

---

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

Rendón Gentil, María del Mar; Sánchez Gijón, María Pilar, dir. Combinació de les eines de reconeixement de veu i de TAO : Anàlisi d'errors i de productivitat. 2020. (1350 Màster en Tradumàtica: Tecnologies de la Traducció)

---

This version is available at <https://ddd.uab.cat/record/249922>

under the terms of the  license

Universidad Autónoma de Barcelona  
Facultad de Traducción e Interpretación  
Trabajo de Fin de Máster  
Máster en Tradumática – Curso 2019-2020

Combinación de herramientas de  
reconocimiento de voz y de TAO:  
Análisis de errores y productividad.



Alumna: M<sup>a</sup> del Mar Rendón Gentil

[mrendongentil@gmail.com](mailto:mrendongentil@gmail.com)

Tutora: Pilar Sánchez Gijón

Barcelona, julio de 2020



**Título:** Combinación de herramientas de reconocimiento de voz y de TAO: Análisis de errores y productividad.

**Resumen:** Los sistemas de reconocimiento de voz han experimentado un auge en los últimos años gracias a los avances que se están llevando a cabo en el ámbito de la Inteligencia Artificial. Sin embargo, su aplicación no goza de mucha popularidad en el contexto de la práctica de la traducción por diversos factores, entre los que se encuentra la falta de precisión. Los principales objetivos de este trabajo se centran en la descripción de la integración de los sistemas de reconocimiento automático del habla (RAH) con herramientas de traducción asistida (TAO), en la valoración de los errores que se produzcan y en la evaluación de la productividad. Para ello, se trabaja con los sistemas de RAH de Windows, Mac y memoQ, junto con las herramientas de TAO OmegaT, Memsorce y memoQ. El plan de trabajo consiste en la traducción de un mismo texto utilizando las herramientas de TAO por medio del teclado y por medio de los sistemas de RAH mencionados para hacer una comparación del tiempo empleado en la traducción de cada segmento y los posibles errores que surjan durante el proceso.

**Palabras clave:** reconocimiento automático del habla, herramienta de traducción asistida por ordenador, OmegaT, Memsorce, memoQ.

---

**Title:** Combination of speech recognition and CAT tools: error analysis and productivity

**Abstract:** Nowadays, automatic speech recognition systems (ASR) have experienced an important growth thanks to the advances in the field of Artificial Intelligence. However, they are not very popular in the context of translation practice due to several factors, including the lack of accuracy. The main objectives of this work are focused on the description of the integration of the ASR systems with computer aided translation tools (CAT), and on the evaluation of the errors produced in the translation process and productivity. To this end, we will use Windows, Mac and memoQ ASR systems along with OmegaT, Memsorce and memoQ. The work plan will include the translation of the same text by using CAT tools with the keyboard and the ASR systems mentioned above to make a comparison of the time spent in the translation of each segment and the possible errors that may arise in the process.

**Keywords:** automatic speech recognition, computer aided translation tools, OmegaT, Memsource, memoQ.

---

**Títol:** Combinació de les eines de reconeixement de veu i de TAO: Anàlisi d'errors i de productivitat

**Resum:** Els sistemes de reconeixement de veu han experimentat un augment durant els darrers anys gracies als avenços que s'han dut a terme en l'àmbit de la Intel·ligència Artificial. Tanmateix, la seva aplicació no gaudeix de gaire popularitat en el context de la pràctica de la traducció per diversos factors, entre els quals es troba la manca de precisió. Els objectius principals d'aquest treball se centren en la descripció de la integració dels sistemes de reconeixement automàtic de la parla (RAP) amb eines de traducció assistida per ordinador (TAO), en la valoració dels errors que es produeixin i en l'avaluació de la productivitat. Per dur-ho a terme, es treballarà amb els sistemes de RAP de Windows, Mac i memoQ, així com amb les eines de TAO: OmegaT, Memsource i memoQ. El pla de treball consistirà en la traducció d'un mateix text utilitzant les eines de TAO mitjançant el teclat i els sistemes de RAP esmentats. Amb els resultats, es farà una comparació del temps invertit en la traducció de cada segment i els possibles errors que hagin sorgit durant el procés.

**Paraules clau:** reconeixement automàtic de la parla, eines de traducció assistida per ordinador, OmegaT, Memsource, memoQ.

*Agradecimientos:*

*A mi tutora, por su paciencia y apoyo durante estos meses tan difíciles que hemos  
vivido.*

*A todos los profesores del máster, porque han sido una fuente de inspiración.*

*A la familia tan bonita que encontré en Barcelona, porque, aunque breve, han hecho  
que sea maravilloso.*

*Y a mi madre, porque, esté donde esté y haga lo que haga, siempre es el pilar más  
importante de mi vida.*

# Índice de contenidos

1. Introducción.....	1
1.1. Motivación y justificación del trabajo .....	1
1.2. Objetivos, hipótesis y planificación del trabajo .....	1
2. Marco teórico .....	3
2.1. Definición de los sistemas de RAH.....	3
2.2. Componentes .....	4
2.3. Funcionamiento .....	5
2.4. Limitaciones del reconocimiento del habla.....	6
2.5. Historia .....	7
2.6. Los sistemas de RAH en el ámbito de la traducción .....	10
3. Metodología.....	12
3.1. Características del texto que se va a traducir.....	15
3.2. Prueba de la memoria de traducción y glosario.....	16
3.3. Integración y creación de comandos .....	19
3.4. Condiciones especiales que se tendrán en cuenta a la hora de traducir.....	21
4. Problemas encontrados.....	22
4.1. Problemas comunes a todas las herramientas.....	22
4.1.1. Problemas de naturaleza técnica.....	22
4.1.2. Problemas de naturaleza lingüística.....	23
4.2. Problemas concretos de cada herramienta.....	25
4.2.1. Sistema de RAH de Windows .....	25
4.2.2. Sistema de RAH de Mac .....	27
4.2.3. Sistema de RAH de memoQ.....	28

5. Resultados.....	29
5.1. Resultados de las herramientas de TAO con voz .....	29
5.2. Resultados de las herramientas de TAO sin voz .....	30
5.3. Resultados de los segmentos sin etiquetas .....	31
5.4. Resultados de los segmentos con TA .....	32
5.5. Resultados de memoQ con y sin voz.....	33
6. Propuestas de mejora.....	36
7. Conclusiones .....	37
8. Bibliografía.....	40

# Índice de ilustraciones

Ilustración 1: Esquema general de un sistema de RAH (Bocos Hernández 2009).....	5
Ilustración 2: Diseño de la primera máquina de RAH (Varona Fernández 1997) .....	7
Ilustración 3: Esquema de sistemas de RAH y herramientas de TAO .....	13
Ilustración 4: Menú del sistema de RAH de la versión macOS Mojave .....	14
Ilustración 5: Menú del sistema de RAH de la versión macOS Catalina .....	15
Ilustración 6: Análisis del archivo en Memsource .....	16
Ilustración 7: Análisis del archivo en OmegaT .....	16
Ilustración 8: Análisis del archivo en memoQ .....	16
Ilustración 9: Propuesta de un término del glosario de OmegaT .....	18
Ilustración 10: Opción de glosario y segmento en Memsource.....	19
Ilustración 11: Opción de glosario y segmento en memoQ.....	19
Ilustración 12: Comandos predeterminados del sistema de RAH de Mac .....	20
Ilustración 13: Comandos personalizados en el sistema de RAH de Mac .....	20
Ilustración 14: Proceso de entrenamiento del reconocimiento de voz de Windows .....	21
Ilustración 15: Segmento en OmegaT en el que se produce un fallo .....	23
Ilustración 16: Segmento en OmegaT del que se ha omitido una parte .....	24
Ilustración 17: Segmento en Memsource en el que aparece un nombre propio mal escrito .....	24
Ilustración 18: Segmento en Memsource en el que se puede ver que "Países Bajos" no está escrito con mayúsculas .....	25

Ilustración 19: Segmento en Memsource en el que aparece "guion" escrito literalmente en lugar de con un signo .....	25
Ilustración 20: Mensaje del sistema de reconocimiento de voz de Windows .....	26
Ilustración 21: Cuadro de edición de texto del sistema de reconocimiento de voz de Windows.....	26
Ilustración 22: Mensaje que muestra el sistema de RAH de Windows cuando no entiende el comando que se le dicta.....	27
Ilustración 23: Respuesta del sistema de RAH de Windows cuando se le indica que inserte un espacio .....	27
Ilustración 24: Segmento de Memsource en el que aparecen los espacios mal .....	28
Ilustración 25: Segmento en Memsource en el que se ven las mayúsculas al insertar una etiqueta .....	28
Ilustración 26: Segmento en OmegaT en el que se muestran los espacios que aparecen al insertar etiquetas .....	28
Ilustración 27: Segmento en memoQ en el que se comprueba cómo el sistema de RAH no entiende palabras en otro idioma.....	29

## Índice de gráficos

Gráfico 1: Porcentaje de segmentos en los que el sistema falla .....	23
Gráfico 2: Porcentaje de segmentos con error homófonas .....	24
Gráfico 3: Comparativa del tiempo empleado con cada herramienta con voz.....	30
Gráfico 4: Comparativa del tiempo empleado con cada herramienta sin voz.....	31
Gráfico 5: Comparativa entre los segmentos más largos sin etiquetas.....	32
Gráfico 6: Comparativa de tiempos con TA y sin TA.....	33
Gráfico 7: Comparativa de tiempos en memoQ con y sin voz.....	33
Gráfico 8: Comparativa de tiempo de los segmentos del tercer grupo en memoQ.....	34
Gráfico 9: Comparativa de tiempos en Memsource y OmegaT con y sin voz.....	35
Gráfico 10: Comparativa de tiempo de los segmentos del segundo grupo en Memsource y OmegaT .....	36

## **1. Introducción**

### *1.1. Motivación y justificación del trabajo*

El uso de las nuevas tecnologías es más evidente cada día en nuestra vida cotidiana. Como es lógico, este uso se ha trasladado a ciertos ámbitos profesionales, aportando una serie de beneficios que se traducen en mayor comodidad o incremento de la productividad. En el campo de la traducción, la utilización de herramientas de traducción asistida por ordenador (en adelante, TAO). podría decirse que está completamente normalizado y sería casi impensable trabajar en muchos contextos sin los recursos de que disponemos actualmente. Sin embargo, hay otros tipos de tecnologías, como sería el caso de los sistemas de reconocimiento automático del habla (en adelante, RAH) que todavía no gozan de la popularidad y el interés que algunos reclaman por las aparentes ventajas que puede ofrecer a la práctica profesional de la traducción. Por este motivo, este trabajo pretende ser una contribución al estudio de los sistemas de RAH en combinación con las herramientas de TAO.

La motivación personal para la realización de este trabajo parte del deseo de profundizar sobre uno de los ámbitos de la tecnología más populares actualmente, pero desde el punto de vista y con aplicación en la traducción. La investigación de herramientas que puedan facilitarnos aún más nuestra labor se presenta como un objetivo tanto productivo, como interesante y motivador. Además, este estudio no solo pretende aportar resultados dirigidos al ámbito académico o profesional, sino que también es el resultado de la culminación del máster que hemos cursado y la puesta en práctica de muchos de los recursos y conocimientos que se han adquirido a lo largo de estos meses.

### *1.2. Objetivos, hipótesis y planificación del trabajo*

A continuación, se exponen los objetivos específicos de este TFM, que serán la guía que nos marcará muchas de las decisiones que se tomen durante el desarrollo del estudio:

- Describir cómo es la integración del reconocimiento de voz y las herramientas de TAO en la actualidad
- Comprobar la posibilidad de utilizar el reconocimiento de voz en la posesición y su eficacia a la hora de dictar posibles órdenes
- Analizar los posibles efectos del uso del reconocimiento de voz en combinación con herramientas de TAO sobre la productividad
- Evaluar los fallos y cómo afectan a la labor traductológica
- Proponer posibles soluciones y mejoras

Se parte de la hipótesis de que el uso de sistemas de RAH puede ser un valor añadido en la práctica de la traducción, ayudando a incrementar la productividad. De esta forma, se pretende investigar en qué contexto se consigue este aumento de productividad, con qué tipo de textos, qué sistema presenta menos problemas y de qué forma se produce la integración de los sistemas de RAH en las herramientas de TAO que no lo tengan integrado de forma predeterminada. Asimismo, también se quieren obtener conclusiones que puedan ser el comienzo de un futuro estudio basado en los fallos que presentan los sistemas, tanto los propios como los derivados de su aplicación en la práctica de la traducción.

El contenido del trabajo puede dividirse en cinco partes. Una primera parte teórica que expone, de forma general, conceptos sobre los sistemas de reconocimiento de voz (componentes, funcionamiento e historia) y, más específicamente, su aplicación en el ámbito de la traducción (ver apartado 2). La segunda parte conformaría el marco práctico del trabajo y presentaría el escenario en el que se desarrolla este estudio, describiendo el plan que se va a seguir, qué pruebas se van a realizar, así como las diferentes adaptaciones que hay que hacer para llevarlo a cabo (ver apartado 3). La tercera parte incluiría la exposición de los problemas encontrados durante la aplicación de los sistemas de RAH en la práctica traductológica y los datos obtenidos en cuanto a la velocidad, tanto con la voz como con el teclado (ver apartados 4 y 5). Por último, se proponen una serie de mejoras y se presentan las conclusiones extraídas de todo el proceso (ver apartados 6 y 7).

Por último, puntualizamos que la naturaleza de este trabajo y las condiciones en las que se desarrolla pueden alterar ligeramente los resultados finales. Debido a que las pruebas van a ser realizadas únicamente por la autora del trabajo, esta conoce a la perfección el texto con el que se va a trabajar. De este modo, se pretende partir del mismo punto cada vez que se utilice una herramienta de TAO o un sistema de RAH diferente. Aun así, al tratarse de una sola persona, los resultados son simplemente una aproximación y un primer acercamiento a un posible estudio futuro que contemple otras variables y que tenga en cuenta a un mayor número de traductores para, de esa forma, poder elegir diferentes textos y que el escenario del experimento sea más realista. Por estos motivos, somos conscientes de que será inevitable que se produzcan errores cuyo origen sea algún fallo al traducir o en el uso del teclado. Son fallos menores que no se tendrán en cuenta de forma concluyente en la lectura general de los resultados.

## **2. Marco teórico**

En este apartado se presentará una visión teórica de la principal herramienta sobre la que versa este trabajo, los sistemas de RAH. En concreto, se ofrecerá una definición de los sistemas de RAH y se explicarán sobre sus componentes, cómo funciona, su historia y su implicación en el ámbito de la traducción.

### *2.1. Definición de los sistemas de RAH*

Los sistemas de reconocimiento de voz son aquellos que decodifican las señales acústicas emitidas por el locutor y lo transcriben o lo utilizan como orden para otro programa. Están estrechamente relacionados con los sistemas de síntesis, conformando a su vez lo que se conoce como procesamiento de voz. A pesar de que los objetivos de ambas ramas son diferentes, comparten conceptos, formulaciones y técnicas.

A su vez, se incluirían en el ámbito de las tecnologías del habla, donde se encuadran también los sistemas de diálogo, en los que la interacción no se produce en una sola dirección, sino que existe un intercambio entre la máquina y el usuario (Llisterra 2009). Estos sistemas pueden clasificarse en base a diferentes características. Así, podríamos encontrar sistemas dependientes o independientes del locutor, atendiendo a si es necesario

entrenarlos o no; de habla continua o palabras aisladas, o según el tamaño del vocabulario de entrada (Calvo Lance 2010).

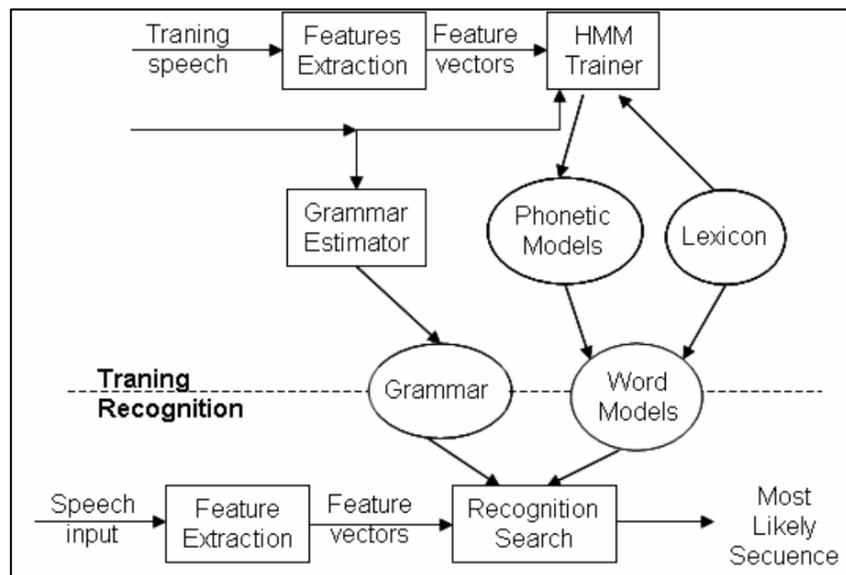
## 2.2. Componentes

Un sistema de RAH se conforma de:

- Un modelo acústico-fonético, que muestra las características acústicas de los fonemas de la lengua en cuestión. A partir de estos modelos y de una señal de entrada, se obtiene una hipótesis de lo que el locutor ha dicho. El modo de representación más común es el modelo oculto de Markov (HMM, Hidden Markov Model). Estos modelos presentan una representación de la pronunciación y mediante un proceso matemático de probabilidad en el que comparan con los patrones de un alfabeto fonético, devuelven la opción más probable (Bocos Hernández 2009). El funcionamiento de estos modelos se basa en las cadenas de Markov, que se trata de un proceso estocástico que no observa ningún evento pasado, sino el estado actual para hacer suposiciones. Este proceso de observación recibe el nombre de *decoding* o *interference* y el algoritmo que se utiliza durante ese proceso es el de Viterbi (Jurafsky y Martin 2009).
- Un diccionario de pronunciación, en el que se especifican los sonidos que componen una palabra y se representan con símbolos para que el sistema pueda usarlos a la hora de realizar la transcripción.
- Un modelo de lenguaje o gramática, que indica la relación entre las palabras. La importancia de este modelo dependerá de si nuestro sistema reconoce palabras aisladas o habla continua. Es decir, si el objetivo es reconocer una secuencia de palabras, el modelo de lenguaje será una parte clave porque mostrará la relación entre palabras que el sistema comprobará posteriormente (Bocos Hernández 2009). Este modelo es el encargado de aportar las condiciones que hacen que una secuencia de palabras sea gramaticalmente coherente. Es decir, realiza predicciones sobre la secuencia de palabras que tiene más probabilidades de ocurrir (Lleida y Ortega 2016).

### 2.3. Funcionamiento

Durante el reconocimiento del habla, los sistemas de RAH desarrollan un proceso de clasificación de patrones que se puede dividir en tres fases: adquisición de datos, preproceso de datos y clasificación por decisión. En la primera fase, se obtienen datos acústicos analógicos y se transforman a un formato digital para que, durante la segunda fase, se analicen y se decida qué representación es la más adecuada. Finalmente, en la tercera fase se termina de decidir la clasificación que se hará, en este caso, de cada uno de los fonemas.



*Ilustración 1: Esquema general de un sistema de RAH (Bocos Hernández 2009)*

Básicamente, el sistema devuelve la secuencia de palabras con mayor probabilidad teniendo en cuenta las características acústicas que se le han presentado de entrada. Así, gracias a los algoritmos utilizados, se calcula la secuencia más probable de entre todas las palabras que tiene como referencia en su vocabulario (Lleida y Ortega 2016).

Existen una serie de fenómenos que afectan al proceso de reconocimiento de voz (Varona Fernández 1997):

- La continuidad, es decir, si se hacen pausas entre sílabas o palabras.
- Las repeticiones que se efectúan en el transcurso de una conversación natural.

- La variedad de hablantes, entonación, timbre, forma de expresión, etc.

#### *2.4.Limitaciones del reconocimiento del habla*

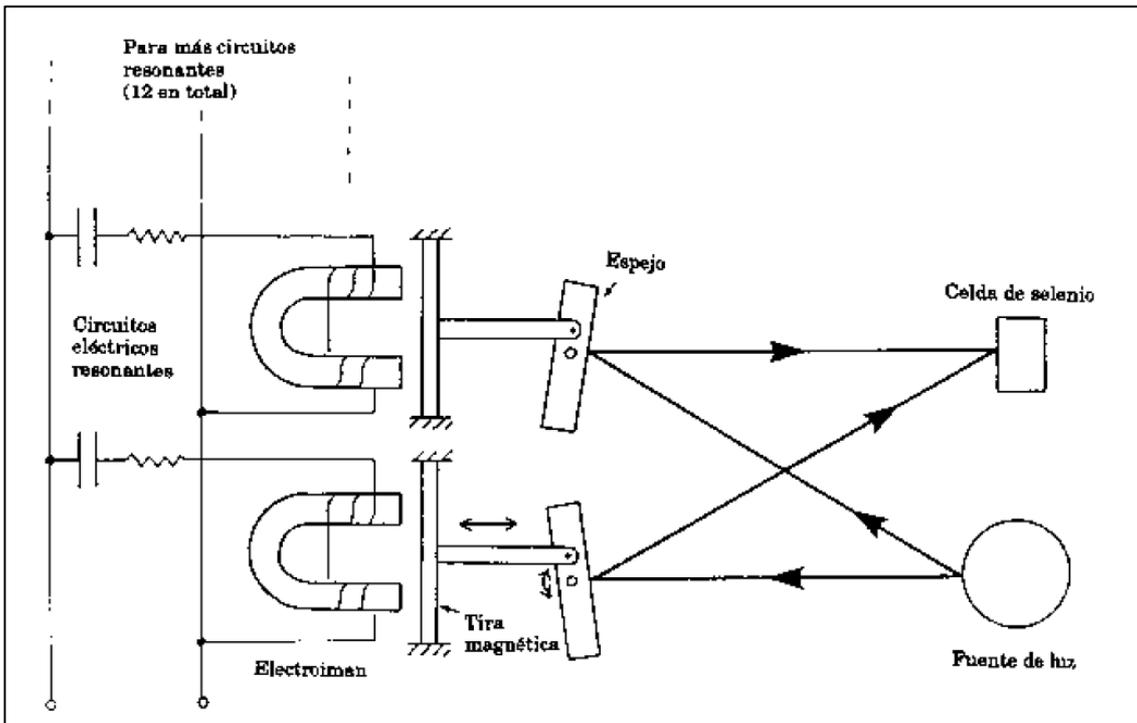
El reconocimiento del habla es una herramienta muy útil que permite la comunicación con las máquinas. Sin embargo, aún tiene que hacer frente a una serie de limitaciones que lo diferencian de la exclusiva capacidad humana y genera errores. Así, encontramos problemas relacionados con la continuidad, ya que, al hablar, no se respetan los espacios del lenguaje escrito; relacionados con las modificaciones que sufre un fonema en contacto con su fonema anterior o posterior; las diferencias del habla en función de cada locutor (acento, sexo, identidad); la necesidad de contar con una gran cantidad de espacio para poder almacenar y procesar datos (Díaz 2002).

Asimismo, en su utilización, se presentan otra serie de inconvenientes entre los que se incluyen la dificultad añadida al tener que especificar los signos de puntuación, situación a la que no estamos acostumbrados porque no es necesario al hablar o la necesidad de encontrarse en un lugar silencioso, ya que el sistema puede no distinguir entre la voz emitida por el usuario y la emitida por otras personas. Esta condición la convierte, en ocasiones, en una herramienta exclusiva en el ámbito laboral que solo puede ser utilizada por aquellos trabajadores que dispongan de un despacho privado. Además, al dictar, no es posible mantener la misma velocidad que se tiene al hablar y hay que tener siempre en cuenta que se trata de una herramienta a la que hay que dedicarle mucho tiempo, tanto en el proceso de familiarización y de entrenamiento, como a la hora de revisar aquello que se ha escrito (Gohil 2014).

Por otro lado, se ha demostrado que las nuevas tecnologías de reconocimiento de voz establecen diferencias entre sexos y procedencias. Esta situación es resultado de un problema de base social, ya que, como los motores de reconocimiento de voz se alimentan de conversaciones o discursos pronunciados en su mayoría por hombres blancos, los resultados que ofrece son más favorecedores para este perfil de persona (Tatman 2017).

## 2.5.Historia

El gran avance de los sistemas de RAH se produjo durante los años 70 con el desarrollo de los ordenadores, aunque se construyeron algunos sistemas eléctricos a principios del siglo XX. Probablemente, J.B. Flowers fue la primera persona en diseñar una máquina para transcribir voz en 1916. Inventó lo que se conocería como alfabeto fonográfico, gracias a sus conocimientos sobre la transmisión de mensajes por cable entre submarinos. Gracias a esta máquina, Flowers convertía los sonidos de las letras en ondas sin usar ningún componente eléctrico, sino dos electroimanes y condensadores (Varona Fernández 1997).



*Ilustración 2: Diseño de la primera máquina de RAH (Varona Fernández 1997)*

A partir de este momento, podrían diferenciarse cuatro generaciones de los sistemas de RAH (Lleida y Ortega 2016). La primera generación se desarrollaría desde los años 30 a los años 50 aproximadamente, período durante el que se fabricaron los primeros dispositivos que reconocían sonidos mediante sistemas electrónicos analógicos. En los próximos años, se confirman los primeros intentos de lo que conoceríamos hoy como un sistema de RAH, aunque su alcance y uso eran muy limitados.

En 1962, IBM presenta la máquina Shoebox, capaz de reconocer 16 palabras en inglés y que marcaría, junto con el uso de los primeros ordenadores, el inicio de la segunda generación. La siguiente generación surgiría a raíz del programa DARPA-SUR (*Defense Advanced Research Projects Agency-Speech*) del Ministerio de Defensa de EE. UU., a partir del cual se comenzarían a crear sistemas basados en el reconocimiento de patrones, con vocabulario de tamaño medio y dependientes del hablante (Ibídem).

No es hasta los años 80 cuando esta área de investigación empieza a cobrar mucha más importancia y dos grandes empresas, como son IBM y AT&T Bell Laboratories, comienzan a investigar sobre las aplicaciones comerciales de estos sistemas (Juang y Rabiner 2005). Aquí comenzaría la cuarta generación, caracterizada por el asentamiento de las bases matemáticas de los sistemas de RAH, con el uso de los modelos ocultos de Markov y el uso de gramáticas de n-gramas.

Descubrir e implementar los modelos ocultos de Markov en los sistemas de RAH supuso un punto decisivo en la historia de esta tecnología. Dragon Systems nació en 1982 y trabajó en pequeños proyectos privados y financiados por el gobierno hasta que en 1997 presentaron su Dragon Naturally Speaking, un software para ordenadores de uso personal que contaba con un vocabulario de 23.000 palabras. Al mismo tiempo, IBM presentó su sistema de reconocimiento de voz para dictado, ViaVoice. En los siguientes meses, siguieron a estas empresas Microsoft y Apple. Sin embargo, estos nuevos sistemas no gozaron del éxito que se esperaba por una serie de razones. En primer lugar, los usuarios habían interiorizado las acciones de teclear y hacer clic y les resultaba más difícil hablar para emitir una orden. Además, podía ser poco práctico dependiendo del lugar de trabajo, ya que se requieren condiciones especiales como la falta de ruido externo. Por otro lado, se demostró que los usuarios eran reacios a entrenar los sistemas leyendo diferentes párrafos, gracias a los cuales los sistemas se adaptarían a sus voces y alcanzarían un nivel de rendimiento satisfactorio.(Bikel y Zitouni 2012).

A partir de los 90, diversas empresas, como AT&T's Bell Laboratories, IBM's T. J. Watson Research Center y Texas Instruments' laboratories., Microsoft o Apple empezaron a invertir e investigar aún más. En Europa, había programas financiados por la Unión Europea en las que trabajaban empresas como Siemens y Philips, organizaciones nacionales de telecomunicaciones como France Telecom, Telecom Italia, Telefónica e

instituciones académicas de casi todos los países del oeste de Europa. En Japón, los estudios se llevaron a cabo por parte de Nippon Telephone Telegraph (NTT) y Mitsubishi, entre otras empresas (Ibídem).

Desde entonces, estos sistemas han seguido siendo desarrollados y las mejoras que han experimentado han conseguido su introducción en el mercado con grandes empresas. Estas mejoras se han producido gracias a la inteligencia artificial (en adelante, IA). Acuñada en 1956 por John McCarthy, la inteligencia artificial es definida por Marvin Minsky, uno de los padres de la IA, como «la ciencia de construir máquinas para que hagan cosas que, si las hicieran los humanos, requerirían inteligencia» (Escolano et al. 2003, 4). Se trata de la capacidad de las máquinas para usar algoritmos y, a través de datos, poner en práctica lo aprendido para tomar decisiones tal y como lo haría un ser humano. Esta tecnología se aplica a muchos ámbitos siendo el reconocimiento de voz uno de ellos. Además de la IA, también encontramos uno de los enfoques que la conforman, el aprendizaje automático (o *machine learning*). Se definiría como la capacidad de los ordenadores o máquinas de aprender sin estar programados para ello a partir de algoritmos de los que obtienen patrones de datos. Y, dentro del aprendizaje automático, encontramos el aprendizaje profundo (o *deep learning*), utilizado para desempeñar acciones que requieren el tratamiento de grandes cantidades de datos. Este tipo de aprendizaje utiliza redes neuronales que se organizan y reconocen patrones (Rouhiainen 2018).

Las nuevas investigaciones, dedicadas a la reducción de errores que presentan los sistemas de RAH, siguen demostrando que estos sistemas no son tan precisos como un humano. Por este motivo, diseñar y desarrollar técnicas para el reconocimiento del habla que sean mucho más precisas y eficientes se ha convertido en un auténtico desafío. Así, encontramos que el enfoque de la IA podría considerarse como una de las mayores oportunidades de conseguirlo (Haridas 2018).

Las grandes empresas como Google han implementado estas nuevas tecnologías para mejorar las funciones o aplicaciones de los sistemas de RAH que utilizan. De hecho, el software de reconocimiento de voz de Google es capaz de entender nuestro idioma hasta con un 95% de exactitud (Ponsoda 2017). Esta situación está produciendo un auge en la aplicación de sistemas de RAH. Tanto es así que, la firma de investigación Reasearch and

Markets prevé que el mercado del reconocimiento del habla crecerá de 969 millones de dólares en 2018 a 3505 millones de dólares en 2024<sup>1</sup>.

Aunque aún existen limitaciones considerables, estamos un paso más cerca de conseguir esa relación hombre-máquina que se busca desde hace tanto tiempo. Gracias al rápido desarrollo de los dispositivos móviles, Nuance se encuentra en la lista de las empresas líderes en el reconocimiento de voz y se cree que es esta misma empresa la que presta sus servicios para desarrollar a Siri (Knight 2012). Además, actualmente también gozan de gran popularidad los asistentes personales como Siri, Cortana, Google Now y Amazon Echo, aunque también podemos mencionar otros proyectos públicos, como Common Voice<sup>2</sup> de Mozilla, dedicados a crear una base datos de voces de código abierto y al desarrollo de Deep Speech, un motor de reconocimiento de voz de código abierto.

## *2.6. Los sistemas de RAH en el ámbito de la traducción*

La idea de utilizar la voz para reproducir textos es tan antigua como la propia idea de los ordenadores (Liyanapathirana et al. 2019). En este sentido, al igual que se han introducido los ordenadores y el uso de diferentes herramientas para mejorar la calidad y la productividad en la práctica de la traducción, también ha existido un interés por desarrollar y utilizar sistemas en los que la voz sea el medio de entrada o, incluso, el de entrada y el de salida. Interés que podía verse propiciado por la forma en la que se concebían (y se conciben) los ordenadores en la sociedad:

It is now clear that regardless of what people believe or know about the inner workings of computers, they talk about them and interact with them as social entities. People act toward computers as if they were people; they are polite to them, treat them as team members, and expect among other things that computers should be able to understand their needs, and be capable of interacting with them naturally (Jurafsky y Martin 2009, 8).

---

<sup>1</sup> Disponible en: <https://www.researchandmarkets.com/reports/4749380/speech-and-voice-recognition-market-by-technology>

<sup>2</sup> Disponible en: <https://voice.mozilla.org/es/about>

En el contexto de la traducción, se popularizó el uso de dictáfonos en las organizaciones internacionales durante los años 60 y 70 y, con frecuencia, los traductores colaboraban con transcritores para dictar sus traducciones (Liyanapathirana et al. 2019). A partir de los años 90, los investigadores comienzan a estudiar la combinación de los sistemas de RAH en la traducción y se empiezan a realizar los primeros estudios centrados en minimizar la tasa de error de los sistemas de RAH mediante la combinación de los primeros con modelos probabilísticos de TA (Zapata 2016).

Desde entonces, los estudios realizados han permitido sacar ciertas conclusiones a este respecto, aunque, si bien es cierto, no ha supuesto un área que haya despertado mucho interés entre los profesionales de la traducción (Zapata 2012). De entre los temas tratados podemos encontrar las ventajas e inconvenientes de utilizar los sistemas de RAH en el ámbito de la traducción y propuestas para una mejor integración (Dragsted et al. 2011), análisis de los errores que se producen con la utilización de dichos sistemas y posibles formas de corregirlos automáticamente (Errattahi et al. 2015), el uso de estos sistemas en la posesión en el seno de organizaciones internacionales (Mesa-Lao 2014, Liyanapathirana et al. 2019), el desarrollo de otras técnicas, como serían el *speech to speech* con el uso del RAH para transcribir y traducir los discursos del Parlamento Europeo (Stüker et al. 2007), así como la evaluación de la experiencia del uso de RAH por profesionales y estudiantes (Ciobanu 2014, 2016; Zapata 2016; Dragsted et al. 2011).

En general, los resultados obtenidos de dichos estudios son favorables con respecto a la introducción de los sistemas de RAH en el flujo de trabajo de los traductores. Además, en los casos en los que los participantes no lo habían probado en el ámbito profesional normalmente comenzaban las pruebas con una postura negativa. Esta percepción cambia al final del estudio haciéndolos concluir, en la mayoría de los casos, que sí harían uso de los sistemas de RAH de forma combinada con otras herramientas o con el teclado (m, Zapata 2016, Dragsted et al. 2011). De esta forma, se llega incluso a concluir que “ASR has the potential to increase the productivity and creativity of the translation act, but the advantages can be overshadowed by a reduction in translation quality unless thorough revision processes are in place” (Ciobanu 2016, 124).

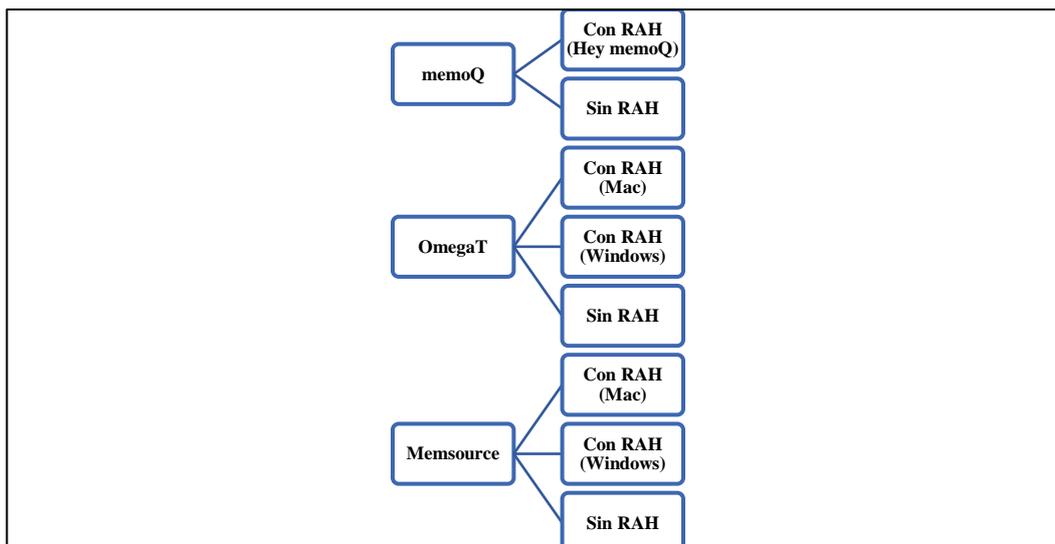
No obstante, los resultados de dichos estudios e investigaciones también dejan al descubierto las carencias o fallos que presentan los sistemas de RAH cuando se aplican

en el campo de la traducción. Entre ellos se encuentran la sustitución, eliminación o inserción de términos, lo que genera una gran falta de precisión en el texto traducido, la imposibilidad de acceder a otras funciones que ofrecen las herramientas de TAO o la incompatibilidad con las mismas, o la inversión económica necesaria para adquirir material, entre otros.

En definitiva, nos encontramos ante un escenario que, en general, presenta ventajas para la práctica de la traducción, como las mencionadas anteriormente o las relacionadas con la prevención o el alivio de ciertas afecciones médicas derivadas de la profesión (Ciobanu 2014). Sin embargo, todavía es necesario seguir investigando e invirtiendo para poder conseguir sistemas cuyos beneficios superen los fallos que puedan presentar.

### **3. Metodología**

Para la realización de este trabajo se utilizarán los sistemas de RAH de Windows, Mac y memoQ, juntos con las herramientas de TAO OmegaT, memoQ y Memsources. Los dos primeros sistemas son externos a las herramientas de TAO, mientras que la segunda está integrada. En cuanto al texto con el que se va a trabajar, este habrá sido analizado y tratado de antemano. El motivo principal de este procedimiento reside en el objetivo de anular la variable de la dificultad traductológica. Es decir, se pretende partir del punto en el que el proceso de traducción como tal (dudas de vocabulario, de género, etc.) no va a influir, por ejemplo, en el tiempo que se tarde en traducir cada segmento. De esta forma, el esquema de trabajo sería así:



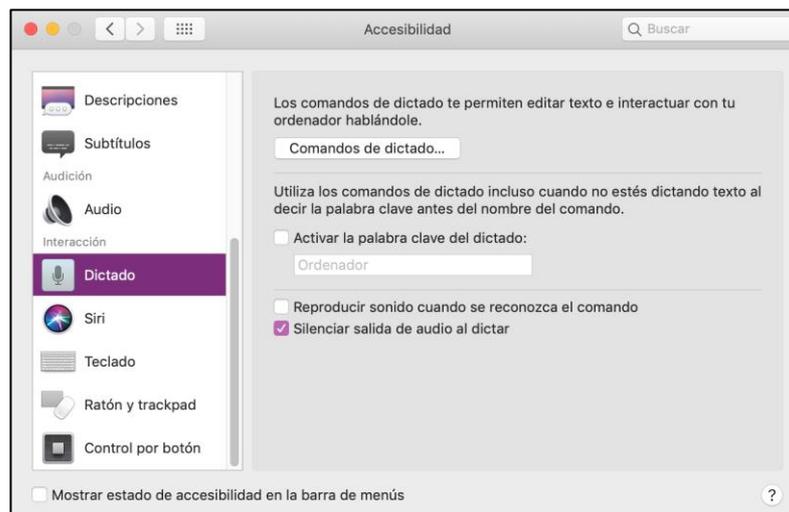
*Ilustración 3: Esquema de sistemas de RAH y herramientas de TAO*

El estudio comenzará con una previa familiarización y preparación de los diferentes sistemas y herramientas. La utilización de estos sistemas requiere la interiorización de los comandos e instrucciones propios de cada uno de ellos para que el proceso de traducción se desarrolle de la forma más fluida posible. Tras esto, se realizará una prueba para valorar si se utilizarán otras funciones disponibles en las herramientas de TAO, como la memoria de traducción (en adelante, MT) y el glosario (ver apartado 3.2). Una vez se lleven a cabo las traducciones con los diferentes sistemas y herramientas, se recogerán los datos relativos al tiempo empleado en cada segmento y se realizarán cálculos, como la distancia de edición o distancia entre palabras con la fórmula de Levenshtein. Además, se tendrán en cuenta los problemas que se hayan ido observando para hacer un posterior análisis de estos, así como las diferencias que se detecten entre los archivos traducidos con voz y sin voz con la opción de comparación de documentos de Word. Finalmente, se presentarán los resultados obtenidos y se analizarán elaborando una serie de conclusiones.

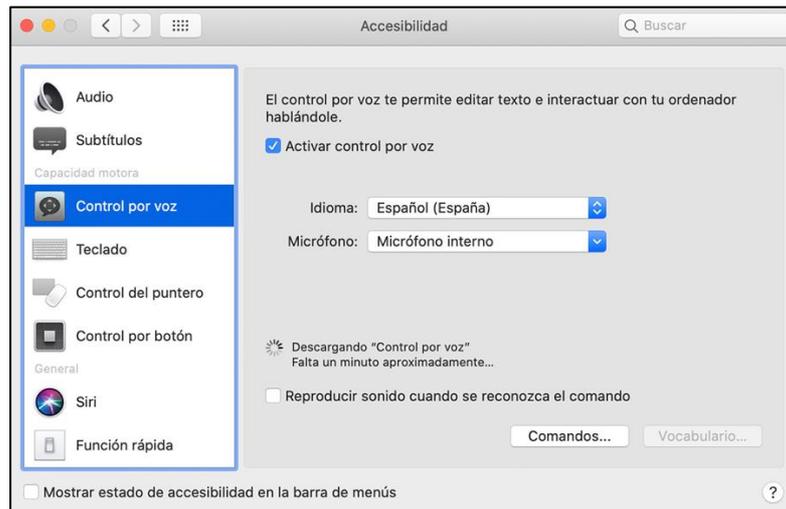
Es necesario señalar que la versión utilizada del software de Mac ha sido macOS Mojave, debido a que se han experimentado problemas con una versión más actualizada, macOS Catalina. Durante el proceso de elaboración del trabajo, se actualizó a la siguiente versión de software, lo que derivó en la incompatibilidad del sistema de RAH con las herramientas de TAO. En este caso, durante el proceso de conocimiento de las herramientas y sistemas, se contaba con la versión macOS Mojave, cuyo sistema de RAH

sí que era compatible tanto con OmegaT como con Memsource. Sin embargo, se actualizó la siguiente versión del software y se comprobó que el sistema de RAH había cambiado.

En primer lugar, la aplicación ya no se llama “dictado”, tal y como se muestra en la Ilustración 4, sino que había cambiado a “control por voz” (Ilustración 5). Por otro lado, también se comprobó que se habían incluido nuevos comandos predeterminados, añadiéndole, de esta forma, nuevas funciones. Sin embargo, al proceder a la integración con las herramientas, aunque si escribía lo que se dictaba, el sistema de RAH no respondía a los comandos personalizados (ver apartado 3.3). Por este motivo, fue necesario hacer un *downgrade* de la versión del software y recuperar la que se tenía en primer lugar. Esta situación deja las puertas abiertas a un posible estudio que compruebe en más profundidad cómo sería el funcionamiento del sistema de RAH de Mac en una versión diferente de software.



*Ilustración 4: Menú del sistema de RAH de la versión macOS Mojave*



*Ilustración 5: Menú del sistema de RAH de la versión macOS Catalina*

### *3.1. Características del texto que se va a traducir*

En este estudio, se ha utilizado un archivo .html de una página del sitio web del Consejo Europeo<sup>3</sup> en la que se expone información sobre la respuesta de la Unión Europea a la crisis del coronavirus. El archivo cuenta con una media de 2000 palabras y 280 segmentos aproximadamente (dependiendo de la herramienta de TAO que lo analice y sin tener en cuenta las repeticiones, tal y como se muestra en las imágenes que aparecen a continuación). Debido a su propia naturaleza, no se considerarán todos los segmentos a la hora de extraer resultados. Es decir, aquellos que contengan texto que no se localiza o que estén formados por una o dos palabras solamente y no aporten datos muy relevantes no se tendrán en cuenta.

---

<sup>3</sup> Disponible en: <https://www.consilium.europa.eu/en/policies/covid-19-coronavirus-outbreak-and-the-eu-s-response/european-solidarity-in-action/>

	Segments	Pages	Words	Characters	%
<b>Net Rate</b>	<b>230</b>	<b>6.36</b>	<b>1786</b>	<b>9880</b>	
All	253	6.59	1847	10249	100
Repetitions	8	0.06	16	103	0.9
101%	8	0.09	24	145	1.3
100%	5	0	0	0	0
95%-99%	5	0.12	31	185	1.7
85%-94%	2	0.04	11	65	0.6
75%-84%	2	0.08	21	125	1.1
50%-74%	34	0.87	240	1352	13
0%-49%	189	5.33	1504	8274	81.4

*Ilustración 6: Análisis del archivo en Memsource*

Estadísticas del proyecto					
	Segmentos	Palabras	Caracteres (sin espacios)	Caracteres (con espacios)	#Archivos
<b>Total:</b>	501	2,809	22,386	24,006	1
<b>Pendientes:</b>	501	2,809	22,386	24,006	1
<b>Únicos:</b>	398	2,383	17,969	19,467	1
<b>Pendientes únicos:</b>	398	2,383	17,969	19,467	1

*Ilustración 7: Análisis del archivo en OmegaT*

Tipo	Segmentos	Palabras de origen	Caracteres de origen	Etiquetas de origen	Porcentaje
Todos	270	1865	10134	619	100,00
Pretraducido basado en documentos/Contexto doble	0	0	0	0	0,00
Repetición	7	40	188	1	2,14
101%	0	0	0	0	0,00
100%	0	0	0	0	0,00
95%-99%	0	0	0	0	0,00
85%-94%	0	0	0	0	0,00
75%-84%	0	0	0	0	0,00
50%-74%	0	0	0	0	0,00
Sin coincidencias	263	1825	9946	618	97,86

*Ilustración 8: Análisis del archivo en memoQ*

### *3.2. Prueba de la memoria de traducción y glosario*

Con el fin de poder analizar las funciones disponibles en las herramientas de TAO, se han preparado un archivo .html con 12 segmentos (de diferente tamaño: 1-5 palabras, 7-15 palabras y +15 palabras), junto con una pequeña MT y un glosario. Con esta prueba se pretende comprobar la operabilidad de los sistemas de RAH con algunas de las funciones que ofrecen las herramientas de TAO y si existe un aumento de la productividad. De esta forma, servirá para decidir si, a la hora de realizar el estudio, se introducirán las propuestas de la MT y del glosario o si, por el contrario, se traducirá sin insertar ninguna de las opciones que aparezcan.

En este caso, la prueba consistía en traducir dos veces los segmentos del archivo creado con cada herramienta utilizando la voz: la primera vez insertando las opciones propuestas y editándolas, en caso de que fuera necesario y, la segunda vez, sin insertar ninguna de las propuestas, sino que simplemente dictando la traducción utilizando como referencia las propuestas de la memoria y el glosario.

En la elaboración del archivo .html que se iba a utilizar para la prueba, se han tenido en cuenta diferentes supuestos que se pueden dar cuando se insertan las opciones propuestas por la MT. Entre estos supuestos, encontramos:

- El segmento contiene algún término que aparece en el glosario.
- La memoria de traducción propone un *fuzzy match* para un segmento largo.
- La memoria de traducción propone un *fuzzy match* para un segmento corto.
- La memoria contiene propone la traducción correcta del segmento en su totalidad, pero sin incluir las etiquetas.

Es necesario señalar que se ha omitido el supuesto en el que la MT contiene la traducción correcta del segmento completo, ya que se da por hecho que el incremento de productividad es evidente al no tener que editar la propuesta.

Los resultados de la prueba son diferentes en función de las funciones de cada herramienta, así como de la longitud y las características de los segmentos que se van a traducir. En primer lugar, ninguna de las herramientas experimenta una reducción de tiempo tan significativa comparado con el aumento de las probabilidades de que se produzca un fallo. Es decir, el proceso de insertar la propuesta de la MT y editarlo conlleva la utilización de muchos más comandos con los que es más probable que haya algún error. Esto es, aquellos segmentos que son largos y solo necesitan la inserción de etiquetas si que presentarían, por ejemplo, una reducción del tiempo empleado en su traducción. Sin embargo, aquellos segmentos cortos que necesitan la edición de alguno de sus términos presentarían unos resultados contrarios a los mencionados. Además, para que los resultados influyeran (de manera positiva o negativa) realmente en la

productividad, sería necesario contar con un mayor número de segmentos en la MT y, por tanto, un archivo que contara con más texto.

Finalmente, para disminuir los posibles problemas ocasionados y poder cumplir con los objetivos principales del estudio, se decidió no recurrir a la inserción de las propuestas de la MT. No obstante, gracias a la prueba, se han podido comprobar ciertos aspectos que parecen relevantes a la hora de hablar sobre la integración y compatibilidad de los sistemas de RAH y las herramientas de TAO.

En lo que respecta a OmegaT, la opción del glosario propone los términos a medida que se van escribiendo las palabras. Por lo tanto, sería necesario escribir la primera letra del término en cuestión para que la herramienta lo sugiriese y poder introducirlo. En cuanto a la utilización de esta función con el sistema de RAH, nos encontramos con el problema de que la mayoría de las veces la propuesta ni siquiera aparece porque las palabras se escriben directamente, sin dejar que el glosario haga alguna propuesta. Y, por otro lado, en los casos en los que sí aparecía alguna opción propuesta (porque el término estaba formado por dos o más palabras y aparecía la propuesta al escribir la primera palabra), es imposible seleccionar la opción deseada ya que no existe un comando para ello.



*Ilustración 9: Propuesta de un término del glosario de OmegaT*

Esto no ocurre por ejemplo en Memsourc y en memoQ, donde tanto las propuestas de la memoria de traducción como las del glosario pueden elegirse utilizando comandos, tal y como se muestra en la Ilustración 10 y la Ilustración 11 a continuación.

CAT			
1	COVID-19	TB	COVID-19
2	Foreign Affairs	TB	Asuntos Exteriores
COVID-19, Foreign affairs & international relations		COVID-19, Asuntos Exteriores y relaciones internacionales	

*Ilustración 10: Opción de glosario y segmento en Memsource*

COVID-19, Foreign affairs & international relations	COVID-19, asuntos exteriores y relaciones internacionales
<div style="border: 1px solid gray; padding: 5px;"> <p>Resultados</p> <p>COVID-19 1 COVID-19</p> </div>	

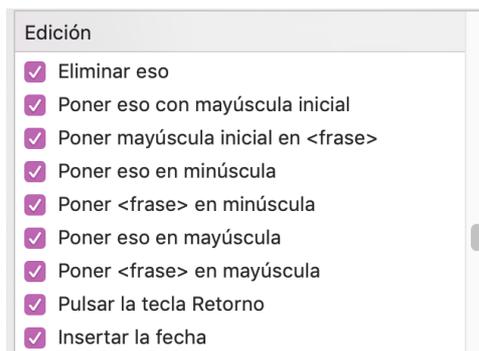
*Ilustración 11: Opción de glosario y segmento en memoQ*

En el caso de las pruebas que se realizarán con TA, se seguirá la misma dinámica que la mencionada. Es decir, no se insertará la propuesta de traducción, sino que se recurrirá a ella como referencia, a excepción de los casos en los que no se necesite ningún tipo de edición. En este último caso, se añadirá la propuesta directamente, confirmando que la traducción es correcta.

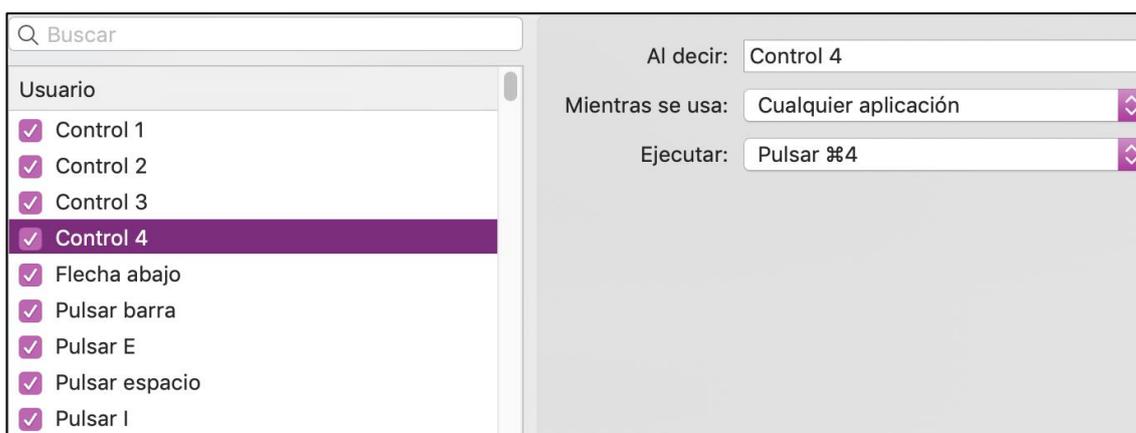
### *3.3. Integración y creación de comandos*

La integración de los diferentes sistemas de RAH con las herramientas de TAO es diferente y necesita de condiciones específicas en cada caso.

En cuanto a Mac, se han incorporado como comandos algunos de los atajos de teclado que se utilizan durante el proceso de traducción. Esto ocurre porque el sistema no incluye ningún comando para pulsar teclas específicas, lo que sería necesario si se quiere pasar al siguiente segmento, introducir etiquetas, introducir el segmento original, etc. La Ilustración 12 muestra algunos de los comandos que incluye el propio sistema (aunque no todos son compatibles con las herramientas de TAO, como se explica en el apartado 4.1.2) y, en la Ilustración 13, se muestran los comandos que han sido creados específicamente para este trabajo.



*Ilustración 12: Comandos predeterminados del sistema de RAH de Mac*



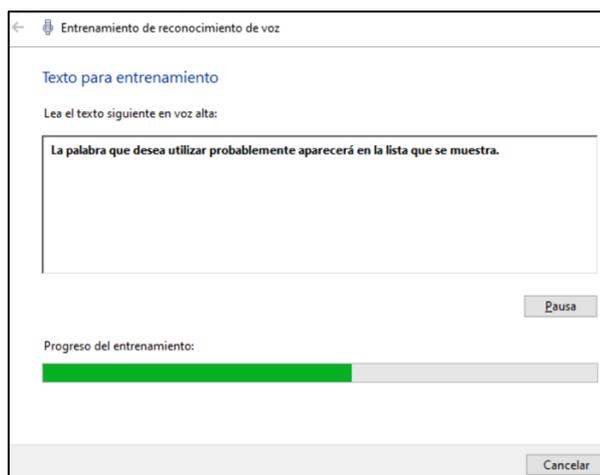
*Ilustración 13: Comandos personalizados en el sistema de RAH de Mac*

Los comandos personalizados son los siguientes:

- “Pulsar e”, para introducir una etiqueta.
- “Pulsar l”, para introducir como traducción el segmento original.
- “Pulsar intro”, para confirmar el segmento y pasar al siguiente.
- “Control 1/2/...”, para introducir las opciones que ofrece la herramienta de TAO, tanto del glosario, como de la MT o la TA.
- “Pulsar barra”, para introducir un espacio de no separación.

Estos comandos han sido los mismos tanto para Memsorce, como para OmegaT, ya que lo único que se ha modificado ha sido la combinación de teclas a la que iban asociada. A este respecto, habría que señalar que la versión web de Memsorce no permite modificar sus atajos de teclado, a diferencia de su versión de escritorio. Por este motivo, ha sido necesario realizar las traducciones en la aplicación de escritorio Memsorce Editor.

Por su parte, Windows permite utilizar las combinaciones de teclas como si se utilizara el teclado simplemente diciendo “pulsar tecla(s)”, aunque, a diferencia de Mac, recomienda realizar un proceso de entrenamiento del sistema, como se muestra en la siguiente imagen, para que se acostumbre a la voz y la forma de hablar del locutor. En cuanto a los comandos que incluye este sistema, pueden consultarse en la página de ayuda al usuario<sup>4</sup>



*Ilustración 14: Proceso de entrenamiento del reconocimiento de voz de Windows*

En el caso de memoQ, al ser un sistema que integra la propia herramienta, no necesita de ninguna preparación previa. Sin embargo, el dictado se hace por medio de una aplicación, Hey memoQ<sup>5</sup>, que solo es compatible con teléfonos iPhone, por lo que es necesario contar con uno si se quiere utilizar esta función. Por lo que respecta a los comandos, también pueden consultarse en su respectiva página de ayuda al usuario<sup>6</sup>.

### *3.4. Condiciones especiales que se tendrán en cuenta a la hora de traducir*

Después de completar el proceso de familiarización y aprendizaje con los sistemas y herramientas, se tomaron una serie de decisiones que condicionarán el estudio y los resultados obtenidos. Así pues, no se corregirá ningún fallo que pueda aparecer en los

---

<sup>4</sup> Disponible en: <https://support.microsoft.com/es-es/help/12427/windows-speech-recognition-commands>

<sup>5</sup> Disponible en: <https://apps.apple.com/us/app/hey-memoq/id1440587736>

<sup>6</sup> Disponible en: <https://docs.memoq.com/current/en/hey-memoQ/hey-memoq.html>

segmentos traducidos, a excepción de los casos en los que el sistema de RAH falle y deje de funcionar. En estos casos, concretamente, se procederá a borrar lo escrito y comenzar de nuevo o reiniciar el sistema para poder continuar, valorando la opción más adecuada según la situación. En cuanto a las etiquetas y los signos ortográficos, se ha observado que el sistema no introduce un espacio cuando se inserta alguno de ellos. Por este motivo, se ha decidido invertir un poco más de tiempo para especificar la presencia de un espacio siempre entre el final de una palabra y un signo ortográfico o una etiqueta. Esta última situación solo ocurre con Memsources y OmegaT, ya que en memoQ no existe problemas con los espacios porque no permite insertar etiquetas (ver el apartado 4.2.2).

## **4. Problemas encontrados**

En este apartado procederemos a comentar los problemas, tanto comunes como propios de cada sistema, que se han ido presentando a lo largo del estudio.

### *4.1. Problemas comunes a todas las herramientas*

En primer lugar, los problemas comunes en todas las herramientas podrían dividirse en técnicos y lingüísticos. Dentro de este apartado cabe destacar que, aunque los problemas son clasificados como comunes, ello no quiere decir que ocurran en los mismos casos o con la misma frecuencia en cada herramienta. Estos problemas se dividen, a su vez, en técnicos y lingüísticos:

#### **4.1.1. Problemas de naturaleza técnica**

El principal problema técnico que se presenta ocurre cuando el sistema no responde a un comando y hay que repetirlo de nuevo, cuando borra lo que ha escrito y hay que volver a dictarlo o cuando deja de funcionar y hay que reiniciarlo. Los tres casos se dan en todas las herramientas, aunque no con la misma frecuencia, tal y como se muestra en el siguiente gráfico, que refleja que Memsources es la herramienta que presenta más problemas de este tipo.

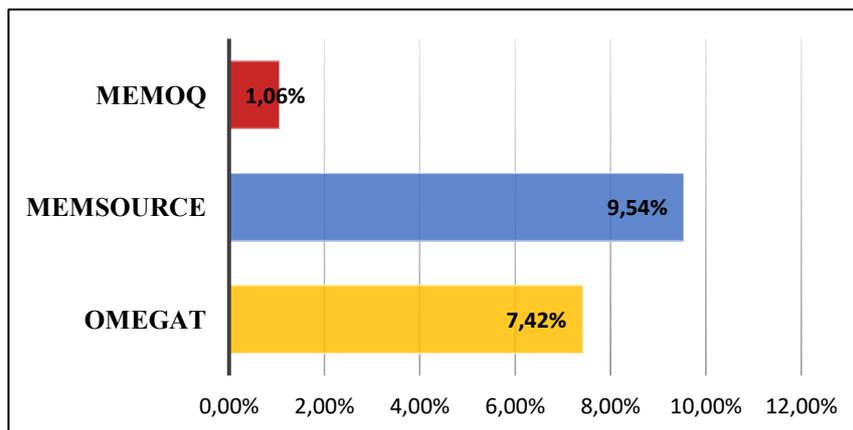


Gráfico 1: Porcentaje de segmentos en los que el sistema falla

La Ilustración 15 refleja lo que le sucede a OmegaT cuando el sistema de RAH se bloquea y deja de funcionar. Suele subrayar una parte de la frase y, o bien es imposible continuar y hay que reiniciarlo, o bien borra esa parte que está subrayada de forma inesperada mientras el locutor sigue dictando, haciendo que se pierda parte del segmento traducido.

The <s0>General Secretariat of the Council</s0> is a body of staff responsible for assisting the European Council and the Council of the EU.  
 La <s0>secretaría general del Consejo</s0> es un órgano administrativo<segmento 0256 ¶>

Ilustración 15: Segmento en OmegaT en el que se produce un fallo

En estos casos, el problema se resuelve fácilmente reiniciando la herramienta y continuando el dictado por el lugar donde se dejó o, si existen muchos errores u omisiones, borrando el texto y comenzando de nuevo.

#### 4.1.2. Problemas de naturaleza lingüística

Dentro de este apartado encontramos varios tipos de errores. En primer lugar, se encontraría la confusión de palabras homófonas o *speaks*<sup>7</sup>. Tal y como muestra el gráfico, este problema no se da tampoco con la misma frecuencia entre memoQ, OmegaT y Memsources, pero sí entre estas dos últimas.

<sup>7</sup> Se trata del término con el que uno de los participantes en el estudio de Ciobanu (2014) designa a los errores producidos entre palabras homófonas o como fruto de dudas o vacilaciones, ruido ambiental o cualquier fallo del hardware o software.

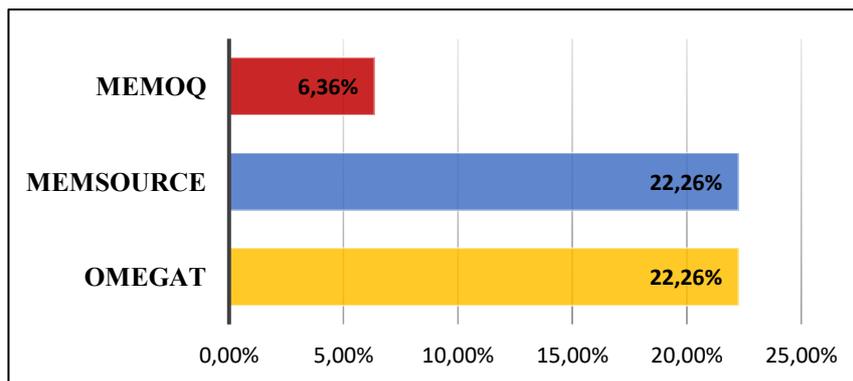


Gráfico 2: Porcentaje de segmentos con error homófonas

Algunas de estos términos serían:

Plurianual	Pluri ya no más
UE	QT
Respiradores	Tres tiradores
Codecisión	Con decisión
Viseras	Vísceras
Luxemburgo	Los engordo

En segundo lugar, encontramos segmentos en los que el sistema ha omitido parte del texto, tal y como se muestra en la siguiente imagen:

**The EU and member states make over €20 billion available for partners**  
 La UE y los estados miembros ponen a disposición de sus socios más de euros.  
 <segmento 0398 ¶>

Ilustración 16: Segmento en OmegaT del que se ha omitido una parte

Por otro lado, también se comprueba que no se transcriben correctamente los nombres propios.

201	Josep Borrell, EU High Representative for Foreign Affairs and Security Policy	Yo se Borrell, alto representante de la unión para asuntos exteriores y política de seguridad	✓
-----	---	---	---

Ilustración 17: Segmento en Memsorce en el que aparece un nombre propio mal escrito

Además, no es posible marcar las mayúsculas de los términos que las lleven (a excepción de las siglas que pueden dictarse letra por letra), por lo que, si el sistema no identifica que una palabra se escriba con mayúscula inicial, no la introduce. En el caso del sistema de RAH de Mac, aunque contempla entre sus comandos originales el poder escribir una palabra con mayúscula inicial, la realidad es que no responde a tal orden. (ver Ilustración 12).

175	Luxembourg and Germany are sending their air ambulances, pilots and specialised staff to help Italy, France and the Netherlands. 1 in their struggle to save lives. 1	Los engordo y Alemania Están enviando sus aviones sanitarios, pilotos y personal especializado para ayudar a Italia, Francia y los países bajos. 1 En su lucha por salvar vidas 1.	✓
-----	---	--	---

*Ilustración 18: Segmento en Memsorce en el que se puede ver que "Países Bajos" no está escrito con mayúsculas*

Por último, en varias ocasiones el sistema de reconocimiento de Mac escribe los signos ortográficos de forma literal.

160	Timeline--Council actions on COVID-19	Cronología Guión Actuaciones del Consejo en relación con la COVID-19	✓
-----	---------------------------------------	--	---

*Ilustración 19: Segmento en Memsorce en el que aparece "guion" escrito literalmente en lugar de con un signo*

#### 4.2. Problemas concretos de cada herramienta

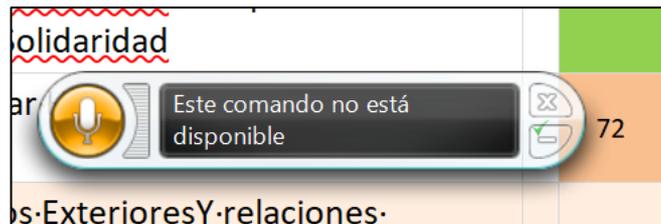
En este apartado se exponen los problemas que presenta cada herramienta de forma individual y que,

##### 4.2.1. Sistema de RAH de Windows

En cuanto al uso del sistema de reconocimiento de voz de Windows, es necesario mencionar la incompatibilidad parcial a la hora de combinarlo con OmegaT y Memsorce. En lo que respecta a la primera herramienta, el sistema es capaz de interpretar las órdenes de los diferentes comandos, pero no escribe cuando se le dicta. En

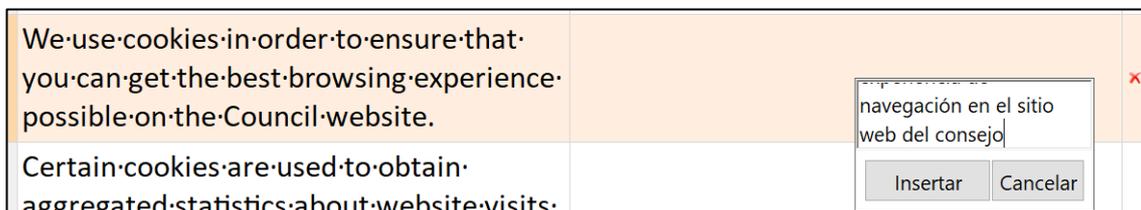
el caso de Memsource, se presentan una serie de problemas que incapacitan el desarrollo del proceso de traducción por completo.

En primer lugar, el sistema permite seleccionar el texto de un segmento, pero no responde cuando se le ordena que lo borre, mostrando un mensaje que indica “Este comando no está disponible”.



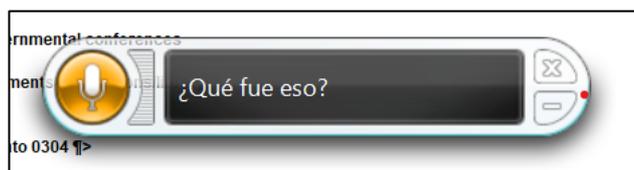
*Ilustración 20: Mensaje del sistema de reconocimiento de voz de Windows*

Por otro lado, cuando se dicta un texto, la respuesta del sistema debería ser mostrar un cuadro de edición de texto en el que va apareciendo el texto dictado para poder insertarlo posteriormente, como el que puede verse en la Ilustración 21.



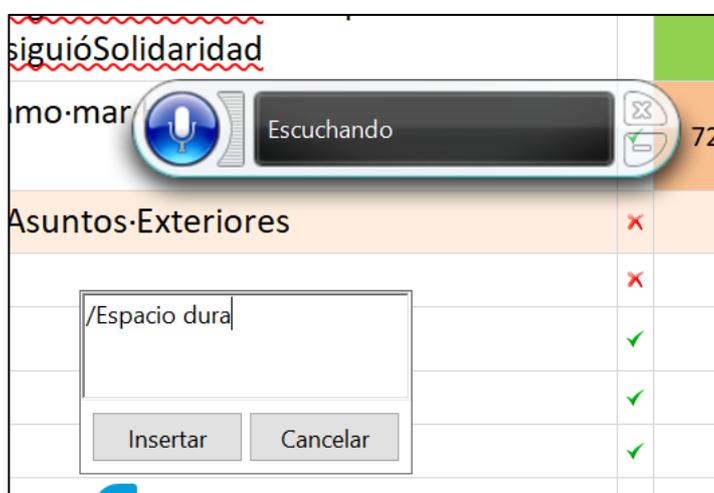
*Ilustración 21: Cuadro de edición de texto del sistema de reconocimiento de voz de Windows*

El problema reside en que, en ocasiones, el sistema no reconoce el segmento como un lugar de edición y no muestra el cuadro de edición de texto para poder dictar la traducción. Cuando esto ocurre, aparece un mensaje, como el que se ve en la Ilustración 22, que indica al usuario que el sistema no ha entendido la última orden o comando que se le ha dicho. Otras muchas veces, el sistema muestra el cuadro de edición, pero no es capaz de captar todo el texto que se le dicta.



*Ilustración 22: Mensaje que muestra el sistema de RAH de Windows cuando no entiende el comando que se le dicta*

Además, al intentar introducir algún signo ortográfico o, como en este caso, la barra espaciadora, no lo reconoce y lo intenta escribir literalmente, como puede verse en la Ilustración 23.



*Ilustración 23: Respuesta del sistema de RAH de Windows cuando se le indica que inserte un espacio*

Estos problemas de incompatibilidad nos han llevado a descartar al sistema de RAH de Windows como objeto de estudio para este trabajo, ya que imposibilitan la práctica de la traducción de manera fluida.

#### 4.2.2. Sistema de RAH de Mac

En el caso de Memsorce, lo que ocurre es que, si no se dicta el texto muy rápido y el sistema interpreta una pausa, por pequeña que sea, escribe las palabras sin espacios y con mayúscula inicial. En el caso de OmegaT, también ocurre en algunos casos, aunque no de forma sistemática como en Memsorce, sino que suele observarse sobre todo con las conjunciones “y”, “e” y “o”.

14	With your permission, we will use AT internet cookies to produce aggregated, anonymous data about our visitors' browsing and behaviour on our website.	Con su autorización. Como, Utilizaremos cookies de AT internet Para generar datos agregados y anónimos. Es una relación Y comportamiento en el.	✓
----	--	---	---

*Ilustración 24: Segmento de Memsorce en el que aparecen los espacios mal*

Por otro lado, al insertar las etiquetas, se observa otro error en ambas herramientas, aunque no de la misma forma. Esto es, ninguna de las dos introduce un espacio entre el final de una palabra y la etiqueta. En el caso de Memsorce, siempre se escriben con mayúscula inicial las palabras que siguen a la etiqueta (ver Ilustración 25), mientras que, en OmegaT, se añade un espacio justo detrás de la etiqueta de apertura (Ilustración 26).

The press office holds 1 press events 1, offers 2 audiovisual coverage 2 of major events and provides 3 facilities for journalists 3.	El servicio de prensa celebra 1 Ruedas de prensa 1. Ofrece 2 Cobertura audiovisual 2. De los actos importantes y proporciona 3 Instalaciones para los periodistas 3.
---	--

*Ilustración 25: Segmento en Memsorce en el que se ven las mayúsculas al insertar una etiqueta*

<p><b>You can get in contact to arrange a &lt;s0&gt;visit&lt;/s0&gt;, ask &lt;s1&gt; questions &lt;/s1&gt; about the work of both institutions, and request a &lt;s2&gt; document, &lt;/s2&gt; among other services.</b></p> <p>Puede ponerse en contacto con la sede para concertar una &lt;s0&gt; visita &lt;/s0&gt;, formular &lt;s1&gt; preguntas &lt;/s1&gt; sobre la labor de ambas instituciones. O solicitar un &lt;s2&gt; documento &lt;/s2&gt;, entre otros servicios. &lt;segmento 0336 ¶&gt;</p>
--

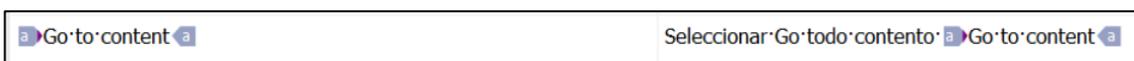
*Ilustración 26: Segmento en OmegaT en el que se muestran los espacios que aparecen al insertar etiquetas*

### 4.2.3. Sistema de RAH de memoQ

En lo que respecta a este sistema, presenta una serie de problemas relacionados con acciones básicas que dificultan la traducción de textos que contienen marcas de formato y etiquetas, como es el caso de este estudio.

En primer lugar, el sistema no permite insertar las etiquetas en el segmento de destino. Se podría pensar que una solución válida sería copiar el segmento de origen para insertarlas y, a continuación, reemplazar el texto origen por su traducción. Sin embargo, esto no es posible porque el sistema no reconoce el texto origen en muchos casos. Como se muestra en la Ilustración 27, se insertó el segmento original y luego se intentó seleccionarlo para borrarlo, pero no fue posible. Por eso mismo, puede leerse “Go todo contento”, porque el sistema no ha reconocido las palabras en inglés. Además, ocurriría

lo mismo con los espacios de no separación. Esto también ocurre de igual modo con los algunos signos ortográficos, el espacio normal o el espacio de no separación.



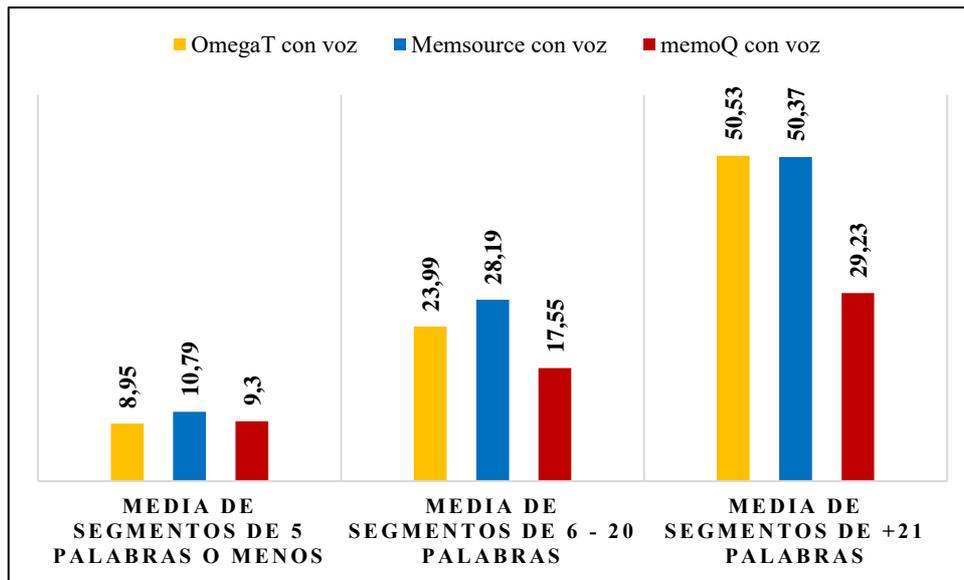
*Ilustración 27: Segmento en memoQ en el que se comprueba cómo el sistema de RAH no entiende palabras en otro idioma*

## **5. Resultados**

En este apartado se comentarán los resultados que se han obtenido durante las pruebas relativos al tiempo empleado en la traducción de cada segmento. Para ello, se dividirán los segmentos en tres grupos atendiendo a su longitud (de 0-5 palabras, de 6-20 palabras y +21 palabras). De esta forma, en primer lugar, se comentarán los resultados de forma general, haciendo una comparación entre los sistemas y las herramientas utilizadas y también en aquellos casos en los que no se ha usado ningún sistema de RAH. Después, se comentarán los resultados relativos a los segmentos sin etiquetas, para determinar si es rentable utilizar los sistemas de RAH en este tipo de textos. Por otro lado, se analizarán también los datos relativos a la TA, para poder concretar la utilidad de los sistemas de RAH en combinación con la TA. Por último, se analizarán los datos de memoQ, por un lado, y Memsource y OmegaT, por otro, tanto con voz como sin ella, ya que no se ha utilizado el mismo sistema de RAH y creemos necesario establecer comparaciones entre las herramientas para conocer los fallos y los posibles niveles de productividad.

### *5.1. Resultados de las herramientas de TAO con voz*

Como puede verse en el Gráfico 3, la diferencia entre las herramientas varía en función de la longitud de los segmentos. Sin embargo, tanto OmegaT como Memsource siguen prácticamente la misma línea, mientras que memoQ parece ser la herramienta que emplea menos tiempo.

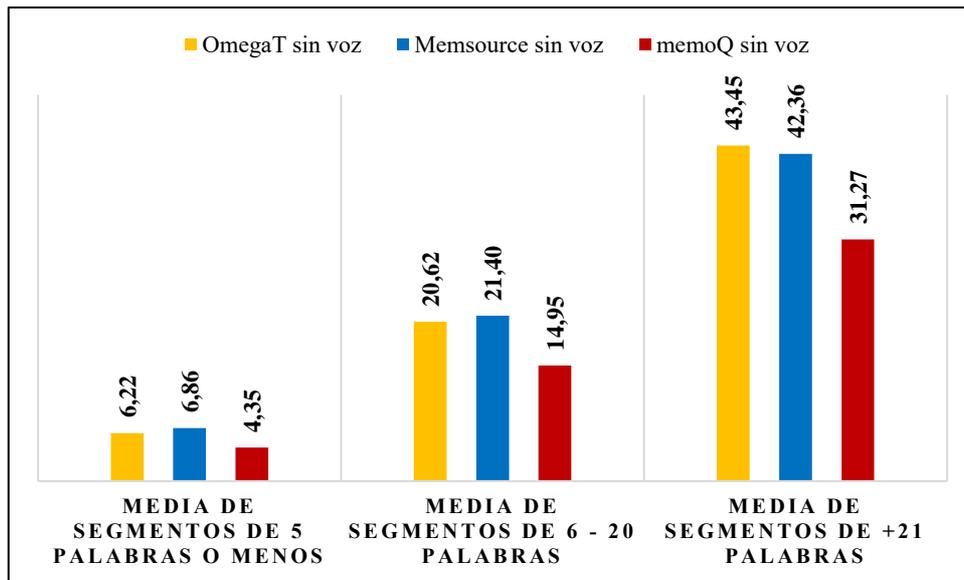


*Gráfico 3: Comparativa del tiempo empleado con cada herramienta con voz*

No obstante, habría que aclarar ciertos aspectos sobre este resultado. Si bien está claro que memoQ es la herramienta cuyo sistema de RAH emplea menos tiempo en traducir cada segmento, también habría que tener en cuenta las carencias que van a contener los mismos. Es decir, la herramienta permite que se dicte haciendo menos pausas sin el riesgo de que el sistema de RAH se bloquee o borre lo que ha escrito (situaciones que se dan en las otras dos herramientas y que se ha comentado en el apartado 4.1.1), pero no permite, por ejemplo, añadir etiquetas o espacios de no separación. Por ese motivo, encontramos las mayores diferencias de tiempo en el tercer grupo, ya que está compuesto por segmentos más largos, que precisan de más pausas y que contienen etiquetas.

### *5.2. Resultados de las herramientas de TAO sin voz*

En este caso, también destacamos la diferencia de memoQ por encima del resto de herramientas. Puede verse cómo Memsorce y OmegaT se encuentran prácticamente igualados. A pesar de estas diferencias, se trata de un caso bastante curioso y es necesario aclarar que esta diferencia en los resultados entre herramientas no responde a ningún motivo en concreto.

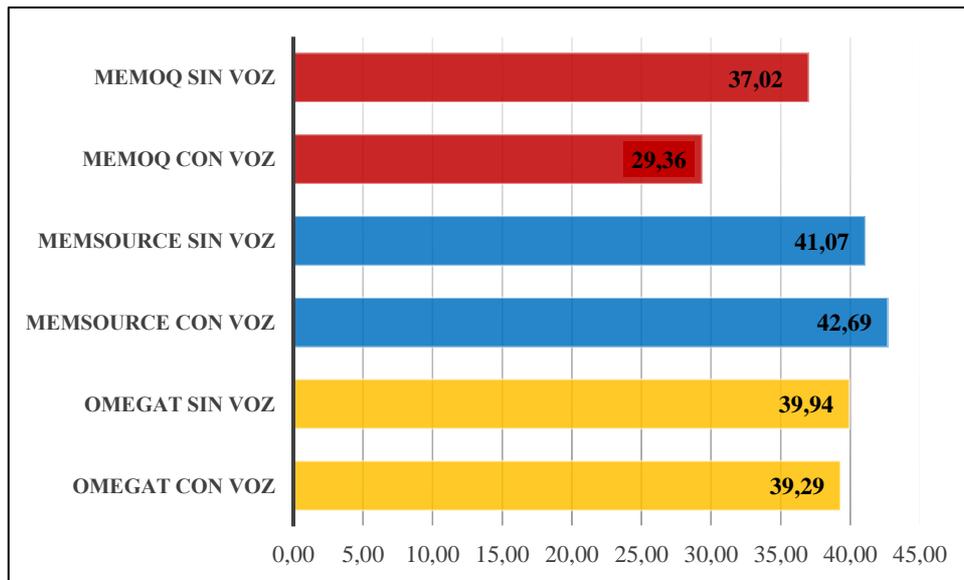


*Gráfico 4: Comparativa del tiempo empleado con cada herramienta sin voz*

Es decir, las condiciones en las que se realizaron las pruebas fueron las mismas y no se dio ninguna situación que pudiera originar un resultado diferente entre herramientas. Sin embargo, tal y como se explicó al comienzo del trabajo, la naturaleza del mismo y el contexto en el que se desarrolla tienen unas características que podrían provocar la alteración parcial de algunos resultados, como sería el caso. Siendo conscientes de estos detalles previamente, comentamos este dato en concreto para aclarar posibles dudas, pero sin considerarlo como concluyente.

### *5.3. Resultados de los segmentos sin etiquetas*

La siguiente gráfica muestra la comparación del tiempo empleado en los segmentos de más de 21 palabras sin etiquetas. De nuevo, puede observarse cómo memoQ es la herramienta con la que se tarda menos tiempo en traducir los segmentos de dichas características, utilizando tanto el teclado como el sistema de RAH como medio de entrada.



*Gráfico 5: Comparativa entre los segmentos más largos sin etiquetas*

Como ya se ha mencionado anteriormente, el sistema de RAH con el que cuenta memoQ permite que se dicte texto sin realizar tantas pausas como en Memsorce u OmegaT. Esta característica unida a la imposibilidad de insertar etiquetas, hacen que este tipo de segmentos (sin etiquetas, marcas de formato y de longitud considerable) sean los óptimos si se busca un incremento de la productividad mediante el uso de la voz.

#### *5.4. Resultados de los segmentos con TA*

El siguiente gráfico refleja cómo el empleo de la TA reduce el tiempo de traducción de cada segmento en los tres grupos en todas las herramientas. En este caso, cabe mencionar que, dentro del grupo de segmentos de 5 palabras o menos, la reducción del tiempo empleado en cada uno es de prácticamente el doble al usar la TA. Esto se debe a que aquellos segmentos cuya propuesta de la TA no necesita ningún tipo de modificación conforman el grupo de menor número de palabras. En los otros dos grupos de segmentos puede comprobarse cómo la dinámica sigue siendo la misma, aunque la diferencia no es tan notable, ya que los segmentos son más largos, por lo que la mayoría de las propuestas de la TA son también dictadas realizando las modificaciones pertinentes. En este caso en concreto, se han valorado los resultados de los segmentos que no contenían etiquetas ni marcas de formato para que, de esa forma, todos estuvieran en igualdad de condiciones.

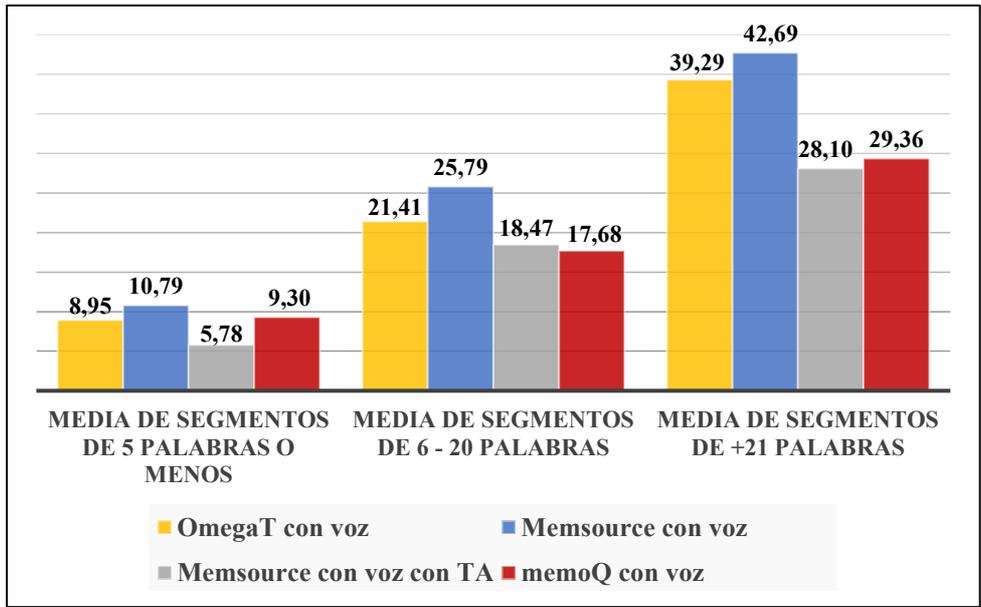


Gráfico 6: Comparativa de tiempos con TA y sin TA

### 5.5. Resultados de memoQ con y sin voz

En este punto, analizaremos los datos relativos al uso de memoQ con y sin voz. El Gráfico 7 nos muestra que se produce una reducción del tiempo en los grupos de 6 a 20 palabras y de más de 20 palabras al usar la voz con esta herramienta.

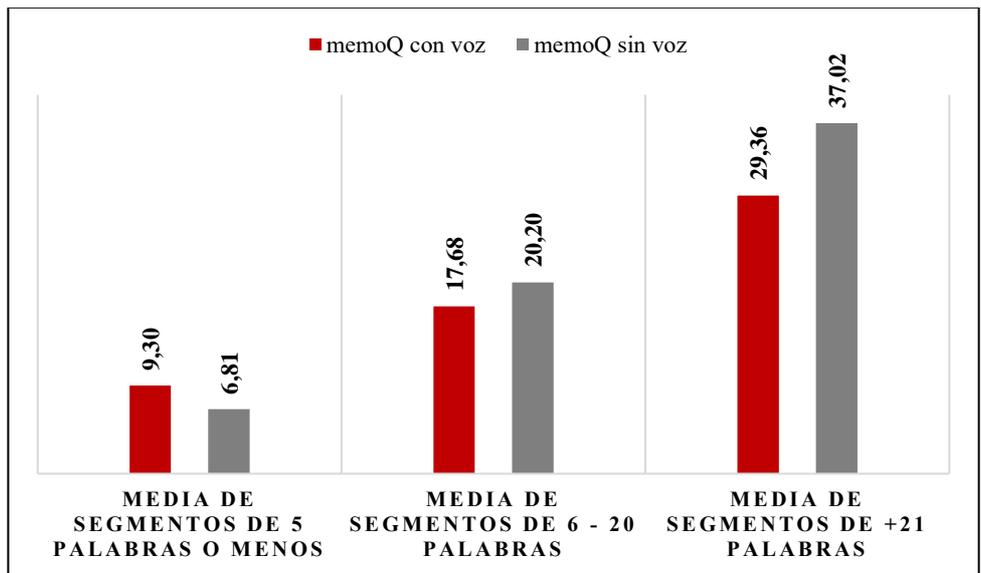


Gráfico 7: Comparativa de tiempos en memoQ con y sin voz

A continuación, pasamos a analizar con más profundidad el último grupo de segmentos, ya que es el que presenta una diferencia más alta. En este caso se puede comprobar que la tendencia es bastante estable y el motivo de esta superioridad en el uso de la voz con respecto al teclado se debe, como se ha mencionado anteriormente, a las características propias de la herramienta (no presenta problemas al dictar un texto muy largo sin pausas), y del texto (se trata de un texto con etiquetas que la herramienta no puede insertar con la voz, pero que sí se hará con el teclado, lo que se traduce por un aumento en el tiempo empleado a traducir).

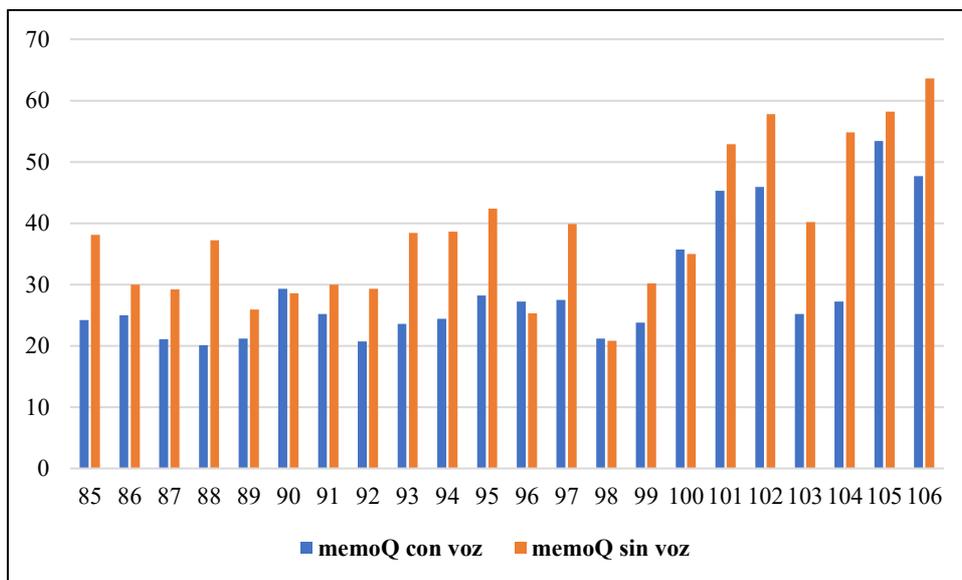
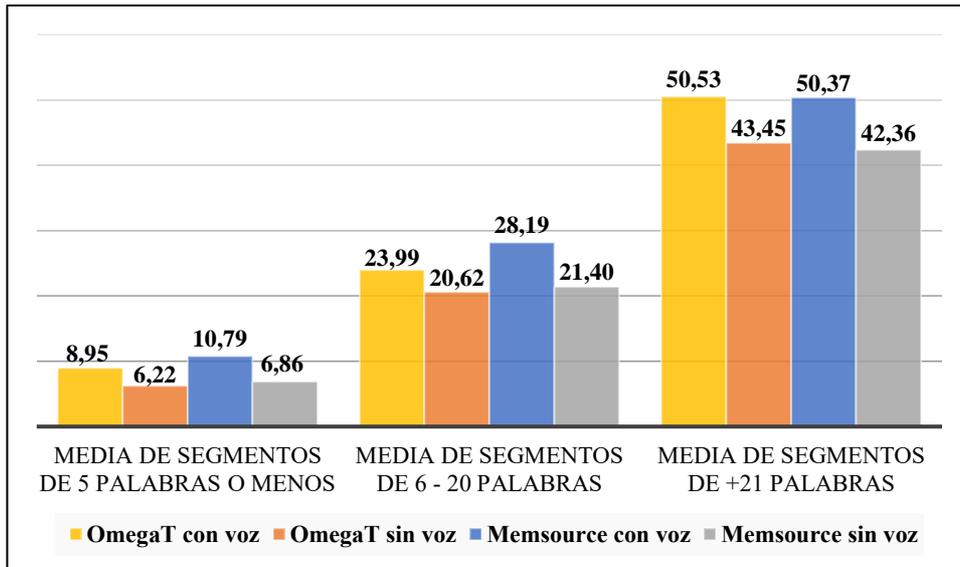


Gráfico 8: Comparativa de tiempo de los segmentos del tercer grupo en memoQ

### 5.6. Resultados de OmegaT y Memsource con y sin voz

Por último, pasamos a comentar los datos sobre Memsource y OmegaT con y sin voz. El Gráfico 9 indica que la mayor diferencia de tiempo entre la voz y el teclado en ambas herramientas se da en los segmentos más largos. Por otro lado, los segmentos de entre 6 y 20 palabras presentan la mayor diferencia de tiempo al usar la voz.



*Gráfico 9: Comparativa de tiempos en Memsorce y OmegaT con y sin voz*

El Gráfico 10 muestra de forma más detallada los segundos invertidos en la traducción de los segmentos de 6 a 20 palabras. En la mayoría de ellos, OmegaT dedica menos tiempo, aunque hay algunos casos en los que ocurre lo contrario, como en el segmento 81. En este segmento en concreto, el sistema de RAH de Mac se bloqueó mientras se utilizaba con OmegaT, lo que provocó ese aumento tan evidente de tiempo.

Esta tendencia que se observa en Memsorce de emplear más tiempo en cada segmento se debe a que con esta herramienta se producen más fallos relacionados con la repetición de comandos o el bloqueo del propio sistema de RAH, en comparación con OmegaT (aunque el sistema de RAH utilizado es el mismo en ambos).

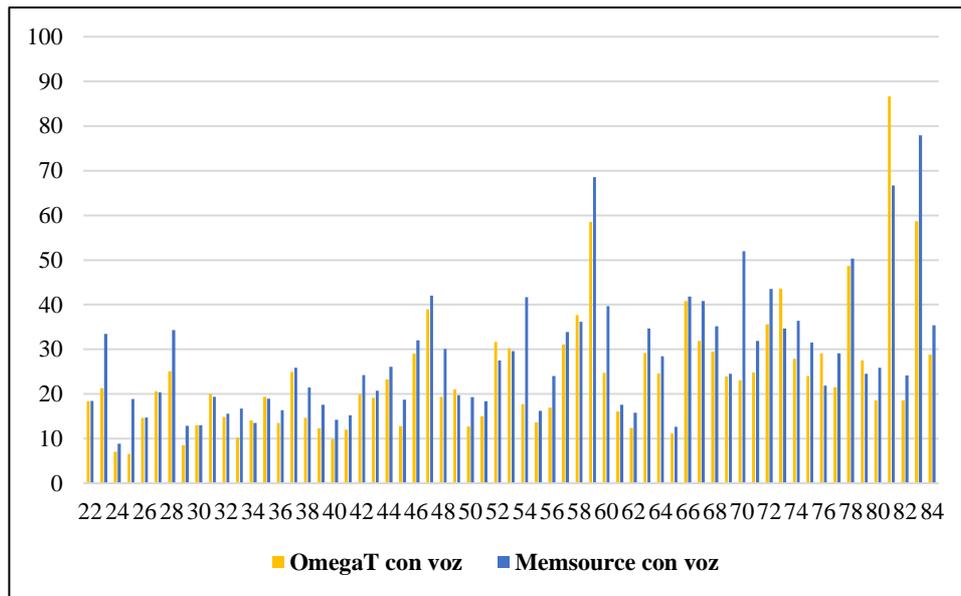


Gráfico 10: Comparativa de tiempo de los segmentos del segundo grupo en Memsorce y OmegaT

## 6. Propuestas de mejora

Una vez analizados los resultados que se han obtenido, es posible proponer algunas mejoras en relación con el funcionamiento de los sistemas de RAH y su utilización en la práctica de la traducción.

En primer lugar, sería necesario revisar la integración del sistema de RAH de Windows con Memsorce y OmegaT. Este sistema presenta una serie de problemas a la hora de trabajar con ambas herramientas que podrían mejorarse si se atendiera específicamente a su compatibilidad.

Esta revisión de la compatibilidad también puede trasladarse al sistema de RAH de Mac, ya que, aunque entre sus comandos predeterminados incluye muchas funciones y órdenes, no todas responden desde las herramientas de TAO con las que se han probado. Alguna de estas órdenes sería, por ejemplo, poder escribir las palabras con mayúscula inicial. Además, sería necesario que el sistema pudiera detectar de alguna forma la coherencia (a nivel ortográfico) de la frase que está escribiendo, para así incluya los espacios necesarios entre palabras e introduzca las mayúsculas de forma apropiada y no prácticamente en cada palabra. Por último, podría ser bastante productivo el hecho de incorporar un proceso de entrenamiento, parecido al de Windows, que permita al sistema familiarizarse con la

voz y el acento del usuario que lo va a usar y que permitiera añadir de forma manual algunos términos especializados o nombres propios que aparezcan en el texto.

En el caso de memoQ, sería imprescindible que se incluyera la posibilidad de añadir etiquetas o marcas de formato. Por otro lado, aunque no se haya probado en este trabajo, sería también interesante que mejorara la movilidad en el interior del segmento, añadiendo la opción de eliminar la parte que se seleccione con vistas a poder poseerla.

## **7. Conclusiones**

Tras la realización de este TFM se han extraído una serie de conclusiones que serán comentadas en este apartado.

En primer lugar, se han alcanzado los objetivos propuestos al inicio del trabajo, tanto a nivel académico como a nivel personal. Estos objetivos estarían relacionados con la descripción de la integración de los sistemas de RAH con las herramientas de TAO, la comprobación de la posibilidad de utilizar el reconocimiento de voz para poseerla, el análisis del tiempo destinado en la traducción de cada segmento, la evaluación de los fallos producidos con cada sistema de RAH y su impacto en el desarrollo de la labor traductológica y, finalmente, la reflexión sobre posibles mejoras.

Además, queda patente la necesidad de investigar e invertir para poder implementar este tipo de tecnologías en el flujo de trabajo de un traductor. Todavía existen ciertos aspectos que habría que mejorar para que suponga un valor añadido en su totalidad y se convierta en un recurso en el que los traductores profesionales quieran invertir, no solo su dinero, sino también su tiempo.

No debemos olvidar que se trata de sistemas que exigen de una inversión tanto en materiales (micrófonos o licencias de ciertos programas), como en tiempo, ya que requieren un periodo de familiarización para efectuar las adaptaciones pertinentes. Además, se trata de un método que exige de un riguroso proceso de revisión para corregir los posibles errores que se produzcan. No obstante, hemos podido comprobar cómo, en función del texto que se vaya a traducir, podría presentar una disminución del tiempo dedicado a la traducción de cada segmento, incrementando así la productividad.

En este caso, se ha demostrado que traducir por medio de los sistemas de RAH evaluados no es un método recomendable para textos que incluyan etiquetas, ya que el tiempo invertido, la mayoría de las veces, es mucho mayor que si se hiciera por medio del teclado. Sin embargo, la situación cambia si se habla de textos que no las contienen. En este caso, el tiempo invertido es prácticamente el mismo utilizando la voz y el teclado, tanto en Memsorce como en OmegaT, en la mayoría de los casos. Sin embargo, en memoQ sí que se registra una reducción de tiempo con el uso de la voz, ya que su propio sistema de RAH es capaz de reconocer las palabras dictadas sin necesidad de hacer tantas pausas.

Por otro lado, también se ha comprobado que el sistema de RAH de Mac presenta un mejor funcionamiento con OmegaT que con Memsorce. Existe una diferencia menor de tiempo invertido en cada segmento, surgen menos problemas relacionados con el bloqueo del sistema o la repetición de comandos y la calidad de los segmentos traducidos es mejor (en cuanto a la correcta colocación de espacios o el uso de mayúsculas).

En definitiva, este trabajo pretende ser una contribución al estudio de estas tecnologías, pero, al mismo tiempo, puede servir como base de futuras investigaciones de mayor envergadura que cuenten con la colaboración de más profesionales para poder extraer datos más realistas y esclarecedores. Asimismo, contar con más usuarios colaboradores ampliaría el abanico de opiniones y experiencias sobre el uso de estos sistemas en el flujo de trabajo de un traductor. Por otro lado, también ha supuesto el resultado de un reto personal, no solo por la necesidad de aplicar todos los conocimientos aprendidos hasta ahora, sino también por la necesidad de hacer frente a diferentes complicaciones que han surgido a lo largo del estudio, como la incompatibilidad del sistema de RAH de Windows o de la nueva versión del software de Mac.

Por último, nos gustaría volver a remarcar la presencia de las tecnologías y su influencia en todos los aspectos de nuestra vida diaria. Tenemos a nuestra disposición una gran variedad de herramientas que pueden ser muy útiles en el campo de la traducción ampliando nuestra capacidad de trabajo y mejorando la calidad de nuestras producciones. En este caso en concreto, la implementación de sistemas de RAH puede llegar a ser un valor añadido cuando se utiliza en los casos con las características mencionadas

anteriormente y sería interesante continuar el desarrollo de estos sistemas para que, en un futuro, pudiera extenderse su aplicación a todo tipo de textos.

## 8. Bibliografía

Bikel, Daniel, y Zitouni Imed. Multilingual natural language processing applications: from theory to practice. IBM Press, 2012.

Bocos Hernández, Marta. “Sistema de Reconeixement de Veu per a Domòtica” Sabadell: Universitat Autònoma de Barcelona, Escola Universitària d’Informàtica, 2009. [https://ddd.uab.cat/pub/trerecpro/2011/hdl\\_2072\\_151974/BocosHernandezMartaR-ETISa2008-09.pdf](https://ddd.uab.cat/pub/trerecpro/2011/hdl_2072_151974/BocosHernandezMartaR-ETISa2008-09.pdf) [Última consulta: junio 2020].

Calvo Lance, Marcos. “Desarrollo de un sistema de reconocimiento automático del habla”. Valencia: Universidad Politécnica de Valencia, 2010 <https://riunet.upv.es/bitstream/handle/10251/10172/memoriaProyecto.pdf> [Última consulta: junio 2020].

Ciobanu, Dragos. “Of Dragons and Speech Recognition Wizards and Apprentices”. *Revista Tradumàtica: tecnologies de la traducció*, Núm. 12, Traducció i qualitat (2014): 524-38.

Deng, Li, et al. "Recent advances in deep learning for speech research at Microsoft." 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2013.

Díaz, Jesús, José Carlos Segura Luna y Antonio Rubio Ayuso. *Reconocimiento de voz continua: aproximaciones basadas en HMM y en redes neuronales recurrentes*. Departamento de Electrónica y Tecnología de Computadores, Universidad de Granada, 2002.

Dragsted, Barbara, Inger Margrethe e Inge Hansen. “Speaking your translation: students’ first encounter with speech recognition technology”. *The International Journal for Translation & Interpreting Research*, Vol 3, No 1 (2011): 10-43.

Errattahi, Rahhal; El Hannani, Asmaa y Ouahmane, Hassan. “Automatic speech recognition errors detection and correction: A review” *Procedia Computer Science*, vol. 128, (2018): 32–37.

Escolano, Francisco et al. *Inteligencia artificial: modelos, técnicas y áreas de aplicación*. Editorial Paraninfo, 2003.

Gohil, Manish. "Artificial Intelligence for Speech Recognition". *International Multidisciplinary. Research Journal*, Vol 1, No 2 (2014).

Haridas, Arul Valiyavalappil, Ramalatha Marimuthu, y Vaazi Gangadharan Sivakumar. "A critical review and analysis on techniques of speech recognition: The road ahead." *International Journal of Knowledge-based and Intelligent Engineering Systems* 22.1 (2018): 39-57.

Jurafsky, Daniel y Martin, James. *Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. 2nd Edition. New Jersey: Pearson Education, 2009.

Juang, Biing-Hwang y Rabiner, Lawrence. "Automatic Speech Recognition—A Brief History of the Technology Development". Georgia Institute of Technology. Atlanta Rutgers University and the University of California. Santa Barbara 1 (2005): 67.

Liyanapathirana, Jeevanthi, Pierrette Bouillon, and Bartolomé Mesa-Lao. "Surveying the potential of using speech technologies for post-editing purposes in the context of international organizations: What do professional translators think?" *Proceedings of Machine Translation Summit XVII. Volume 2: Translator, Project and User Tracks* (2019): 149-158. <https://archive-ouverte.unige.ch/unige:123740> [Última consulta: junio 2020].

Lleida, E. y Ortega, A. "Reconocimiento del lenguaje hablado". Ariel, S.A. y Fundación Telefónica, (Eds.). *Tecnologías del lenguaje en España. Comunicación inteligente entre personas y máquinas*, 2016 1-17.

Llisterri, Joaquim. "Les tecnologies de la parla". *Revista de Sociolingüística de la Universitat de Barcelona*, Monogràfic: Llengües minoritzades i tecnologia lingüística: perspectives de futur, Núm. 7, (2009): 11-19.

Mesa-Lao, Bartolomé. "Speech-Enabled Computer-Aided Translation: A Satisfaction Survey with Post-Editor Trainees." *Proceedings of the EACL 2014 Workshop on Humans and Computer-assisted Translation*. 2014.

*No tienes escapatoria, Google te entiende.* Ponsoda, Jordi Gisbert. 2017. Andro4all. Junio 2020. <https://andro4all.com/2017/06/google-entiende-95-lenguaje-humano>

Rodríguez, Luis Javier. y Amparo Varona. “Reconocimiento automático del habla: perspectiva desde un grupo de investigación del País Vasco”. Eusko Ikaskuntza. XIII Congreso de Estudios vascos, 1996. 317-323.

Rouhiainen, Lasse. "Inteligencia artificial." Madrid: Alienta Editorial (2018).

Stuker, Sebastian, et al. "Speech translation enhanced ASR for european parliament speeches-on the influence of ASR performance on speech translation." *2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07*. Vol. 4. IEEE, 2007.

Tatman, Rachael. "Gender and dialect bias in YouTube's automatic captions." *Proceedings of the First ACL Workshop on Ethics in Natural Language Processing*. 2017.

Varona Fernández, Amparo “Antecedentes y desarrollo de los sistemas actuales de reconocimiento automático del habla”. Fórmula. Cuadernos de Ciencias Físicoquímicas y Matemáticas, Núm. 4, (1997): 321-346.

*Where Speech Recognition Is Going.* Knight, Will. 2012. MIT Technology Review. Junio 2020. <https://www.technologyreview.com/2012/05/29/255893/where-speech-recognition-is-going/>. [Última consulta: junio 2020].

Zapata Rojas, Julián. “Traduction dictée interactive : intégrer la reconnaissance vocale à l’enseignement et à la pratique de la traduction professionnelle” M.A. thesis, University of Ottawa, 2012. [http://www.ruor.uottawa.ca/en/bitstream/handle/10393/23227/Zapata\\_Rojas\\_Julian\\_2012\\_these.pdf?sequence=1](http://www.ruor.uottawa.ca/en/bitstream/handle/10393/23227/Zapata_Rojas_Julian_2012_these.pdf?sequence=1) [Última consulta: junio 2020].

Zapata Rojas, Julián. “Translators in the Loop: Observing and Analyzing the Translator Experience with Multimodal Interfaces for Interactive Translation Dictation Environment Design”. 2016. M.A. thesis

[https://ruor.uottawa.ca/bitstream/10393/34978/1/Zapata\\_Rojas\\_Julian\\_2016\\_thesis.pdf](https://ruor.uottawa.ca/bitstream/10393/34978/1/Zapata_Rojas_Julian_2016_thesis.pdf) [Última consulta: junio 2020].