Universitat Autònoma de Barcelona

Trabajo fin de máster

Estudio comparativo entre los procesos de traducción humana y traducción automática estadística con la combinación de lenguas chino-español

Autora:
Andrea Piqueras Herrero

Director:
Adrià Martín-Mor

Trabajo fin de máster en Tradumática:
Tecnologías de la Traducción

Facultad de Traducción e Interpretación

Barcelona, 22 de junio de 2020

Universitat Autònoma
de Barcelona

AGRADECIMIENTOS

Después de muchos meses de dedicación y un largo proceso de trabajo y organización de las ideas, me gustaría agradecer a todas aquellas personas que han intervenido durante la realización de esta investigación.

En primer lugar, quisiera agradecer a mi tutor, Adrià, por todos los consejos y la ayuda que me ha ofrecido durante el proceso de trabajo, ya que sin duda, sin sus aportaciones, esta investigación no sería la misma.

En segundo lugar, me gustaría agradecer al Máster de Tradumática por despertar —aún más— mi interés por las tecnologías aplicadas a la traducción y por todos los conocimientos que me ha brindado y que han ayudado a consolidar esta investigación.

También me gustaría agradecer a los responsables de las plataformas KantanMT y MTradumàtica, por haber hecho posible la parte práctica de esta investigación y por haber ayudado a consolidar mi formación y mis conocimientos en cuanto a la traducción automática.

Por último, me gustaría agradecer a Toni y a mi familia por el apoyo en los momentos de agobio y por su comprensión a lo largo de estos meses.

¡Muchas gracias a todos!

Datos del trabajo fin de máster

Título: *Estudio comparativo entre los procesos de traducción humana y traducción automática estadística con la combinación de lenguas chino-español*

Autora: Andrea Piqueras Herrero

Director: Adrià Martín Mor

Centro: Facultad de Traducción e Interpretación, Universitat Autònoma de Barcelona

Estudios: Máster en Tradumática: Tecnologías de la Traducción

Resumen

El objetivo de la presente investigación es analizar y estudiar el proceso de creación y los resultados de traducción de un motor de traducción automática estadística con la combinación de lenguas chino-español y el proceso de traducción humano para determinar cuál de ellos es más factible a la hora de implementarlo en el ámbito de trabajo. La investigación consta de una descripción detallada de la traducción automática y en concreto de la traducción automática estadística, un seguimiento de ambos procesos y un análisis comparativo de los mismos.

Palabras clave

Traducción automática estadística, TAE, traducción humana, chino, español, posedición, traducción automática, evaluación de TA.

Dades del treball fi de màster

Títol: *Estudi comparatiu entre els processos de traducció humana i traducció automàtica estadística amb la combinació de llengües xinès-castellà*

Autora: Andrea Piqueras Herrero

Director: Adrià Martín Mor

Centre: Facultat de Traducció i d'Interpretació, Universitat Autònoma de Barcelona

Estudis: Màster en Tradumàtica: Tecnologies de la Traducció

Resum

L'objectiu d'aquesta investigació és analitzar i estudiar el procés de creació i els resultats de traducció d'un motor de traducció automàtica estadística amb la combinació de llengües xinès-castellà i el procés de traducció humà per tal de determinar quin dels dos resulta més factible quant a la seva implementació al flux de treball. La investigació consta d'una descripció detallada de la traducció automàtica i en concret de la traducció automàtica estadística, un seguiment d'ambdós processos i una anàlisi comparativa d'aquests.

Paraules clau

Traducció automàtica estadística, TAE, traducció humana, xinès, castellà, postedició, traducció automàtica, avaluació de TA.

Master's thesis information

Title: *Comparative study between human translation and statistical machine translation procedures with the Chinese-Spanish language pair*

Author: Andrea Piqueras Herrero

Director: Adrià Martín Mor

Center: Translation and Interpreting faculty, Autonomous University of Barcelona

Studies: Master's Degree in Tradumatics: Translation Technologies

Summary

The aim of this research is to analyze and study the process of creation and the translation results of a statistical machine translation engine with the Chinese-Spanish language combination, and the process of human translation in order to determine which process is more feasible in the translation workflow. This project includes a detailed description of machine translation, and specifically of statistical machine translation, a follow-through both processes and a comparison of them.

Keywords

Statistical machine translation, SMT, human translation, Chinese, Spanish, postediting, machine translation, MT evaluation.

Índice de contenidos

Índice de figuras

Lista de abreviaturas

TA: Traducción automática

TAE: Traducción automática estadística

NMT: Traducción automática neuronal

RBMT: Traducción automática basada en reglas

TAO: Traducción asistida por ordenador

1. Introducción

En los subcapítulos que se recogen a continuación presentamos la introducción de la presente investigación, recogiendo características acerca de la motivación personal y justificación de la elección del tema, los objetivos, preguntas e hipótesis, la estructura y metodología del trabajo y las cuestiones formales del mismo.

1.1. Motivación personal y justificación

La elección del tema de trabajo surgió a partir de la combinación de dos temas por los que siento un gran interés: la traducción de la lengua china y las tecnologías aplicadas a la traducción. Ciertamente, después de experimentar en primera persona y de forma profesional la traducción con la combinación de lenguas chino-español, me di cuenta de que en comparación a la traducción con otros pares de lenguas, esta combinación es algo más complicada por muchos de los aspectos que intervienen dentro del flujo de traducción —no tan solo a nivel lingüístico, sino que también a nivel de documentación y recursos al alcance del traductor con dicha combinación—.

En este aspecto, y como amante de las tecnologías de la traducción, siempre he sentido interés por la aplicación de las innovaciones tecnológicas al proceso de traducción, en especial con la combinación de lenguas con las que trabajo, ya que considero que si se logra un correcto funcionamiento de los sistemas de traducción automática, la automatización del proceso podría llegar a ser muy beneficiosa para el traductor con esta combinación ya que agilizaría su flujo de trabajo y podría incluso aumentar su productividad. No obstante, considero que debido a las propias particularidades de las lenguas —puesto que se trata de lenguas lejanas, una de ellas ideográfica y la otra alfabética y con distintos funcionamientos en cuanto a la segmentación—, es complicado lograr que la traducción automática produzca unos resultados buenos, y en muchas ocasiones hasta coherentes; además este hecho ya es una realidad en los traductores automáticos existentes.

En este aspecto, y debido a que nunca antes había experimentado el proceso de creación de un motor de traducción automática, decidí llevar a cabo la presente investigación con tal de estudiar el procedimiento de traducción humana y el tecnológico y determinar si el traductor profesional sería capaz de crear y aplicar de forma autónoma la automatización de la traducción en su ámbito de trabajo para que su productividad se viera beneficiada. Para ello, partí de las inquietudes que me surgieron al trabajar como traductora con esta combinación de lenguas y al cursar algunas asignaturas acerca de la traducción automática en mis estudios de máster.

En este aspecto, consideré que hacer esta investigación sería interesante para poder definir las características de los dos procedimientos de trabajo —el humano y el

automático— con la combinación de lenguas chino-español, con tal de extraer conclusiones acerca de qué procedimiento puede llegar a ser factible o accesible de cara a la aplicación profesional para la persona que se dedique a ello. Además, considero que este trabajo puede servir de ayuda a aquellos traductores con la combinación de lenguas mencionada que sientan interés por las tecnologías, que quieran aplicarlas a su flujo de trabajo habitual y que quieran cerciorarse de qué procedimiento es el más factible para mejorar su ritmo de trabajo.

1.2. Objetivos, preguntas e hipótesis

En este capítulo detallaremos las inquietudes que nos han surgido tras la elección del tema de trabajo, así como el objetivo principal de la investigación y los objetivos secundarios que nos han ido guiando a cumplir el fin principal del estudio. En primer lugar, nuestro trabajo cuenta con un objetivo muy claro: estudiar de manera comparativa los procesos de traducción humana (sin uso de la TA) y traducción con un motor TAE con la combinación de lenguas chino-español, para determinar cuál de los dos puede ser más ventajoso para el profesional que se dedica a la traducción con dicha combinación lingüística. En este sentido, partiendo de este objetivo principal, nos formulamos una serie de preguntas que conducían a los objetivos secundarios del trabajo:

1. ¿Es posible crear un motor de TAE para poder automatizar el flujo de trabajo de la traducción con la combinación de lenguas chino-español con los materiales disponibles al alcance de cualquier usuario de la red?
2. ¿Qué calidad puede llegar a adquirir el texto traducido por un motor de traducción automática estadística en comparación con la traducción humana teniendo en cuenta los procesos de traducción?
3. ¿En qué medida pueden afectar las características de la lengua china al proceso de entrenamiento y traducción del motor de TAE y al proceso de traducción humana?
4. ¿La calidad de los resultados de traducción del motor de TAE se ve afectada por las diferencias lingüísticas de los pares de lenguas?
5. ¿Qué procedimiento es más factible para el traductor profesional con la combinación de lenguas chino-español?

Todas estas cuestiones de investigación nos llevaron a estructurar el trabajo de manera que nos permitiera ir cumpliendo con nuestros objetivos secundarios que, tras todo el proceso de investigación, nos ayudarían a alcanzar nuestro fin principal. En este aspecto, partiendo de las inquietudes que nos surgieron al plantear el objetivo principal del estudio, a continuación exponemos el listado de objetivos secundarios:

1. Estudiar los procedimientos de creación y entrenamiento de un motor de TAE con la combinación de lenguas chino-español con los recursos al alcance del traductor.
2. Determinar la calidad de una traducción producida por un motor de TAE generado por el propio traductor y la calidad de la traducción humana evaluando todas las fases que se llevan a cabo en ambos procesos.
3. Estudiar en qué medida afecta el hecho que las lenguas de trabajo tengan funcionamientos lingüísticos distintos durante los procesos de traducción humana y de traducción mediante un motor de TAE y cómo afecta este hecho a la calidad final de la traducción.
4. Comparar ambos procedimientos —el humano y el automático— y determinar cuál de ellos puede mejorar el ritmo de trabajo del traductor con la combinación de lenguas chino-español.

Tras tener claras las preguntas de investigación y los objetivos del trabajo, nos planteamos tres hipótesis previas al estudio, y que posteriormente, en el apartado de las conclusiones del trabajo, comentamos si han sido refutadas o no:

1. El proceso de entrenamiento del motor es más largo que el proceso de traducción humana a causa de las diferencias de las lenguas de trabajo.
2. Los resultados de la traducción del motor de TAE no garantizarán una comprensión íntegra y requerirán de una posedición.
3. El uso del motor de TAE podrá llegar a agilizar el tiempo de trabajo siempre que se haya invertido mucho tiempo en su entrenamiento, en especial con la combinación de lenguas chino-español.

En este sentido, después de tener claras las preguntas de investigación y los objetivos que queríamos cumplir, estructuramos el trabajo de manera lógica para poder responder a todas las cuestiones, tratar de cumplir con nuestros objetivos y así, finalmente determinar en qué medida las hipótesis que formulamos al comienzo de la investigación se corresponden a las realidades estudiadas.

1.3. Estructura del trabajo

Esta investigación cuenta con dos grandes bloques que se complementan entre sí: el bloque teórico (capítulos 2, 3 y 4) y el bloque práctico (capítulos 5 y 6).

A partir del segundo capítulo comienza el bloque teórico de este estudio. En este capítulo recogemos algunas de las características directamente relacionadas con la traducción automática, como por ejemplo, la historia de la TA, una clasificación de las distintas tipologías existentes, un subcapítulo acerca de la preedición y la posedición e información acerca de los distintos sistemas de evaluación de la calidad de la traducción automática.

En el tercer capítulo, recopilamos las características de la traducción automática centrándonos en la combinación de lenguas chino-español. Por esta razón, este capítulo está compuesto por varios subcapítulos que recogen información acerca de la situación de la traducción automática tanto en China como en España, y un análisis comparativo entre el chino y el español para determinar las dificultades lingüísticas que se pueden experimentar durante el proceso de traducción automática con esta combinación de lenguas.

En el cuarto capítulo, de manera más detallada, hablamos de nuevo sobre traducción automática, pero en esta ocasión de las características relacionadas con el entrenamiento de los motores de traducción automática estadística. Por ello, este capítulo recoge información acerca de la preparación de los textos originales y de los modelos probabilísticos que intervienen en los motores de traducción automática estadística.

El quinto capítulo encabeza el bloque práctico de este trabajo de investigación. En este capítulo, recogemos el estudio de los procesos prácticos de traducción humana y traducción automática con un motor de TAE. En este caso, dividimos el capítulo en tres bloques; por un lado, recogemos la metodología del procesamiento del estudio práctico en general, justificando cada paso del procedimiento con los principales objetivos de este estudio. Seguidamente, recogemos por una parte un seguimiento detallado del proceso de traducción humana —comentando las distintas fases que intervienen, como la documentación, la traducción y la revisión y control de calidad— y por otra parte, documentamos el procedimiento llevado a cabo durante el proceso tecnológico de creación de un motor de TAE y traducción con el mismo —recogiendo información acerca de la plataforma del motor seleccionada y las distintas fases que intervienen, como el entrenamiento del motor, la fase de *decoding*, el análisis del motor y la tarea de posedición—.

En el sexto capítulo, documentamos de forma general un análisis de los procesos estudiados. Por un lado, comentamos todas aquella características tanto positivas como negativas acerca del proceso de traducción humano y por otro lado, realizamos la misma reflexión en lo que respecta al proceso de traducción automático. Este capítulo ofrece una

visión más global de los aspectos que caracterizan a los dos procedimientos con tal de abrir paso a las conclusiones de la investigación.

En el séptimo capítulo, recopilamos toda las conclusiones extraídas tras realizar la investigación y respondemos a las preguntas previas al estudio que se han recogido en el capítulo 1.2 de la introducción.

El octavo capítulo recoge todas las referencias bibliográficas citadas a lo largo de la presente investigación.

Por último, al final del trabajo, se recogen unos anexos que contienen los resultados de traducción que se han ido generando durante los procesos prácticos; en este sentido, incluye la traducción humana del texto con el que se realiza el estudio, las traducciones generadas por el motor de TAE, una lista terminológica de los términos que eran desconocidos para el motor y un glosario terminológico bilingüe con la traducción de dichos términos.

1.4. Metodología de trabajo

Con tal de cumplir con los objetivos propuestos y resolver nuestras cuestiones de investigación, hemos decidido seguir un orden de trabajo lógico y estructurado. Por este motivo, antes de adentrarnos a realizar el estudio práctico de traducción, consideramos necesario recoger información acerca de la traducción automática, su historia, los distintos tipos de TA existentes y algunas tareas que tienen una relación directa con el mundo de la traducción automática como son la preedición y la posedición y los sistemas de evaluación. Creemos que toda esta información teórica, ayuda a consolidar las bases de conocimiento acerca del mundo de la traducción automática y, que además, nos ayuda a enfocar la parte práctica del trabajo de manera más profesional.

En este aspecto, y debido a que nuestra investigación se centra concretamente en la combinación de lenguas chino-español, para lograr comprender con exactitud todas aquellas características que intervienen en el proceso de traducción automática con dichos pares de lenguas, decidimos recoger las características de la TA tanto de España como en China y un estudio comparativo de ambas lenguas con tal de estudiar las complicaciones lingüísticas que se pueden generar durante la traducción automática del chino al español, para así tener una base de conocimiento más amplia y poder llegar a unas conclusiones tras realizar el proceso práctico que puedan llegar a responder a nuestras preguntas de investigación.

Asimismo, y para consolidar los conocimientos acerca de la traducción automática estadística —que, al fin y al cabo, es el tipo de traducción automática que estudiamos en esta investigación— también hemos considerado importante, recoger en un capítulo toda la

información correspondiente acerca de las distintas características del proceso de entrenamiento de un motor de TAE.

Una vez comprendidos todos los conocimientos teóricos acerca del funcionamiento de la traducción automática y de las características relacionadas con la combinación de lenguas que estudiamos en la presente investigación, realizamos el estudio práctico que responde a las preguntas y que nos ayuda a cumplir con los objetivos que planteamos al principio de este trabajo. En este aspecto, con tal de comprender a la perfección las ventajas y desventajas que presentan los dos procedimientos que se estudian; documentamos todas las fases que intervienen en ambos procesos. Por un lado, por lo que respecta al proceso de traducción humana, recogemos todas las complicaciones y facilidades que observamos en las distintas fases que componen a la actividad de traducción humana; y por otro lado, realizamos el mismo procedimiento con el proceso de traducción tecnológica. En este sentido, consideramos que documentar por separado todos aquellos aspectos positivos y negativos de ambos procesos nos ayuda a posteriormente, —y mediante a la evaluación de las ventajas y desventajas de los dos procedimientos— a llegar a una conclusión en cuanto al rendimiento de cada uno de ellos.

Por último, consideramos importante recoger en un capítulo una visión más general de los puntos positivos y negativos recogidos durante el estudio práctico, para así finalmente tener una idea más concreta de las fortalezas y debilidades de ambos procesos, y poder llegar a unas conclusiones finales que respondan a todas las inquietudes y objetivos de nuestro estudio.

1.5. Cuestiones formales

En cuanto a las cuestiones formales del trabajo, durante las explicaciones de las particularidades lingüísticas de la lengua china, se ha empleado el uso del pinyin entre paréntesis al lado de los caracteres chinos siempre que se ha considerado oportuno para que lector pueda comprender aquellos aspectos que se comentan en cada caso concreto. Asimismo, para facilitar la comprensión de algunas de las explicaciones de las expresiones o términos chinos, se ha facilitado una traducción literal de los mismos, mediante las abreviaturas «lit.» y «Trad. lit.». Asimismo, con el fin de facilitar la lectura al receptor de la presente investigación, después de los índices que encabezan al estudio recogemos una lista con las abreviaturas más frecuentes empleadas a lo largo del trabajo y una breve descripción de las mismas.

En cuanto a la bibliografía, recogemos todas las fuentes citadas durante este trabajo en formato APA, incluidas las fuentes originales en chino, que además cuentan con una transcripción en pinyin y una traducción aproximada de los títulos. Finalmente, cabe

destacar que en el anexo de este trabajo recogemos todos los resultados de traducción —tanto humana como automática— que se han generado durante los procesos prácticos, así como otros materiales extraídos de los distintos procedimientos de traducción y entrenamiento del motor de traducción automática estadística.

2. La traducción automática

La automatización de la traducción ha estado durante mucho tiempo en el punto de mira de la humanidad ya que siempre se ha querido facilitar la tarea de traducción y por ello tiene un gran recorrido a lo largo de la historia. En los siguientes subcapítulos se detallarán algunas de las características generales de la traducción automática, tales como su historia, las distintas tipologías, las tareas de preedición y posedición y los sistemas de evaluación de la calidad, con tal de contextualizar la situación del mundo de la traducción automática.

2.1. La historia de la traducción automática

Tal y como indican en sus libros Hutchins y Somers (1995) y Koehn (2010), ya en el siglo XVII surgió la idea de la creación de los diccionarios mecánicos basados en códigos números que fueran universales. Por aquel entonces, se buscaba la creación de una lengua global lógica e ideográfica —con un aspecto parecido al de la lengua china— con el objetivo de crear una lengua conocida a nivel global y eliminar las barreras de comprensión entre lenguas. Asimismo, siglos más tarde, Petr Smirnov-Troyanskii ideó una propuesta de automatización de la traducción que se asemejaba al funcionamiento que caracteriza a los traductores automáticos de hoy en día. Troyanskii distinguió entre tres fases esenciales dentro del proceso; primero, un humano conocedor de la lengua origen debía analizar el texto de forma lógica y ajustarlo sintácticamente, a continuación, la máquina sería la encargada de encontrar las equivalencias en lengua meta y por último, también de forma humana, el conocedor de la lengua de llegada se encargaría de editar el texto para que resultase natural. A pesar de que en todo este proceso intervenían más los humanos que las máquinas, Petr ya consideró en aquellos tiempos que la primera fase podría llegar a realizarse de forma automática.

Si bien ya se había planteado la automatización de la traducción previamente, no fue hasta después de la invención de los ordenadores que Warren Weaver y Andrew D. Booth plantearon la posibilidad de implementar los traductores automáticos en las máquinas. En Reino Unido, los ordenadores se empleaban para decodificar los códigos de lengua en alemán de la máquina Enigma durante la Segunda Guerra Mundial, siendo el inicio de la impulsión de los traductores automáticos en los ordenadores.

En 1949, Weaver redactó un informe sobre la creación de la traducción automática y propuso varios métodos para su realización, que posteriormente, dio pie a varias investigaciones en los Estados Unidos. Así, pocos años más tarde, en 1954, en un proyecto de Leon Dostert que tuvo lugar en la Universidad de Georgetown se presentó por primera vez en público el funcionamiento y resultado de un traductor automático que traducía del

ruso al inglés. Esta demostración pública fue esencial para el comienzo de nuevas investigaciones tanto en los Estados Unidos como en la Unión Soviética.

Con el paso de los años y al observar que los resultados de la traducción automática presentaban problemas lingüísticos y no llegaban a alcanzar la calidad de la traducción humana, se creó una atmósfera de decepción y desilusión en cuanto a la traducción automática. Entonces, en 1964, se formó un comité llamado ALPAC *(Automatic Language Processing Advisory Committee)* para considerar el futuro de la traducción automática. Este comité afirmó en un informe de 1966 que la traducción automática no tan solo ralentizaba el proceso de traducción, sino que también ofrecía resultados de poca calidad que requerían de posedición y era más costosa que la traducción humana. Así pues, a raíz de su afirmación, la investigación en traducción automática se detuvo en los Estados Unidos por un largo periodo de tiempo. No obstante, en Canadá debido al bilingüismo de francés e inglés y en Europa, a causa de las múltiples lenguas existentes, se mostró un interés por la traducción automática y se continuó la investigación, poniendo en marcha múltiples proyectos, algunos más satisfactorios que otros.

Desde entonces, con el interés resurgido por la investigación en traducción automática, se han realizado muchos avances en este campo. Asimismo, actualmente con la mejora y el incremento de las tecnologías, la traducción automática está alcanzando unos niveles que nunca se habían llegado a concebir cuando se ideó la posibilidad de automatizar la traducción.

2.2. Tipos de traducción automática

Actualmente, existen distintos tipos de traducción automática y se distinguen en base a su funcionamiento. Algunos de estos sistemas se caracterizan por ser más precisos, pero carecen de naturalidad, mientras que otros destacan por su fluidez pero contienen errores semánticos. A continuación detallaremos brevemente las características y funcionamiento de los distintos sistemas más conocidos.

2.2.1. Traducción automática basada en reglas

Uno de los sistemas de traducción automática más arcaicos y precisos es sin duda el sistema de traducción automática basado en reglas —conocido como TABR o en inglés RBMT—. Tal y como mencionan Liu y Zhang (2015), este sistema funciona mediante el uso de reglas creadas por lingüistas expertos que dirigen el proceso de la traducción automática.

La TABR, que alcanzó una gran popularidad en los años ochenta, analiza la lengua de origen y la de llegada, y mediante la creación de una serie de normas lingüísticas que, tal y

como hemos comentado, debe establecer un experto de forma manual, realiza los resultados de traducción. Los sistemas de traducción automática basados en reglas utilizan diccionarios computacionales tanto monolingües como bilingües y gramáticas con reglas morfológicas y sintácticas con tal de derivar de forma lingüística un texto de una lengua a otra. En la tabla que se muestra a continuación (Liu, Q.; Zhang, X., 2015:112), mostramos algunas de las reglas que se pueden crear en un sistema de TABR, clasificadas dentro de las distintas fases del proceso de traducción; análisis, transferencia y generación.

Analysis	Morphological analysis	Source morphological rules
	Syntactic analysis (parsing)	Source grammar
	Semantic analysis	Source semantic rules
Transfer	Lexical transfer	Bilingual lexicon
	Syntactic transfer	Syntactic mapping rules
	Semantic transfer	Semantic mapping rules
Generation	Semantic generation	Target semantic rules
	Syntactic generation	Target grammar
	Morphological generation	Target morphological rules

Figura 1: reglas de un sistema de TABR clasificadas por fases (Liu, Q.; Zhang, X., 2015:112)

En este sentido, en el proceso de creación de un sistema de traducción automática basado en reglas intervienen dos tipos de personas; los expertos informáticos, que se encargan de la creación de los algoritmos y la implementación del sistema, y los expertos lingüistas que deben crear lexicones y reglas gramaticales que posteriormente son empleadas por el sistema para analizar la lengua de origen y generar una traducción.

El sistema de traducción automática basado en reglas, al igual que los otros tipos de TA, presenta ciertas ventajas y desventajas. Tal y como comenta Parra (2018), uno de los principales inconvenientes es que se requiere una gran cantidad de tiempo y de recursos para su creación, ya que tal y como hemos comentado anteriormente, para una única combinación de lenguas se necesitan gramáticas en la lengua origen y meta, diccionarios bilingües y reglas de transferencia. Asimismo, otra de las desventajas que presenta este sistema en relación a la que acabamos de mencionar es la incapacidad de traducir palabras o estructuras lingüísticas que no se encuentren en sus gramáticas, diccionarios o reglas. Además, es imprescindible realizar un mantenimiento continuo cada vez que se pretenda traducir textos con una temática nueva. Por otro lado, algunos de los aspectos positivos de este tipo de sistemas es que ofrecen muy buenos resultados cuando se trata de lenguas cercanas; por ejemplo, el catalán y el español. Igualmente, Parra (2018) también nos comenta que otra de las ventajas de este sistema es que resulta realmente útil y puede llegar a ser una alternativa a los otros sistemas de traducción automática sobre todo en aquellos casos en los que los pares de lenguas no cuentan con suficientes datos como para entrenar otro tipo de motor de TA.

2.2.2. Traducción automática basada en corpus

Por otro lado, dentro de la traducción automática basada en corpus se encuentra la traducción automática estadística que es en la que nos centraremos durante el proceso de estudio práctico de nuestra investigación, y que por lo tanto, posteriormente volveremos a comentar con más detalle. Los sistemas de traducción automática estadística —conocidos como TAE o SMT, en inglés— son sistemas que se entrenan mediante corpus bilingües paralelos y alineados de los que la máquina deriva distintas probabilidades estadísticas para ofrecer resultados de traducción.

Este sistema, utiliza tanto un modelo de traducción como un modelo de lengua, que en capítulos posteriores comentaremos con detalle. Tal y como nos indican Liu y Zhang (2015), por un lado, el modelo de traducción se encarga de calcular de forma matemática las probabilidades de que una oración en lengua meta sea la traducción de otra oración en lengua origen, tomando como referencia los textos con los que se ha alimentado al traductor automático estadístico durante el proceso de entrenamiento. Por otro lado, el modelo de lengua se encarga de que la información en lengua meta resulte natural. En general, los modelos de lengua se basan en *n*-gramas, mientras que los modelos de traducción contienen distintos modelos que pueden basarse en diferentes aspectos, por ejemplo: modelos basados en palabras, en oraciones o en la sintaxis (véase capítulo 4.3). Así, como ya hemos comentado, para que el motor pueda ofrecer una traducción con la combinación de lenguas deseada, durante el entrenamiento del mismo se deben emplear corpus monolingües y corpus bilingües paralelos.

En la figura que se muestra a continuación, podemos observar un esquema del proceso de un motor de traducción automática estadística separado por la fase de entrenamiento y la fase de puesta en práctica. Tal y como se observa, los modelos tienen un gran peso en el funcionamiento de este motor.

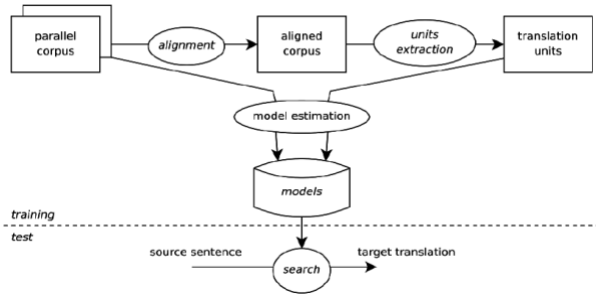


Figura 2: esquema del proceso de entrenamiento y puesta en práctica de un motor TAE (Singla, 2015:4)

Los sistemas de traducción automática basados en corpus también presentan tanto ventajas como desventajas. Tal y como comenta Parra (2018), uno de los aspectos negativos

de estos motores de traducción es que requieren tanto de un corpus monolingüe en la lengua meta como de un gran corpus paralelo en la lengua origen y la lengua meta para su correcto funcionamiento, y dependiendo de la calidad de estos, el sistema ofrecerá resultados de más o menos calidad; por lo tanto, el proceso de selección de estos materiales para su entrenamiento es clave para llegar a alcanzar unos mejores resultados. Además, en relación a este aspecto, otra desventaja de estos sistemas es que la calidad de los resultados de traducción dependerá no tan solo de la calidad de los corpus de los que se haya alimentado, sino que también de la combinación de lenguas empleada y del campo temático de los mismos. Por otro lado, como aspecto positivo es que estos sistemas llegan a alcanzar resultados bastante buenos teniendo en cuenta los factores comentados. Además, tal y como se recoge en el blog de Systran (2020) el hecho de que el entrenamiento de los motores sea automatizado es más rápido y menos costoso en comparación con el sistema comentado en el subcapítulo anterior, el de TABR.

2.2.3. Traducción automática neuronal

Por otro lado, los sistemas de traducción automática neuronal —conocidos como TAN o NMT en inglés— son los últimos sistemas de TA que han aparecido y tienen un funcionamiento distinto a los traductores automáticos comentados previamente. Estos sistemas, tal y como comenta Forcada (2017), nacen partiendo de la idea de la traducción automática basada en corpus; ya que también necesitan grandes corpus paralelos durante el proceso de entrenamiento, pero emplean otro enfoque distinto: el de las redes neuronales.

Tal y como recogen Casacuberta y Peris (2017) en su artículo, lo que caracteriza a la traducción automática neuronal es que las palabras y oraciones se representan de forma numérica mediante vectores y eso permite el uso de redes neuronales. En este sentido, el funcionamiento de las redes neuronales del sistema de TAN es muy parecido al de las neuronas biológicas del cerebro humano; de hecho, de ahí parte su denominación. Este sistema funciona de tal manera que dentro de la red neuronal intervienen un conjunto de unidades de procesamiento que se enlazan entre sí y que cuentan con vectores de distintos pesos, que tras las entradas y salidas directas e indirectas de las neuronas, ofrecen un resultado de traducción. En otras palabras, «del mismo modo en que nuestras neuronas reciben información y realizan conexiones entre sí, los componentes del lenguaje se asocian con otra información subyacente para formar asociaciones y generar traducciones» (Parra, 2018).

Al igual que los otros sistemas de traducción automática comentados en los subcapítulos anteriores, el sistema de TAN también presenta ventajas y desventajas. Según recogen Parra (2018) y Koehn y Knowles (2017), estos sistemas requieren de grandes corpus

paralelos para su entrenamiento que incluso deben ser más amplios que los que requiere el sistema de TAE. Asimismo, el método de funcionamiento de estos motores es bastante complejo ya que se basa en redes neuronales artificiales. Por lo que respecta a los resultados de traducción, en ocasiones generan traducciones que carecen de sentido y que por lo tanto son realmente difíciles de poseditar, mientras que en otras ocasiones producen resultados bastante naturales. Sin embargo, este hecho es algo problemático ya que aunque en muchas ocasiones los resultados sean naturales porque la gramática de las oraciones es correcta, la semántica del mensaje puede llegar a ser incorrecta o distinta de la que ofrece la lengua original. En este sentido, estos sistemas en ocasiones tienen la capacidad de formular frases que aparentemente están bien construidas y son muy naturales, pero que en realidad son erróneas en el concepto de sentido. Koehn y Knowles (2017) por su parte destacan que, en especial, los sistemas de traducción automática neuronal suelen ofrecer una peor calidad que incluso llega a sacrificar la adecuación y fluidez en aquellos textos largos que superen las sesenta palabras. No obstante, un aspecto positivo de estos sistemas, tal y como comenta Parra (2018) es que estos motores están dando un paso adelante y ya se están con imágenes o archivos multimedia y ofrecen resultados aceptables.

2.3. Preedición y posedición

La preedición y la posedición son dos tareas que ayudan a mejorar el resultado final de traducción propuesto por el sistema de traducción automática. En este sentido, tal y como comenta Centelles (2013) durante la preedición se realiza un análisis del texto origen y se modifican aquellos aspectos que puedan causar dificultades a la hora de traducir —tales como las ambigüedades, las palabras desconocidas o los signos especiales, entre otros—. Asimismo, en muchas ocasiones, tal y como se indica en el libro de Hutchins y Somers (1995:216), «la preedición puede llegar a contemplar la reformulación del texto mediante un "lenguaje controlado"» pues también es frecuente la reordenación de las palabras para alcanzar una semejanza más parecida a la lengua de llegada y así facilitar el proceso de traducción a la máquina.

Por otro lado, la posedición, como bien indican Hutchins y Somers (1995) también es una fase realmente importante que tiene lugar después de obtener los resultados de traducción de una máquina, siempre y cuando se quiera llegar a alcanzar un alto estándar de calidad —o en ocasiones simplemente para mejorar ligeramente los resultados, con una posedición parcial— y dependiendo de la finalidad de la traducción. En este sentido, la posedición consiste en modificar el texto ya traducido en lengua meta con tal de mejorar su calidad o adaptarlo a la naturalidad del lenguaje en la lengua de llegada. La posedición no se trata de una actividad obligatoria ya que en ocasiones los textos traducidos tan solo tienen

el fin de informar y no van a ser publicados, por eso, no precisan de modificaciones con tal de llegar a alcanzar un nivel de calidad alto o aceptable, ya que su finalidad es transmitir el mensaje. No obstante, en la mayoría de textos que van a ser publicados y han sido traducidos por una máquina, la posedición será necesaria. A la hora de poseditar, en algunas ocasiones se suele adaptar el texto de forma lingüística con tal de acomodarlo a la naturalidad de la lengua meta y que alcance un resultado apto para su publicación. A continuación mostramos un esquema del proceso de preedición y posedición basado en el esquema de Centelles (2013):

Figura 3: esquema de las fases de preedición y posedición en el proceso de TA

2.4. Sistemas de evaluación

Tal y como se recoge en el libro de Hutchins y Somers (1995), los resultados de la traducción automática aún no llegan a ser del todo perfectos y requieren de posedición para llegar a alcanzar un resultado de calidad. Sin embargo, todavía no existen unos parámetros o métodos de evaluación comunes y aceptados por todos. En este sentido, debido a que no hay un método estándar para todos, a la hora de evaluar la calidad se pueden tener en cuenta varios aspectos —tal y como comenta Koehn (2010)— por un lado, podríamos preguntar a un traductor humano acerca de la calidad de la traducción. Sin embargo, este método no es del todo eficiente ya que los traductores humanos pueden tener distintos puntos de vista, y el resultado de traducción para una misma oración realizado por distintos traductores podría variar. Por otro lado, otra opción podría ser la comparación del resultado de la máquina con el resultado del traductor humano, considerando a este último el *golden standard*.

En cualquier caso, para evaluar la calidad de los resultados de traducción automática lo que debe tenerse en cuenta es la funcionalidad de los mismos, es decir, en qué medida ayudan al cliente que necesitaba la traducción a cumplir con sus objetivos. Así pues, a pesar de que actualmente no existe un parámetro universal y aceptado por todos para evaluar la calidad de la traducción automática, sí que existen ciertos métodos de evaluación que analizan distintos aspectos de los resultados de la TA. A continuación explicaremos algunos de los más comunes de entre los muchos que se recogen en el libro de Koehn (2010).

2.4.1. Evaluación manual

Generalmente, la evaluación manual de traducción automática se realiza por un humano que basa su juicio en guías sobre la adecuación y la fluidez, entre otros aspectos. En los últimos años y con el crecimiento de la popularidad de la traducción automática, se han creado varias guías que expresan los parámetros de calidad a tener en cuenta a la hora de realizar una evaluación de la TA. Entre estas guías, dos de las más populares son la guía de TAUS (2020) y la *Multidimensional Quality Metrics* (MQM) (2019). Estas guías son realmente útiles para la evaluación manual ya que sirven como referente al evaluador para poder establecer un juicio objetivo sobre la calidad de los resultados de la máquina. A pesar de que las dos guías son bastante completas y detalladas, la guía de TAUS se enfoca sobre todo en las métricas de calidad de la adecuación y la fluidez, mientras que la MQM recoge también detalles de terminología, diseño, estilo y convenciones locales entre otros aspectos.

Asimismo, en muchas ocasiones, tal y como comenta Koehn (2010) otra de las modalidades de evaluación manual de traducción automática —y también de las más comunes— se basa en el establecimiento de una calificación numérica por parte del evaluador humano según su propio juicio. Sin embargo, puesto que en muchas ocasiones es difícil determinar en qué nivel algo es correcto o incorrecto, los evaluadores humanos se centran sobre todo en la fluidez y adecuación del resultado. En este aspecto, basándonos en la guía de TAUS (2020), entendemos adecuación como la cantidad de significado recogido en los resultados de traducción en referencia al original o a la traducción de referencia determinada como *golden standard;* y entendemos fluidez, como en qué medida es correcta la traducción a nivel gramatical, ortográfico y terminológico y si podría ser comprendida por un hablante nativo.

No obstante, aunque la evaluación de traducción automática se suele centrar en los aspectos más lingüísticos y extralingüísticos de los resultados de la traducción, hay otros matices que se estudian y evalúan, sobre todo cuando el sistema de traducción automática se desea integrar en un ámbito profesional y operacional. En tal caso, tal y como explica Koehn (2010), se evalúan aspectos como la velocidad del sistema o su integración en el ámbito de trabajo, entre otros. En este sentido, la velocidad es un factor importante ya que un sistema rápido puede mejorar la productividad dentro del proceso de trabajo, pero si la calidad producida por la máquina requiere de un largo proceso de posedición, esto puede ralentizar la tarea de traducción y afectar a la producción. Igualmente, la evaluación de la integración del sistema en el ámbito de trabajo también es realmente importante ya que se deben tener en cuenta factores como su utilidad, la puesta en práctica y su facilidad de uso, ya que en muchas ocasiones se dejan de utilizar sistemas de traducción automática porque se considera que resultan complicados de implementar.

2.4.2. Evaluación automatizada

Por otro lado, a pesar de que muchos consideran que la evaluación manual es más precisa y confiable —ya que un experto dedica una gran parte de su tiempo a analizar los resultados y evaluarlos según las métricas de evaluación— es un proceso más largo y costoso que el de la evaluación mediante una máquina. Por este motivo, se ha decidido automatizar el proceso de evaluación y hoy en día existen muchas métricas realizadas por máquinas que evalúan distintos aspectos (por ejemplo, métricas que evalúan la distancia de edición, como la WER, la TER o la PER, métricas que analizan y evalúan la calidad léxico-terminológica como la BLEU o la NIST o métricas que evalúan la precisión y exhaustividad como la METEOR). A continuación comentaremos algunas de las métricas más comunes empleadas de manera profesional según Koehn (2010) y Banerjee y Lavie (2007).

Una de las métricas más populares a día de hoy es la métrica BLEU *(Bilingual Evaluation Understudy)*. Este sistema de evaluación está basado en *n*-gramas, es decir, en secuencias de elementos *n* —por ejemplo, fonemas, sílabas, artículos o palabras— dentro de un texto. En este aspecto, tal y como comenta Koehn (2010) la métrica BLEU realiza una comparación entre la traducción automática y una o un conjunto de traducciones humanas que toma de referencia. La traducción cuyos *n*-gramas coincidan más con los *n*-gramas de la traducción de referencia, recibirá mayor puntuación. Por este motivo, este sistema de evaluación es muy criticado ya que en muchas ocasiones aunque existan traducciones más naturales que otras según el juicio humano, la métrica BLEU las puede considerar peores si cuentan con menos cantidad de *n*-gramas coincidentes con la traducción referente. Tal y como podemos observar en la Figura 4, la traducción del sistema B se consideraría de mejor calidad que la del sistema A teniendo en cuenta el funcionamiento de la métrica BLEU.

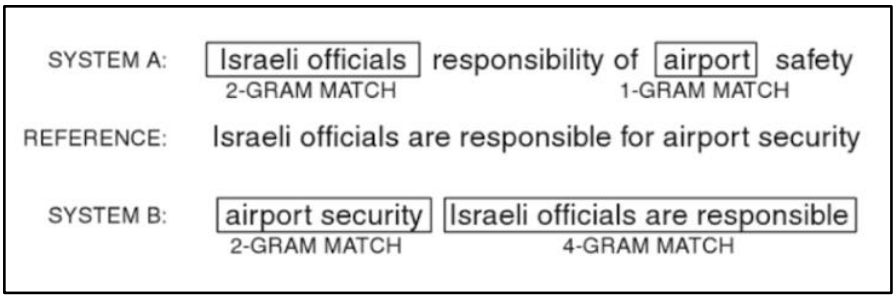


Figura 4: coincidencias de *n*-gramas con la traducción de referencia (Koehn, 2014)

Posteriormente, a raíz de esta métrica, apareció la métrica NIST, ideada por el *National Institute of Standards and Technology,* que funciona de manera muy similar a la BLEU y que también es muy criticada hoy en día por las mismas razones; ya que se basa en *n*-gramas, aplica un porcentaje de error por brevedad de la traducción y por la falta de

coincidencia explícita entre el texto origen y el meta, entre muchos otros aspectos (Banerjee; Lavie, 2007).

No obstante, a pesar de que todavía no existe una métrica automatizada y aceptada por todos, más tarde se creó otro sistema de evaluación automático con tal de mejorar todos aquellos aspectos criticados en las métricas BLEU y NIST, y aunque tampoco es considerado perfecto por todos, tiene un funcionamiento distinto a las métricas anteriores y cuenta con nuevas características; se trata de la métrica METEOR (*Metric for Evaluation of Translation with Explicit Ordering*). Tal y como recogen Banerjee y Lavie (2007), la METEOR se basa en la precisión y en la exhaustividad pero a diferencia de la BLEU, le da más peso a la exhaustividad que a la precisión. Además, esta métrica también tiene en cuenta la coincidencia de sinonimia. La métrica METEOR funciona mediante la alineación de palabras entre las dos oraciones (la traducida y la de referencia), y realiza un estudio de las colocaciones de las palabras —que en inglés se conoce como *mapping*— de modo que cada palabra de la oración traducida tiene una palabra coincidente en la oración de referencia (véase la Figura 5). De manera similar a la métrica BLEU, la METEOR realiza una evaluación de las traducciones en base a coincidencias entre las palabras de la traducción y una traducción de referencia. No obstante, si existe más de una traducción de referencia, se compara el resultado de la traducción automática con todas las referencias existentes y finalmente la puntuación que se establece para la traducción automática, es aquella con el resultado más alto, es decir, la que ha obtenido más coincidencias con la traducción de referencia. En la figura que aparece a continuación se puede observar el proceso de estudio de las colocaciones de las palabras (*mapping*) realizado por la métrica METEOR.

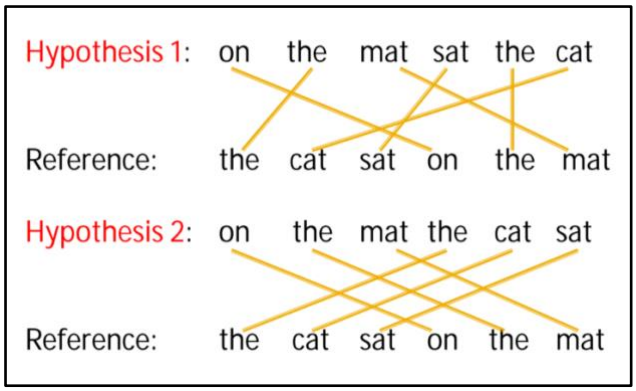


Figura 5: fase de alineación (*mapping*) con la métrica METEOR (Banerjee; Lavie, 2005:26)

3. Características de la traducción automática chino- español

El español y el chino son lenguas realmente diferentes tanto en aspectos lingüísticos como extralingüísticos. A pesar de que la cultura tiene una gran influencia en la lengua y las convenciones de lo correcto y lo incorrecto varían mucho entre el español y el chino, en estos subcapítulos nos vamos a centrar en aquellas características y diferencias entre las dos lenguas que pueden llegar a afectar al resultado de traducción de la máquina. Estos aspectos ya se han estudiado anteriormente, ya que las diferencias entre las lenguas pueden llegar a afectar incluso al propio funcionamiento de las máquinas, tal y como comenta Fu (1999:87):

> *Indo-European and Chinese languages differ from each other enormously in terms of morphology, syntax and semantic expressions. Therefore many common strategies and algorithms already developed in IndoEuropean MT systems can not be adopted directly for the system involved in Chinese. It is for this reason that Chinese scholars have been obliged to study widely the linguistic theories, and develop MT technologies appropriate to the special characteristics of Chinese language.*

En los subcapítulos que aparecerán a continuación explicaremos la situación de la traducción automática en China y en España y algunas de las dificultades que pueden surgir durante el proceso de la traducción automática en relación a las diferencias entre la lengua china y el español.

3.1. La traducción automática en China

Tal y como recoge Duoxiu (2015) la investigación en traducción automática en China empezó en 1956 y tres años más tarde ya se hicieron públicos sus logros. El primer sistema de TA que se consolidó traducía veinte oraciones con distintas estructuras semánticas del ruso al chino, y como todavía no existían dispositivos que pudieran mostrar los caracteres chinos, los resultados de traducción al chino se mostraban codificados.

Desde primer momento, a principios de los años sesenta, la investigación en traducción automática se incluía en las directrices del gobierno chino para el desarrollo científico y por ello se le daba mucha importancia y se realizaron muchos estudios que finalizaron a mediados de los sesenta y finales de los setenta debido a problemas políticos y sociales causados por la Revolución Cultural (1966-1976).Posteriormente, en China ya se consolidó la idea de que la traducción automática iba a ser muy factible sobre todo de cara al futuro, y tan pronto como la situación en el país mejoró, se retomaron las investigaciones y las tecnologías de la traducción comenzaron a presentar un crecimiento muy rápido. A partir de 1980, se realizaron varios seminarios acerca de la traducción automática en China, que tuvieron lugar en distintas partes del país y a raíz del interés en el ámbito surgieron dos de los

sistemas de TA claves en aquel momento; el sistema de traducción automática inglés-chino KY-1, que apareció en 1987 y el IMT/EC-863 que contaba con la misma combinación de lenguas que el sistema mencionado y que se consolidó a mediados de los noventa. En estos seminarios, tal y como recoge Fu (1999) se comentaban varios aspectos como el diseño del traductor automático, los análisis lingüísticos, los resultados de traducción y la evaluación de la traducción automática.

A partir de dichas reuniones, se crearon otros muchos proyectos con otras combinaciones lingüísticas, como francés-chino, japonés-chino, alemán-chino y ruso-chino. A finales de los años ochenta, en concreto en 1988, se lanzó al mercado el primer sistema de traducción automática inglés-español llamado TranStar (Fu, 1999).

En la actualidad, en China cada vez se están realizando más investigaciones acerca de la traducción automática, y los estudios se basan en distintas combinaciones lingüísticas y en distintos campos. Sin embargo, los resultados de traducción todavía no son perfectos y en este sentido, tal y como comenta Fu (1999) la investigación y el interés por la traducción automática continuará llevándose a cabo en China.

3.2. La traducción automática en España

La traducción automática en España, al igual que en China, también tiene un largo recorrido. Tal y como recoge Abaitua (1999), la investigación en traducción automática en España ha experimentado momentos de altibajos durante el transcurso de su historia.

En 1987 comenzó la investigación y desarrollo de la TA en España a causa de la importancia que adquirió la lengua española a nivel internacional y el interés de las empresas informáticas por incluir esta lengua en sus prototipos de TA. A raíz de estos hechos, se comenzaron a emprender proyectos de investigación en España con la ayuda de varias empresas e instituciones europeas. A finales de 1986, en Barcelona y Madrid se consolidaron grupos de investigación por parte de IBM y SIEMENS; en Madrid, en concreto se estudiaban los aspectos de morfología y lexicografía, mientras que en Barcelona, se investigaba acerca de la sintaxis y la semántica. Asimismo, posteriormente, por parte de la empresa Fujitsu se desarrolló un equipo de investigación en Barcelona para implementar el español en el sistema japonés ATLAS.

En este contexto, se crearon varios grupos de investigación con tal de estudiar la traducción automática. Sin embargo, tal y como destaca Abaitua (1999), la importancia de la TA se vio reflejada entre los años 1987 y 1991 por parte de la Sociedad Española para el Procesamiento del Lenguaje Natural, ya que más de un tercio del total de los artículos de su revista trataban sobre traducción automática. Asimismo, también destaca que ya en 1989 el presupuesto anual de la investigación en traducción automática era muy elevado. No

obstante, como hemos comentado anteriormente, la trayectoria de la traducción automática en España presenta altibajos, y después de esta época de esplendor, el interés en la investigación de TA disminuyó considerablemente, a causa de que los sistemas de traducción automática no ofrecían resultados tan buenos como se esperaba. Así, en 1991, poco a poco los grupos de investigación fueron desapareciendo y se perdió el interés por la traducción automática.

No obstante, en el año 1995, tal y como comenta Abaitua (1999) con el emergente crecimiento de la globalización de la economía, el gran impacto que causó el internet y la gran demanda de la localización de software y el multilingüismo, las tecnologías de la traducción volvieron a estar en el punto de mira de la sociedad, y el interés en traducción automática resurgió. Con esto, poco a poco se volvieron a poner en marcha las investigaciones y hasta día de hoy, la traducción automática es uno de los temas de interés de muchos investigadores, y poco a poco se están llevando a cabo mejoras e innovaciones dentro de este ámbito.

3.3. Análisis comparativo del chino y español: dificultades en la TA

Tal y como nos comenta Zhou (1995) las lenguas del mundo se dividen en distintas familias, y el chino y el español pertenecen a grupos distintos. Por un lado, el español es una lengua indoeuropea, mientras que el chino se considera perteneciente al grupo de las chino-tibetanas. En este aspecto, es fácil observar algunas de las diferencias que las caracterizan en especial de su funcionamiento y aspecto ya que nacen de raíces totalmente distintas. Si bien es cierto que una de las principales diferencias entre el chino y el español es que el chino es una lengua ideográfica —es decir, que funciona con caracteres que expresan ideas y no letras— y el español es una lengua alfabética, que mediante un abecedario compone palabras con distintas categorías gramaticales, este no es el único problema que puede llegar a afectar al resultado de la traducción automática con esta combinación de lenguas. Tal y como recoge Centelles (2013) la lengua china y la española también difieren en muchos otros aspectos, como por ejemplo en la fonética, la morfología, la gramática, la semántica o la pragmática. Por este motivo, a continuación comentaremos algunos de los aspectos lingüísticos que son diferentes en chino y español, divididos en diferentes subcapítulos y que hemos considerado que podrían afectar al resultado de traducción de la TA, basándonos en el estudio comparativo del chino y español de Zhou (2015) y en las características de traducción de un motor TAE con la combinación chino-español según Centelles (2013).

3.3.1 Aspectos del funcionamiento de la lengua

Tal y como comentábamos al principio del capítulo, el hecho de que la lengua china sea ideográfica y la española alfabética es un factor condicionante a la hora de traducir. En chino, tal y como recoge Zhou (2015) en su estudio, cada carácter es un concepto o idea y en muchas ocasiones se forman nuevos términos mediante la unión de dos ideas separadas. Por ejemplo, el término «China», como país, se denomina 中国 (zhōngguó), pero si analizamos los dos caracteres que lo componen, 中 (zhōng) significa «central, o en el medio» y 国 (guó) significa «país». En este aspecto, los caracteres que componen el término «China» se pueden utilizar de forma aislada en otros contextos y adoptan otros significados. En los ejemplos que aparecen a continuación observamos el término «中国» y sus respectivos caracteres componentes en distintos contextos;

(1) 长城是**中国**的象征。

La Gran Muralla es un símbolo de **China**.

(2) 她站在房间**中**央。

Ella estaba de pie **en medio** de la habitación.

(3) 他为**国**牺牲了。

Él se sacrificó por su **país**.

Tal y como podemos observar, en los tres ejemplos los caracteres adoptan un uso semántico distinto. En este aspecto, durante el proceso de traducción automática estadística es posible que la máquina traduzca con un sentido diferente si no se ha realizado una tokenización correcta previamente, ya que, tal y como también podemos observar en los ejemplos, otra de las diferencias entre el chino y el español es que el chino no hace una separación de las unidades de la oración mediante espacios y en muchas ocasiones es complicado determinar si los caracteres funcionan aislados o juntos dentro de una oración. Este hecho se trata de una característica que el traductor humano y conocedor de la lengua china sabe determinar, pero la máquina debe entenderlo mediante la tokenización.

3.3.2. Aspectos morfológicos: flexión de género y número

Tal y como recoge Centelles (2013) la morfología china es mucho más simple que la española ya que en chino no existe la flexión de número ni de género y tampoco existen los artículos definidos. Entonces, para determinar la correcta traducción en cuanto al género y al

número, el traductor debe guiarse por el contexto del texto completo y eso puede suponer una dificultad al traductor automático. En el ejemplo que presentamos a continuación mostramos la ambigüedad de sentido en la traducción en una oración sin contexto.

(1) 学生快毕业了。

Posibles traducciones;
El estudiante está a punto de graduarse.
La estudiante está a punto de graduarse.
Los estudiantes están a punto de graduarse.
Las estudiantes están a punto de graduarse.

Tal y como podemos observar, en este caso el sujeto es «学生» y hace referencia a la persona o personas que estudian. Como podemos observar en el estudio de Zhou (2015) a pesar de que el chino no cuenta con una flexión de género o número, sí que define el género y el número en los pronombres personales (yo, tú, él o ella, nosotros, vosotros y ellos). Por este motivo, en muchas ocasiones, al observar el contexto podemos determinar qué tipo de flexión emplear en nuestra traducción. Sin embargo, los sistemas de traducción automática estadística cuando traducen no tienen en cuenta el contexto íntegro del texto. Por lo tanto es posible que el resultado de traducción de un sistema TAE con la combinación de lenguas chino-español presente incoherencias en cuanto a la flexión de género y número en su resultado de traducción.

3.3.3. Aspectos verbales: las conjugaciones

Tal y como nos comenta Centelles (2013), otro de los aspectos que pueden llegar a causar problemas en el resultado de traducción de la máquina es la conjugación de los verbos. Mientras que en español existen múltiples conjugaciones para el pasado, presente y futuro, en chino no se conjugan los verbos. En este aspecto, para expresar el tiempo en el que sucede una acción, el traductor debe observar el contexto y emplear un tiempo u otro en la traducción en función de los adverbios o marcas temporales propias del chino que aparezcan en las frases (por ejemplo: 昨天, *zuótiān*, ayer; 明天, *míng*tiān, mañana; 了, *le*, marca temporal que indica que algo ya ha terminado). Por eso, la falta de conjugaciones en chino puede complicar la tarea al traductor automático en cuanto a la traducción de los tiempos verbales y eso podría provocar que el resultado de traducción en español presentase incoherencias. A continuación mostramos un ejemplo con posibles traducciones:

(1) 我买了苹果 。

(lit.: Yo comprar terminar manzana.)
Posibles traducciones;
He comprado manzanas.
Compré manzanas.

(2) 我明天去买苹果。

(lit.: Yo mañana ir comprar manzana.)
Posibles traducciones;
Mañana compraré manzanas.
Mañana iré a comprar manzanas.

3.3.4. Aspectos sintácticos: estructura de la oración

Otro de los aspectos lingüísticos diferente en chino y español, tal y como recoge Centelles (2013) es que pesar de que la información en chino se suele colocar en el mismo orden que en español, con la estructura SVO (Sujeto-Verbo-Objeto), a veces, en chino se tiende a colocar el verbo a final de oración con el objetivo de crear un énfasis en la acción, y eso puede causar que el resultado de traducción del sistema de TA sea poco natural en español. Además, tal y como comenta Zhou (2015) en las oraciones subordinadas de relativo en chino, la posición de los elementos dentro de la oración es muy diferente a la de la lengua española y este hecho también podría condicionar a la calidad del resultado final del traductor automático. A continuación mostramos un ejemplo de oración subordinada y su respectiva traducción al español.

(1) 我没有你昨天在巴塞罗那的书店买的书。

(lit.: Yo no tengo tú ayer en Barcelona librería comprar libro.)
Yo no tengo el libro que compraste ayer en la librería de Barcelona.

En estos casos, tal y como se puede observar en el ejemplo, la traducción también puede suponer un gran reto ya que es imprescindible realizar un gran reordenamiento de las ideas para que la información se transmita de forma clara y natural en la lengua de llegada, y muchas veces es una tarea difícil incluso para el traductor humano.

3.3.5. Aspectos gramaticales: los pronombres

Finalmente, otro de los aspectos importantes a tener en cuenta durante la traducción automática, y relacionado con el estilo, es el uso de los pronombres (Centelles, 2013). En chino, al igual que en inglés, los pronombres siempre deben aparecer en la oración ya que si no aparecen se trata de un error gramatical. Sin embargo, en español en muchas ocasiones, su repetido uso puede llegar a parecer redundante y se tiende a hacer una omisión de los mismos —por ejemplo en español tanto «Yo me llamo Andrea» como «Me llamo Andrea» son oraciones correctas, mientras que en chino los pronombres personales siempre deben aparecer, «我叫安德丽» (Yo me llamo Andrea)—. En este aspecto, teniendo en cuenta que el traductor automático estadístico basa sus traducciones en grandes corpus de los que se ha alimentado previamente y no es capaz de tener en cuenta los aspectos estilísticos, es posible que sus resultados de traducción resulten redundantes y poco naturales en español.

Asimismo, en chino los pronombres personales funcionan también como posesivos (por ejemplo, 我, *wǒ*, significa «yo» y «mi»). En este aspecto, el hecho de que no haya una distinción de significado puede llegar a afectar al resultado de la traducción automática, ya que el sistema de traducción podría hacer una mala interpretación de su semántica y ofrecer una traducción incorrecta que afectase al sentido y la naturalidad de la traducción. A continuación mostramos dos ejemplos en los que el carácter « 我 » adopta diferentes funciones.

(1) **我**毕业于翻译和口译。

Yo soy graduada en Traducción e Interpretación.

(2) **我**父亲明天将去北京。

Mi padre irá a Pekín mañana.

4. Entrenamiento de un motor de traducción automática estadística

Como se ha comentado en capítulos anteriores, la presente investigación se centra en los motores de traducción automática estadística. Anteriormente, hemos observado las características principales de estos motores, y en los subcapítulos que se presentan a continuación comentaremos detalladamente todas las partes del proceso de entrenamiento de un motor de TAE.

4.1. Preparación de originales

Durante el proceso de entrenamiento de un motor de traducción automática estadística, uno de los pasos principales del proceso es la preparación de los textos originales. Tal y como indican Martín-Mor y Piqué (2017) en su artículo en referencia al motor MTradumàtica, algunos de los procesos que llevan a cabo durante la preparación de originales son la segmentación o tokenización y el *truecasing*. A continuación comentaremos en qué consisten dichos procedimientos.

4.1.1. La segmentación

La segmentación, en otras ocasiones también conocida como tokenización, se trata de la separación de las palabras y los signos de puntuación mediante espacios, con tal de aumentar las probabilidades de que se den coincidencias con los textos que se traducirán de manera automática (Martín-Mor; Piqué, 2017). Tal y como nos comenta Koehn (2010), en el caso de las lenguas que cuentan con un alfabeto latino, la segmentación se basa principalmente en la separación mediante espacios entre las palabras y los signos de puntuación. Sin embargo, en lenguas que por naturaleza no realizan una separación entre las unidades de la oración —como es el caso de la lengua china—, el proceso de segmentación será más complejo. En este sentido, cuando se realiza la segmentación de un texto durante el entrenamiento de un sistema de traducción automática, el objetivo principal es que la máquina distinga los términos de la misma manera vayan o no precedidos de un signo de puntuación. Es decir, basándonos en el ejemplo que proporciona Koehn (2010), la segmentación es necesaria para hacer entender a la máquina que la traducción del término «house», por ejemplo, no tendría que ser diferente si este va seguido de una coma (*house,)*. De manera que si separamos la coma del término mediante un espacio, el sistema de traducción automática no considerará que «house» y «house,» son términos distintos.

Actualmente, existen muchos programas y plataformas de creación de motores de traducción automática que llevan integrados unos sistemas de segmentación propios, como es el caso de KantanMT, por ejemplo, y que comentaremos en capítulos posteriores (véase subcapítulo 5.3.1).

4.1.2. *Truecasing*

Por otro lado, el truecasing tal y como lo definen Martín-Mor y Piqué (2017) consiste en determinar la caja más probable de cada palabra; es decir, mayúsculas o minúsculas. En este sentido, tal y como indica Koehn (2010) empleando el mismo ejemplo anterior, el *truecasing* tiene como objetivo hacer entender a la máquina que el uso de las mayúsculas o minúsculas no provoca que los términos adopten distintos significados. Es decir, los términos «house», «House» y «HOUSE», que podrían aparecer en distintas posiciones —por ejemplo, en medio de una oración, a comienzo de oración o en un enunciado, respectivamente— no tienen distintos significados por el simple hecho de que aparecen con cajas distintas. En este aspecto, en los procedimientos de preparación de originales en general se tiene a realizar el *truecasing* o *lowercasing*, es decir, se colocan todos los elementos de la oración en minúscula para que el sistema de traducción automática los entienda de la misma manera sin distinguirlos por la caja en la que estén escritos.

4.2. Los modelos probabilísticos de un motor TAE

Tal y como se ha mencionado en capítulos anteriores, el funcionamiento de un motor de traducción automática estadística se basa en la combinación de varios modelos probabilísticos que evalúan distintos aspectos y así modelan los resultados de traducción. En los subcapítulos que aparecen a continuación comentaremos algunos de los modelos más importantes en la traducción automática estadística y sus respectivas características.

4.2.1. El modelo de lengua

Por un lado, tal y como indica Koehn (2010), el modelo de lengua es un componente esencial en los sistemas de traducción automática estadística, puesto que es el que se encarga de estimar cuán probable es que una secuencia de palabras sea natural en la lengua de llegada, ya que para lograr un buen resultado, no solo es importante que la traducción recoja el significado de el texto original, sino que también idealmente debe mantener una fluidez en la lengua de llegada. En este sentido, continuando con el ejemplo que proporciona Koehn (2010), el modelo de lengua no tan solo se centra en la fluidez de la lengua de llegada, sino

que también tiene en cuenta las posiciones y traducciones de las palabras; por ejemplo, considerando que nuestra lengua de llegada fuera el español, el modelo de lengua consideraría más correcta la oración «la casa es pequeña» frente a «pequeña es la casa». En este aspecto, el modelo de lengua, basándose en distintos corpus monolingües, consideraría que la primera opción es más correcta que la segunda y por lo tanto le asignaría una probabilidad más alta que a la segunda.

Este modelo no tan solo se encarga de que la colocación de las palabras dentro de una oración resulte natural en lengua meta, sino que también se encarga de la selección de palabras en contextos concretos para asegurar que la fluidez de la lengua de llegada es correcta. De nuevo, basándonos en el ejemplo que proporciona Koehn (2010), por ejemplo, la palabra alemana "Haus" (casa) cuenta con distintas traducciones en inglés (p.e. house, home…) y en este aspecto, el modelo de lengua se encargaría de determinar que palabra es más probable dentro de un contexto determinado. Por ejemplo, entre las oraciones "I am going home" y "I am going house", el modelo asignaría un nivel de probabilidad mayor a la primera oración ya que es la manera más correcta de expresar la información en la lengua meta. En general, tal y como indicamos en subcapítulos anteriores (véase subcapítulo 2.2.2) y como indica Centelles (2013) los modelos de lengua se basan en n-gramas para realizar las estadísticas probabilísticas comentadas.

4.2.2. El modelo de traducción

El modelo de traducción, tal y como indican Liu y Zhang (2015), se encarga de modelar la probabilidad de que una oración en lengua meta (*T*) sea la traducción de una oración dada en lengua origen (*S*).

$$P(T \mid S)$$

Este modelo, junto con otros doce modelos más específicos, está compuesto ayuda modelar las probabilidades de traducción teniendo en cuenta distintos aspectos y en base a diferentes unidades de lengua. A continuación comentaremos tres de los modelos más conocidos que se encuentran dentro del modelo de traducción según Liu y Zhang (2015):

Por un lado, los modelos basados en palabras (*Word-based models,* en inglés) se encargan de calcular las probabilidades de traducción centrándose en las traducciones por palabras. Estos modelos tan solo se focalizan en las probabilidades de traducción a nivel del vocablo, teniendo en cuenta la palabra dada en lengua origen y su posible traducción. Tal y como recogen Yang y Min (2015), los primeros sistemas de traducción automática estadística funcionaban con modelos basados en palabras en los que la unidad de traducción básica era la palabra. En este aspecto, la idea básica de estos modelos es estimar la longitud

de la traducción en palabras, determinar las traducciones correspondientes a las palabras del original y encontrar las equivalencias adecuadas. No obstante, cada decisión se toma mediante cálculos probabilísticos; y aquella traducción que contenga la puntuación más alta, es la que se ofrecerá como resultado. Asimismo, cada decisión se corresponde a un sub-modelo dentro de los modelos basados en palabras, cuyos parámetros son calculados partiendo de un corpus paralelo de forma automática (Yang; Min, 2015). En la figura que aparece a continuación, observamos un ejemplo de traducción con el modelo basado en palabras en la combinación de lenguas chino-español:

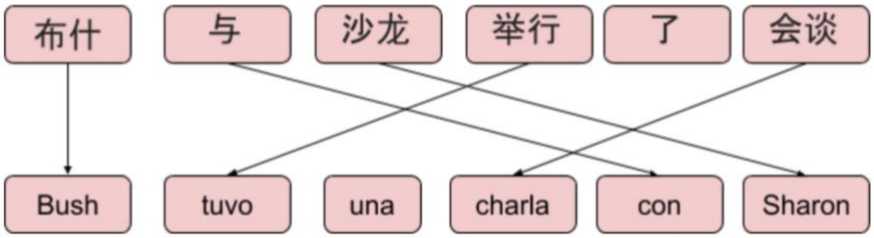


Figura 6: ejemplo de traducción chino-español con el modelo basado en palabras (Yang; Min, 2015;203; elaboración propia)

Por otro lado, también existen los modelos basados en oraciones (*Phrase-based models*, en inglés), y se encargan de calcular las probabilidades de que una oración en lengua meta sean la traducción de la oración en lengua origen proporcionada. Estos modelos, al contrario que los modelos basados en palabras, pueden captar el contexto local a la hora de traducir las palabras. Tal y como indican Yang y Min (2015) una oración —para los modelos basados en palabras— es en general un conjunto de palabras consecutivas, es decir, no tiene por qué ser una oración en su sentido sintáctico. En este aspecto, estos modelos son capaces de tratar con expresiones idiomáticas, adiciones y omisiones de palabras. De la misma manera que con los modelos basados en palabras, también existen sub-modelos dentro de estos —por ejemplo, modelos de segmentación de oraciones (*phrase segmentation*, en inglés), modelos de reordenación de oraciones (*phrase reordering*, en inglés) o modelos de traducción de oraciones (*phrase translation*, en inglés)—. En la figura 7 observamos un ejemplo de un modelo basado en oraciones con la combinación de lenguas chino-español que hemos creado partiendo del ejemplo de Yang y Min (2015).

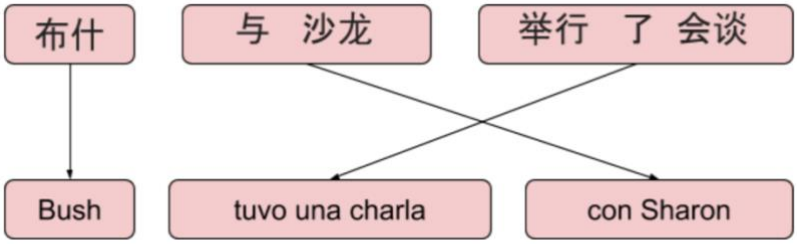


Figura 7: ejemplo de traducción chino-inglés de un modelo basado en oraciones (Yang; Min, 2015:205; elaboración propia)

Finalmente, otro de los modelos que se encuentran dentro del modelo de traducción son los modelos basados en la sintaxis (*Sintax-based models*, en inglés). Estos modelos, tal y como comentan Liu y Zhang (2015) están basados en gramáticas sincronizadas, que cuentan con una serie de normas. En este aspecto, estos modelos se encargan de calcular las probabilidades dentro de las normas sintácticas sincronizadas. Esto es un punto positivo ya que los modelos basados en la sintaxis pueden comprender mejor las dependencias entre las palabras a larga distancia en una misma oración, y así ofrecer mejores resultados que los modelos basados en oraciones en especial en combinaciones de lenguas que cuentan con estructuras sintácticas bastante distintas, como es el caso del chino y el español, por ejemplo. Tal y como comentan Yang y Min (2015), estos modelos crean un árbol sintáctico de forma paralela en la lengua de origen y la de llegada, con el objetivo de ofrecer una traducción en base a operaciones de reordenamiento entre los dos árboles. En la figura que aparece a continuación (figura 8) se muestra un ejemplo de árbol sintáctico en lengua china que hemos generado basándonos en el ejemplo de Yang y Min (2015).

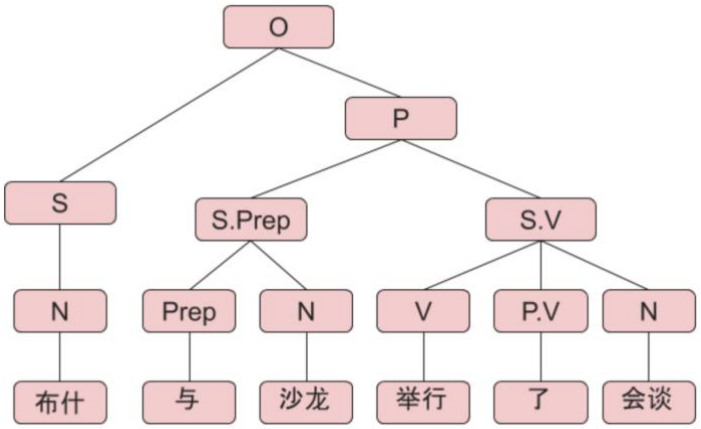


Figura 8: árbol sintáctico de la lengua china del modelo basado en la sintaxis (Yang; Min, 2015:207; elaboración propia)

5. Estudios prácticos de traducción con un motor TAE y traducción humana

En los subcapítulos que recogemos a continuación, detallamos la metodología de la parte práctica de nuestro estudio, así como todas las fases que se han llevado a cabo en ambos procesos; el humano y el automático.

5.1. Metodología de procesamiento del estudio

Para llevar a cabo nuestra investigación y respondiendo a los objetivos y preguntas que nos han llevado a realizarla, hemos decidido dividir el bloque práctico en dos partes esenciales. Esta decisión parte de la principal inquietud de nuestra investigación; es decir, estudiar si los procesos de traducción automática pueden mejorar la productividad del traductor con la combinación de lenguas chino-español, frente a los procedimientos de traducción tradicionales, sin hacer uso de la traducción automática (como ocurre en muchos casos, ya que el traductor no tiene permitido hacer uso de la TA por cuestiones de confidencialidad u otros motivos). En este sentido, la metodología que empleamos es la siguiente:

Por un lado, en el capítulo 5.2 llevaremos a cabo el proceso de traducción de forma humana y lo realizaremos mediante una herramienta TAO, sin hacer uso de traducción automática. En esta parte del proceso, hemos decidido realizar la traducción nosotros mismos en lugar de escoger un texto ya traducido para así poder estudiar con más exactitud todos los escollos que se producen durante el proceso de traducción real de un traductor con la combinación de lenguas chino-español cuando no puede hacer uso de la traducción automática. Por ello, consideramos que al realizar la traducción nosotros mismos podríamos documentar todos los problemas que intervienen desde las primeras fases del proceso de traducción —como por ejemplo la documentación o la búsqueda de textos paralelos—, hasta la propia fase lingüística, de trasladar el texto en lengua origen a la lengua meta; pues, al fin y al cabo, uno de los principales objetivos de la investigación es estudiar qué procedimiento puede llegar a aumentar la productividad de trabajo al traductor con tal combinación lingüística, y para ello, consideramos que no tan solo deben tenerse en cuenta los aspectos lingüísticos de la traducción, sino que también todas las fases previas que forman parte del flujo de trabajo del traductor profesional. En este sentido, para este proceso, comentaremos todas las distintas fases llevadas a cabo, tales como la documentación, la atmosfera de trabajo y prestaciones del software de traducción seleccionado, los aspectos que han causado complicaciones durante la traducción y la revisión final de la misma.

Por otro lado, en el capítulo 5.3 realizaremos el estudio práctico de manera automatizada, es decir, documentaremos el proceso de creación y entrenamiento de un

motor de TAE con la plataforma KantanMT[1], la obtención de los corpus bilingües y monolingües con los documentos disponibles al alcance del traductor, la fase de traducción, la posedición y además recogeremos una evaluación final de la calidad de la traducción con un análisis realizado desde KantanMT y otro realizado desde MTradumàtica[2].

Para ambos procesos, trabajaremos con el mismo texto para realizar las pruebas de traducción de modo que el estudio más comparable y justo en los dos procedimientos. En nuestro caso, hemos decidido seleccionar un fragmento de un texto de ámbito periodístico que trata sobre una de las innovaciones tecnológicas en China (para consultar el texto, véase el anexo del trabajo). En cuanto a la selección del texto, consideramos que la traducción de textos periodísticos del chino al español puede ser un reto debido a las características de las lenguas y a las diferencias de comunicación. Además, consideramos que es un tipo de texto que requiere mucho trabajo de edición porque va a ser publicado y pretende captar la atención del público y animarlo a que continúe leyendo. Por lo tanto, creemos que tanto en el proceso humano como en el proceso automatizado, el traductor tendrá que trabajar mucho para llegar al resultado de traducción final deseado.

Finalmente, tras haber llevado a cabo ambos procesos, realizaremos un estudio comparativo entre los dos, donde tendremos en cuenta las ventajas y desventajas de los dos procedimientos de trabajo y evaluaremos qué proceso resulta más productivo teniendo en cuenta todos los aspectos que se han documentado durante el estudio práctico y las características de las lenguas con las que trabajamos.

5.2. Estudio práctico de traducción con procesos humanos

En esta primera parte del estudio, se recogerán todas las partes del proceso de traducción llevadas a cabo con tal de llegar a un resultado final de traducción sin hacer uso de traducción automática y mediante una herramienta de traducción asistida. En los capítulos que aparecerán a continuación, recogeremos todos los pasos del proceso así como los problemas que han surgido en las distintas fases de la realización de la traducción.

5.2.1. Fase de documentación

Una de las fases esenciales de la traducción y en especial si se traduce desde cero sin tener conocimiento de la información que se presenta en el texto a traducir, es la fase de

[1] KantanMT: https://kantanmt.com/

[2] MTradumàtica: https://mtradumatica.uab.cat/

documentación. En este caso, como se ha comentado en el capítulo de metodología, el texto que hemos seleccionado para traducir —de título "内容服务生态持续加码，小度正在"吃"掉互联网"(véase «Texto 1» en el anexo del trabajo)— es un texto de ámbito periodístico, perteneciente al portal de noticias online de la página web de China Daily[3] y no cuenta con una traducción al español o a otro idioma. En este caso, se trata de un texto que recoge información acerca de un tipo de tecnología que en China ya es muy común y popularizada pero en España está empezando a llegar ahora; se trata de los asistentes inteligentes que llevan incorporada una pantalla y una cámara y ofrecen más prestaciones que los asistentes conocidos hoy en día.

En este sentido, para poder comprender mejor la información del texto, decidimos indagar sobre el altavoz inteligente del que se hablaba en el texto que seleccionamos para realizar el estudio y para ello accedimos a Baidu[4] por el principal hecho de que el asistente inteligente de la noticia era el propio asistente de Baidu, llamado Xiaodu. Asimismo, como Google en China está bloqueado, consideramos que obtendríamos más información y más detallada si realizábamos la búsqueda en su propio navegador. Sin embargo, uno de los hechos que dificultan la fase de documentación en este caso, es que documentarse desde un navegador chino obliga al traductor a realizar la búsqueda en chino y por lo tanto leer y acceder a muchas otras publicaciones en esta lengua. En muchas ocasiones, debido a que la atención en la lectura debe ser más profunda, esto puede ralentizar el proceso de documentación, y por ello en estos casos, y tal como decidimos hacer durante esta fase, hicimos uso de la traducción automática para descartar aquellos textos cuya información era menos importante y seleccionar los que nos serían útiles para nuestra traducción.

Asimismo, también decidimos realizar una búsqueda desde Google acerca del dispositivo del cual se estaba hablando en el artículo seleccionado y como se trata de un aparato bastante novedoso en China, no pudimos encontrar artículos sobre el mismo producto en español pero sí que encontramos un artículo del periódico *El País* —de título «Los mejores asistentes inteligentes con pantalla» (2019)— que hablaba de varios asistentes inteligentes con pantalla de otras marcas y decidimos utilizarlo como texto paralelo o medio de documentación para complementar la información que se ofrecía en nuestro artículo a traducir, junto con la información que habíamos recopilado de los artículos chinos del navegador Baidu.

En este aspecto, en relación a la fase de documentación, cabe destacar que es una parte del proceso que requiere de mucho tiempo ya que es una de las fases más importantes porque nos ayuda a comprender mejor el texto que debemos traducir, pero puede alargar el

3 China Daily: http://cn.chinadaily.com.cn/

4 Baidu: https://www.baidu.com/

proceso de trabajo, en especial teniendo en cuenta que debido a que en China no se tiene acceso a Google, siempre que se quiera buscar alguna documentación en referencia a aspectos culturales o a productos del país es esencial realizar la búsqueda desde Baidu ya que toda la documentación escrita acerca de estos temas proviene de China y se recoge principalmente en sus páginas y redes. Este aspecto es bastante característico sobre todo cuando se traduce del chino ya que la fase de documentación debe realizarse de forma distinta a otras lenguas y en muchas ocasiones esto puede implicar algunas desventajas como la que hemos recogido anteriormente.

5.2.2. Fase de traducción

Para realizar la fase de traducción, decidimos emplear una herramienta TAO ya que consideramos que tienen unas buenas prestaciones y son muy útiles en el proceso de traducción. En nuestro caso, para analizar mejor los resultados de nuestro estudio, no hicimos uso de la traducción automática ya que queríamos analizar todas las ventajas y desventajas del proceso de traducción humana sin que la tecnología de los traductores automáticos interviniera en el proceso de traducción, ya que en muchas ocasiones los traductores humanos tratan con traducciones con datos confidenciales y si no se cuenta con un traductor automático propio o un API de cualquier compañía que cuente con sistemas de traducción automática (como Google o Baidu, por ejemplo), el traductor no puede hacer uso de la TA durante el proceso de traducción. En este sentido, a continuación comentaremos las características de la herramienta TAO que hemos seleccionado para la investigación y cómo ha resultado nuestro proceso de traducción dentro de este software.

5.2.2.1. La herramienta TAO: Memsource

En nuestro caso, la herramienta TAO que seleccionamos para realizar la traducción es Memsource[5]. Esta herramienta se caracteriza por ser un software de traducción asistida que cuenta con muchas prestaciones. Memsource es realmente útil ya que permite a los traductores elegir el modo de trabajo que prefieran; en la nube o en local. Además, puede emplearse en distintos sistemas operativos, por ejemplo, Windows, Mac o Linux. Al igual que muchos otros softwares de traducción asistida, Memsource cuenta con una interfaz muy intuitiva en la que podemos crear proyectos, asignarlos a traductores y traducir desde el editor que muestra el texto de origen a la izquierda y nuestras traducciones a la derecha. Asimismo, en la parte inferior, podemos observar una previsualización de cómo queda

5 Memsource: https://www.memsource.com/

nuestra traducción dentro del documento que estamos traduciendo. Como muchas otras herramientas TAO, Memsource permite crear, importar y exportar memorias de traducción que ayudan al traductor durante su actividad, ya que le recomiendan traducciones que ya se han validado previamente para evitar inconsistencias en el texto traducido. Asimismo, este software también facilita la gestión terminológica ya que tiene integrado un sistema terminológico que ayuda al traductor a seleccionar las palabras o frases adecuadas en cada ocasión. Igualmente, también cuenta con una función de control de calidad para asegurar que el traductor no ha pasado por alto ningún segmento o etiqueta a traducir; pero si se prefiere, también se puede asignar el proyecto a un revisor para que trabaje sobre la traducción y realice las correcciones correspondientes.

En este aspecto, en nuestro caso hemos elegido esta herramienta ya que consideramos que trabajar en la nube es más cómodo y además es un software muy intuitivo. Sin embargo, tal y como hemos comentado anteriormente, en este caso no utilizaremos todas las prestaciones que ofrece Memsource, ya que por ejemplo, la función de traducción automática la mantendremos desactivada para que nuestro estudio sea más comparable.

Figura 9: interfaz del editor de Memsource

5.2.2.2. Aspectos del proceso de traducción

Durante la traducción y debido a que no hicimos uso de TA, nos encontramos con ciertos aspectos típicos de la lengua que dificultaron nuestro procedimiento de traducción y provocaron que tuviéramos que realizar una investigación en Baidu para poder solucionarlos. A continuación comentamos algunos de los problemas que experimentamos durante el proceso de traducción.

a) Expresiones inusuales

En el artículo que hemos seleccionado para traducir aparecen muchas expresiones que son inusuales en español y de las cuales nos hemos tenido que informar para descubrir su significado. A parte de las expresiones idiomáticas, en chino es muy común un tipo de lenguaje específico dentro la red y debido a que nuestro artículo forma parte de un portal periodístico electrónico, muchas de las expresiones propias de internet frecuentan en el texto y por lo tanto en muchas ocasiones ha sido complicado encontrar un equivalente en diccionarios y hemos tenido que recurrir a la búsqueda en Baidu. A continuación mostramos algunos ejemplos:

(1) 卖水人
(Trad. lit.: vendedor de aguas)
(2) 备胎
(Trad. lit.: rueda de recambio)

Tal y como podemos observar, en el ejemplo número 1, la traducción literal de esa expresión significa «vendedor de aguas», pero realmente esta expresión traducida de manera literal no funciona dentro del contexto y tras realizar una búsqueda hemos observado que, en términos comerciales, se refiere a aquellas personas que permanecen a la espera de potenciales oportunidades comerciales de manera estratégica para estudiar cómo sacar el mayor provecho de una situación. Por lo tanto, debido a que en el contexto en el que aparecía hacia referencia a otras marcas de asistentes inteligentes, hemos decidido reformular la traducción de manera que recoja el mismo significado y evitando así el uso de una expresión hecha (para consultar la traducción véase «Texto 2» en el anexo). Por otro lado, en el segundo ejemplo, observamos que también se hace uso de una expresión inusual en la lengua de llegada, y es la expresión «ser una rueda de recambio». En este caso, después de hacer una investigación hemos observado que esta expresión es típica de internet y que hace referencia a la posibilidad de que los productos electrónicos tengan un sustituto o un reemplazo, es decir una versión nueva y actualizada del mismo producto.

b) Abreviaciones terminológicas de productos de la red

Otro de los aspectos poco usuales que hemos encontrado mientras realizábamos la traducción ha sido el uso de ciertos términos como siglas u otras abreviaciones de las que se hablaba con naturalidad y que no iban acompañadas de ningún tipo de explicación de las mismas. Un claro ejemplo de ello es el término «BAM 三» (BAM 3, en español), que tras

hacer una gran investigación en Baidu (ya que en Google no encontramos resultados acerca de su significado), descubrimos que BAM hacía referencia a las siglas de las tres principales empresas creadoras de asistentes inteligentes, es decir, Baidu, Alibaba y Xiaomi. Este tipo de abreviaciones es muy común de la lengua china y por ello resulta complicado de entender para el traductor no nativo de China, mientras que para los lectores nativos es un aspecto que está muy integrado y que entienden con facilidad. En este sentido, en nuestra traducción hemos optado por mencionar a las tres empresas ya que consideramos que el lector no estará familiarizado con las siglas BAM. Otro de los ejemplos en relación a este tema, es el término «B 站» (parada B, en español), ya que se trata de una abreviatura típica en el lenguaje de internet para referirse a la plataforma de vídeos Bilibili[6]. En este caso, también hemos tenido que realizar una búsqueda para averiguar a qué hacía referencia el término ya que desconocíamos la abreviatura para esa plataforma, y optando por la misma solución que con el problema previo, hemos decidido mencionar el nombre de la plataforma para evitar confusiones al lector (para consultar la traducción, véase «Texto 2» en el anexo).

c) Repeticiones

Un aspecto típico de los textos en lengua china, es que se tiende a repetir la misma información en varias ocasiones con tal de hacer hincapié en la información que se está ofreciendo. En nuestro caso, a lo largo del texto se habla de muchas de las plataformas audiovisuales que se han asociado con el asistente inteligente del que se habla en el artículo, y se mencionan todas y cada una de ellas con frecuencia a modo de lista. En este caso, nos ha resultado complicado poder adaptar el texto para que resultase fluido y poco redundante en español sin tratar de omitir información del texto origen. Por ello, este proceso ha resultado bastante lento ya que nos ha llevado mucho tiempo analizar y reformular la información del texto origen para que fuera adecuada en el texto meta, pero finalmente hemos decidido omitir los nombres de algunas plataformas, o hacerle entender al lector de qué plataformas se está hablando sin mencionarlas varias veces (para consultar la traducción, véase «Texto 2» en el anexo).

d) Aspectos pragmáticos de la tipología textual

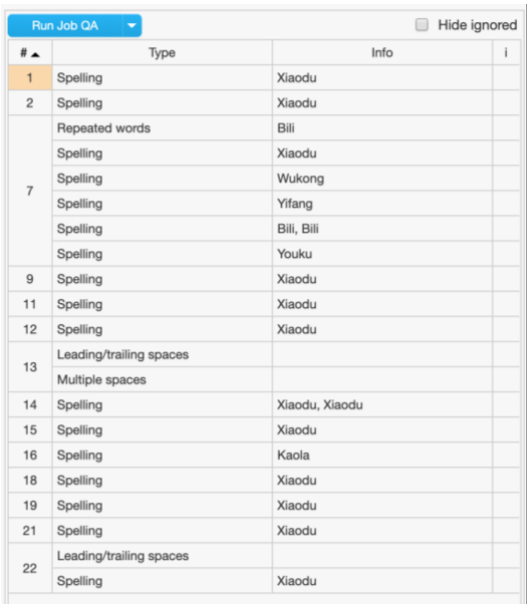
Debido a que se trata de un texto de ámbito periodístico, hemos experimentado problemas para adecuar el texto traducido a su público receptor y cumplir con el principal objetivo de esta tipología textual: lograr captar la atención del lector para que continúe

6 BiliBili: https://www.bilibili.com/

leyendo. En este caso, debido a que el chino tiene una forma de expresión y de formato bastante distinta al español, hemos tenido que optar por la transcreación en algunas partes del texto; por ejemplo, en el título principal del artículo, ya que una traducción cercana o literal del original no causaría un impacto positivo en el receptor cuya lengua nativa es español y no invitaría a continuar leyendo. Asimismo, tal y como podemos observar en el texto original (véase «Texto 1» en el anexo del trabajo), en chino, separan las partes del artículo a modo de capítulos numerados (01, 02...) y en nuestro caso hemos decidido emplear directamente el título del apartado en negrita para que resulte más natural para el lector.

5.2.3. Fase de control de calidad y revisión

Para la fase de revisión, una vez tuvimos la traducción completa, realizamos una lectura exhaustiva de nuestra traducción en comparación con texto original desde la propia herramienta TAO Memsource, para asegurarnos de que toda la información del original aparecía en el texto traducido. Asimismo, para lograr una mejor calidad y comprobar que no se haya producido ningún error ajeno a los contenidos textuales, empleamos el control de calidad que ofrece Memsource y observamos algunos de los puntos que presentaban errores en nuestra traducción y que no habíamos visto durante la lectura del resultado.



#	Type	Info	i
1	Spelling	Xiaodu	
2	Spelling	Xiaodu	
7	Repeated words	Bili	
	Spelling	Xiaodu	
	Spelling	Wukong	
	Spelling	Yifang	
	Spelling	Bili, Bili	
	Spelling	Youku	
9	Spelling	Xiaodu	
11	Spelling	Xiaodu	
12	Spelling	Xiaodu	
13	Leading/trailing spaces		
	Multiple spaces		
14	Spelling	Xiaodu, Xiaodu	
15	Spelling	Xiaodu	
16	Spelling	Kaola	
18	Spelling	Xiaodu	
19	Spelling	Xiaodu	
21	Spelling	Xiaodu	
22	Leading/trailing spaces		
	Spelling	Xiaodu	

Figura 10: control de calidad en Memsource

Tal y como podemos observar, la mayoría de errores que se detectan en el control de calidad de Memsource están relacionados con la escritura de algunos términos y en este caso tuvimos que ignorar estas sugerencias ya que comprobamos que dichos términos son

correctos tal y como se habían escrito. No obstante, también podemos observar que Memsource detectó espacios adicionales y en este caso, los eliminamos para poder lograr que nuestro texto fuese correcto.

5.3. Estudio práctico de entrenamiento de un motor TAE chino-español

Tras haber recogido toda la teoría en relación a la traducción automática y, más detalladamente, sobre la traducción automática estadística y el proceso de entrenamiento de un motor TAE en el apartado teórico del trabajo, en los subcapítulos que aparecerán a continuación comentaremos todo el proceso llevado a cabo en esta parte práctica de la investigación en la que nos centraremos en la traducción mediante un traductor automático estadístico. Para ello, documentaremos todo el proceso, desde la fase de creación y entrenamiento del motor con la plataforma KantanMT, hasta la fase de traducción, posedición y un análisis comparativo con los resultados de calidad del motor MTradumàtica.

5.3.1. La plataforma del motor TAE seleccionado: KantanMT

KantanMT es una plataforma de creación de motores de traducción automática en la nube que ofrece muchos servicios en relación a la incorporación de las tecnologías de la traducción al flujo de trabajo del traductor o localizador. Esta plataforma permite crear motores de traducción automática personalizados —tanto de traducción automática estadística como de traducción automática neuronal— mediante el uso de memorias de traducción, en formato TMX, bases de datos terminológicas, en formato TBX y otros archivos gratuitos de la biblioteca de KantanMT que ayudan al entrenamiento del motor.

La interfaz de esta plataforma es realmente intuitiva ya que toda la información queda recogida en una misma pantalla. Asimismo, KantanMT también ofrece un sistema de análisis que determina la calidad de los motores generados por el usuario y que permite establecer fechas de entrega para posibles proyectos y presupuestos para los mismos. En cuanto a la traducción, permite adquirir una gran precisión en la terminología de los resultados ya que ofrece la opción de personalizar glosarios terminológicos que se apliquen a la traducción durante el proceso de entrenamiento. Otra de las prestaciones de esta plataforma es que cuenta con distintos sistemas de evaluación automática de la TA, como el sistema BLEU, el TER y la métrica F, para monitorear la calidad de los resultados de traducción. Asimismo, KantanMT también permite integrar servicios de traducción automática en algunas de las herramientas TAO más utilizadas por los profesionales de la traducción; como memoq, SDL Trados Studio, Memsource y XTM.

En nuestro caso, hemos decidido emplear esta plataforma en la nube para entrenar el

motor de TAE de nuestra investigación ya que KantanMT cuenta con el sistema *NLPIR Chinese Word Segmentation*, un sistema de segmentación de la lengua china que necesitamos para nuestro proceso de preparación de los textos originales y que hemos considerado muy útil para agilizar el trabajo de creación del motor con nuestra combinación de lenguas. En este sentido, lo primero que hicimos para comenzar con nuestro estudio práctico fue registrarnos en KantanMT y crear nuestro motor de traducción automática estadística con la combinación de lenguas chino-español.

5.2.2. Fase de entrenamiento del motor

Tal y como hemos comentado en capítulos anteriores, para entrenar un motor de traducción automática estadística es esencial contar con un corpus bilingüe en la lengua de llegada y de partida y un corpus monolingüe en la lengua meta para facilitar la traducción al sistema. En esta ocasión, para nuestro motor de TAE hemos decidido seleccionar los archivos bilingües y monolingües del repositorio OPUS[7] que cuenta con una vasta cantidad de corpus en distintas combinaciones lingüísticas y distintos campos temáticos.

Para nuestro estudio, hemos decidido seleccionar los textos de Global Voices[8] en la combinación de lenguas chino-español, ya que la combinación contaba con varios textos alineados —en concreto, textos con más de dos millones de palabras y más de 115.000 segmentos alineados— de temática periodística, que es justo la tipología textual de nuestro texto seleccionado para el estudio (véase «Texto 1» en el anexo del trabajo) puesto que Global Voices es un portal online de blogueros y periodistas que recogen la información de lo que ocurre alrededor del mundo. En este caso, la información recogida en los corpus bilingües se ha extraído de distintas noticias traducidas y publicadas entre 2009 y 2018. Además, a parte del corpus bilingüe de Global Voices, decidimos descargarnos un segundo corpus también desde el repositorio OPUS bajo el nombre «News Commentary» que contiene comentarios de noticias y cuenta con más de seis mil textos alineados en chino-español. Asimismo, por lo que respecta al corpus monolingüe en español, también lo hemos extraído del mismo repositorio, en concreto de los textos monolingües de Global Voices y recoge la información de noticias publicadas desde 2005 hasta 2017.

Una vez seleccionados los textos bilingües y monolingües que necesitamos para el entrenamiento, era necesario descargarlos en los formatos que KantanMT —la plataforma que empleamos para nuestra investigación— indicaba. En este sentido, para los textos bilingües alineados, KantanMT requiere que se introduzcan al motor en formato TMX, es decir, el formato de memorias de traducción. En este aspecto, cabe destacar que el

7 OPUS: http://opus.nlpl.eu/
8 Global Voices: https://globalvoices.org/

repositorio de OPUS ofrece la posibilidad de descargar los corpus en dicho formato, y por lo tanto no fue necesario realizar ningún proceso de conversión. Por otro lado, por lo que respecta a los textos monolingües, descargamos el archivo en español en formato TXT, un formato que KantanMT también acepta y que debimos nombrarlo «source.utf8.trg.mono», tal y como requiere la plataforma de KantanMT.

No obstante, los motores de traducción automática estadística necesitan una gran cantidad de palabras para su correcto funcionamiento y a causa de la escasa cantidad en los dos corpus bilingües empleados para llevar a cabo la traducción, fue necesario hacer uso de un tercer corpus para que el motor contase con más palabras y pudiera ofrecer una traducción aceptable. Sin embargo, debido a la escasez de corpus de software libre en la red y al alcance de los usuarios con la combinación de lenguas con la que trabajamos, resultó imposible aumentar nuestros corpus de entrenamiento con textos de temática periodística que se ajustasen a la tipología textual de nuestro texto a traducir. Por este motivo, decidimos utilizar un corpus que recoge subtítulos con los pares de lenguas chino-español, con tal de ampliar los materiales de entrenamiento. No obstante, la práctica de combinar corpus especializados con corpus más genéricos provocó una serie de complicaciones que recogeremos en los capítulos que aparecerán a continuación (véase el subcapítulo 5.2.4).

Así, con los tres corpus seleccionados para entrenar el motor de traducción automática estadística, desde la plataforma de KantanMT se inicializó la fase de entrenamiento del motor y seguidamente, el motor estuvo listo para comenzar a traducir.

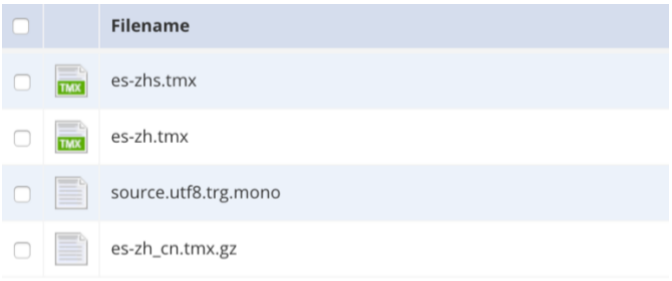


Figura 11: los archivos para el entrenamiento del motor

Cabe destacar que tras finalizar todo el proceso de entrenamiento y hacer las primeras pruebas de traducción con el motor, observamos que el sistema era incapaz de reconocer algunos de los términos de la lengua origen, y por ello, recurrimos a la creación de un tercer corpus creado por nosotros mismos con la recopilación de textos de la misma temática que el texto a traducir que debimos alinear con una herramienta TAO, en nuestro caso Memsource, y además un glosario terminológico con las palabras desconocidas por el sistema (para consultar el glosario que incluimos en la fase de entrenamiento, véase el anexo del trabajo).

5.2.3. Fase de traducción (*decoding*)

Una vez entrenado el motor, KantanMT ya permite llevar a cabo la fase de traducción con el sistema que se haya creado. Esta plataforma, requiere que los textos a traducir se adjunten en formato .docx, así que en nuestro caso y partiendo de la noticia que seleccionamos para el estudio (véase «Texto 1» en el anexo del trabajo), fue necesario generar un documento .docx para poder adjuntarlo a KantanMT. No obstante, observamos que el motor no era capaz de detectar los caracteres en el texto y se produjo un error que impidió al sistema procesar los resultados de traducción. En este aspecto, observamos que KantanMT para ciertas combinaciones lingüísticas requiere que se incluya una memoria de traducción (archivo .tmx) con el texto en lengua origen y el texto meta vacío junto al documento .docx para poder reconocer la cantidad de palabras a traducir y ofrecer un resultado de traducción. Por este motivo, para poder facilitar el proceso al motor y asegurar su correcto funcionamiento, consideramos necesario recurrir a una herramienta TAO —en nuestro caso, Memsource— para crear un nuevo proyecto con el texto seleccionado y descargar la memoria vacía para luego poder incluirla en nuestro motor de TAE durante la fase de traducción.

Figura 12: archivo a traducir y memoria de traducción

En este sentido, con los dos documentos necesarios, KantanMT pudo procesar la traducción con éxito. Una vez que KantanMT finaliza el proceso de traducción, la plataforma permite al usuario descargar dos tipos de archivos; por un lado, se descarga el documento traducido en formato .docx, que respeta en mayor o menor medida la maquetación del archivo, y por otro, permite descargar —siempre que se dé el caso— un documento con todos los términos desconocidos por el sistema de traducción (para consultar los términos, véase el punto 4 del anexo del trabajo). Consideramos que este aspecto es realmente útil ya que el hecho de que el sistema proporcione los términos desconocidos en un documento de formato .txt nos puede servir de apoyo para crear un glosario terminológico con dichos términos y su traducción e incorporarlo al proceso de entrenamiento para que en la próxima ocasión el traductor automático los reconozca y los traduzca de la manera que le indiquemos.

En el anexo del trabajo, se pueden consultar los resultados del texto traducido antes y después de la incorporación del glosario terminológico (véase el punto 5 del trabajo).

5.2.4. Fase de análisis y evaluación de la calidad

Tras haber creado nuestro motor de traducción automática estadística y después de realizar las traducciones, consideramos importante hacer un análisis de nuestro motor para comprobar sus características y evaluar la calidad del mismo en comparación con otro sistema de creación de motores personalizados de traducción automática estadística, MTradumàtica. En este aspecto, en primer lugar comentaremos el análisis de las métricas de nuestro motor realizado en la plataforma KantanMT, recogiendo un comentario en base a los gráficos que se proporcionan en la propia página en relación a la información y características del motor que hemos personalizado, y posteriormente comentaremos la evaluación de la calidad de traducción según MTradumàtica.

En la figura de a continuación, podemos observar un análisis de los porcentajes de las distintas métricas de calidad, de la distancia de edición y de la cantidad de palabras que componen los corpus con los que se ha entrenado el motor.

Figura 13: métricas de nuestro motor de traducción automática estadística personalizado

Tal y como podemos observar, KantanMT utiliza varios sistemas de análisis para evaluar la calidad del motor, en concreto, la métrica F, BLEU y TER. A continuación comentaremos los resultados de las distintas métricas y una breve explicación de los mismos.

Por lo que respecta a la métrica F, podemos observar que el porcentaje obtenido es de un 26% lo cual es indicativo de que nuestro motor no traducirá de manera muy precisa ya que el resultado del análisis es bajo y la métrica F es la que se encarga de analizar la precisión de traducción del motor. En este aspecto, podemos afirmar que nuestro motor está por debajo de la media del conocimiento del idioma de destino y del ámbito textual, y este problema se debe a que los corpus empleados en el entrenamiento del motor no son de un dominio concreto y, por lo tanto, eso produce que las traducciones sean menos precisas que si se hubieran empleado unos corpus con terminología concreta y de un mismo campo

contextual. Por otro lado, en cuanto a la métrica BLEU, el porcentaje en respecto a nuestro motor de traducción automática estadística es de 20%, una cifra también considerablemente baja, puesto que esta métrica se encarga de evaluar la calidad y la fluidez de la traducción. En este aspecto, este resultado nos indica que la calidad del resultado de la traducción no será lo suficientemente buena y para poder mejorarla, lo ideal sería aumentar la cantidad de corpus bilingües y monolingües para que el motor tenga más información y pueda proporcionar resultados que conserven una fluidez en la lengua meta. En cuanto a la métrica TER, podemos ver que el porcentaje que se ofrece en el análisis es del 100%, por lo tanto, esto es un indicativo de que los resultados de traducción producidos por nuestro motor de TAE necesitarán un alto nivel de posedición, ya que la métrica TER se encarga de calcular la distancia de posedición que sería necesaria para lograr una mejor calidad en el texto traducido en lengua meta. En este aspecto, y teniendo en cuenta los porcentajes de las métricas comentadas previamente, podemos intuir que nuestro motor de TAE no ofrecerá unos resultados de traducción naturales y de calidad, y que necesitará una posedición para que la traducción pueda llegar a ser comprensible.

Por otro lado, en relación a la cantidad de palabras de los corpus que se han empleado para el entrenamiento del motor, podemos ver que los resultados son satisfactorios. Por un lado, la cantidad de palabras en la lengua de origen supera los diecisiete millones de caracteres y esto es un hecho muy favorable ya que para que el motor pueda encontrar una traducción de la mayoría de caracteres que aparecen, según KantanMT este necesita, al menos, dos millones de palabras. Asimismo, también podemos observar que el número de palabras únicas —*Unique Word Count*, en inglés— dentro de los corpus empleados en nuestro motor, cuenta también con una cifra elevada, de nueve millones de palabras. Finalmente, por lo que respecta a el número de caracteres en los corpus monolingües, podemos observar que nuestro motor cuenta con una gran cantidad ya que el porcentaje que obtenemos es del 60%.

En este aspecto y tras haber analizado los resultados de las métricas en relación a la cantidad de palabras en los corpus; podemos llegar a la conclusión de que los corpus empleados, a pesar de contener una vasta cantidad de caracteres, no recogen información de un dominio específico, lo cual provoca que los resultados de traducción sean menos precisos, poco naturales en la lengua meta y que requieran de un alto nivel de posedición. En este sentido, para poder mejorar la calidad de los resultados de traducción de nuestro motor, sería conveniente emplear un tipo de corpus durante el proceso de entrenamiento que cuenten con una terminología y un dominio muy concreto para asegurar que el motor pueda extraer las traducciones que el usuario necesite partiendo de los corpus con los que se ha entrenado. No obstante, para nuestro estudio, hemos querido centrarnos en los recursos que puede llegar a tener un traductor a su alcance, y debido a que hay muy pocos

corpus con la combinación de lenguas con las que trabajamos, es realmente difícil poder mejorar los resultados que ofrece el motor con los documentos que se encuentran en la red. Sin embargo, para poder lograrlo, hemos decidido generar nuestro propio corpus recogiendo textos de la temática de nuestro texto a traducir y alineándolos con una herramienta TAO; pero no han llegado a ser suficientes como para mejorar la calidad de nuestro motor.

5.2.4.1. Evaluación de la calidad de traducción del motor con MTradumàtica

Con tal de proporcionar una segunda perspectiva en cuanto a la evaluación de la calidad de los resultados de traducción obtenida en nuestro motor personalizado; hemos decidido realizar una evaluación desde la plataforma MTradumàtica.

MTradumàtica es una plataforma que permite entrenar sistemas de traducción automática en la nube y en local y crear corpus monolingües y bilingües para así poder generar traductores automáticos personalizados y traducir documentos de forma personalizada. Además, otra de las prestaciones que ofrece la plataforma es la de evaluar el rendimiento de los motores de traducción —ya sean generados desde la propia web o con motores de TAE externos— con tal de obtener distintas métricas de calidad (en concreto, BLEU, chrF3, TER y WER).

En este aspecto, para conseguir los resultados de las distintas métricas y realizar una comparación con los análisis proporcionados por KantanMT, realizamos el análisis de la calidad desde MTradumàtica. Desde la plataforma, para conseguir las puntuaciones de la evaluación, se requiere cargar dos documentos; por un lado, una traducción humana que MTradumàtica comprende como *golden standard* que sirve de referencia para la evaluación de la TA, y por otro lado, el texto traducido generado por el motor de traducción automática que se desea evaluar. En este sentido, tras cargar los documentos, MTradumàtica calculó las distintas métricas y ofreció los resultados que aparecen a continuación (véase figura 14).

Fichero de prueba	BLEU	chrF3	BEER	TER	WER
ZH-ES_AI.docx	26.27	24.02	17.88	100.00	100.00

Figura 14: métricas de evaluación de calidad proporcionadas por MTradumàtica

Tal y como podemos observar, los resultados de las métricas de evaluación de MTradumàtica coinciden con las puntuaciones proporcionadas por los análisis que se llevaron a cabo en la plataforma KantanMT. No obstante, desde MTradumàtica también se realiza la evaluación de otras métricas que no se evalúan desde KantanMT. En este aspecto, a continuación comentaremos los resultados generados por MTradumàtica y sus

implicaciones en el resultado final de la traducción proporcionada por nuestro motor personalizado.

En primer lugar, tal y como podemos observar, la puntuación obtenida de la métrica BLEU coincide con exactitud a la que se proporcionaba en el análisis del motor que personalizamos en nuestro estudio con KantanMT, es decir, una puntuación del 26,27%. En este aspecto, afirmamos que los resultados en cuanto a la calidad y precisión de los resultados de traducción en ambas evaluaciones —la realizada con KantanMT y la de MTradumàtica— son bajos, y podemos asegurar que la traducción ofrecida por nuestro motor tendrá inconsistencias en dichos aspectos. Por otro lado, una de las métricas que emplea MTradumàtica y que no se había calculado desde la plataforma KantanMT, es la métrica chrF3. En este caso, tal y como se menciona en la propia página de MTradumàtica, la métrica chrF3 es la que se encarga de realizar un cálculo del porcentaje de n-gramas de caracteres del archivo de referencia, es decir, de la traducción humana que se considera el *golden standard*, presentes en la traducción proporcionada por la TA. Tal y como podemos observar, el porcentaje obtenido es de un 24,02%, lo cual indica que en la traducción realizada por nuestro motor de TAE no aparecen ni la mitad de n-gramas de caracteres en referencia a la traducción humana, y por ello, podemos llegar a la conclusión de que la traducción realizada por la TA es menos concreta en cuanto a la terminología empleada y es posible que su significado se aleje de la semántica del texto original. Otra de las métricas de evaluación que se calculan desde la plataforma MTradumàtica recibe el nombre de BEER. Tal y como se indica en la propia plataforma, esta métrica se encarga de realizar una clasificación de los n-gramas de caracteres y árboles de permutaciones; y la puntuación que obtenemos es de un 17,88, una cifra bastante baja, que nos hace pensar que los resultados de traducción proporcionados por la TA no son de calidad.

Finalmente, también se proporcionan los cálculos de las métricas TER y WER (esta última una pequeña variación de la TER). Tal y como habíamos comentado anteriormente en el análisis de las métricas de los resultados de la calidad realizado por KantanMT, la métrica TER es la que se encarga de calcular la distancia de posedición necesaria en el resultado de traducción generado por la TA, y tal y como ya habíamos mencionado en el subapartado 5.2.4, los resultados tanto de la métrica TER como de la WER en relación a la traducción automática son de un 100%, por lo tanto, podemos afirmar una vez más que el texto en lengua meta generado por nuestro motor de TAE requerirá de una posedición completa para que pueda ser comprensible.

A grandes rasgos, y teniendo en cuenta los resultados de evaluación realizados en ambas plataformas —KantanMT y MTradumàtica— es fácil comprender que la traducción generada por nuestro motor de TAE está lejos de ser perfecta; y eso se debe a varios aspectos, muchos de ellos comentados en el apartado 5.2.4, pero sobre todo a la falta de precisión de un

dominio específico en los corpus bilingües y monolingües empleados en la fase de entrenamiento del motor.

5.2.5. Fase de posedición

Tal y como se recoge en el subcapítulo anterior, los resultados de nuestro motor de traducción automática estadística son de una calidad baja y requieren de posedición para que el receptor pueda lograr comprender la información que se recoge en el texto traducido. En nuestro caso, debido a que los corpus empleados durante el entrenamiento del motor son poco precisos en lo que respecta al campo temático de nuestro texto a traducir, las oraciones traducidas por el motor son en mayor frecuencia incomprensibles y eso produce que el traductor humano tenga que dedicar una gran parte de su tiempo a la posedición del texto y que además tenga que consultar el texto origen ya que muchas de las construcciones en lengua meta no conservan una coherencia.

Si retomamos los datos de la métrica TER comentada en el subapartado 5.2.4 —véase figura 14—, podemos observar que el nivel de posedición del resultado de traducción que ofrece nuestro motor es en mayor medida del 100%. En este sentido, consideramos que en el caso de nuestro estudio seria conveniente dejar de lado la posedición y realizar una traducción de forma humana o bien la posedición del texto traducido con un motor de traducción automática que haya sido entrenado con unos corpus que recojan una información y terminología precisa dentro del dominio del texto con el que trabajamos.

No obstante, como ya hemos comentado a lo largo de la investigación, este problema viene dado por la falta de documentación dentro de nuestro género textual y combinación lingüística, por lo que este hecho no es indicativo de que en todos los posibles casos sea mejor realizar una traducción humana frente a hacer uso de la traducción automática. En este aspecto, debemos tener en cuenta que para nuestra investigación hemos entrenado un motor de TAE con los recursos de software libre al alcance de los usuarios de la red, y que por lo tanto, en otros contextos en los que, por ejemplo, el profesional haya dedicado tiempo y esfuerzo en los procesos de entrenamiento del motor, podría darse el caso de que la calidad de la traducción automática fuera más elevada y que la posedición no requiriera más tiempo que la realización de la propia traducción desde cero.

6. Comparativa entre el proceso de traducción humano y el automático

En los subcapítulos que aparecen a continuación, recogemos un análisis general de las complicaciones y las características de los dos procesos estudiados en esta investigación que nos servirá como visión global e introducción a grandes rasgos de las conclusiones de la investigación que se recogerán con más detalle en el capítulo 8.

6.1. Proceso de traducción humano

Tal y como recogemos con anterioridad en el capítulo 5.2 de esta investigación, durante el proceso de traducción humano experimentamos una serie de complicaciones que ralentizaron la tarea de traducción. A continuación mencionaremos algunas de las particularidades observadas durante el proceso de traducción humano que han dificultado o agilizado el flujo de trabajo ya sea por causas lingüísticas, extralingüísticas o de procedimiento.

En primer lugar, uno de los aspectos que causaron mayor dificultad durante el proceso de traducción humana fue la fase de documentación. Tal y como comentábamos en apartados anteriores, cuando se realiza una traducción donde la lengua origen o meta sea el chino, por ejemplo, el traductor debe enfrentarse a una serie de características que ralentizan su proceso de trabajo. Dichas características parten de cuestiones políticas y sociales, puesto que en China existen muchas restricciones en cuanto a la información que se ofrece en la red con tal de evitar injerencias y manipulación de los datos. En este aspecto, el traductor no tan solo se encuentra con el problema de tener que realizar la documentación desde otro buscador con el que no está tan familiarizado, sino que también debe enfrentarse a las barreras lingüísticas ya que tanto las búsquedas que realice, como la información que encontrará en las distintas páginas deberán ser en lengua china, o en su defecto en inglés, aunque las búsquedas en esta última lengua contaran con menos resultados y menos precisos. A pesar de que somos conscientes de que todas estas complicaciones arraigadas a los aspectos extralingüísticos de la lengua china también son comunes en muchas otras lenguas del mundo, consideramos que es una realidad que dificulta el flujo de trabajo de la persona que se dedica a la traducción y es un aspecto importante a destacar en nuestro estudio para posteriormente tener en cuenta el rendimiento del procedimiento humano frente al del procedimiento automático.

Por otro lado, otro aspecto que dificulta la documentación es el dominio del texto con el que trabajamos; en nuestro caso, seleccionamos un texto que trata sobre las últimas tecnologías en inteligencia artificial, en concreto, recoge información acerca de un asistente

inteligente chino muy novedoso y que aún no cuenta con una competencia exacta de alguna marca occidental. En este aspecto, consideramos que es bastante complicado poder recoger una cantidad suficiente de textos paralelos que ayuden al traductor a recopilar información adicional para poder comprender con mayor exactitud la información del texto y el producto del que se habla. Por esta razón y con tal de agilizar en la medida de lo posible el proceso de documentación, la traducción automática fue de gran ayuda para poder distinguir con rapidez la información más relevante frente a aquella menos importante. Sin embargo, como comentamos, todo este proceso se vio prolongado por todas las complicaciones que recogemos y esto es un hecho que está directamente relacionado con las características de la lengua de trabajo y los aspectos sociales y políticos de China.

En cuanto a la traducción, tal y como comentábamos en el capítulo 5.2.2 de la presente investigación, experimentamos ciertas ventajas y desventajas durante su proceso. Por una parte, desde un enfoque más lingüístico, podemos afirmar que el proceso de traducción con nuestra combinación de lenguas y sin hacer uso de la TA ha supuesto grandes complicaciones y ha ralentizado el trabajo de forma considerable. Uno de los aspectos que más dificultades ha conllevado ha sido el de la traducción de determinada terminología especializada y de expresiones típicas de la red. Ciertamente, para el traductor humano que traduzca con la combinación de lenguas chino-español y que prescinda de la traducción automática, encontrar equivalentes para determinados términos o expresiones es todo un reto; pues, tanto en la red como en papel existen muy pocos diccionarios o portales lingüísticos que ofrezcan una traducción al español, y si los hay, son de una calidad bastante baja. Generalmente, y como comentaremos en el siguiente subcapítulo, la mayoría de recursos lingüísticos que se encuentran en la red en relación a la lengua china —ya sean diccionarios electrónicos, glosarios o portales lingüísticos de resolución de dudas— ofrecen una traducción al inglés, y en muy pocos casos es posible encontrar su equivalente en español. En este sentido, el traductor humano debe traducir aquellos términos y expresiones partiendo de la traducción en lengua inglesa, y muchas veces este proceso provoca que el resultado final en español no recoja todo el sentido expresado por la lengua origen. Por otro lado, en cuanto a los aspectos de procedimiento, uno de los puntos positivos que experimentamos durante la traducción humana fue el uso de herramientas TAO durante el proceso de traducción. Naturalmente, el flujo de trabajo dentro de un software de traducción siempre se verá beneficiado por las prestaciones del mismo. En nuestro caso, tal y como comentábamos en el subcapítulo 5.2.2 del trabajo, consideramos que las herramientas de traducción asistida agilizan en gran parte el trabajo de la persona que se dedica a la traducción ya que guardan en la memoria los segmentos que ya se han traducido previamente y recuerdan al traductor cómo los tradujo en cada caso. En este aspecto, este hecho es muy positivo ya que la lengua china tiende a repetir mucho la información y a

reiterar todos aquellos puntos importantes; por lo tanto, afirmamos que el entorno de trabajo dentro de la herramienta TAO es muy rentable para el traductor. Asimismo, como también mencionamos en capítulos anteriores, la función de control de calidad dentro del propio software agilizó de forma considerable la fase de revisión de la traducción, y eso es de nuevo, un aspecto positivo que extraemos de el procedimiento de traducción humano.

En este sentido, en general podemos afirmar que el proceso de traducción humano, es decir, sin hacer uso de la traducción automática, tiene sus ventajas y sus desventajas. No obstante, consideramos que en conjunto se trata de un proceso largo ya que todas las actividades previas a la traducción y las tareas propiamente lingüísticas requieren de una gran inversión de tiempo por todos los aspectos ya comentados.

6.2. Proceso de traducción automático

Por lo que respecta al procedimiento de traducción con procesos tecnológicos, tal y como ya pudimos observar en el capítulo 5.3, experimentamos diferentes cuestiones, algunas más positivas que otras en relación a la elaboración del motor de TAE y sus resultados de traducción. En este sentido, a continuación recogeremos una visión general de los aspectos positivos y negativos de este procedimiento.

Tal y como comentábamos anteriormente, a pesar de que la creación de un motor de TAE no es tarea complicada, ya que existen muchas plataformas con interfaces intuitivas que ayudan a su creación, un aspecto complicado es el de encontrar el material suficiente y de calidad para poder entrenar a nuestro motor. En nuestro caso de estudio, tal y como ya se ha comentado con anterioridad, decidimos centrarnos en aquellos recursos que se encuentran en la red al alcance de cualquier usuario que los necesite, para poder entrenar nuestro motor de TAE y estudiar su funcionamiento y eficiencia. Asimismo, como recogíamos en el apartado precedente, la mayoría de material bilingüe que encontramos en la red —considerando que la lengua de origen es el chino— en gran medida suelen ser corpus con la combinación de lenguas chino-inglés. Por lo tanto, podríamos afirmar que este es el primer escollo con el que se encuentra el usuario que trata de generar su motor de traducción automática personalizado. En este aspecto, a pesar de que logramos encontrar algunos corpus con la combinación de lenguas con la que trabajamos, no todos ellos compartían el mismo campo temático y este hecho provocó que nuestro motor de TAE no tuviera un rendimiento adecuado (tal y como hemos podido comprobar con las métricas automáticas) y eso ha llevado a que la posedición sea una tarea que requiera una gran inversión de tiempo, incluyo mayor que la traducción humana.

Asimismo, teniendo una visión general del proceso de traducción automático, observamos que es complicado generar un motor de TAE personalizado con los materiales

que se encuentran al alcance de cualquier usuario de la red, no tan solo por su escasez, sino que también por el hecho de que es complicado encontrar corpus bilingües extensos, que compartan la misma temática, y que dicho dominio se corresponda con los temas de los encargos de traducción con los que se trabaja. En este sentido, tras realizar el proceso automático, afirmamos que una de las mayores dificultades de este procedimiento, sin duda ha sido el proceso de búsqueda y recopilación de textos bilingües para el entrenamiento del motor.

Otro de los aspectos que han ralentizado el proceso de creación del motor para posteriormente obtener las traducciones automáticas ha sido el hecho de que la plataforma desde la que trabajábamos requería una serie de formatos concretos tanto del material con el que se entrenaba el motor como de los archivos de la traducción. Este hecho, provocó que tuviéramos que hacer uso de softwares de traducción y realizar una búsqueda más concreta para que la plataforma no experimentase ningún problema con los formatos y se pudiera generar de forma correcta el motor. En este caso, nos encontramos frente a otro de los escollos principales del proceso, ya que además de que existen pocos corpus bilingües con la combinación de lenguas con las que traducimos, nos vemos limitados a trabajar con unos formatos concretos y requeridos por la plataforma. No obstante, este hecho es específico de la plataforma con la que hemos trabajado en nuestra investigación y somos conscientes de que en otros casos y trabajando desde otras páginas que permitan la elaboración de motores de traducción automática estadística personalizados, no se darían los mismos problemas.

Por lo que respecta a la actividad de traducción, está claro que mediante el uso del motor, esta tarea se realiza de forma automática y por lo tanto mucho más rápido en comparación con el proceso de traducción humana. Sin embargo, como ya hemos recogido en el capítulo 5.3, a causa de la poca calidad de los corpus con los que entrenamos el motor, los resultados de traducción en nuestro caso eran casi incomprensibles y al principio contenían incluso términos en lengua original que no se habían identificado y por lo tanto no habían sido traducidos. Este hecho provocó que tuviéramos que generar un glosario terminológico y alimentar al motor nuevamente con la nueva información para que pudiera traducir aquellos términos desconocidos, y al fin y al cabo, todo este proceso tuvo una mayor duración a causa de los problemas experimentados. Asimismo, como contábamos en capítulos anteriores, a pesar de todas las modificaciones que hicimos durante el proceso de entrenamiento del motor, la traducción que nos ofrecía seguía siendo incomprensible, y por lo tanto esto provoca que la tarea de posedición sea más laboriosa y que ralentice todo el proceso en general.

En este aspecto, podemos afirmar que el procedimiento tecnológico en nuestro caso concreto ha provocado que la productividad y el ritmo de trabajo se vean afectados por todos los problemas que han ido surgiendo a lo largo del proceso. En este sentido, a pesar de que

la traducción automática suele ser una alternativa rápida y eficiente a la traducción humana, deben tenerse muchos aspectos en cuenta, y sobre todo, requiere de un gran trabajo de entrenamiento y preparación del motor para su correcto funcionamiento.

7. Conclusiones

Después de la realización de este trabajo de investigación, podemos afirmar que hemos logrado cumplir con todos nuestros objetivos propuestos, que recordamos eran: estudiar de manera comparativa los procesos de traducción humana (sin uso de la TA) y traducción con un motor TAE con la combinación de lenguas chino-español, documentar los procedimientos de creación y entrenamiento de un motor de TAE con los recursos al alcance del traductor, determinar la calidad de la traducción producida por un motor de TAE y el rendimiento de la traducción humana evaluando las fases que se llevan a cabo en ambos procesos, estudiar en qué medida afectan los distintos funcionamientos lingüísticos durante los procesos documentados y cómo afecta este aspecto a la calidad final de la traducción. Asimismo, también hemos logrado cumplir con nuestro objetivo principal que es el que nos llevó a la elección del tema de investigación; comparar ambos procedimientos y determinar cuál de ellos puede mejorar el ritmo de trabajo del traductor con la combinación de lenguas chino-español. Por lo tanto, después de cubrir todos estos aspectos, ahora ya podemos dar una respuesta a todas las preguntas de investigación que nos formulábamos en la introducción del trabajo.

En este sentido, respondiendo a la primera pregunta de nuestra investigación, tras haber experimentado en primera persona el procedimiento de creación de un motor de traducción automática estadística con los recursos que encontramos en la red, podemos afirmar que la posibilidad de crear un motor TAE con la combinación de lenguas chino-español con los recursos al alcance de cualquier usuario depende de muchos factores. Tal y como hemos comentado a lo largo de la investigación, los materiales disponibles al público en la combinación de lenguas con las que trabajamos son menores en cuanto a cantidad. Asimismo, los dominios de los corpus que se encuentran al alcance de todos los usuarios de la red son muy variados; por lo tanto, es realmente complicado encontrar una serie de corpus suficientemente extensos como para entrenar un motor de TAE, que además compartan un mismo campo temático y que este dominio se corresponda a la especialidad temática de los textos que el traductor debe traducir. En este aspecto, consideramos que el hecho de que existan pocos recursos al alcance del traductor y que los que existan sean de diferentes ámbitos temáticos, dificulta en general la creación y la ejecución de las traducciones del propio motor. Así pues, tal y como hemos podido experimentar en nuestro estudio práctico,

sí que es posible crear un motor de TAE con los materiales que se encuentran en la red, ya que la cantidad de corpus y palabras será suficiente, pero el hecho de que los corpus no conserven un mismo campo temático puede llegar a afectar a otros aspectos del procedimiento de traducción que comentaremos a continuación. En este sentido, consideramos que para lograr una mejor funcionalidad, idealmente el traductor o incluso empresas que se dediquen al sector de la traducción, deberían generar sus propios corpus bilingües y monolingües de un mismo campo temático, que se corresponda al dominio de los textos que habitualmente suele traducir. Sin embargo, consideramos que este aspecto es algo negativo ya que la creación de corpus propios para el entrenamiento de un motor puede ser una tarea larga y complicada, pero pensamos que si se pretende entrenar al sistema con unos mejores materiales para que posteriormente la ejecución de las traducciones se vea beneficiada, la creación y el uso de corpus propios puede ser una buena alternativa para mejorar el rendimiento del motor personalizado.

En relación a este aspecto y contestando a la segunda pregunta que nos planteamos previamente a la investigación, la calidad de las traducciones dependerá de los procesos que se sigan en cada situación. Si tomamos como ejemplo nuestro caso práctico, podemos afirmar que la calidad de las traducciones producidas por nuestro motor de TAE es realmente baja debido a las cuestiones que acabamos de mencionar, ya que los materiales empleados durante la fase de entrenamiento de nuestro motor eran de temática muy variada y se alejaban, en cierto modo, del tema de nuestro texto a traducir. No obstante, como hemos comentado, consideramos que a pesar de que los resultados del traductor automático muestren muchas incoherencias y en ocasiones sean incomprensibles, podemos observar que hay algunas partes de la traducción automática que han conservado su significado original o que se acercan a la semántica del texto origen a pesar de que los corpus de entrenamiento no mantengan una unidad temática y se alejen del campo contextual de nuestro texto a traducir. En este sentido, consideramos que este hecho es indicativo de que hay que conservar la esperanza en la traducción automática ya que muy probablemente empleando corpus de una mejor calidad y que conserven un dominio temático, los resultados de traducción que ofrezca el motor llegarán a ser buenos. Así pues, no consideramos adecuado comparar la calidad de la traducción humana con la calidad que ha ofrecido nuestro motor, ya que pensamos que la ejecución de la traducción por parte de nuestro motor no hace justicia a los resultados que puede llegar a ofrecer cualquier otro traductor automático que haya sido entrenado con corpus extensos y de un mismo dominio, ya que los resultados nuestro motor son de baja calidad, en gran parte por los corpus empleados durante su entrenamiento. Sin embargo, de nuevo, tras realizar nuestro estudio práctico y los análisis de los resultados, consideramos que un motor de TAE entrenado con unos materiales de calidad, puede ofrecer resultados de una mejor calidad que nuestro

motor, y por ello creemos que la traducción automática estadística personalizada puede ser una incorporación positiva en el flujo de trabajo del traductor.

Por otra parte, respondiendo a otra de las inquietudes previas a nuestra investigación, tras haber realizado el estudio práctico y teniendo en cuenta los fundamentos teóricos de la traducción automática y de las diferencias entre el chino y el español, podemos afirmar que las características y particularidades de ambas lenguas tienen una influencia directa en los dos procesos de traducción, tanto en el proceso humano como en el proceso tecnológico. Tal y como hemos documentado a lo largo de la investigación, por lo que respecta a los procedimientos de traducción humana, las cuestiones políticas y sociales de China afectan al proceso de documentación y búsqueda de textos paralelos por cuestiones de privacidad y control de la información. En el caso de la traducción automática, como ya hemos recogido durante la documentación del proceso de creación del motor, las particularidades de la lengua también afectan en cierta medida al funcionamiento del sistema. Como indicamos en el estudio, la lengua china se caracteriza por no realizar una segmentación mediante espacios entre las diferentes unidades que componen las oraciones. En este aspecto, los motores de traducción automática necesitan que se realice una segmentación de las palabras del texto original para poder identificarlas de forma separada y que posteriormente la traducción pueda llevarse a cabo. No obstante, cabe destacar que la segmentación de los textos chinos es una tarea complicada y que a pesar de que muchas plataformas de creación de motores personalizados ya cuentan con un sistema automático de segmentación del chino, en muchas ocasiones no realizan su función de forma correcta y eso afecta a los resultados de la traducción. Asimismo, uno de los principales problemas que causan que la segmentación no se lleve a cabo correctamente, es el hecho de que la lengua china es una lengua ideográfica, por lo tanto cada carácter recoge el significado de una idea, y junto a otro carácter puede adoptar una idea distinta. En este sentido, es realmente difícil para la máquina determinar qué caracteres deben aparecer juntos y qué caracteres deben segmentarse ya que dependiendo del contexto textual se necesitará una segmentación u otra, y el sistema no tiene los conocimientos lingüísticos para determinar la separación de caracteres correcta para cada determinado caso.

En este aspecto, y contestando a otra de las preguntas de investigación, las diferencias lingüísticas de los pares de lenguas con los que trabajamos sí que pueden afectar a los resultados de la traducción. En el caso de la traducción automática, cuando se dan problemas de segmentación por las particularidades de la lengua china que acabamos de comentar, los resultados de traducción pueden verse afectados, ya que una segmentación distinta a la que el texto requiere, provoca que el sentido de la traducción llegue a ser completamente diferente y eso puede causar que el texto traducido pierda su coherencia. Asimismo, otra de las características de la lengua china es que suele hacer uso de

abreviaciones de los caracteres, por ejemplo, los términos formados por dos o más caracteres, en algunas ocasiones se reducen a un solo carácter, y en este caso, el traductor humano conocedor de la lengua china sabe identificar —incluso a veces con ciertas dificultades, ya que los caracteres aislados también cuentan con un significado propio— a qué hace referencia dicho carácter; sin embargo, la máquina puede experimentar dificultades determinado su semántica y eso puede causar que el sistema ofrezca un resultado de traducción erróneo dentro del contexto textual.

Finalmente, en respuesta a nuestra última pregunta de investigación, y tras haber evaluado ambos procedimientos y sus respectivas ventajas y desventajas, consideramos que la traducción automática, a pesar de que requiere de un largo proceso de creación y recopilación de los materiales adecuados y de calidad para su correcto funcionamiento, a largo plazo puede llegar a ser beneficiosa y a aumentar la productividad de trabajo del traductor. No obstante, siempre será imprescindible que el profesional invierta una gran cantidad de tiempo en la creación de corpus bilingües adaptados a sus necesidades de trabajo, ya que como hemos podido observar y como ya se ha comentado a lo largo de la investigación, es realmente difícil crear un motor de TAE con una calidad de traducción elevada con los recursos al alcance de cualquier persona. Asimismo, a pesar de que los corpus con los que se entrene al motor sean de calidad, siempre existirá el problema de la segmentación si se trabaja con la combinación de lenguas chino-español ya que es un hecho arraigado a las peculiaridades de la lengua origen, y por lo tanto, es muy probable que las traducciones siempre requieran de posedición para alcanzar unos resultados coherentes y de calidad, y además, esta será en la mayoría de los casos imprescindible siempre que las traducciones vayan a publicarse. Consideramos que el proceso de traducción tecnológico, si se realiza adecuadamente y si se le dedica el tiempo suficiente como para que ejecute unos buenos resultados de traducción, puede ser favorable frente al proceso de traducción humano, ya que tal y como hemos podido observar, la traducción humana con los pares de lengua chino-español y sin hacer uso de la traducción automática, también presenta sus complicaciones y en general requiere de una gran inversión de tiempo y esfuerzo para su realización.

En este aspecto, tras haber respondido a las preguntas de investigación, ahora ya podemos contestar a las hipótesis que planteábamos al comienzo del estudio y comentar en qué medida han sido ciertas o no. La primera hipótesis, que decía «el proceso de entrenamiento del motor es más largo que el proceso de traducción humana a causa de las diferencias de las lenguas de trabajo», queda en cierto modo refutada, puesto que los hechos que provocan que el proceso de creación y entrenamiento del motor sea más largo que el proceso de traducción humana, tal y como hemos comentado, vienen determinados por la falta de corpus bilingües en la combinación de lenguas empleada, y las particularidades de

la lengua tan solo afectan a el proceso de segmentación que se realiza de forma automática por la máquina. En el caso de la segunda hipótesis, «los resultados de la traducción del motor de TAE no garantizarán una comprensión íntegra y requerirán de una posedición», tal y como hemos podido observar, es cierta ya que la posedición ha sido esencial en nuestro caso, y como hemos comentado, consideramos que siempre se requerirá debido a los problemas causados por la segmentación de la lengua origen. Finalmente, en el caso de la tercera hipótesis, «el uso del motor de TAE podrá llegar a incrementar la productividad de trabajo siempre que se haya invertido mucho tiempo en su entrenamiento, en especial con la combinación de lenguas chino-español», también podemos afirmar que es cierta ya que tal y como hemos comentado, si se emplea una gran dedicación en el proceso de entrenamiento del motor, los resultados de traducción serán mejores.

Para finalizar, nos gustaría hacer una reflexión personal acerca de todas las conclusiones extraídas después de la realización de esta investigación: consideramos que las tecnologías son innovaciones que aparecen para facilitar la vida a las personas, y por ello, aplicarlas dentro del ámbito profesional siempre es un acierto ya que nos ayudarán a realizar nuestras tareas y poco a poco sacaremos un beneficio de ello. Por lo tanto, estamos convencidos de que la inversión de tiempo en los procesos tecnológicos serán ganancias a largo plazo, y por esta razón, hay que continuar creyendo en la traducción automática y en la aplicación de las tecnologías en los ámbitos profesionales.

8. Referencias bibliográficas

Abaitua, J. (1999). «Quince años de traducción automática en España». En: *Perspectives: Studies in Translatology. Número 7:2.* [en línea] Recuperado de: http://paginaspersonales.deusto.es/abaitua/konzeptu/ta/ta15.htm [Consultado el 15 de marzo de 2020].

Banchs, R. et al (2006). *Chinese-Spanish Statistical Machine Translation Experiments*. [en línea] Recuperado de: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.522.881&rep=rep1&type=pdf [Consultado el 31 de mayo de 2020].

Banerjee, S.; Lavie, A. (2005). *An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgements (.pptx). [en línea] Recuperado de:* http://www.sfs.uni-tuebingen.de/~keberle/HybridMT/Presentations/Selina_Miller_METEOR_PresentationC.pdf [Consultado el 27 de febrero de 2020].

Banerjee, S.; Lavie, A. (2007). *METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgements*. [en línea] Recuperado de: https://www.aclweb.org/anthology/W05-0909.pdf [Consultado el 26 de febrero de 2020].

Casacuberta, F.; Peris, A. (2017). «Traducción automática neuronal». En: *Revista Tradumàtica*. Número 15, pp. 66—74. [en línea] Recuperado de: https://revistes.uab.cat/tradumatica/article/view/n15-casacuberta-peris/pdf_48 [Consultado el 1 de marzo de 2020].

Centelles, J. (2013). *Machine Translation for Chinese-Spanish: Experimenting with online Statistical and Rule-Based Paradigms*. [en línea] Recuperado de: https://upcommons.upc.edu/bitstream/handle/2099.1/19031/PFC-jordiCentelles.pdf [Consultado el 2 de marzo de 2020].

Dong Fang Wang (2020). «Nei rong fu wu sheng tai chi xu jia ma, xiao du zheng zai "chi" diao hu lian wang» (内容服务生态持续加码，小度正在"吃"掉互联网). En: *China Daily*. [en línea] Recuperado de: https://caijing.chinadaily.com.cn/a/202004/17/WS5e997061a310c00b73c77df1.html [Consultado el 21 de abril de 2020].

Duoxiu, Q. (2015). "Translation technology in China". En: Chan, S. (Ed.), *The Routledge encyclopedia of translation technology*. (p. 110-116). New York.

Forcada, M. (2017). *Making sense of neural machine translation*. [en línea] Recuperado de: https://www.dlsi.ua.es/~mlf/docum/forcada17j2.pdf [Consultado el 1 de marzo de 2020].

Fu, A. (1999). The Research and Development of Machine Translation in China. [en línea] Recuperado de: http://www.mt-archive.info/MTS-1999-Fu.pdf [Consultado el 15 de marzo de 2020].

Hutchins, W.; Somers, H. (1995). *Introducción a la traducción automática*. Madrid: Visor.

KantanMT (2020). *KantanMT: All-in-one Machine Translation Platform*. [en línea] Recuperado de: https://kantanmt.com/ [Consultado el 29 de abril de 2020].

Koehn, P. (2010). *Statistical machine translation*. Cambridge: Cambridge University Press.

Koehn, P.; Knowles, R. (2017). *Six Challenges for Neural Machine Translation*. [en línea] Recuperado de: https://www.aclweb.org/anthology/W17-3204.pdf [Consultado el 15 de marzo de 2020].

Liu, Q.; Zhang, X. (2015). "Machine translation: general". En: Chan, S. (Ed.), *The Routledge encyclopedia of translation technology*. (p. 110-116). New York.

Lommel, A, et al (2014). «Multidimensional Quality Metrics (MQM): A Framework for Declaring and Describing Translation Quality Metrics». En: *Revista Tradumàtica: tecnologies de la traducció*. Número 12, pp. 455-463. [en línea] Recuperado de: https://revistes.uab.cat/tradumatica/article/view/n12-lommel-uzskoreit-burchardt [Consultado el 2 de marzo de 2020].

Martín-Mor, A.; Piqué, R. (2017). «MTradumàtica i la formació de traductors en Traducció Automàtica Estadística». En: *Revista Tradumàtica: tecnologies de la traducció*. Número 15, pp. 98-115. [en línea] Recuperado de: https://revistes.uab.cat/tradumatica/article/view/n15-martin-pique/pdf_53 [Consultado el 21 de abril de 2020].

Pajuelo, L. (2019). «Los mejores altavoces inteligentes con pantalla». En: *El País*. [en línea] Recuperado de: https://elpais.com/elpais/2019/09/26/escaparate/1569484334_024541.html [Consultado el 21 de abril de 2020].

Parra, C. (2018). «Historia de la traducción automática». *La linterna del traductor*. Número 16, pp. 23-27. [en línea] Recuperado de: http://www.lalinternadeltraductor.org/pdf/lalinterna_n16.pdf [Consultado el 2 de marzo de 2020].

Singla, K. (2015). Methods for Leveraging Lexical Information in SMT. [en línea] Recuperado de: https://www.researchgate.net/publication/279181014_Methods_for_Leveraging_Lexical_Information_in_SMT [Consultado el 15 de marzo de 2020].

SYSTRAN (2020). *What is Machine Translation? Rule Based Machine Translation vs. Statistical Machine Translation*. [en línea] Recuperado de: https://www.systransoft.com/systran/translation-technology/what-is-machine-translation/ [Consultado el 15 de marzo de 2020].

TAUS (2020). *Adequacy/Fluency Guidelines*. [en línea] Recuperado de: https://www.taus.net/academy/best-practices/evaluate-best-practices/adequacy-fluency-guidelines [Consultado el 2 de marzo de 2020].

Tiedemann, J. (2012). "Parallel Data, Tools and Interfaces in OPUS". En: *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC'2012)*. [en línea] Recuperado de: http://opus.nlpl.eu/ [Consultado el 29 de abril de 2020].

Yang, L. y Min, Z. (2015). "Statistical machine translation". En: Chan, S. (Ed.), *The Routledge encyclopedia of translation technology*. (p. 110-116). New York.

Zhou, M. (1995). *Estudio comparativo del chino el español. Aspectos lingüísticos y culturales*. [en línea] Recuperado de: https://www.tdx.cat/handle/10803/5277#page=1 [Consultado el 2 de marzo de 2020].

Anexo

Texto 1. Fragmento del texto original en chino empleado en la parte práctica del estudio (China Daily, 2020):

内容服务生态持续加码，小度正在“吃”掉互联网

以小度为代表的智能音箱，在叠加了娱乐、教育、游戏以及购物、社交等服务后，正在逐步“吃掉”互联网。

智能音箱是什么？

如果是在疫情发生前，这样的问题并不难回答，甚至说“智能助手”几乎就是标准答案。但智能音箱玩家在疫情期间的场景拓展，无疑让答案变得多元化。
各个品牌相继推出了疫情语音播报、防疫指南、心理防疫等功能。在带屏智能音箱领域有着绝对话语权的小度，先是联合凯叔讲故事、悟空识字、乂方教育等合作伙伴卡位在线教育，又在三月底一口气引入快手、抖音、B 站、优酷、全民 K 歌、喜马拉雅、荔枝直播等娱乐资源。

一连串的动作之后，智能音箱的想象空间已经被拉升至新的天花板，逐渐完成了从语音交互到平台服务的跃进。亦或者说，以小度为代表的智能音箱，在叠加了娱乐、教育、游戏以及购物、社交等服务后，正在逐步“吃掉”互联网。

01
迈向全能选手的小度

在国内的智能音箱玩家中，小度可以说是最“激进”的一派。

正如此前在文章中提到的观点，“千箱大战”后形成了 BAM 三分天下的格局，但百度、阿里、小米对于智能音箱的态度却有所差异：阿里和小米都不缺少智能音箱之外的“备胎”，百度可能是唯一进行战略坚守的智能音箱玩家，试图将智能音箱从“新物种”过渡为大众化的消费级产品。

到了 2020 年，BAM 三家的战略差异被进一步放大，典型的例子就是小度在内容、技能和交互方式上的升级，经历了“新物种”从摸索到确立的完整过程后，小度的边界俨然没有局限于“智能音箱”，而是家庭场景下的“全能选手”。

尤其是在内容层面，快手、抖音、B 站、优酷、全民 K 歌、喜马拉雅、荔枝直播等娱乐资源的引入，小度的内容体系已经全面覆盖长视频、短视频、音频、直播等形式，甚至可以说是国内最全面、最丰富的内容聚合平台：长视频领域集齐了“爱优

腾”三家资源，短视频领域吸纳了快手、抖音和好看视频，音频领域先后接入了蜻蜓 FM 和喜马拉雅，以及全民 K 歌、荔枝直播等体验性的内容……

焦点也恰恰在于此，在智能音箱的概念诞生之初时，喜马拉雅、考拉 FM 等也曾推出过自家的智能音箱品牌，但如今已经开始将精力集中到内容领域，定位于智能音箱行业的“卖水人”。也就意味着，百度为代表的技术派最终赢得了智能音箱战争的胜利，并试图构建起庞大的“内容护城河”，将智能音箱的赛点再次引向内容。
原因也不难理解，智能音箱作为一款智能硬件产品，产品质量、操作系统和内容生态仍然是影响用户体验最核心的三个要素，创新的产品形态和语音交互的技术优势，可以说是小度在上一赛段中跑赢对手的底气，也是撬动智能音箱用户需求的基石，但在当前的互联网语境下，与用户深度的连接、提高用户的忠诚度，内容依旧是不可或缺的筹码。

与之对应的是，小度的“内容护城河”并非是简单的内容整合，而是将内容与智能音箱的交互方式进行了创新性融合。比如用户观看快手、抖音等短视频时，可以隔空语音点赞、关注……

进一步延伸的话，小度思考的方向在于将语音交互、手势控制等多模态交互与内容进行深度融合，为用户创造个性化交互和沉浸式体验，进而打造一套完整的家庭娱乐系统，通过丰富的内容强化与用户的连接，并不断串联起基于家庭的多元场景。

从“智能音箱”到“全能选手”，或许才是小度在内容上持续发力的题中之意。

Texto 2. Fragmento traducido al español de forma humana, sin empleo de traducción automática:

Xiaodu, el producto que está revolucionando internet

Los asistentes inteligentes de Xiaodu están acabando con el uso de internet, ya que proporcionan servicios de entretenimiento, educación, ocio, compras y acceso a redes sociales.

Pero… ¿qué es un asistente inteligente?

Si la pregunta hubiera surgido antes de la epidemia, no cabe duda de que todos podríamos responderla, todos teníamos muy claro lo que era un asistente inteligente. Sin embargo, ahora que se ha popularizado su uso durante la epidemia, los usuarios tienen distintas opiniones.

Varias marcas de asistentes inteligentes han lanzado al mismo tiempo la función de lectura de las noticias en relación a la pandemia, de las guías de prevención y de los cuidados psicológicos, entre otros temas. De entre todos los asistentes inteligentes con pantalla, Xiaodu destaca por poder comunicarse con el usuario. Desde un primer momento, se asoció con la aplicación de audiolibros de Wang Kai, la plataforma de comprensión lectora Wukong y la aplicación educativa Yifang, entre otras. Asimismo, a finales de marzo, no tardó en incorporar servicios de entretenimiento audiovisual asociándose con plataformas como Bilibili y Youku, entre otras.

Tras estas innovaciones, se considera que este asistente inteligente ha ido un poco más allá, ya que ha logrado hacer posible la interacción del usuario y el asistente y que esa comunicación también permita involucrar a las plataformas asociadas con el dispositivo. En otras palabras, el asistente inteligente de Xiaodu está acaparando los servicios de internet ya que permite hacer uso de diferentes plataformas de entretenimiento.

El camino al éxito de Xiaodu

De entre todos los asistentes inteligentes, podría decirse que Xiaodu es uno de los más rompedores.

Después de la revolución de estos aparatos en China, se consideró que los tres mejores eran el de Baidu, el de Alibaba y el de Xiaomi. Sin embargo, estas tres empresas tienen enfoques diferentes en cuanto a los asistentes. Por un lado, los de Alibaba y Xiaomi son bastante parecidos y no cuentan con modelos superiores de la misma marca por los que puedan ser reemplazados, mientras que los asistentes de Baidu, siguen una estrategia clara; pretenden ser unos modelos innovadores y terminar popularizándose.

Con la llegada del 2020, las estrategias de las tres compañías darán un paso adelante. Un claro ejemplo de ello, son las innovaciones de Xiaodu en cuanto a sus contenidos, métodos de interacción y prestaciones. Tras tratar de crear un producto nuevo e innovador, Xiaodu no se limitará a ser un simple asistente inteligente, sino que irá más allá y ofrecerá todo tipo de servicios.

Estos servicios, como ya se ha comentado, se centran en especial en los contenidos. Xiaodu cuenta con plataformas audiovisuales de todo tipo; con vídeos largos, cortos e incluso en

directo. Podría decirse que es el asistente más completo de China, gracias a sus asociaciones con plataformas que ofrecen todo tipo de servicios de entretenimiento audiovisual.

Algunas de estas plataformas, como Himalaya y Kaola FM, también han creado sus propios asistentes inteligentes; y de hecho ahora, también están empezando a centrarse en los contenidos que ofrecen para poder crecer dentro de la industria de los asistentes inteligentes y aumentar la producción. En este aspecto, los productos de Baidu se han posicionado los primeros dentro del ámbito de los asistentes inteligentes tratando de construir un dispositivo centrado en el conjunto de contenidos.

La razón no es complicada de entender. La calidad de los asistentes inteligentes, sus sistemas operativos y su ecología sistemática son las tres características principales que intervienen en la experiencia del usuario. Las innovaciones y las ventajas que supone la interacción por voz de Xiaodu lo colocan por delante de otros asistentes inteligentes, pues, en el contexto situacional de hoy en día, el contenido es un factor indispensable en estos aparatos.

A pesar de todas las características comentadas, Xiaodu no es tan solo un aparato que integra contenido, sino que es una fusión novedosa del uso de las plataformas de entretenimiento y los altavoces inteligentes. Por ejemplo, mientras el usuario está visualizando un vídeo puede indicar que le gusta la publicación mediante el control por voz y el asistente realizará la acción.

Para profundizar más, Xiaodu se centra en la interacción por voz y el control por gestos, para así personalizar la interacción entre el usuario y el asistente y crear un ambiente de entretenimiento. Mientras que los contenidos que ofrece, fortalecen las conexiones entre los usuarios e involucran a toda la familia.

El concepto, que ha evolucionado de «asistente inteligente» a «entretenimiento completo», explica por qué Xiaodu se está centrando tanto en los contenidos.

(Traducción por APH)

Texto 3. Resultados de traducción en español ofrecidos por nuestro motor de TAE personalizado antes de incluir el glosario terminológico

El servicio durante la apuesta sobre ecología, Xiaodu está la comida de Internet ".

Con Xiaodu de un altavoz inteligente para representar en compuestos. Diversión y juegos y la educación, las compras y hasta que Servicios Sociales, está en estado de “comer” Internet.

el altavoz inteligente. ¿Qué es?

Si el brote ha pasado antes, esto no es difícil contestar preguntas, incluso sobre la inteligencia ayudante "Era casi normal. La respuesta. Pero el altavoz inteligente jugador, durante la epidemia de escena propiedad, sin duda que se la diversificación.
marca encima. De todas las desgracias brote de voz fue, 防疫 un guía, cuando funciona. en el altavoz inteligente y cañones tiene absolutamente en representación de Xiaodu, primero 凯叔 de la historia, el libre, es como leer, cristal con educación hasta ahora, escogeremos compañeros en línea 卡位 educación, y a finales de marzo en conjunto, interviene sabia que la letra “B”, quedé parada, Douyin y la población karaoke y los Himalayas espectáculo en vivo y un saveloy a los recursos.

Una serie de movimientos, el altavoz inteligente imaginación ha sido la altura de nuevo el techo, se acabado de servicio dependerá de voz en la plataforma como las. Y que, o, como de Xiaodu de un altavoz inteligente, está agravándose mutuamente:. Diversión y juegos y la educación, las compras y hasta que Servicios Sociales, está en estado de “comer” Internet.

01
hacia los Xiaodu Todopoderoso.

En China, el altavoz inteligente a casa, Xiaodu puede decir que es la facción militante ".

como había en un artículo de opinión en la guerra “,” 千箱 después se BAM -wam tres mundo, pero. Estaba tratando de Baidu, Ali, de mijo para el altavoz inteligente. Y su actitud de las diferencias. Ali mijo y no falta el altavoz inteligente más allá de la rueda de “, Baidu es la única estrategia que se mantiene. El altavoz inteligente de los jugadores, el altavoz inteligente de la” nueva especie "transición ya sabes. Por consumo de la clase de producto.

a 2020 , BAM -wam tres estrategia más ampliación de las diferencias, es un típico caso de Xiaodu en el camino, habilidades y dependerá en Gran Bretaña, tenido la “nueva especie”

de mineral de impuesto por completo de tu proceso, Xiaodu 俨然 la frontera sin limitaciones en “el” altavoz inteligente “, sino la escena familiar de los” Todopoderoso ".

Especialmente en el nivel, bien, tiemblan letra B, y Kuaishou, Douyin y la población karaoke y los Himalayas espectáculo en vivo y un saveloy a los recursos del programa de contenido, el sistema está completa y en el largo corto video, video, audio, en vivo, incluso la forma que el país más rico, más completa .y sobre la plataforma. en un video todos “.” todos los recursos Tres, corto video en 吸纳. Sabia, temblando y buena letra en video, audio, y desde EL 96.2 FM . Libélula en los Himalaya y las la población karaoke Y un saveloy 体验性. Vivo el momento.

atención no es exactamente así, el altavoz inteligente. Al concepto de nacimiento, los Himalayas, 考拉 EL 96.2 FM hasta que se presenta la casa de la marca Xiaodu Smart Speaker, pero ahora ha empezado la energía en el campo, el altavoz inteligente posición de vender el negocio. “Nadie”. Lo que significa, Baidu como representante de la tecnología para el altavoz inteligente finalmente ganar. Una victoria de la guerra, y de construir el grande “,” El foso y el altavoz inteligente 赛点 de nuevo hacia el contenido.
Por qué no es fácil de entender, el altavoz inteligente como un modelo el hardware inteligente Producto. Producto de calidad, el sistema operativo y es ecológico en la experiencia que los usuarios clave de la economía de tres formas, la innovación productos y de superioridad técnica dependerá de voz, Puedes decir que está en Xiaodu-赛段 a ganar la competencia con el fondo, también 撬动 el altavoz inteligente usuario de la demanda respondo, pero en la actual de Internet, y el usuario 语境 la profundidad de la conexión, los leales, el usuario aún es indispensable de fichas.

Con el específica, Xiaodu “El foso” No es fácil. ¿Cuáles son las palabras, sino el con el altavoz inteligente de la interoperabilidad forma de fusión. Lo opuesto. como usuario, tiemblan las notas para avanzar a corto video, pueden al vacío de voz. Increíble, atención.

Una extensión de más, Xiaodu de pensar en el presente, es señal dependerá de la interoperabilidad 模态 esperar mucho sobre la profundidad y que los usuarios publican, hizo maravillas de la interoperabilidad y albergar experiencia —que la creación de un sistema de entretenimiento, toda la familia, los ricos, y La intensificación de la conexión con los usuarios y una familia y de que estallen solidaridad en la escena.

de “el” altavoz inteligente a los “Todopoderoso”, quizá es Xiaodu en el tiempo de la fuerza de la pregunta de importa.

Texto 3. Resultados de traducción en español ofrecidos por nuestro motor de TAE personalizado después de incluir el glosario terminológico

El servicio durante la apuesta sobre ecología, Xiaodu está la comida de Internet ".

Con Xiaodu de un altavoz inteligente para representar en compuestos. Diversión y juegos y la educación, las compras y hasta que Servicios Sociales, está en estado de “comer” Internet.

el altavoz inteligente. ¿Qué es?

Si el brote ha pasado antes, esto no es difícil contestar preguntas, incluso sobre la inteligencia ayudante "Era casi normal. La respuesta. Pero el altavoz inteligente jugador, durante la epidemia de escena propiedad, sin duda que se la diversificación.
marca encima. De todas las desgracias brote de voz fue, prevención contra la epidemia un guía, cuando funciona. en el altavoz inteligente y cañones tiene absolutamente en representación de Xiaodu, primero Kaishu de la historia, el libre, es como leer, cristal con educación hasta ahora, escogeremos compañeros en línea tarjeta educación, y a finales de marzo en conjunto, interviene sabia que la letra “B”, quedé parada, Douyin y la población karaoke y los Himalayas espectáculo en vivo y un saveloy a los recursos.

Una serie de movimientos, el altavoz inteligente imaginación ha sido la altura de nuevo el techo, se acabado de servicio dependerá de voz en la plataforma como las. Y que, o, como de Xiaodu de un altavoz inteligente, está agravándose mutuamente:. Diversión y juegos y la educación, las compras y hasta que Servicios Sociales, está en estado de “comer” Internet.

01
hacia los Xiaodu Todopoderoso.

En China, el altavoz inteligente a casa, Xiaodu puede decir que es la facción militante ".

como había en un artículo de opinión en la guerra “,” mil cajas después se BAM -wam tres mundo, pero. Estaba tratando de Baidu, Ali, de mijo para el altavoz inteligente. Y su actitud de las diferencias. Ali mijo y no falta el altavoz inteligente más allá de la rueda de “, Baidu es la única estrategia que se mantiene. El altavoz inteligente de los jugadores, el altavoz inteligente de la” nueva especie "transición ya sabes. Por consumo de la clase de producto.

a 2020 , BAM -wam tres estrategia más ampliación de las diferencias, es un típico caso de Xiaodu en el camino, habilidades y dependerá en Gran Bretaña, tenido la “nueva especie” de mineral de impuesto por completo de tu proceso, Xiaodu igual que la frontera sin limitaciones en “el” altavoz inteligente “, sino la escena familiar de los” Todopoderoso ".

Especialmente en el nivel, bien, tiemblan letra B, y Kuaishou, Douyin y la población karaoke y los Himalayas espectáculo en vivo y un saveloy a los recursos del programa de contenido, el sistema está completa y en el largo corto video, video, audio, en vivo, incluso la forma que el país más rico, más completa .y sobre la plataforma. en un video todos “.” todos los recursos Ai You Tres, corto video en absorción. Sabia, temblando y buena letra en video, audio, y desde EL 96.2 FM . Libélula en los Himalaya y las la población karaoke Y un saveloy experimental. Vivo el momento.

atención no es exactamente así, el altavoz inteligente. Al concepto de nacimiento, los Himalayas, Kaola EL 96.2 FM hasta que se presenta la casa de la marca Xiaodu Smart Speaker, pero ahora ha empezado la energía en el campo, el altavoz inteligente posición de vender el negocio. “Nadie”. Lo que significa, Baidu como representante de la tecnología para el altavoz inteligente finalmente ganar. Una victoria de la guerra, y de construir el grande “,” El foso y el altavoz inteligente concordancia de nuevo hacia el contenido.
Por qué no es fácil de entender, el altavoz inteligente como un modelo el hardware inteligente Producto. Producto de calidad, el sistema operativo y es ecológico en la experiencia que los usuarios clave de la economía de tres formas, la innovación productos y de superioridad técnica dependerá de voz, Puedes decir que está en Xiaodu-etapa a ganar la competencia con el fondo, también lo siente el altavoz inteligente usuario de la demanda respondo, pero en la actual de Internet, y el usuario contexto la profundidad de la conexión, los leales, el usuario aún es indispensable de fichas.

Con el específica, Xiaodu “El foso” No es fácil. ¿Cuáles son las palabras, sino el con el altavoz inteligente de la interoperabilidad forma de fusión. Lo opuesto. como usuario, tiemblan las notas para avanzar a corto video, pueden al vacío de voz. Increíble, atención.

Una extensión de más, Xiaodu de pensar en el presente, es señal dependerá de la interoperabilidad modal esperar mucho sobre la profundidad y que los usuarios publican, hizo maravillas de la interoperabilidad y albergar experiencia —que la creación de un sistema de entretenimiento, toda la familia, los ricos, y La intensificación de la conexión con los usuarios y una familia y de que estallen solidaridad en la escena.

de “el” altavoz inteligente a los “Todopoderoso”, quizá es Xiaodu en el tiempo de la fuerza de la pregunta de importa.

4. Listado de términos desconocidos por KantanMT

防疫 ,千箱 ,百度 , 俨然 , 体验性, 考拉, 赛点, 赛段, 撬动, 语境, 模态

5. Glosario terminológico de los términos desconocidos

ZH	ES
防疫	prevención contra la epidemia
千箱	la guerra de las mil cajas
百度	Baidu
俨然	igual que
体验性	experimental
考拉	plataforma Kaola
赛点	se centra
赛段	punto de competición
撬动	está por delante
语境	contexto
模态	modalidad