Estudio de viabilidad para la implantación de la traducción automática en la empresa VITAE

Equipo responsable del estudio:

Plataforma del Español

Iris Córdoba Mondéjar

Universidad Europea de Madrid

Celia Rico Pérez

Universidad Antonio de Nebrija

María Ortiz Jiménez

Anna Doquin de Saint Preux

Hermes Traducciones y Servicios Lingüísticos, S.L.

Juan José Arevalillo Doval

Manuel Arcedillo Jiménez

EDTConsulting

Jorge Cabero Zumalacárregui

Índice

1. Introducción 3

2. Resumen ejecutivo 4

3. Viabilidad técnica y de los procesos 5

3.1. Tipo de motor: sistemas basados en reglas y sistemas estadísticos 5

3.2. Identificación de flujos de trabajo 6

3.3. Aplicación de estándares 7

3.4. Interoperabilidad 7

3.5. Seguridad de los datos 8

3.6. Lista de verificación para determinar la viabilidad técnica en la implantación de un motor de traducción automática 9

4. Evaluación de la calidad de los motores de traducción automática mediante métricas 12

4.1. Especificaciones técnicas 13

4.2. Descripción de motores de traducción automática y métricas de evaluación 13

4.3. Metodología 15

4.4. Limitaciones 15

4.5. Análisis de resultados 16

4.5.1. Análisis de los resultados de la fase 1 16

4.5.2. Análisis de resultados de la fase 2 17

4.5.3. FASE 1: Resultados comparados por lenguas 19

4.5.4. FASE 2: Evaluación lingüística y de productividad de dos textos especializados 23

4.5.5. Resultados comparados de la fase 2 24

4.6. Conclusiones generales de la evaluación 25

5. Viabilidad económica y modelo de negocio: estudio de caso 26

5.1. Cuestiones técnicas y de procesos 26

5.2. Evaluación de la calidad 30

5.3. Modelo económico 31

6. Conclusiones 34

"El avance tecnológico no ha eliminado la capacidad humana de cometer errores".

Noah Gordon

1. Introducción

El sector de la traducción aporta un valor añadido en la internacionalización de las empresas y, con un crecimiento anual estimado de 5,13 % en todo el mundo, necesita el apoyo de la tecnología para ofrecer un servicio innovador que permita a las empresas de traducción competir en condiciones de igualdad en el ámbito internacional.

Desde el Cluster Plataforma del Español entendemos que **la traducción automática es ya una realidad** y, ante la clara evidencia de que los resultados que se obtienen son aceptables para determinados contextos de uso, nos planteamos la necesidad de realizar el presente estudio, en cuyos contenidos se incluyen un análisis de la viabilidad técnica y de los procesos, una evaluación de la calidad, mediante una comparativa entre diferentes sistemas y motores, y un análisis de la viabilidad económica y modelo de negocio, para poder obtener en su conjunto una serie de conclusiones y recomendaciones para su implantación en las empresas.

Así, en el marco de las ayudas del Ministerio de Industria, Energía y Turismo a las Agrupaciones Empresariales Innovadoras, y, confiando siempre en favorecer e impulsar el sector de la industria del español, y en este caso concreto al de la traducción, solicitamos financiación para el desarrollo de este proyecto, en el que han participado cuatro entidades socias de **la Plataforma del Español, la Universidad Europea, la Universidad Antonio de Nebrija, Hermes Traducciones y Servicios Lingüísticos y EDTConsulting**.

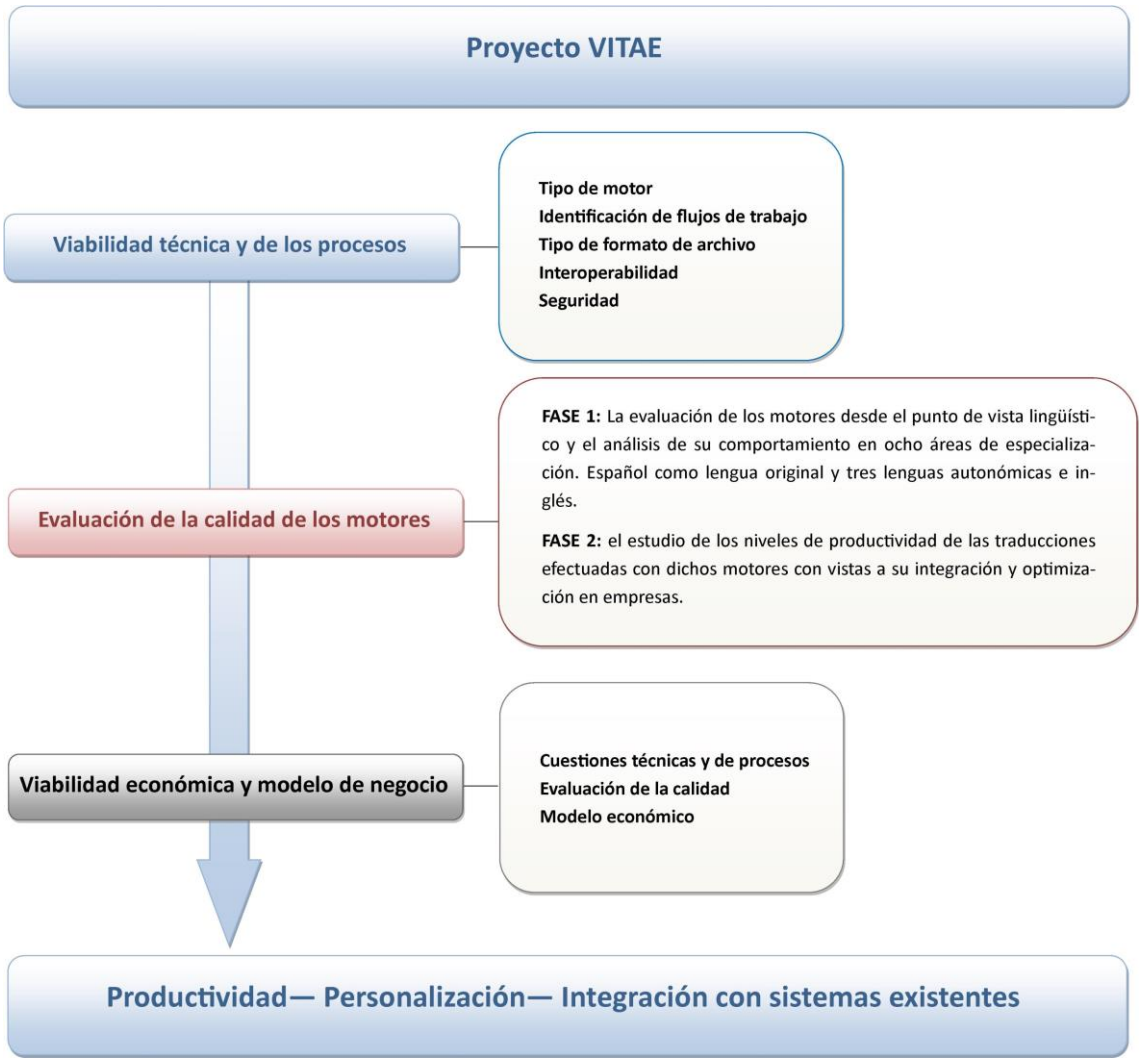
Es necesario destacar que en el desarrollo del estudio no se ha buscado determinar cuál es la mejor solución única con vistas a su incorporación a un proceso de traducción profesional, sino que el objetivo final ha sido el de delimitar tendencias y recomendaciones según los rendimientos, dejando claro que no se trata simplemente de incorporar una nueva herramienta a las que ya puedan existir en la empresa, sino que implica una revisión en profundidad de los procesos con los que se trabaja y determinar, además, con qué objetivo queremos introducir un nuevo elemento al flujo de trabajo de una entidad.

En este sentido, como se puede observar a lo largo de los contenidos del estudio y de sus conclusiones, queda bastante claro que cualquier empresa deberá evaluar las prestaciones del sistema de traducción automática que quiera implantar, realizar un análisis comparativo de diferentes motores antes de decidirse por uno en concreto, personalizar dicho sistema, mediante la creación o adaptación de los diccionarios que utiliza, y realizar su integración con el resto de herramientas de traducción asistida implementadas para poder obtener un alto grado de rendimiento.

"El avance tecnológico no ha eliminado la capacidad humana de cometer errores".

Noah Gordon

2. Resumen ejecutivo



3. Viabilidad técnica y de los procesos

En esta sección se identifican las necesidades para la implantación de la traducción automática desde el punto de vista de su *viabilidad técnica*, con respecto a los siguientes aspectos:

1. Tipo de motor: sistemas basados en reglas y sistemas estadísticos.
2. Identificación de flujos de trabajo.
3. Tipo de formato de archivo.
4. Interoperabilidad.
5. Seguridad.

3.1. Tipo de motor: sistemas basados en reglas y sistemas estadísticos

A efectos del presente estudio de viabilidad no se hace necesario realizar una descripción exhaustiva de la diferencia entre un motor de traducción automática basado en reglas y otro estadístico[1], aunque sí resulta pertinente conocer los fundamentos de uno y otro sistema ya que esto nos permitirá determinar las posibilidades de adaptación del motor al tipo de texto (o campo de especialidad) con el que vamos a tratar.

Los sistemas basados en reglas (*RBMT*, por sus siglas en inglés, *Rule-based Machine Translation*), codifican a priori toda la información lingüística así como el conocimiento enciclopédico necesario para realizar la traducción. Esta codificación se lleva a cabo mediante diccionarios, reglas gramaticales, reglas léxicas y semánticas. Por su parte, los sistemas estadísticos (*SBMT*, por sus siglas en inglés, *Statistically-based Machine Translation*) *aprenden* a traducir a partir de modelos estadísticos de traducciones realizadas previamente, almacenadas como corpus paralelos.

Desde este punto de vista, se puede formular entonces **una serie de preguntas que nos ayuden a determinar las posibilidades de adaptar el sistema de traducción automática (TA) al tipo de texto que se va a procesar** (o al campo de especialidad):

a) *Motores basados en reglas*: creación y adaptación de diccionarios
   - ¿Permite el sistema identificar fácilmente un listado de palabras no traducidas?
   - ¿Se genera automáticamente un archivo con este listado?
   - ¿Se puede crear fácilmente un diccionario de usuario con este listado de palabras?
   - ¿Qué grado de complejidad tienen las entradas del diccionario?
   - ¿Se pueden adaptar los diccionarios que ya existen en el sistema?

b) *Motores basados en reglas*: creación y adaptación de reglas
   - ¿Se pueden crear nuevas reglas léxicas, sintácticas o semánticas en función del tipo de texto?

[1] El lector interesado en conocer más a fondo el funcionamiento de los diferentes motores de traducción automática puede consultar http://www.erudit.org/revue/meta/1997/v42/n4/003822ar.pdf

- ¿Es posible adaptar las reglas del sistema?
- ¿Qué grado de complejidad tiene la codificación de nuevas reglas?

c) *Motores estadísticos*: corpus de entrenamiento

- ¿Es posible entrenar el sistema para adaptarlo a diferentes textos?
- ¿Qué textos paralelos existen en el campo de especialidad de que se trate para crear un corpus de entrenamiento?
- ¿Qué grado de dificultad presenta el proceso para entrenar el motor?
- Disponibilidad de textos paralelos para crear el corpus de entrenamiento

3.2. Identificación de flujos de trabajo

Con independencia de que el proceso de traducción implantado en la empresa esté basado en herramientas de escritorio o en la nube (o esté concebido de forma más o menos tradicional) es una cuestión clave **identificar el flujo que sigue el texto que se va a traducir desde que llega el encargo del cliente hasta que se realiza la entrega final del trabajo.** Este análisis nos permitirá determinar las posibilidades de implantar un sistema de TA que vaya acorde con los procesos ya implantados o, en caso contrario, identificar aquellos que deban modificarse antes de implantar el sistema.

A modo de ejemplo, en la figura 1 se muestra un flujo de trabajo en el que se incorporan herramientas TAO y TA.

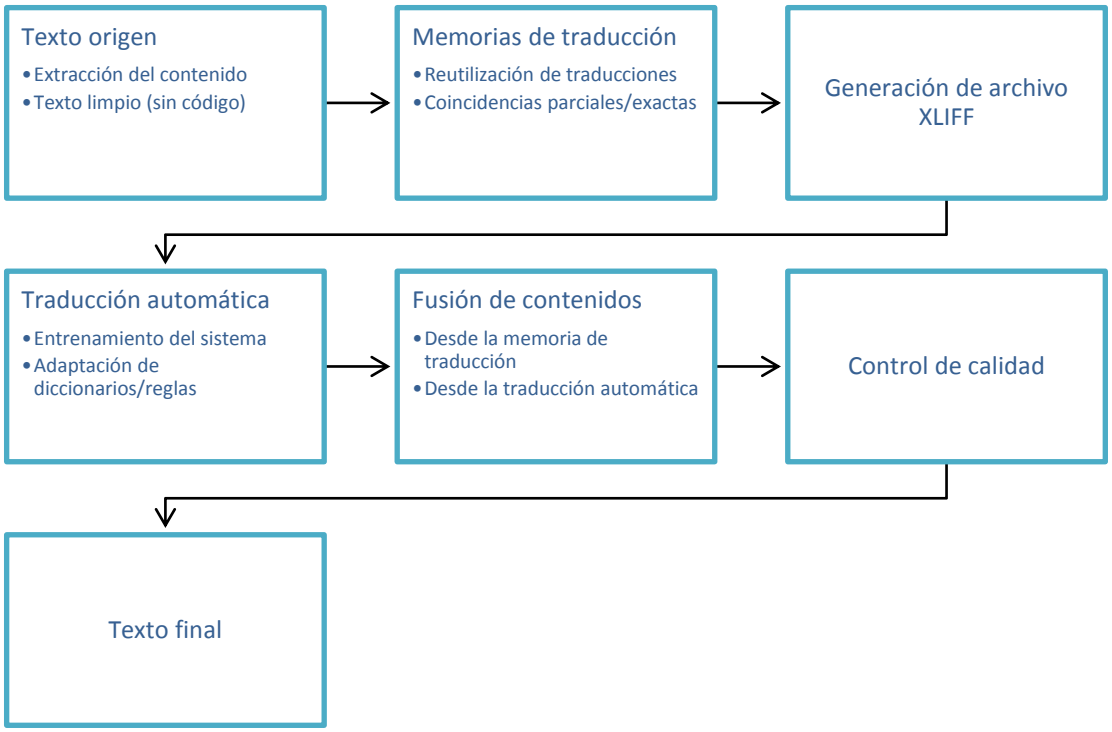


Figura 1. Ejemplo de flujo de una traducción

Plataforma del Español · GOBIERNO DE ESPAÑA · MINISTERIO DE INDUSTRIA, ENERGÍA Y TURISMO · AEI minetur · agrupaciones empresariales innovadoras · FEDER Fondo Europeo de Desarrollo Regional · UNIÓN EUROPEA "Una manera de hacer Europa"

En este contexto, las cuestiones que deberían tenerse en cuenta son las siguientes:

- ¿Contempla el flujo de traducción ya implantado en la empresa la incorporación de nuevas herramientas?
- ¿Con qué facilidad puede incorporarse la TA a ese flujo de trabajo?
- ¿Es necesario cambiar el sistema de producción?
- ¿Tiene el sistema herramientas para la preparación/extracción del texto?
- ¿Es necesario realizar tareas de importación de formato?
- ¿Contiene el sistema herramientas de revisión o posedición?
- ¿Se pueden preparar los archivos dentro del propio sistema?
- ¿Se puede hacer un seguimiento del estado de los archivos?

3.3. Aplicación de estándares

La industria de la traducción y, por extensión, la de la generación de contenidos digitales multilingües ha desarrollado una serie de estándares con el objetivo de unificar y armonizar los diferentes formatos de archivo que se emplean. Entre ellos destacan los siguientes:

- para la generación de contenidos: SRX, W3C ITS, Unicode;
- para el intercambio de datos en memorias de traducción: TMX, GMX, XLIFF.

Por otra parte, la mayoría de los contenidos que se manejan en la industria de la traducción utiliza formatos propietarios entre los que destacan doc, txt, pdf, xls. Así, con independencia del formato que se utilice, es clave comprobar que el sistema de TA que vamos a implantar puede procesarlo o, en caso contrario, que permite la conversión. Las preguntas que surgen entonces son las siguientes:

- ¿Qué formato tienen los contenidos que se deben traducir?
- ¿Qué formatos acepta el sistema de TA?
- ¿Hay posibilidad de convertir los formatos que no acepta?
- ¿Qué grado de complejidad supone esta conversión?

3.4. Interoperabilidad

La integración de diferentes herramientas en el entorno de traducción tiene como objetivo **facilitar el proceso, de modo que la información o los contenidos generados en una aplicación durante una de las fases del proyecto puedan utilizarse en la siguiente fase o herramienta.** De manera general, se pueden establecer tres grandes categorías de *software* de traducción: los sistemas de gestión de la traducción, las herramientas TAO y la TA.

Figura 2. Interoperabilidad

A la hora de integrar la traducción automática con todas las herramientas que ya existen en la empresa, es imprescindible comprobar la interoperabilidad entre todas ellas para que los archivos y proyectos que se generan en cada una puedan ser utilizados por el resto. De lo contrario, no sería eficaz tener un sistema de TA aislado del resto de *software* de traducción y que genera ineficiencias en el flujo de los contenidos.

3.5. Seguridad de los datos

La seguridad de los datos depende directamente del sistema de traducción automática que queramos implantar. En el caso de que se utilice una herramienta de escritorio, instalada en modo local en los servidores de nuestra empresa, la seguridad de los datos quedará establecida por los protocolos que tengamos implantados. Por otra parte, si utilizamos un servicio de *software* en la nube (SaaS*, Software as a Service*, por sus siglas en inglés) será imprescindible cerciorarse del tipo de tratamiento que reciben por parte del proveedor del servicio: dónde quedan almacenados, qué uso se les va a dar, qué opciones de confidencialidad permiten, qué permisos de acceso se pueden establecer, etc.

3.6. Lista de verificación para determinar la viabilidad técnica en la implantación de un motor de traducción automática

Análisis de necesidades técnicas	Observaciones
Motores basados en reglas: creación y adaptación de diccionarios	
• ¿Permite el sistema identificar fácilmente un listado de palabras no traducidas?	
• ¿Se genera automáticamente un archivo con este listado?	
• ¿Se puede crear fácilmente un diccionario de usuario con este listado de palabras?	
• ¿Qué grado de complejidad tienen las entradas del diccionario?	
• ¿Se pueden adaptar los diccionarios que ya existen en el sistema?	
Motores basados en reglas: creación y adaptación de reglas	
• ¿Se pueden crear nuevas reglas léxicas, sintácticas o semánticas en función del tipo de texto?	
• ¿Es posible adaptar las reglas del sistema?	
• ¿Qué grado de complejidad tiene la codificación de nuevas reglas?	
Motores estadísticos: corpus de entrenamiento	
• ¿Es posible entrenar el sistema para adaptarlo a diferentes textos?	
• ¿Qué textos paralelos existen en el campo de especialidad de que se trate para crear un corpus de entrenamiento?	
• ¿Qué grado de dificultad presenta el proceso para entrenar el motor?	
• Disponibilidad de textos paralelos para crear el corpus de entrenamiento	

Análisis del flujo de trabajo	Observaciones
• ¿Contempla el flujo de traducción ya implantado en la empresa la incorporación de nuevas herramientas?	
• ¿Con qué facilidad puede incorporarse la TA a ese flujo de trabajo?	
• ¿Es necesario cambiar el sistema de producción?	
• ¿Tiene el sistema herramientas para la preparación/extracción del texto?	
• ¿Es necesario realizar tareas de importación de formato?	
• ¿Contiene el sistema herramientas de revisión o posedición?	
• ¿Se pueden preparar los archivos dentro del propio sistema?	
• ¿Se puede hacer un seguimiento del estado de los archivos?	
Aplicación de estándares	**Observaciones**
• ¿Qué formato tienen los contenidos que se deben traducir?	
• ¿Qué formatos acepta el sistema de TA?	
• ¿Hay posibilidad de convertir los formatos que no acepta?	
• ¿Qué grado de complejidad supone esta conversión?	
Interoperabilidad	**Observaciones**

• ¿Con qué herramientas de traducción se cuenta ya en la empresa?
  ○ ¿Sistema de gestión de traducciones?
  ○ ¿Sistema de gestión de clientes/proveedores?
  ○ ¿Herramientas TAO?	

• ¿Qué posibilidades de interacción existen entre las diferentes herramientas?
  ○ ¿Existe la posibilidad de establecer un flujo automático de los datos de una herramienta a otra?
  ○ ¿Existe alguna API (*Application Programming Interface)* que permita conectar una herramienta con otra?	

Seguridad de los datos	Observaciones

• ¿Qué tipo de sistema de traducción automática se ha implantado?
  ○ ¿En servidor local?
  ○ ¿Cómo servicio en la nube?	

• ¿Qué protocolos se utilizan para la protección de los datos?	
• ¿Qué permisos de acceso a los datos se pueden establecer?	

4. Evaluación de la calidad de los motores de traducción automática mediante métricas

En esta sección se procederá a la descripción del análisis de diversos motores de traducción automática disponibles (tanto gratuitos en su versión en línea como residentes y entrenados), no para determinar cuál de ellos sería el mejor desde un punto de vista global, con vistas a una selección única como herramienta auxiliar también única para su implantación en los procesos de traducción de una empresa del sector, sino para comprobar el rendimiento de todos ellos en determinados entornos de especialización y combinaciones de lenguas, ya que, como se demostrará más adelante, estos factores son determinantes en el rendimiento y resultados de todos ellos. Por lo tanto, nuestro objetivo principal es la identificación de tendencias sobre la base de los resultados obtenidos mediante el análisis de calidad de varias métricas. El estudio se ha llevado a cabo en dos fases diferenciadas atendiendo a la tipología de textos, lenguas, evaluación lingüística y evaluación de productividad:

- **FASE 1:** se centra en la evaluación de los motores desde el punto de vista lingüístico y el análisis de su comportamiento en ocho áreas de especialización (textos de tipo administrativo, cultural, financiero, jurídico, sanitario, social, medioambiental y técnico) con una combinación de lenguas con el español como lengua original y tres lenguas autonómicas e inglés como lenguas meta.

- **FASE 2:** comprende el estudio de los niveles de productividad de las traducciones efectuadas con dichos motores con vistas a su integración y optimización en empresas de servicios lingüísticos a través del análisis de las traducciones automáticas de dos textos pertenecientes a sendas áreas de especialización del inglés al español: un texto jurídico (un contrato) y un texto científico-técnico (el manual de un antivirus).

Los análisis correspondientes a las dos fases se realizaron mediante el procesamiento de las traducciones automáticas generadas por los distintos motores con cuatro métricas especializadas en la valoración lingüística (fase 1: métricas BLEU, NIST, TER y WER) y la incorporación de una quinta adicional orientada a la evaluación de la productividad (fase 2: además de las cuatro métricas ya mencionadas, se añade la función de comparación de traducciones del programa Rainbow, perteneciente al paquete Okapi Framework).

4.1. Especificaciones técnicas

La totalidad de las pruebas se efectuó con *software* de código abierto bajo sistema operativo Linux con el fin de lograr una mayor integración de todo el proceso y evitar en la medida de lo posible eventuales fallos por incompatibilidad de sistemas, bien en lo referente al funcionamiento de las herramientas en línea seleccionadas para la evaluación de las métricas, o bien del *software* relacionado residente en el ordenador mismo. En el cuadro siguiente se especifican los datos técnicos del equipo utilizado.

Sistema operativo: Linux Edubuntu
Ofimática: LibreOffice para Linux
Navegador: Mozilla Firefox para Linux
***Software* asociado**: Okapi Framework para Linux, SSPS
Plataforma de evaluación de métricas: Asiya (http://asiya.cs.upc.edu/demo/asiya_online.php)

Tras un período inicial de ensayos llevado a cabo en ordenadores con sistema operativo Windows y la utilización de máquinas virtuales Linux en ellos, donde pudieron observarse varios problemas de rendimiento de los sistemas, en especial en el funcionamiento de las métricas, finalmente se optó por la unificación de criterios en lo relativo a la selección del sistema operativo y el tipo de *software* empleado mediante la personalización de un ordenador equipado con sistema operativo Linux Edubuntu y diversas aplicaciones de código abierto (tanto en línea como residentes) compatibles con el sistema de evaluación Asiya, plataforma en línea desarrollada por la Universidad Politécnica de Barcelona, que ofrece al usuario la posibilidad de evaluar la corrección lingüística de sus traducciones automáticas mediante la aplicación simultánea de un amplio catálogo de métricas. Según las recomendaciones de esta plataforma se seleccionó el navegador Mozilla Firefox con el fin de lograr una mayor integración, y tanto los textos originales, las traducciones de referencia y las traducciones automáticas se han manipulado y procesado como archivos de textos sin formato (txt).

4.2. Descripción de motores de traducción automática y métricas de evaluación

La selección de los **motores de traducción automática** se ha llevado a cabo atendiendo a diversos factores como sus respectivas tecnologías, la disponibilidad de lenguas para las traducciones en distintas direcciones, el acceso a los distintos motores (en línea o instalados en ordenadores) o la posibilidad de que dichos motores puedan ser entrenados (esto es, ser personalizados para la especialización en la que vayan a utilizarse mediante la instalación de los diccionarios adecuados, etc.). En una primera clasificación en lo referente a su tecnología, se ha procurado que en cada caso haya al menos un representante de cada tipo de motor, que podríamos resumir de la forma siguiente:

- **Motores basados en reglas:** realizan transformaciones a partir del texto original, reemplazando palabras por el equivalente más adecuado en la lengua meta y aplicando las reglas gramaticales y diccionarios propios de cada idioma interviniente en el proceso de traducción. Los motores de este tipo seleccionados para el estudio han sido **Apertium** (diseñado para la traducción automática de lenguas próximas) y **OpenTrad,** que combina el motor Apertium y el sistema Matxin para pares de lenguas menos comunes y con disponibilidad más reducida, como el español–euskera.
- **Motores estadísticos:** este tipo de motores aplica un modelo estadístico a partir del análisis de corpus de textos bilingües y paralelos, optimizando recursos y el uso del lenguaje natural. Los motores seleccionados para el estudio han sido **Google Translate** y **Microsoft Translator**.
- **Motores híbridos:** aúnan las características propias de los dos anteriores en un intento de ofrecer traducciones automáticas de mejor calidad. El motor elegido para este estudio es **SYSTRAN**, tanto en su versión general en línea como en la entrenada residente.

En lo que respecta a las métricas, y tal y como ya se ha mencionado con anterioridad, de entre las disponibles en la plataforma Asiya se han seleccionado las cuatro que mejor evalúan la adecuación lingüística de las traducciones automáticas generadas y que pasamos a describir brevemente:

- **BLEU**: compara las traducciones automáticas con una de referencia (realizada por un traductor) considerada como correcta. En sus mediciones, la traducción automática sería de mayor calidad en cuanto más se asemejara a la de referencia en un nivel de frases. En sus puntuaciones, la traducción es mejor cuanta más alta sea su puntuación.
- **NIST**: esta métrica está basada en BLEU, aunque a diferencia de esta, otorga un mayor valor a las coincidencias correctas. Al igual que BLEU, las puntuaciones elevadas determinan la buena calidad de la traducción.
- **TER**: evalúa las traducciones automáticas según el mínimo número de ediciones necesario para su correspondencia con la traducción de referencia. Los resultados son negativos, entendiéndose como valores óptimos aquellos más cercanos a cero.
- **WER**: mide el índice de error por palabra, aunque no proporciona detalles sobre la naturaleza de los errores de traducción. Al igual que en el caso anterior, un índice cercano a cero indica una traducción de buena calidad.

Además, para la evaluación de la productividad en los textos más largos de la fase 2 se ha añadido la función de **comparación de traducciones** de la aplicación **Rainbow** integrada en el paquete **Okapi Framework**, que mide la similitud de las traducciones automáticas con respecto a la de referencia basándose por un lado en coincidencias parciales y por otro en la denominada *distancia de Levenshtein*, o distancia entre palabras y el número mínimo de operaciones necesarias para la transformación de una cadena de caracteres en otra mediante inserción, eliminación o sustitución de un carácter.

4.3. Metodología

Para las dos fases del estudio se han utilizado dos tipos de textos diferenciados. En la fase 1 se han empleado ocho textos pequeños originales en español de otras tantas áreas de especialización, traducidas a cuatro lenguas meta (traducciones de referencia): catalán, euskera, gallego e inglés. En la fase 2, se han elegido dos textos especializados más extensos (uno científico —el manual de un antivirus— y otro jurídico —un contrato—), de unas 1000 palabras cada uno y en inglés para su traducción al español. Al igual que en la fase 1, también se han aportado sus correspondientes traducciones de referencia, pero, a diferencia de esta, además de los motores ya especificados se ha incorporado al estudio una versión del texto científico adicional, traducida automáticamente con un motor SYSTRAN entrenado para poder contrastar sus resultados específicos en lo que a productividad se refiere con respecto a los motores generales.

Todos los textos (originales, traducciones de referencia y traducciones automáticas) fueron segmentados previamente para hacer coincidir el número de sus respectivas unidades de traducción (condición indispensable para el correcto funcionamiento de las métricas) y así facilitar la ejecución de todo el proceso de evaluación.

4.4. Limitaciones

Durante el estudio (y especialmente en la fase 1) hemos podido comprobar que no todas las combinaciones de lenguas y direcciones de traducción están disponibles en todos los motores, siendo el gallego y el euskera las lenguas más afectadas con dos motores cada una. Mostramos la disponibilidad de motores en la siguiente tabla:

	Google T	Microsoft T	Apertium/OpenTrad	SYSTRAN
Español	√	√	√	√
Inglés	√	√	√	√
Catalán	√	√	√	
Euskera	√		√	
Gallego	√		√	

Nota: Apertium y OpenTrad son el mismo motor (a excepción de la TA bidireccional del euskera). Por ello aparecen en la misma columna.

No obstante, y aunque el número de motores de traducción automática sea el mínimo requerido para las combinaciones español-gallego y español-euskera, hemos conseguido que haya un representante de cada tecnología de traducción (reglas y estadístico) en todos los casos. Por su parte, se aplican todas las métricas seleccionadas a la totalidad de las traducciones generadas de forma automática.

4.5. Análisis de resultados

4.5.1. Análisis de los resultados de la fase 1

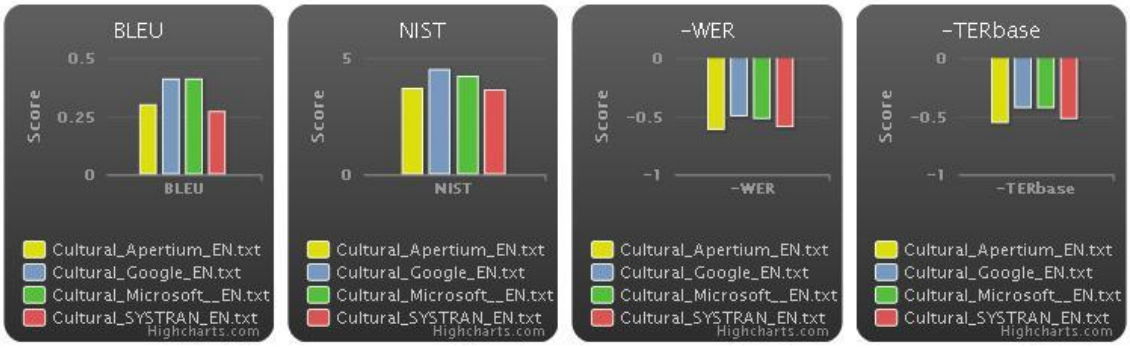
Para el análisis de resultados de la fase 1, se han creado unos formularios individuales para cada texto especializado en las distintas lenguas analizadas con la siguiente información:[2]

- datos del texto original en español: especialización, número de palabras, número de segmentos y formato del archivo;
- datos de la traducción de referencia (no automática): especialización, lengua meta, número de segmentos y formato del archivo;
- datos generales de las traducciones automáticas: especialización, lengua meta, número de segmentos, formato de los archivos y motores de traducción automática utilizados.

A continuación aportamos los resultados globales de las cuatro métricas filtrados por documentos (tabla numérica) y sistemas (gráficos). En este punto debemos añadir que, durante la prueba, también recogimos los resultados filtrados por segmentos, pero consideramos que de cara a su interpretación y a la consulta del público en general, era mejor prescindir de estos últimos, ya que las puntuaciones se presentaban de una forma más confusa, reiterativa y menos concluyente. Todos los cuadros de resumen de puntuaciones de métricas y los gráficos proceden de los resultados obtenidos por las diversas traducciones en la plataforma Asiya:

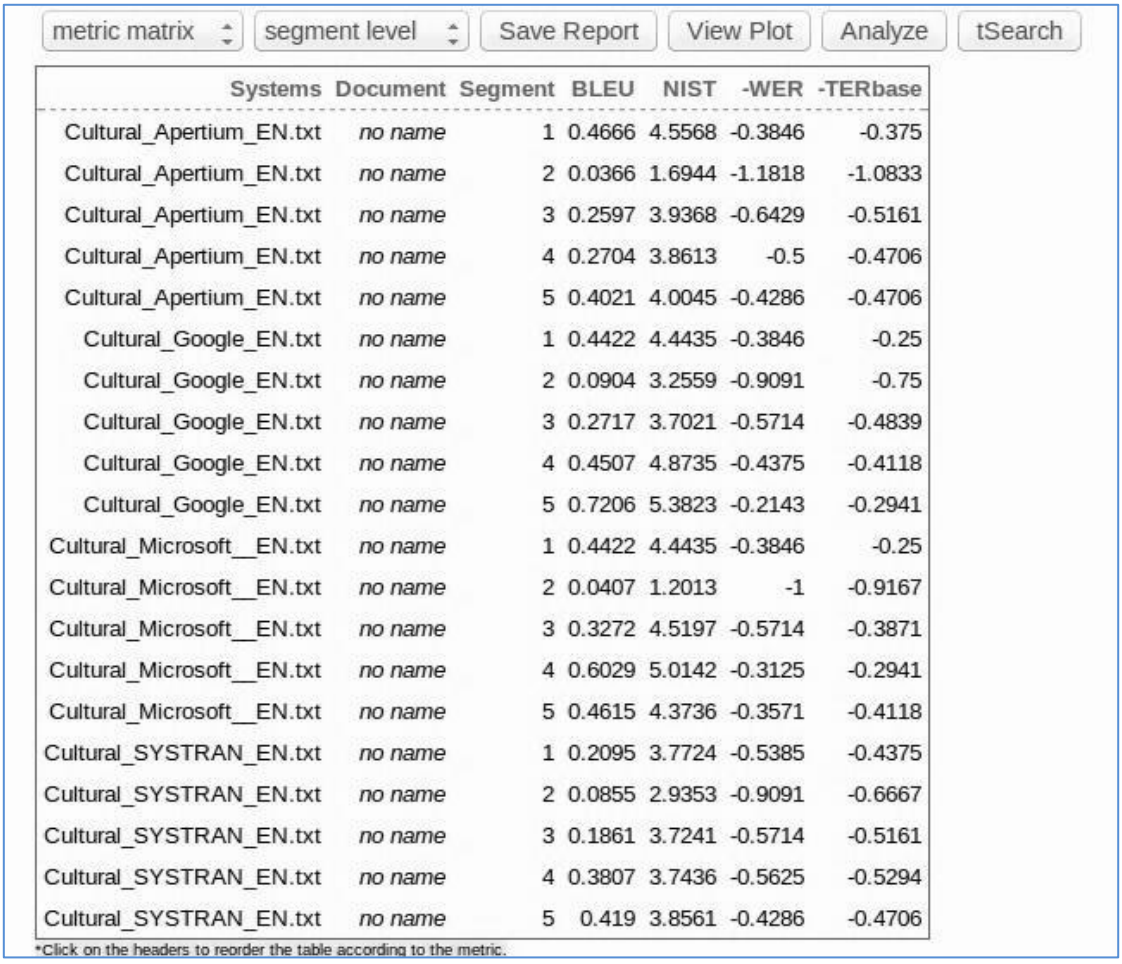
Systems	Document	BLEU	NIST	-WER	-TERbase
Cultural_Apertium_EN.txt	*no name*	0.3001	3.7096	-0.6276	-0.5484
Cultural_Google_EN.txt	*no name*	0.4108	4.5074	-0.5034	-0.4301
Cultural_Microsoft__EN.txt	*no name*	0.4073	4.2218	-0.5251	-0.4194
Cultural_SYSTRAN_EN.txt	*no name*	0.2681	3.6572	-0.602	-0.5161

Figura 3. Ejemplo de cuadro de resumen de resultados por documento



[2] Los datos completos con los resultados parciales y globales de las fases 1 y 2 pueden consultarse en el anexo.

Figura 4. Ejemplo de gráfico de resumen de resultados por sistema



Systems	Document	Segment	BLEU	NIST	-WER	-TERbase
Cultural_Apertium_EN.txt	*no name*	1	0.4666	4.5568	-0.3846	-0.375
Cultural_Apertium_EN.txt	*no name*	2	0.0366	1.6944	-1.1818	-1.0833
Cultural_Apertium_EN.txt	*no name*	3	0.2597	3.9368	-0.6429	-0.5161
Cultural_Apertium_EN.txt	*no name*	4	0.2704	3.8613	-0.5	-0.4706
Cultural_Apertium_EN.txt	*no name*	5	0.4021	4.0045	-0.4286	-0.4706
Cultural_Google_EN.txt	*no name*	1	0.4422	4.4435	-0.3846	-0.25
Cultural_Google_EN.txt	*no name*	2	0.0904	3.2559	-0.9091	-0.75
Cultural_Google_EN.txt	*no name*	3	0.2717	3.7021	-0.5714	-0.4839
Cultural_Google_EN.txt	*no name*	4	0.4507	4.8735	-0.4375	-0.4118
Cultural_Google_EN.txt	*no name*	5	0.7206	5.3823	-0.2143	-0.2941
Cultural_Microsoft__EN.txt	*no name*	1	0.4422	4.4435	-0.3846	-0.25
Cultural_Microsoft__EN.txt	*no name*	2	0.0407	1.2013	-1	-0.9167
Cultural_Microsoft__EN.txt	*no name*	3	0.3272	4.5197	-0.5714	-0.3871
Cultural_Microsoft__EN.txt	*no name*	4	0.6029	5.0142	-0.3125	-0.2941
Cultural_Microsoft__EN.txt	*no name*	5	0.4615	4.3736	-0.3571	-0.4118
Cultural_SYSTRAN_EN.txt	*no name*	1	0.2095	3.7724	-0.5385	-0.4375
Cultural_SYSTRAN_EN.txt	*no name*	2	0.0855	2.9353	-0.9091	-0.6667
Cultural_SYSTRAN_EN.txt	*no name*	3	0.1861	3.7241	-0.5714	-0.5161
Cultural_SYSTRAN_EN.txt	*no name*	4	0.3807	3.7436	-0.5625	-0.5294
Cultural_SYSTRAN_EN.txt	*no name*	5	0.419	3.8561	-0.4286	-0.4706

*Click on the headers to reorder the table according to the metric.

Figura 5. Ejemplo descartado de cuadro de resumen de resultados por segmentos

Al final del análisis de cada texto breve se aporta una tabla resumen con los archivos que han conseguido las mejores puntuaciones en las métricas, acompañada de un breve comentario de valoración parcial. Por último, como conclusión a los análisis de la fase 1 se incluye un resumen general en el que se detallarán los mejores comportamientos de los motores según los pares de lenguas intervinientes en las traducciones y su área de especialización.

4.5.2. Análisis de resultados de la fase 2

En la fase 2, donde se han analizado las traducciones al español de los dos textos más extensos (uno científico-técnico y otro jurídico) con sus respectivas traducciones automáticas, además de proceder de la misma forma que en la fase 1 con las cuatro métricas de análisis lingüístico, también se ha llevado a cabo la comparación de sus respectivas traducciones mediante el programa Rainbow del paquete Okapi Framework. En esta ocasión, y en lo que al texto científico-técnico se refiere, también se ha analizado una traducción adicional llevada a cabo

con un motor entrenado residente (SYSTRAN). Al igual que en la fase 1, se aportan las cuatro valoraciones de las métricas lingüísticas en el mismo tipo de tablas y gráficos, a los que hay que añadir los resultados de la función de comparación de traducciones de Rainbow, que aportan los valores de correspondencia parcial o *fuzzy match* (FM) y distancia de edición (ED) tanto por segmento como por palabra:

Total Number of Segments:	61
Total Number of Words:	1138
Average word count per segment:	18,66
Average ED-Score (by segment):	Referencia científico to Systran pre-edit = 81,89, Referencia científico to Systran entrenado = 78,87, Systran pre-edit to Systran entrenado = 92,87
Average FM-Score (by segment):	Referencia científico to Systran pre-edit = 87,41, Referencia científico to Systran entrenado = 85,93, Systran pre-edit to Systran entrenado = 96,67
Average ED-Score (by word):	Referencia científico to Systran pre-edit = 80,34
Average FM-Score (by word):	Referencia científico to Systran pre-edit = 87,57
Edit Effort Score:	16,04

Figura 6. Resultados de correspondencia parcial y distancia de edición de Rainbow

Los resultados de esta segunda fase, además de presentarse de una forma similar a los de la fase 1 para constatar los motores con un mejor rendimiento en cada caso, contarán con una tabla de resumen al final de la sección que ilustre las conclusiones globales.

A continuación procederemos a presentar los resultados parciales de las fases 1 y 2, así como las conclusiones generales de las evaluaciones de calidad.

4.5.3. FASE 1: Resultados comparados por lenguas

Presentamos los resultados comparados de los cuatro motores para cada género textual y clasificados por lenguas meta. Para ello hemos considerado los motores que han obtenido las puntuaciones máximas en las métricas seleccionadas (BLEU, NIST, -TER –WER).

1. **Catalán** (comparación de tres motores)

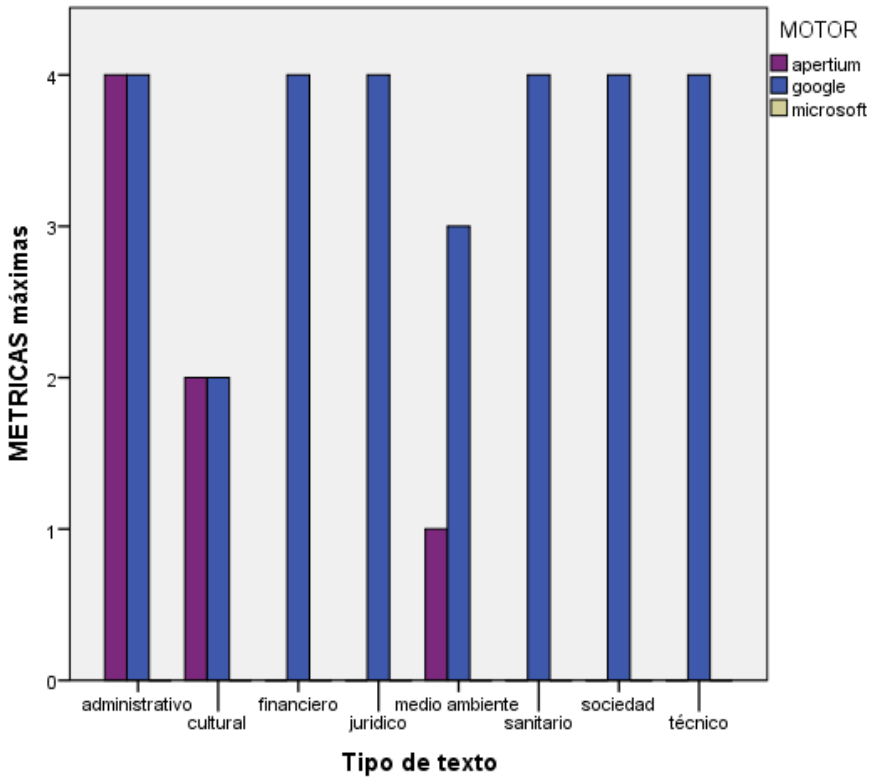


Figura 7. Comparación de tres motores: catalán

Para la lengua catalana podemos concluir que el motor que mejor rendimiento ha mostrado es Google Translate, ya que obtiene los mejores resultados en cada tipo de texto.

En seis de las ocho especializaciones (administrativo, financiero, jurídico, sanitario, sociedad, técnico) Google obtiene la puntuación máxima en las cuatro métricas. El motor Apertium obtiene buenos resultados en textos administrativos e iguala a Google en textos culturales (dos puntuaciones máximas). Podemos observar también que el motor de Microsoft para la lengua catalana no consigue obtener ninguna puntuación máxima en ningún tipo de texto.

En resumen, en el caso de la traducción español-catalán, y si tenemos en cuenta las puntuaciones máximas, el motor que mejor funciona es Google. No obstante, conviene

recordar, tal y como puede comprobarse en los resultados parciales de cada texto, que el motor por reglas Apertium ha ocupado la segunda posición en la mayoría de los casos y en muchas ocasiones con valores casi idénticos a los de Google, por lo que no debería desecharse para trabajos de traducción automática con esta combinación de lenguas.

2. **Euskera** (comparación de dos motores)

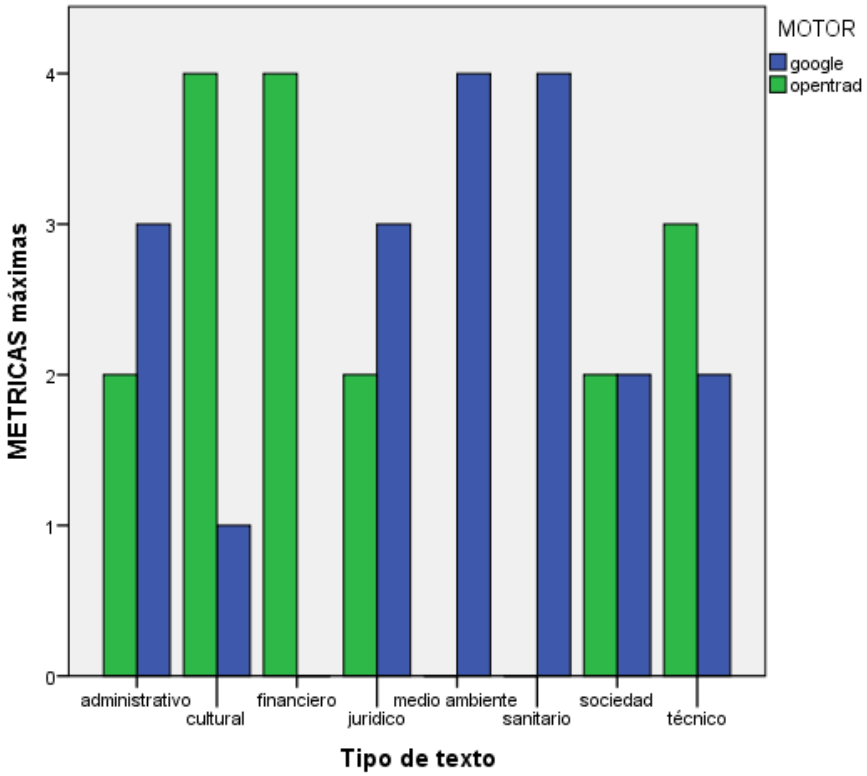


Figura 8. Comparación de dos motores: euskera

Para las traducciones con la lengua vasca, los dos motores que se han probado (Google y OpenTrad) obtienen resultados comparables y equivalentes: ambos obtienen la unanimidad de las cuatro métricas en dos tipos de texto (cultural y financiero para OpenTrad y medio ambiente y sanitario para Google). Obtienen una media de puntuaciones máximas similares, aunque la de Google es ligeramente superior a la de OpenTrad (Google: 2,3, OpenTrad: 2,1). En este caso podemos afirmar que ambos motores funcionan bien y obtienen resultados satisfactorios salvo en textos medioambientales y sanitarios, donde OpenTrad no obtiene ninguna puntuación máxima.

3. **Gallego** (comparación de dos motores)

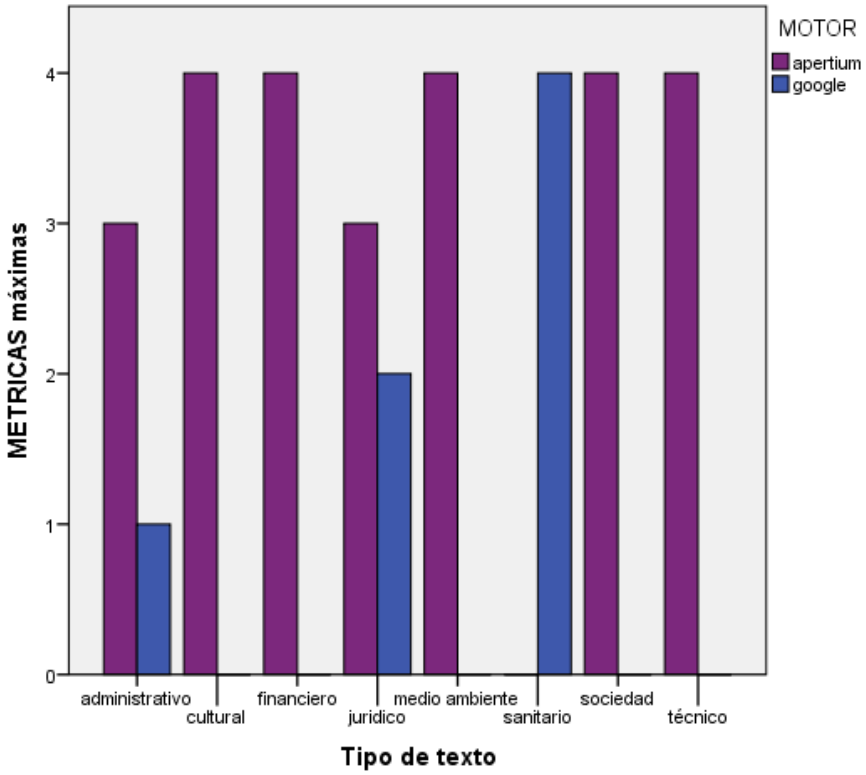


Figura 9. Comparación de dos motores: gallego

En cuanto a las evaluaciones de las traducciones a la lengua gallega, Google ya no presenta una posición tan *hegemónica* en los resultados. En este caso es el motor Apertium el que obtiene la unanimidad en las cuatro métricas en seis tipos de textos (culturales, financieros, medioambientales, sociales y técnicos) y puntuaciones superiores a las de Google en los demás (administrativos, jurídicos), a excepción de los textos sanitarios, único caso en el que Google lo supera.

Plataforma del Español · GOBIERNO DE ESPAÑA · MINISTERIO DE INDUSTRIA, ENERGÍA Y TURISMO · AEI minetur · agrupaciones empresariales innovadoras · FEDER Fondo Europeo de Desarrollo Regional · UNIÓN EUROPEA "Una manera de hacer Europa"

4. **Inglés** (Comparación de cuatro motores)

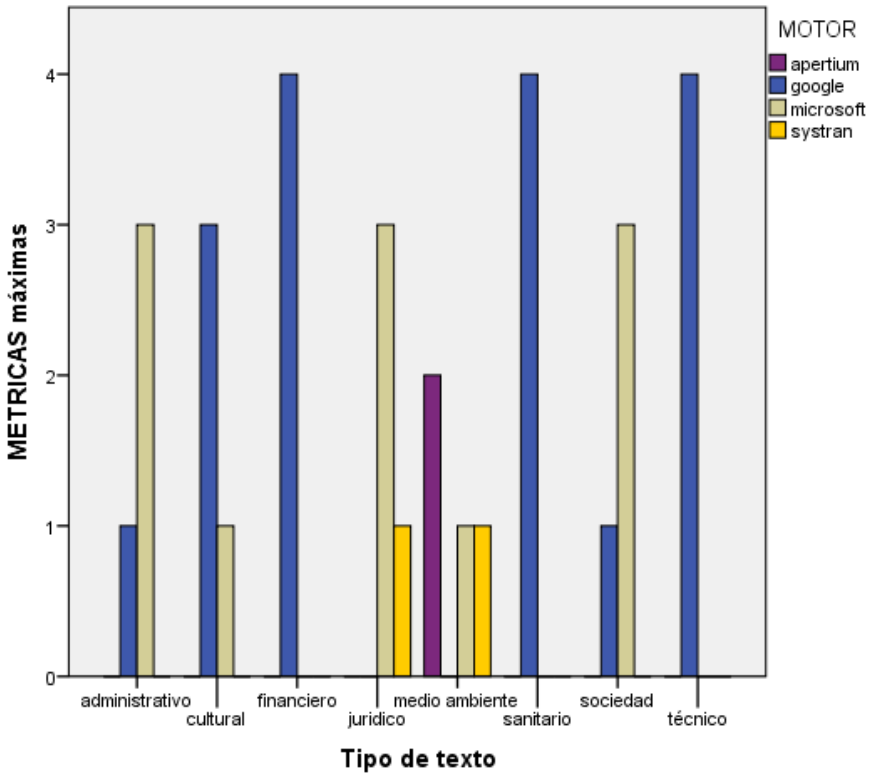


Figura 10. Comparación de cuatro motores: inglés

En lo que respecta a las traducciones español-inglés, de los cuatro motores que se han probado, destaca una vez más el motor Google Translate, que obtiene la unanimidad de máximos resultados positivos en tres ocasiones (textos financieros, sanitarios y técnicos) cuando ninguno de los tres restantes lo consiguen en ningún género, si bien en numerosas ocasiones le siguen a corta distancia. Tal y como consta en los gráficos de los resultados parciales, debemos recordar que la misma regularidad puede apreciarse en la clasificación de los motores que ocupan la segunda y tercera posición, esto es, el traductor de Microsoft y SYSTRAN. Dicha regularidad revela a todas luces la superioridad de los motores estadísticos por encima de los de reglas en las traducciones al inglés, a pesar de que los resultados conseguidos por los motores híbridos no son desdeñables en absoluto.

Plataforma del Español · GOBIERNO DE ESPAÑA · MINISTERIO DE INDUSTRIA, ENERGÍA Y TURISMO · AEI minetur · agrupaciones empresariales innovadoras · FEDER Fondo Europeo de Desarrollo Regional · UNIÓN EUROPEA "Una manera de hacer Europa"

4.5.4. FASE 2: Evaluación lingüística y de productividad de dos textos especializados

Una de las mayores diferencias de esta segunda fase con respecto a la primera reside en que la longitud de los textos sometidos a análisis permite interpretar mejor el alcance de las puntuaciones de las métricas, ya que intervienen otros factores que también deben tenerse en cuenta a la hora de determinar el rendimiento real de las traducciones automáticas generadas por los motores seleccionados: en el caso del texto jurídico, dos motores estadísticos, uno híbrido y otro basado en reglas, todos ellos generalistas, y en el texto científico-técnico, todos los anteriores y un motor entrenado *offline* (SYSTRAN de escritorio con diccionarios especializados instalados), el análisis de su calidad lingüística, la valoración de su rendimiento según lenguaje más o menos reiterativo o formulario, o el análisis contrastivo de motores generales y entrenados en determinadas áreas de especialización. Por ello, a la hora de determinar los más adecuados para los dos textos propuestos, hemos presentado las máximas puntuaciones de las cuatro métricas intervinientes en la fase 1 más las puntuaciones de Rainbow obtenidas en la distancia de edición (ED) y correspondencia parcial o *Fuzzy Match* (FM) tanto en segmentos como en palabras.

Centrándonos ya en el análisis de las traducciones al español del texto jurídico (un contrato), las valoraciones de todas las métricas se decantan por las generadas por Google Translate (estadístico), a pesar de que la valoración de la distancia de edición en Rainbow sea inferior a la del segundo motor (SYSTRAN, híbrido), el cual a su vez obtiene resultados muy próximos a los de Google, aunque a la baja (circunstancia que se repite con Microsoft, lo que confirma la tendencia ya analizada de la solidez y fiabilidad de los motores estadísticos). Por el contrario, Apertium (reglas) no mantiene el nivel de los demás motores en este campo, al conseguir los peores resultados y situarse a una distancia considerable de los otros tres.

En lo que respecta a la evaluación de calidad de las traducciones automáticas al inglés del texto científico-técnico (un manual de usuario de un programa antivirus) se ha incorporado un quinto motor de traducción automática, un programa SYSTRAN para escritorio/servidor entrenado con diccionarios especializados en la materia ya instalados, con el fin de observar si la personalización de dicho tipo de motores conlleva alguna diferencia en los baremos de productividad comparados con los obtenidos por las traducciones del mismo texto en motores convencionales.

El resultado es altamente clarificador, ya que de forma unánime no solo ha destacado como el que mejores traducciones ha generado, al conseguir las mayores puntuaciones en todas las métricas, sino también por el nivel de dichas puntuaciones, claramente más altas que la media conseguidas por los demás motores, en especial en el apartado de correspondencias parciales (FM), con índices superiores a los 85 puntos (en los demás casos, los valores van de 65 a 80 puntos), lo que indica el alto rendimiento de este tipo de sistemas en todos los aspectos: corrección lingüística y productividad.

4.5.5. Resultados comparados de la fase 2

En el cuadro siguiente presentamos los resultados de las valoraciones de todas las métricas aplicadas a las traducciones de los dos textos propuestos según métricas y sistemas de traducción automática:

EN-ES	Archivo	Fuzzy Match (seg/pal)	BLEU	NIST	WER	TER
Científico-técnico	Systran (G)	67,28/67,08	0,2538	4,744	-0,6274	-0,5323
	Systran (E)	85,93/85,84	0,5356	7,1962	-0,3467	-0,2629
	Apertium	66,64/66,83	0,2331	4,8954	-0,5947	-0,5065
	Microsoft	79,70/79,31	0,4704	6,8089	-0,3732	-0,3081
	Google	80,43/80,75	0,4915	6,7707	-0,4179	-0,3129
Jurídico	Systran (G)	60,70/61,49	0,2117	4,9254	-0,6622	-0,5619
	Apertium	22,82/24,69	0,0183	1,076	-0,9703	-0,8885
	Microsoft	60,49/64,39	0,1949	4,785	-0,7158	-0,5592
	Google	62,80/65,73	0,2249	5,0766	-0,7099	-0,5336

A pesar de obedecer a supuestos algo diferentes (en el caso de las traducciones del texto jurídico, el análisis está orientado a la determinación del mejor motor general en la traducción automática de un texto especializado tanto en su rendimiento lingüístico como productivo; en el del texto científico-técnico, los resultados de los motores generales se comparan además con los de un motor residente entrenado), pueden observarse líneas de interpretación secundarias de las evaluaciones obtenidas en las métricas.

En primer lugar, y en lo que a las traducciones del texto jurídico a través de motores generales se refiere, es el motor estadístico el que obtiene las mejores puntuaciones en todas las métricas, incluida la de Rainbow, que computa el porcentaje de correspondencias parciales, seguido de un híbrido no entrenado (SYSTRAN) y de otro estadístico, con valores prácticamente equiparables, por lo que también deberían tenerse en cuenta a la hora de determinar los motores con mejores rendimientos (en ambos casos, Apertium aparece en el puesto más bajo).

En segundo lugar, y atendiendo esta vez a las traducciones del texto científico-técnico, la incorporación al estudio de un motor entrenado *offline* (SYSTRAN) con acceso a diccionarios especializados se ha llevado a cabo con el doble objetivo de determinar las diferencias en productividad con respecto a los motores generales anteriormente citados y su operatividad efectiva dentro de los procesos de traducción como herramienta integrante. Pese a que los resultados de los motores generales repiten el mismo patrón del predominio de los estadísticos sobre los basados en reglas e híbridos, sí es cierto que, tal y como hemos podido comprobar en los resultados parciales, la personalización de motores híbridos residentes según las necesidades específicas de la empresa hace que este motor destaque por encima de los demás como una gran herramienta que debe tenerse en cuenta en la implementación del *software* de los sistemas de traducción en las empresas a tenor de los excelentes resultados obtenidos en el estudio, francamente superiores a los de los demás.

4.6. Conclusiones generales de la evaluación

La presente evaluación de los motores de traducción automática, tanto en línea como en versión residente entrenada, no busca determinar cuál es el mejor como solución única con vistas a su incorporación en un proceso de traducción profesional. Nuestro objetivo ha sido el de delimitar tendencias según sus respectivos rendimientos atendiendo tanto a las lenguas implicadas en el proceso de traducción como a las especializaciones de los textos originales, para de facilitar una elección correcta en cada caso.

Hemos podido concluir que, si tenemos en cuenta dichas tendencias, los motores estadísticos han demostrado ser más eficientes en la práctica totalidad de los textos especializados y en las combinaciones de las lenguas más extendidas, como el catalán y el inglés. Estos resultados resultan hasta lógicos si tenemos en cuenta el volumen de corpus paralelos bilingües que poseen para realizar las traducciones y el gran número de lenguas disponibles para las combinaciones lengua original-lengua meta. Por su parte, los motores basados en reglas han resultado ser la mejor opción en la mayoría de los textos especializados cuando se trata de combinaciones con lenguas cooficiales más minoritarias (es lo que sucede con el euskera y el gallego, y OpenTrad y Apertium, respectivamente), en especial cuando la lengua meta es lingüísticamente próxima a la original, como el gallego. Esta circunstancia nos lleva a pensar que, si bien es cierto que los motores estadísticos (Google en especial) cuentan con un gran abanico de lenguas para determinar las combinaciones intervinientes en las traducciones, la disponibilidad cuantitativa (y cualitativa) de corpus paralelos de las lenguas minoritarias para que estos motores lleven a cabo su análisis estadístico es significativamente más pequeña que con otras más generalizadas como el inglés o el español, por lo que es en estos ámbitos donde los motores basados en reglas ofrecen mejores resultados.

Ahora bien, no debemos olvidar que algunos de dichos motores son susceptibles de ser entrenados con diccionarios y memorias de traducción *ad hoc* según la tipología de los proyectos de traducción de una empresa del sector en sus versiones para su instalación en servidores internos o en ordenadores. Hemos podido comprobar a través de los datos

numéricos de las métricas que el rendimiento de los sistemas de traducción automática en estos casos mejora de forma considerable debido a la evidente adecuación de los sistemas de TA a las demandas de las empresas de traducción desde el punto de vista lingüístico y productivo, por lo que su incorporación a la infraestructura informática debe ser considerada como una herramienta válida para propiciar los índices de mejora en los tiempos de ejecución y la calidad de los proyectos de traducción.

5. Viabilidad económica y modelo de negocio: estudio de caso

El estudio de caso se ha realizado en la sede de la empresa Hermes Traducciones y Servicios Lingüísticos, SL. http://www.hermestrans.com/.

Hermes y Traducciones es una empresa con capital 100 % español fundada en 1991, que cuenta con doble certificación de calidad por la norma europea de traducción EN-15038 y por la ISO-9001. Es, además, la única empresa española que cuenta con el certificado de calidad lingüística de la Fundéu-BBVA. Un importante diferenciador de Hermes Traducciones sobre otras empresas del sector es su capacidad de producción interna y la implantación de un sistema centralizado de gestión de la calidad que permite el control exhaustivo de todos los procesos humanos y técnicos, con más de 150 indicadores de procesos, que permiten localizar patrones de error indebidos para una rectificación casi instantánea. Con especialización en localización, cuenta con tecnología propia para su sistema de gestión del rendimiento (SGR) y desarrollo de aplicaciones internas que redunden en la productividad personal. No solo dedicada a un perfil empresarial, colabora asiduamente con numerosas universidades españolas y extranjeras en todo lo relacionado con el sector de la traducción. Finalmente, pertenece a numerosas organizaciones sectoriales en España y en el mundo, y figura en la lista de empresas de Europa meridional en la posición 15.ª y en la 157.ª del mundo según CSA.

Para elaborar el estudio de caso se han revisado los diferentes aspectos planteados en las secciones precedentes de este informe, con respecto a los siguientes puntos:

- cuestiones técnicas y de procesos;
- evaluación de la calidad;
- modelo económico.

5.1. Cuestiones técnicas y de procesos

El caso que se describe corresponde al uso de un motor de TA basado en reglas (Systran) para la combinación inglés-español. Systran es una solución de escritorio y el proceso que aquí se describe refleja precisamente esta circunstancia. En el caso de motores de TA en la nube, será necesario adaptar el proceso. Para más información, puede verse un catálogo de herramientas completas en EAMT (http://www.eamt.org) TAUS (https://www.taus.net) y sus diferentes módulos complementarios *(plug-ins).*

A continuación, se detalla todo el proceso y se anotan las cuestiones técnicas cuando es necesario.

- Paso 1. El cliente envía el *kit de localización/traducción*, que contiene todos los materiales necesarios para realizar el trabajo (glosarios, memorias, documentación, *software*, imágenes…).

- Paso 2. Se extrae el texto e, independientemente del formato en el que lo envíe el cliente, se genera un archivo XLIFF.

- Paso 3. Se analiza el texto para comprobar el grado de aprovechamiento de la memoria de traducción. En la figura 11 se muestra un ejemplo del análisis de un texto de 3.901 palabras, de las cuales 328 tienen una correspondencia del 100 % en la memoria; 160 están entre 95 % y 99 %; y 187 entre 85 % y 94 %. Normalmente, las correspondencias que están por debajo del 75 % se pasan al motor de TA, con la excepción de traducciones en el par catalán-español en las que el umbral es superior (95 %) dados los buenos resultados que se obtienen con el motor Apertium.

Type	Segments	Words	Characters	Percent
PerfectMatch	0	0	0	0.00%
Context Match	38	249	1544	6.38%
Repetitions	73	304	2913	7.79%
Cross-file Repetitions	1	12	60	0.31%
100%	51	328	2011	8.41%
95% - 99%	17	160	982	4.10%
85% - 94%	29	187	1053	4.79%
75% - 84%	27	201	1198	5.15%
50% - 74%	14	116	634	2.97%
New	203	2344	12791	60.09%
Total	**453**	**3901**	**23186**	**100%**

Figura 11. Análisis de la memoria

El análisis del aprovechamiento de la memoria de traducción tiene tres ventajas principalmente:

a. permite reducir el precio por palabra en aquellos segmentos que se traducen bien con la memoria, bien mediante TA;
b. en el caso de que se utilice la TA como un servicio contratado externamente a la empresa, se reduce también el precio por palabra traducida puesto que únicamente se envían los segmentos que quedan por debajo del umbral previamente definido;
c. se ahorran costes en la creación de glosarios del cliente/proyecto porque no es necesario incluir todas aquellas palabras que están por encima del umbral definido.

- Paso 4. Exportación de los segmentos que están por debajo del umbral del 75 %.

- Paso 5. Conversión de formato. Estos segmentos, que están en formato XLIFF se pasan a texto plano para facilitar su procesamiento en el sistema de TA. En las figuras 12 y 13 puede verse el mismo segmento en formato XLIFF y TXT. El programa que se emplea para realizar la conversión es Rainbow(Okapi).

```
</sdl:seg-defs></trans-unit><trans-unit id="ab5dee1e-c43f-4f23-a1c1-15eddad35b73"><source>
4) For remote installation of Administration Server, please specify the value /v&quot;EULA=1
&quot; for the command line options in the installer, in the Administration Server
installation package&apos;s configuration file, in order to express your acceptance of the
License Agreement.</source><seg-source><mrk mtype="seg" mid="21">4) For remote installation
of Administration Server, please specify the value /v&quot;EULA=1&quot; for the command
line options in the installer, in the Administration Server installation package&apos;s
configuration file, in order to express your acceptance of the License Agreement.
</mrk></seg-source><target/><sdl:seg-defs><sdl:seg id="21" />
</sdl:seg-defs></trans-unit><trans-unit id="aba69a43-cb07-45cc-a021-dc8c4712a58f"><source>
On Microsoft Windows operating systems:</source><seg-source><mrk mtype="seg" mid="22">On
Microsoft Windows operating systems:</mrk></seg-source><target/><sdl:seg-defs><sdl:seg id=
"22" /></sdl:seg-defs></trans-unit><trans-unit id="55cd48b8-5863-4318-a576-7c025e1ed3e2"
><source>On Linux operating systems:</source><seg-source><mrk mtype="seg" mid="23">On Linux
operating systems:</mrk></seg-source><target/><sdl:seg-defs><sdl:seg id="23" />
```

Figura 12. Ejemplo de texto en formato XLIFF

```
4) For remote installation of Administration Server, please specify the value /v"EULA=1" for
the command line options in the installer, in the Administration Server installation
package's configuration file, in order to express your acceptance of the License
Agreement.CRLF
On Microsoft Windows operating systems:CRLF
On Linux operating systems:CRLF
```

Figura 13. El mismo texto en formato plano, listo para pasarlo por la TA

- Paso 6. Se envía el texto para su traducción en Systran. Este sistema permite la adaptación de los glosarios y su posterior selección por proyecto/cliente, con indicación del orden de prioridad a la hora de utilizarlos.

- Paso 7. Evaluación léxica de los resultados. Se analiza la traducción para comprobar qué palabras no se encuentran en los diccionarios (figura 14, palabras marcadas en rojo).

For remote installation of Administration Server, please specify the value /v"EULA=1" for the command line options in the installer, in the Administration Server installation package's configuration file, in order to express your acceptance of the License Agreement.	Para la instalación alejada del servidor de la administración, especifique por favor el valor /v " EULA=1" para la línea de comando opciones en el instalador, en el fichero de la configuración de la administración del servidor del paquete de la instalación, para expresar su aceptación del contrato de licencia.
On Microsoft Windows operating systems:	En los sistemas operativos de Microsoft Windows:
On Linux operating systems:	En los sistemas operativos de Linux:

Figura 14. Resultados de Systran con las palabras que no se encuentran en el diccionario marcadas en rojo

- Paso 8. Se activan los diccionarios personalizados y se comprueban los cambios (figura 15, palabras marcadas en verde).

For remote installation of Administration Server, please specify the value /v"EULA=1" for the command line options in the installer, in the Administration Server installation package's configuration file, in order to express your acceptance of the License Agreement.	Para la instalación remota del Servidor de Administración, especifique por favor el valor /v " EULA=1" para las opciones de la línea de comandos en el instalador, en el archivo de la configuración del paquete de instalación del Servidor de Administración, para expresar su aceptación del contrato de licencia.
On Microsoft Windows operating systems:	En los sistemas operativos de Microsoft Windows:
On Linux operating systems:	En los sistemas operativos de Linux:

Figura 15. Las palabras en verde marcan los resultados de los diccionarios personalizados

- Paso 9. Se exporta la traducción y se aplica una serie de *scripts* de posedición que, de manera automática, activan cambios como los siguientes:

  - eliminar la expresión *por favor*;
  - eliminar los usos innecesarios de *usted*;
  - cambiar la voz pasiva a pasiva refleja;
  - cambiar léxicamente el verbo *querer* por *desear*;
  - corregir el uso incorrecto de preposiciones;
  - corregir el uso incorrecto de comas;
  - corregir el orden sintáctico incorrecto;
  - eliminar el uso incorrecto del guion.

  Una vez que se han aplicado estas reglas (que dependen de cada par de idiomas), se genera, de manera automática, un informe de los cambios realizados de modo que se pueda hacer en todo momento un seguimiento del trabajo.

- Paso 10. Se inserta de nuevo el texto en el archivo XLIFF original de donde se exportó originalmente.

- Paso 11. Se crea una memoria de traducción con todo el proyecto y se pasa al poseditor, con indicación del origen de cada traducción (bien de la propia memoria, bien como resultado de la TA).

- Paso 12. La posedición se lleva a cabo directamente en el entorno de traducción.

- Paso 13. Cierre del proyecto. El poseditor aporta información que servirá para actualizar los glosarios del sistema de TA, realizar análisis de la productividad alcanzada en el proyecto y optimizar los costes.

5.2. Evaluación de la calidad

Dado que el principal objetivo de la TA es el ahorro de costes y reducción de plazos, el indicador más importante es el aumento de productividad logrado por el poseditor, calculado a partir de las horas dedicadas a la posedición frente a las horas estimadas de traducción sin TA. En el caso estudiado, el aumento de productividad fue de un 26 %.

Ahora bien, ese aumento de productividad podría no deberse exclusivamente a la calidad de la traducción automática, sino que pueden haber intervenido las numerosas variables que afectan a la productividad de los traductores. Para lograr una estimación más objetiva de la calidad de la TA que contribuya a elaborar un modelo de predicción de costes y plazos para futuros proyectos similares, se calculan las métricas automáticas mencionadas anteriormente.

Para ello, se genera un documento con el resultado de la TA y otro con los segmentos poseditados. Debido a que los archivos finales contienen una mezcla de coincidencias de la memoria y los propios segmentos poseditados, hay que lograr aislar los segmentos procesados por TA. La exportación inicial de los segmentos con coincidencia inferior a un 75 % permite identificar estos segmentos fácilmente. También pueden utilizarse los atributos que algunos entornos de traducción asistida escriben en archivos XLIFF.

Además de los valores BLEU y TER (calculados mediante Asiya en este caso), también puede calcularse una distancia de edición equivalente a las coincidencias parciales de la memoria. Este valor tiene la ventaja de ser más intuitivo para toda persona familiarizada con los sistemas de traducción asistida. Debido a que no hay un estándar para el cálculo de coincidencias parciales, en este ejemplo se ha usado el valor arrojado por Okapi Rainbow.

Aumento de productividad	Coincidencia parcial	BLEU	TER
26 %	78,61 %	51,10	35,30

Debido a los numerosos factores que afectan a la productividad de cada traductor y a las propias limitaciones de cada métrica, es difícil establecer una norma universal que permita predecir cuantitativamente con total exactitud plazos y costes. Es decir, no puede asegurarse que en futuros proyectos con idéntico valor de BLEU se alcanzará idéntica productividad. Sin embargo, el uso conjunto de estas métricas sí permite obtener una visión cualitativa del rendimiento de la TA más objetivo, lo cual es particularmente útil a la hora de negociar con clientes que intenten imponer descuentos con un TA de baja calidad.

5.3. Modelo económico

A continuación se detallan los datos económicos asociados a la implantación de un modelo de traducción automática en el flujo de trabajo de una empresa de traducción de tamaño medio.

Para ello se han tomado como referencia los datos de costes pertenecientes a una persona dedicada a tiempo completo para las labores de integración del proceso, creación de motores de traducción, preparación de corpus, mantenimiento y entrenamiento de los motores, y procesamiento de archivos de traducción.

Como los datos de salarios mínimos del convenio de Oficinas y Despachos afecto al sector de la traducción presentan un mínimo anual de 16 410 euros brutos para un perfil técnico, que consideramos bajo, se han comprobado diversos estudios de Common Sense Advisory y LISA más específicos del sector de la traducción y localización, si bien estos datos pueden tener cierta obsolescencia por su fecha de publicación. Por lo tanto, se ha consultado a ASPROSET (Asociación Sectorial de Proveedores de Servicios de Traducción) al respecto del salario anual bruto que podría tener un perfil técnico para la ejecución de las tareas arriba mencionadas. A tal respecto, se establece como referencia un salario anual neto, incluidos los costes de Seguridad Social y fiscales, de 36 000 euros. A partir de esta cantidad se tienen en cuenta 1772 horas laborables según el convenio, lo que arroja un precio por hora de 20,32 euros, y a su vez 162,53 euros como coste por día.

El perfil de esta persona debe incluir conocimientos técnicos de programación y un conocimiento del sector de la traducción y los procesos y flujo de trabajo que conllevan un proyecto habitual, puesto que se plantea el uso de la traducción automática como un medio complementario del proceso y no como el fin último. Asimismo, este técnico puede efectuar otras tareas diversas de apoyo relacionadas con las tecnologías aplicadas de la información, pero para este estudio se ha tenido en cuenta una dedicación del 100 %.

En un principio se entiende que el trabajo de este técnico requerirá muchas horas de preparación de todo el entorno, recopilación y optimización de corpus, creación de motores y posterior entrenamiento, por lo que la recuperación inicial de la inversión durante el primer año —o medio año—, sobre todo, será mínima o inexistente. Una vez establecidos los procesos de traducción automática e integrados en el flujo de trabajo específico, empezará a mostrar cierta reducción paulatina de costes.

Como ejemplo de esto último, hemos considerado la comparación de un proyecto concreto de 260 971 palabras con la combinación de francés a español desde tres perspectivas de costes: traducción completa por traductores, uso de programas de traducción asistida, y traducción automática y posedición.

Comparativa de los tres modelos de traducción

La traducción con el modelo tradicional efectuada por traductores sin ninguna herramienta de reaprovechamiento supondría un esfuerzo de 696 horas/persona. Para este cálculo se asume

un ritmo de 375 palabras por hora (equivalente al estándar de 3000 palabras por día). En este caso no se han tenido en cuenta las horas necesarias de revisión por un tercero.

La traducción mediante programas de traducción asistida requirió un examen previo de reaprovechamiento para poder diferenciar entre el número de palabras totales y palabras compensadas,[3] referencia crucial para el cálculo de horas de traducción. Para tal fin, los porcentajes de compensación se han establecido a partir de valores habituales en el sector en proyectos debidamente dimensionados, como se puede ver en el cuadro siguiente:

Franjas de reaprovechamiento	Compensación	Palabras totales	Palabras compensadas
Repeticiones	25 %	133 470	33 367,5
95-99 %	30 %	3860	1158
85-94 %	40 %	5479	2191,6
75-84 %	50 %	14 804	7402
0-74 %	100 %	103 358	103 358
Totales		**260 971**	**147 477**

Con este enfoque de uso de memorias de traducción las horas necesarias fueron 393, lo que supone una significativa reducción en el tiempo de dedicación, a la vez que un ahorro de costes muy importante. Al igual que en el caso anterior con traducción manual, no se han tenido en cuenta las horas de revisión que supondrían un incremento adicional del coste.

Finalmente, la traducción automática requirió 332 horas de posedición, sin añadir las 50 horas correspondientes a labores de ingeniería para la adecuación del motor de traducción y el resto, a las tareas de posedición, que están cubiertas en el coste de inversión de esta solución, pues se entiende que es un coste de infraestructura ya contemplado y no un coste de producción propiamente dicha. En este supuesto la traducción automática tenía una finalidad complementaria, pues se combinó con la traducción asistida.

En el cuadro siguiente puede verse el resumen comparado de las horas (en el caso de la traducción manual y asistida se trata de horas teóricas a partir de los distintos análisis de

[3] Se entiende por palabras *compensadas* el resultado de aplicar a las palabras totales los porcentajes de reaprovechamiento de los distintos tramos reutilizables de la memoria, que siempre dará un número inferior de palabras al total. Se trata de un concepto fundamental en el sector porque se utiliza para los cómputos de productividad y plazos.

volúmenes, mientras que en el caso de la traducción automática son horas reales de las que informaron los poseditores):

Tipo de traducción	Palabras	Horas	% de ahorro sobre traducción manual	% de ahorro sobre traducción asistida (TAO)
Traducción manual	260 971	696	---	---
Traducción con TAO	260 971	393	43,53	---
Traducción con TAO+TA+PE	260 971	331	52,44	15,78

Además, al haber usado un recurso interno para la posedición, las horas ahorradas respecto a los modelos más tradicionales pudieron emplearse para otros proyectos con el consiguiente incremento de productividad y ahorro en costes.

En el caso de posteriores proyectos de tipo similar basados en ese mismo motor, los porcentajes de ahorro con toda seguridad mostrarían incluso mejores resultados por el propio reentrenamiento del motor.

6. Conclusiones

La implantación de la traducción automática no es algo trivial. No se trata simplemente de incorporar una nueva herramienta a las que ya puedan existir en la empresa de traducción, sino que implica una revisión en profundidad de los procesos con los que se trabaja y determinar, además, con qué objetivo queremos introducir un nuevo elemento al flujo de trabajo:

- ¿Queremos aumentar la productividad?
- ¿Queremos mejorar la calidad?
- ¿Queremos diseñar un nuevo modelo de negocio?
- ¿Queremos ofrecer un nuevo servicio?
- ¿Queremos reducir costes?

En este sentido, como hemos podido ver a lo largo del estudio, será necesario **evaluar las prestaciones del sistema de traducción automática que queramos implantar** o, incluso, realizar un análisis comparativo de diferentes motores antes de decidirnos por uno en concreto.

Así, tras desarrollar este estudio particular, entendemos que se deben tener en cuenta las conclusiones siguientes:

- **La integración de la TA en un entorno productivo real no es trivial, puesto que afecta a los procesos empresariales y a los procedimientos operativos, que son el lugar natural de la gestión de proyectos**, pero también precisan el **apoyo de los recursos técnicos** como se ha explicado.

- **Se pueden dar problemas de seguridad y confidencialidad** en el uso de materiales de clientes cuando se utilizan plataformas públicas, por lo que se recomienda encarecidamente que estos contenidos siempre se encuentren en servidores de la instalación interna de la empresa sin sacarlos al exterior.

- **La traducción automática no se trata de una herramienta estándar que puede empezar a usarse según se compra**, puesto que requiere un alto índice de **personalización y mantenimiento continuo**, **adaptándola a las distintas necesidades** de cada dominio o especialización, e incluso cliente específico.

- **La implantación de un modelo de traducción automática y posedición precisa de una gestión eficaz del cambio**, pues el paradigma de trabajo puede variar considerablemente en muchas ocasiones, lo que puede dar problemas internos de aceptación por parte de personas con más experiencia en entornos más tradicionales; esto se acrecienta incluso cuando se cuente con colaboradores externos, dado que existe un rechazo de entrada a este tipo de soluciones, muchas veces basado en el desconocimiento de lo que puede suponer un uso racional como apoyo y no como fin.

- Por la razón anterior, **es conveniente desmitificar los rasgos negativos que comporta la traducción automática en muchos entornos** debido a un uso indebido y generalizado mal entendido.

- **El conocimiento de uso de esta tecnología puede comportar ventajas adicionales**: algunos clientes pueden usar la traducción automática de modo indiscriminado para sus traducciones, por lo que la capacidad de análisis de este tipo de trabajos puede ayudar a una valoración real de la calidad o poca calidad de esos textos, y aceptarlos o rechazarlos según proceda, siempre con datos cuantificables y objetivos.

- **Como ya se ha comentado en varias ocasiones, esta tecnología puede aportar mejoras en la productividad** personal si su implantación se hace de modo racional y personalizado.

Finalmente, podríamos resumir estas conclusiones en tres puntos clave:

- La traducción automática permite mejorar la **productividad.**

- Para obtener unos resultados óptimos **es siempre necesario personalizar el sistema de traducción automática**, bien mediante la creación o adaptación de los diccionarios que utiliza el sistema (en el caso de los basados en reglas) o por la incorporación de corpus de entrenamiento específicos (en el caso de los motores estadísticos).

- **Es imprescindible realizar una integración de la traducción automática** con el resto de herramientas de traducción asistida si se quiere obtener un alto grado de rendimiento.