., Blanchon, H., Seligman, M. y Bellynck, V. (2010). MT on and for the Web. *Proceedings of the 6th International Conference on Natural Language Processing and Knowledge Engineering (NLPKE-2010)*. Congreso llevado a cabo en Beijing, China. En línea. Consultado el 25 de junio de 2018. Recuperado de <http://www-clips.imag.fr/geta/herve.blanchon/Pdfs/NLP-KE-10.pdf>

Casacuberta Nolla, F. y Peris Abril, A. (22 de diciembre de 2017). Traducción automática neuronal. *Tradumática*, 15, 66-74. En línea. Consultado el 17 de mayo de 2018. Doi: <https://doi.org/10.5565/rev/tradumatica.203>

Casacuberta Nolla, F. y Vidal Ruiz, E. (2016). Traducción automática y asistida. En A. Luis Gonzalo (coord.), *Tecnologías del lenguaje en España. Comunicación inteligente entre personas y máquinas* (pp. 86-95). Barcelona, España: Ariel.

Centro Virtual Cervantes (2018). *Diccionario de términos clave de ELE*. En línea. Consultado el 29 de mayo de 2018. Recuperado de https://cvc.cervantes.es/ensenanza/biblioteca_ele/diccio_ele/diccionario/errorpragmatico.htm

Cerdá, R. (1995). *Perspectivas en traducción automática*. Valencia, España: Lynx.

Cheyner, A. (2012, 21 de diciembre). How did Siri get its name?. *Forbes*. En línea. Consultado el 30 de marzo de 2018. Recuperado de <https://www.forbes.com/sites/quora/2012/12/21/how-did-siri-get-its-name/#2cf6ec9b376b>

Chicharro, R. (2017, 5 de junio). Siri se quiere convertir en el nuevo traductor portátil. *Tuexperto*. En línea. Consultado el 30 de marzo de 2018. Recuperado de <https://www.tuexperto.com/2017/06/05/siri-se-quiere-convertir-en-el-nuevo-traductor-portatil/>

Cortez Vásquez, A., Vega Huerta, H. y Pariona Quispe, J. (2009). Procesamiento de lenguaje natural. *Revista de Investigación de sistemas e informática. Volumen 6*, 45-54.

Cruces Colado, S. (2001). El origen de los errores en traducción. En D. Jiménez (Coord.) VIII Coloquio de la Asociación de Profesores de Filología Francesa de la Universidad Española. *Écrire, traduire et représenter la fête*. Congreso llevado a cabo en Valencia, España.

Dean, J. y Ng, A. (2012, 26 de junio). Using large-scale brain simulations for machine learning and A.I.. *Google*. En línea. Consultado el 25 de junio de 2018. Recuperado de <https://googleblog.blogspot.com/2012/06/using-large-scale-brain-simulations-for.html>

Diéguez Morales, M. I. (2001). Aciertos y errores en la traducción automática: metodología de la enseñanza-aprendizaje de la traducción humana. *Onomázein: Revista de Lingüística, Filología y Traducción*. En línea. Consultado el 29 de mayo de 2018. Recuperado de [https://www.researchgate.net/publication/26847440_Aciertos_y_errores_en_la_traducccion_automatica_metodologia_de_la_ensenanza-aprendizaje_de_la_traducccion_humana](https://www.researchgate.net/publication/26847440_Aciertos_y_errores_en_la_traduccion_automatica_metodologia_de_la_ensenanza-aprendizaje_de_la_traducccion_humana)

Elman, J. L. (1990). Finding structure in time. *Cognitive Science. Volumen 14*, 179-211. San Diego, Estados Unidos: Universidad de California. Recuperado de <http://psych.colorado.edu/~kimlab/Elman1990.pdf>

Ericsson (2018). Ericsson Mobility Report June 2018. *Ericsson*. En línea. Consultado el 25 de junio de 2018. Recuperado de <https://www.ericsson.com/en/mobility-report/reports/june-2018>

Escobar, X. (2009). *Traducción automatizada basada en algoritmos de redes neuronales artificiales*. Tarragona, España: Universidad Rovira i Virgili.

Espiñeira Caderno, S. y Caneda Fuentes, L. (1998). El error y su corrección. En T. Jiménez Juliá (Presidencia). Actas del IX Congreso Internacional de ASELE: Español como lengua extranjera: enfoque comunicativo y gramática. Congreso llevado a cabo en Santiago de Compostela, pp. 473-481. En línea. Consultado el 29 de mayo de 2018. Recuperado de: https://cvc.cervantes.es/ensenanza/biblioteca_ele/asele/pdf/09/09_0476.pdf

Finge, R. (2017, 19 de abril). Grote verbetering voor het Nederlands in Google Translate. *Google*. En línea. Consultado el 25 de junio de 2018. Recuperado de <https://nederland.googleblog.com/2017/04/grote-verbetering-voor-het-nederlands.html>

González, A. (2017, 19 de noviembre). Así funciona el traductor de Siri en iOS 11. *Actualidad iPhone*. En línea. Consultado el 30 de marzo de 2018. Recuperado de <https://www.actualidadiphone.com/asi-funciona-traductor-siri-ios-11/>

Google (2010, 22 de agosto). Inside Google Translate. *Google Translate*. En línea. Consultado el 25 de junio de 2018. Recuperado de https://web.archive.org/web/20100822193710/http://translate.google.com/about/intl/en_ALL

Google Cloud (2018). API Translation. *Google Cloud*. En línea. Consultado el 25 de junio de 2018. Recuperado de <https://cloud.google.com/translate/?hl=es>

Google Play (2018). Traductor de Google. *Google Play*. En línea. Consultado el 25 de junio de 2018. Recuperado de <https://play.google.com/store/apps/details?id=com.google.android.apps.translate>

Google Traductor (2018). Traductor, explora el mundo en más de 100 idiomas. *Google*. En línea. Consultado el 25 de junio de 2018. Recuperado de https://translate.google.com/intl/es_ALL/about/languages/

Google Translate Community (2018). Google Translate Community FAQ. *Google Translate Community*. En línea. Consultado el 25 de junio de 2018. Recuperado de https://docs.google.com/document/d/1dwS4CZzgZwmvoB9pAx4A6Yytmv7itk_XE968RMiqpMY/pub

Hernández Mercedes, P. (2002). Entorno a la traducción automática. *Cervantes*. N°2, 101-117.

- Hurtado Albir, A. (1999). *Enseñar a traducir*. Madrid, España: Edelsa.
- Hurtado Albir, A. (2001). *Traducción y Traductología. Introducción a la Traductología*. Madrid, España: Cátedra.
- Hutchins W.J. y Somers H.L. (1995). *Introducción a la traducción automática*. Madrid, España: Visor. Traducción dirigida por Joseba Abaitua.
- Hutchins, W. J., y Somers, H.L. (1992). *An introduction to machine translation*. Londres, Reino Unido: Academic Press.
- IEMD (2016, 18 de noviembre). ¿Qué es el traductor Google Translate y para qué sirve? *IEMD, Instituto Internacional Español de Marketing Digital*. En línea. Consultado el 25 de junio de 2018. Recuperado de <https://iiemd.com/blog/es-traductor-de-google/que-es-google-translate-traductor-de-google>
- Intertext (2017, 29 de noviembre). ¿Cómo afecta a la traducción la inteligencia artificial? *Intertext*. En línea. Consultado el 29 de marzo de 2018. Recuperado de <https://www.intertext.es/como-afecta-a-la-traducion-la-inteligencia-artificial/#inteligencia>
- Jiménez Martín, P. y Sánchez Allende, J. (2015). De Eliza a Siri: La evolución. *Tecnología y desarrollo*, volumen XIII. Madrid. En línea. Consultado el 12 de mayo de 2018. Recuperado de https://revistas.uax.es/index.php/tec_des/article/view/616/572
- Johnson, M., Schuster, M., Le, Q.V., Krikun, M., Wu, Y., Chen, Z.,..., Dean, F. (2016, 14 de noviembre). Google's Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation. Google. En línea. Consultado el 25 de junio de 2018. Recuperado de <https://arxiv.org/pdf/1611.04558v1.pdf>
- Jordan, M. I. (1990). Attractor Dynamics and Parallelism in a Connectionist Sequential Machine. En J. Diederich (Ed.). *Artificial neural networks* (pp. 112-127). Nueva Jersey, Estados Unidos: IEEE Press.
- Koehn, P. (2010). *Statistical Machine Translation*. Cambridge, Reino Unido: Cambridge University Press.
- Krauwier, S. (2003). *Aspecten van het automatisch vertalen*. Utrecht: Utrecht Institute of Linguistics OTS.
- Linguaserve (2018). Traducción Automática y calidad bajo demanda: La pregunta no es si la Traducción Automática es útil o no, sino cuándo y cómo lo es. *Linguaserve*. En línea. Consultado el 28 de marzo de 2018. Recuperado de <https://linguaserve.com/tecnologia-multilingue/tecnologia-linguistica/traducion-automatica/>

López Takeyas, B. (2015). *Introducción a la inteligencia artificial*. Nuevo Laredo, México: Instituto Tecnológico de Nuevo Laredo.

Martínez de Sousa, J. (2014). *Ortografía y ortotipografía del español actual*. Gijón, España: Trea.

Martínez, V. (2017, 17 de noviembre). ¿Cómo traducir idiomas con Siri en iPhone y iPad?. *Paréntesis*. En línea. Consultado el 30 de marzo de 2018. Recuperado de <http://www.parentesis.com/tutoriales/Como traducir idiomas con Siri en iPhone y iPad#>

Mazin, C. (2016, 20 de diciembre). Google limite son outil de traduction gratuit au profit de Cloud Translation, payant. *ActuaLitté*. En línea. Consultado el 25 de junio de 2018. Recuperado de <https://www.actualitte.com/article/lecture-numerique/google-limite-son-outil-de-traduction-gratuit-au-profit-de-cloud-translation-payant/68673>

Mendoza García, M.H. (2017). *La posesión de traducciones de textos técnicos del alemán al castellano* (tesis doctoral). Universitat Autònoma de Barcelona: Barcelona, España.

Microsoft (2018). ¿Qué es una traducción basada en redes neuronales? *Microsoft*. En línea. Consultado el 29 de mayo de 2018. Recuperado de <https://translator.microsoft.com/es/help/articles/neural/#>

Microsoft Translator (2016, 15 de noviembre). Microsoft Translator launching Neural Network based translations for all its speech languages. *Microsoft*. En línea. Consultado el 30 de mayo de 2018. Recuperado de <https://blogs.msdn.microsoft.com/translation/2016/11/15/microsoft-translator-launching-neural-network-based-translations-for-all-its-speech-languages/>

Nieto González, A. (2011, 8 de febrero). ¿Qué es Android? *Xataka*. En línea. Consultado el 25 de junio de 2018. Recuperado de <https://www.xatakandroid.com/sistema-operativo/que-es-android>

Nord, C. (1995). El error en traducción: categorías y evaluación. En A. Hurtado Albir (Ed.). *La enseñanza de la traducción* (pp. 91-107). Castellón de la Plana, España: Universitat Jaume I.

Novet, J. (2017, 30 de marzo). Google now provides AI-powered translations for Arabic and Hebrew. *VentureBeat*. En línea. Consultado el 25 de junio de 2018. Recuperado de <https://venturebeat.com/2017/03/30/google-now-provides-ai-powered-translations-for-arabic-and-hebrew/>

Och, F.J. (2005). *Statistical Machine Translation: Foundations and Recent Advances*. Phuket, Tailandia: Google Inc.

Organización de las Naciones Unidas (2018). Una población en crecimiento. *Naciones Unidas*. En línea. Consultado el 25 de junio de 2018. Recuperado de <http://www.un.org/es/globalissues/population/>

Rus, C. (2017, 9 de marzo). Así es el proceso mediante el cual Siri aprende a hablar un idioma nuevo. *Applesfera*. En línea. Consultado el 30 de marzo de 2018. Recuperado de <https://www.applesfera.com/general/asi-es-el-proceso-mediante-el-cual-siri-aprende-a-hablar-un-idioma-nuevo>

Santiago Pérez, I. (2014). Asistentes Personales por Voz: Cortana vs Siri vs Google Now. *Ignacio Santiago*. En línea. Consultado el 14 de mayo de 2018. Recuperado de <https://ignaciosantiago.com/asistentes-voz-windows-cortana-apple-siri-google-now-android/#siri-ios-apple>

Schuster, M., Johnson, M. y Thorat, N. (2016, 22 de noviembre). Zero-Shot Translation with Google's Multilingual Neural Machine Translation System. *Google AI Blog*. En línea. Consultado el 25 de junio de 2018. Recuperado de <https://ai.googleblog.com/2016/11/zero-shot-translation-with-googles.html>

Souppouris, A. (2012, 26 de enero). Siri voice translation comes to the iPhone 4S via Lingual. *The Verge*. En línea. Consultado el 30 de marzo de 2018. Recuperado de <https://www.theverge.com/2012/1/26/2736060/lingual-for-iphone-4s-brings-voice-translation-to-siri>

Systran (2018). ¿Qué es la traducción automática? Traducción automática basada en reglas vs. traducción automática estadística. *Systran*. En línea. Consultado el 28 de marzo de 2018. Recuperado de <http://www.systran.es/systran/tecnologia-de-traduccion/que-es-la-traduccion-automatica/>

Tanner, A. (2007). Google seeks world of instant translations. *Reuters*. En línea. Consultado el 25 de junio de 2018. Recuperado de <https://www.reuters.com/article/us-google-translate-idUSN1921881520070328>

Tertoolen, R. (2010). *El alcance de la traducción automática, un estudio de la traducción automática del par de lenguas español neerlandés* (trabajo fin de máster). Universiteit Utrecht: Utrech, Países Bajos.

Turovsky, B. (2016, 15 de noviembre). Found in translation: More accurate, fluent sentences in Google Translate. *Google*. En línea. Consultado el 25 de junio de 2018. Recuperado de <https://blog.google/products/translate/found-translation-more-accurate-fluent-sentences-google-translate/>

Turovsky, B. (2017, 25 de abril). Making the internet more inclusive in India. *Google*. En línea. Consultado el 25 de junio de 2018. Recuperado de <https://blog.google/products/translate/making-internet-more-inclusive-india/>

Turovsky, B. (2017, 6 de marzo). Higher quality neural translations for a bunch more languages. *Google*. En línea. Consultado el 25 de junio de 2018. Recuperado de <https://blog.google/products/translate/higher-quality-neural-translations-bunch-more-languages/>

Vázquez Ayora, G. (1977). *Introducción a la traductología*. Georgetown, Estados Unidos: Georgetown University Press.

Wasserman, P.D. (1989). *Neural computing: theory and practice*. Nueva York, Estados Unidos: Van Nostrand Reinhold Co.

Yúbal, F.M. (2017, 30 de noviembre). Así es como las máquinas están aprendiendo a traducir sin diccionarios ni intervención humana. *Xataka*. En línea. Consultado el 19 de mayo de 2018. Recuperado de <https://www.xataka.com/robotica-e-ia/asi-es-como-las-maquinas-estan-aprendiendo-a-traducir-sin-diccionarios-ni-intervencion-humana>

Referencias bibliográficas del corpus

Nos gustaría recalcar que las siguientes referencias se encuentran en el Anexo I junto con el texto al que pertenecen, con el fin de facilitar el acceso al corpus en su documento original:

Abel, C. (14 de mayo de 2018). A Valentine to the Best American Poetry 1993. *Ploughshares at Emerson College*. En línea. Consultado el 15 de mayo de 2018. Recuperado de <http://blog.pshares.org/index.php/a-valentine-to-the-best-american-poetry-1993/>

Adelborg, K., Szépligeti, S.K., Holland-Bill, L., Ehrenstein, V., Horváth-Puhó, E., Henderson, V.W. y Sørensen H.T. (2018). Migraine and risk of cardiovascular diseases: Danish population based matched cohort study. *British Medical Journal*, 360, doi: 10.1136/bmj.k96.

AMI210, User Manual, IBASE Technology Inc., Ilkley, West Yorkshire, Reino Unido, 2016.

Apple Inc. (2016). United States Securities and Exchange Commission Form 10-K. En línea. Consultado el 15 de mayo de 2018. Recuperado de <http://investor.apple.com/secfiling.cfm?filingid=1628280-16-20309&cik=320193>

As the global economy picks up, inflation is oddly quiescent. (4 de noviembre de 2017). *The Economist*. En línea. Consultado el 14 de mayo de 2018. Recuperado de <https://www.economist.com/news/finance-and-economics/21730930-central-banks-are-beginning-raise-interest-rates-anyway-global?zid=307&ah=5e80419d1bc9821ebe173f4f0f060a07>

Bekhrad, J. (11 de enero de 2018). The Rubaiyat: History's most luxurious book of poetry?. *BBC*. En línea. Consultado el 15 de mayo de 2018. Recuperado de <http://www.bbc.com/culture/story/20180111-the-rubaiyat-historys-most-luxurious-book-of-poetry>

Biersdorfer, J.D. (7 de mayo de 2018). Restoring a Lost Recycle Bin. *The New York Times*. En línea. Consultado el 15 de mayo de 2018. Recuperado de https://www.nytimes.com/2018/05/07/technology/personaltech/restoring-lost-recycle-bin.html?rref=collection%2Fsectioncollection%2Fpersonaltech&action=click&contentCollection=personaltech®ion=stream&module=stream_unit&version=latest&contentPlacement=2&pgtype=sectionfront

Boffey, D. (14 de mayo de 2018). EU demands action by Poland's government to protect rule of law. *The Guardian*. En línea. Consultado el 15 de mayo de 2018. Recuperado de <https://www.theguardian.com/world/2018/may/14/eu-demands-action-by-poland-government-to-protect-rule-of-law>

Drugs Act 2005. The Stationery Office, Londres, Reino Unido, 7 de abril de 2005.

Fleur, N. St. (14 de mayo de 2018). Tiny Brains of Extinct Human Relative Had Complex Features. *The New York Times*. Recuperado de <https://www.nytimes.com/2018/05/14/science/homo-naledi-brain.html?rref=collection%2Fsectioncollection%2Fscience&action=click&contentCollection=science®ion=rank&module=package&version=highlights&contentPlacement=5&pgtype=sectionfront>

Gibbs, S. (14 de mayo de 2018). Apple facing class action lawsuit over defective MacBook Pro keyboards. *The Guardian*. En línea. Consultado el 15 de mayo de 2018. Recuperado de <https://www.theguardian.com/technology/2018/may/14/apple-class-action-lawsuit-defective-macbook-pro-keyboards>

Gillers, H. (5 de octubre de 2017). Trump Comments Slam Puerto Rico Debt. *Wall Street Journal*. Recuperado de https://psv4.userapi.com/c816735/u448173336/docs/5aada8a3bff5/The_Wall_Street_Journal_10_05_2017.pdf?extra=pBWgljHoTiMOWDvj7LYrBsixUnr6KTOZ7KQW74y6VSEXmUodpYfIONMrScCPJvWPzx-XIHIC8yU7YK0czLC2H1E68XiznQ3D3d7FPqalTRNaf4aRda6pOy15nADQ

Hou, H., Shoji, T., Zangwill, L.M., Moghimi, S., Saunders, L.J., Hasentab, K.,...Weinreb, R.N. (2018). Progression of primary Open-Angle Glaucoma in Diabetic and Nondiabetic Patients. *American Journal of Ophthalmology*, 189 (mayo 2018), 1-9. doi: 10.1016/j.ajo.2018.02.002.

Keyes, M. (2015). "Friday, 30 May". En *The woman who stole my life* (p. 9). Reino Unido: Penguin Books.

Keyes, M. (2015). "Saturday, 31 May". En *The woman who stole my life* (p. 33). Reino Unido: Penguin Books.

Pearce, A. y Burns, A. (25 de junio de 2018). Democrats Are Turning Out to Vote in the Most Competitive Primary Races. *The New York Times*. En línea. Consultado el 25 de junio de 2018. Recuperado de <https://www.nytimes.com/interactive/2018/06/25/us/politics/midterm-primaries-voter-turnout.html?rref=collection%2Fsectioncollection%2Fpolitics&action=click&contentCollection>

Devlin, H. (11 de mayo de 2018). Scientists to grow 'mini-brains' using Neanderthal DNA. *The Guardian*. En línea. Consultado el 15 de mayo de 2018. Recuperado de <https://www.theguardian.com/science/2018/may/11/scientists-to-grow-mini-brains-using-neanderthal-dna>

R.R. Martin, G. (2015). "Arya". En *Game of thrones: A clash of kings* (p. 32). Estados Unidos: Harpercollins Pub.

R.R. Martin, G. (2015). "Davos". En *Game of thrones: A clash of kings* (p. 741). Estados Unidos: Harpercollins Pub.

The Constitution of the United States of America. U.S. Government Publishing Office, Washington, Estados Unidos, 7 de mayo de 1992 (tras la vigésimoséptima y última enmienda).

TMS320F28335, TMS320F28334, TMS320F28332, TMS320F28235, TMS320F28234, TMS320F28232, Digital Signal Controllers (DSCs) Data Manual, Texas Instruments, Texas, Estados Unidos, 2012.

Vodafone Group (2018). Vodafone Group Results for the year ended 31 March 2018. En línea. Consultado el 15 de mayo de 2018. Recuperado de http://www.vodafone.com/content/dam/vodafone/investors/financial_results_feeds/year-ended-31-march-2018/FY_2018_Presentation_DOWNLOAD.pdf