

Círculo de Lingüística Aplicada a la Comunicación

ISSN: 1576-4737

<https://dx.doi.org/10.5209/clac.77002>

Los motores de traducción automática y su uso como herramienta lexicográfica en la traducción de unidades léxicas aisladas

María Concepción Maldonado González y María Liébana González¹

Recibido: 28 de junio de 2021 / Aceptado: 1 de octubre de 2021

Resumen. Este trabajo ofrece un repaso de algunos sistemas de traducción automática (TA), así como de los problemas más comunes que surgen al hacer uso de la traducción automática en la traducción de unidades léxicas aisladas, fuera de contexto. El objetivo es señalar las limitaciones lingüísticas que presentan los motores de TA y reflexionar sobre la necesidad de incorporar datos lexicográficos a los mismos para mejorar su rendimiento.

Para ello, evaluaremos el comportamiento de Google Traductor, Bing Microsoft Translator y el traductor de DeepL en la traducción de unidades léxicas aisladas. Compararemos los resultados que habríamos obtenido si hubiéramos recurrido a diccionarios bilingües en vez de a motores de TA. Y constataremos que la respuesta generada por los motores de TA resulta insuficiente para solucionar nuestras dudas, y que los datos lexicográficos recogidos en diferentes diccionarios bilingües en línea aportan información lexicográfica más adecuada para satisfacer las necesidades del usuario.

Palabras clave: Traducción, Herramientas de traducción, lexicografía digital, diccionarios bilingües

[en] Machine translation engines and their use as a lexicographic tool to translate self-contained lexical units

Abstract. This paper provides a review of some machine translation (MT) systems, as well as the most common problems we encounter when using machine translation to translate self-contained lexical units out of context.

The focus is to identify the limitations of linguistic MT engines and to consider whether it is necessary to incorporate lexicographic data into these systems, to improve their performance.

To this end, we will take a practical approach and use Google Translator, Bing Microsoft Translator and DeepL to translate self-contained lexical units. We will then evaluate their performance by comparing the results with those we would have obtained had we consulted bilingual dictionaries for the same tasks. We will conclude that the quality of the output generated by MT engines is not sufficiently accurate; in general, the lexicographic data found in many bilingual dictionaries online provides more precise and comprehensive lexicographic information to satisfy the user's communicative needs.

Keywords: Translation, translation tools, electronic lexicography, bilingual dictionaries

Cómo citar: Maldonado, María Concepción; Liébana González, María (2021). Los motores de traducción automática y su uso como herramienta lexicográfica en la traducción de unidades léxicas aisladas. *Círculo de Lingüística Aplicada a la Comunicación* 88, 189-211, <https://dx.doi.org/10.5209/clac.77002>

Índice. 1. Introducción. 2. Herramientas lexicográficas en línea: algunos ejemplos. 3 Los motores de traducción automática. 3.1. Principales sistemas de traducción automática. 3.2. Problemas más comunes de la traducción automática. 3.3. Funcionamiento de Google Traductor, Bing Microsoft Translator y el traductor de DeepL. 4. Análisis de uso de traducción automática en unidades léxicas. 4.1. Una muestra de análisis. 4.2. Valoración de resultados. 5. Conclusiones. Bibliografía.

1. Introducción

La irrupción y el rápido desarrollo de las llamadas Tecnologías de la Información y la Comunicación (TIC) a finales del siglo XX y principios del XXI han tenido un gran impacto en numerosas disciplinas y han modificado muchas de nuestras rutinas: podemos renovar nuestro armario desde la pantalla de nuestro dispositivo móvil, realizar cualquier gestión bancaria sin pisar una sucursal o ver nuestro programa favorito mientras nos dirigimos al trabajo en transporte público.

¹ María Concepción Maldonado González, Universidad Complutense de Madrid. marimald@ucm.es – <https://orcid.org/0000-0003-3543-1778>.
María Liébana González. <https://orcid.org/0000-0002-6413-4548>.

En el caso de la lexicografía, la generalización de internet ha propiciado que los hábitos de consulta de información de los usuarios se caractericen por la necesidad de encontrar respuestas rápidas a las búsquedas realizadas. Como consecuencia, el uso del diccionario en papel se ha visto eclipsado por la creciente demanda de recursos lexicográficos en línea (diccionarios, portales y plataformas multilingües, etc.). Los traductores automáticos se han posicionado como uno de los recursos en línea más demandados por usuarios no profesionales, no solo para traducir textos sino también para realizar consultas bilingües de unidades léxicas. Sin embargo, los motores de traducción automática no son infalibles y, por ello, conocer su comportamiento puede ayudarnos a la hora de hacer uso de este tipo de herramientas.

Este trabajo ofrece un breve repaso de los principales sistemas de traducción automática (TA) que conviven actualmente en el mercado, así como de los problemas más comunes a los que nos exponemos al hacer uso de la traducción automática. El objetivo es señalar, o al menos cuestionar, las limitaciones lingüísticas que presentan los motores de TA y reflexionar sobre la necesidad de incorporar datos lexicográficos a los mismos, con el fin de mejorar su rendimiento. Para ello, evaluaremos, desde un enfoque práctico, el comportamiento de tres herramientas cuyo empleo en la actividad profesional de la traducción en España está más extendido: Google Traductor, Bing Microsoft Translator y el traductor de DeepL.

Diferenciaremos para nuestro estudio dos niveles de uso. Nosotros solo abordaremos en estas páginas el examen de la traducción automática en unidades léxicas y dejaremos para otro estudio el análisis de la traducción automática en textos. En este trabajo que aquí presentamos analizaremos, pues, el funcionamiento de los motores de traducción seleccionados en consultas de unidades léxicas aisladas. Pondremos el foco en la idoneidad de la información lexicográfica en ellos recogida, y compararemos los resultados que habríamos obtenido si, para las mismas tareas, hubiéramos recurrido a diccionarios bilingües en vez de a motores de TA. Constatemos que la respuesta generada por los motores de TA resulta insuficiente para solucionar nuestras dudas y que los datos lexicográficos recogidos en diferentes diccionarios bilingües en línea, por lo general, aportan información lexicográfica más completa y adecuada para satisfacer las necesidades comunicativas del usuario.

Con los datos obtenidos en este trabajo nos atrevemos a afirmar que el rendimiento de los motores de TA mejoraría si se alimentasen no solo de textos paralelos sino también de datos lexicográficos.

2. Herramientas lexicográficas en línea: algunos ejemplos

A raíz de la revolución tecnológica de las últimas décadas, se ha experimentado en las obras lexicográficas una transición en el soporte de sus productos (Maldonado, 2013). Los primeros años, las obras en papel cedieron espacio a aquellas en soporte digital, ya fueran diccionarios electrónicos o diccionarios presentados en CD-ROM. Si bien estos primeros diccionarios electrónicos eran principalmente una versión digital de las obras en papel, abrieron el camino para la verdadera revolución en el campo de la lexicografía: los diferentes recursos lexicográficos en línea.

Las tecnologías han modificado profundamente el uso de los recursos lexicográficos. En la llamada *era digital* nos enfrentamos a un nuevo panorama, donde se observa cómo han cambiado los hábitos de consulta: a día de hoy, los usuarios recurren mayoritariamente a diferentes herramientas en línea para solventar cualquier duda, sea esta lingüística o no. Desde los motores de búsqueda se realizan consultas tan dispares como el resultado del último partido de Rafa Nadal o el plural de *fax*. Es decir, los usuarios recurren a la red para obtener respuestas inmediatas a sus dudas.

Esto no implica que la sociedad actual haya dejado de necesitar diccionarios, sino que los usuarios se han encontrado, de forma más o menos repentina, con una ingente profusión de recursos en línea y recurren a unos u otros indistintamente, sin tener en cuenta sus características.

Por ello, atendiendo a la clasificación de Domínguez (2019), creemos importante diferenciar entre los siguientes recursos:

- Repositorios de recursos lexicográficos: se trata de compendios de diccionarios, que no tienen por qué contar con interconexión ni integración entre sí, y que pertenecen a autores o gestores de contenido diferentes a los del repositorio. Como ejemplo, Domínguez menciona el servicio de diccionarios del periódico español *El Mundo*.
- Agregadores de diccionarios: los metadicionarios o motores de búsqueda seleccionan información en otros recursos (frecuentemente diccionarios bilingües sin interconexión fuera del par de lenguas). Un ejemplo sería *The Free Dictionary*, de Farlex.
- Redes de diccionarios: en las redes de diccionarios, el propietario del portal suele serlo también de los contenidos digitales. El conjunto de diccionarios que ofrece la Real Academia Española ejemplifica este tipo de recursos.
- Portales virtuales: ofrecen la integración de diferentes recursos. Un ejemplo lo constituye *Linguee*, que además de comportarse como diccionario bilingüe, ofrece un gran corpus de traducciones a modo de textos paralelos obtenido de fuentes externas.
- A estos portales y plataformas, consideramos que habría que añadir los traductores automáticos ya que, en los últimos años, como apuntan Domínguez y Valcárcel (2015), se han convertido en uno de los recursos

más utilizados entre los usuarios que acuden a ellos para solventar dudas lingüísticas bilingües, principalmente a la hora de traducir o redactar en otro idioma. Como indican los autores, esto se puede deber principalmente a la tradición en el modo de enseñar las lenguas segundas y extranjeras, donde la traducción de textos y la memorización de glosarios bilingües desempeñan todavía un papel importante.

Una vez realizada esta puntualización sobre las diferentes herramientas en línea, centraremos nuestro trabajo en el análisis de la traducción automática (TA) como herramienta lexicográfica.

3. Los motores de traducción automática

La traducción automática (TA) es el proceso mediante el cual se utilizan herramientas software de computadora para traducir un texto de un lenguaje natural a otro.

La TA no ha nacido con internet (las primeras propuestas sobre sistemas de traducción automática datan de 1933) pero herramientas como Google Traductor, lanzada en 2006 al público general, han disparado su uso.

Afirma Álvarez (2018:212) que «la digitalización [de la profesión del traductor] ha acarreado una “hipercompetitividad”», no siempre basada en mejores aptitudes o calidad. Con la excepción de la traducción jurada, la traducción es una profesión desregularizada y, de un tiempo a esta parte, se observa una tendencia a que exista un cierto *intrusismo* en el mercado. Existe la idea generalizada de que cualquiera que tenga conocimiento (que no dominio) de dos lenguas puede traducir.

Por otro lado, la globalización ha hecho que un gran número de empresas españolas compitan no solo a nivel local, regional o nacional, sino internacional. Esto hace que, como requisito para acceder al mercado laboral, sea necesario el conocimiento de, al menos, un idioma distinto a la lengua materna. Como sabemos, el inglés se ha convertido en el idioma de los negocios, la lengua común en la que se comunican empresas repartidas por diferentes países y continentes. En algunos casos, se podría decir que el inglés se ha convertido en el idioma corporativo; todos los empleados de las oficinas de España, pongamos por caso, han de poder desenvolverse en inglés. Sin embargo, se recibe como una grata sorpresa que algún compañero de las oficinas de Estados Unidos hable español. Estos conocimientos de idiomas implican que mucha de la documentación con la que se trabaja, o la comunicación por escrito entre compañeros o grupos de interés (socios, proveedores, etc.), sea en inglés. Ante esto, muchos profesionales de diferentes sectores se enfrentan a la redacción y traducción de documentación en inglés. Para ayudarse en sus labores, recurren a diferentes herramientas en línea, si bien no siempre tienen las destrezas o el tiempo suficiente para valorar si los resultados de sus pesquisas son óptimos o adecuados.

Por último, entra en la ecuación un tercer factor: la calidad del texto producido. En muchas ocasiones, el destinatario de los textos traducidos o redactados en otro idioma suele asumir sin más la calidad. Nuestra experiencia profesional evidencia que, en empresas ajenas a materias lingüísticas, se tiende a dar más valor a la rapidez que a la calidad del documento, sobre todo, si hablamos de textos técnicos o económicos (mientras los números del proyecto en cuestión sean correctos, el resto de *literatura* del texto se considera secundaria). La calidad deja de ser un factor diferenciador y, por lo tanto, se esfuma el valor añadido que pueden aportar los traductores profesionales. ¿Para qué recurrir a ellos si los motores de traducción automática pueden hacer el trabajo más rápido y con un coste menor?

En las siguientes secciones vamos a abordar cómo funcionan algunos de los recursos más utilizados para la traducción o la redacción de textos por usuarios sin formación lingüística, es decir, usuarios no profesionales. Es evidente que la tecnología ha llegado a nuestras vidas para quedarse y en muchos aspectos ha aportado mejoras en el trabajo y en nuestra vida. Sin embargo, no conviene olvidar que los programas de traducción automática no son infalibles y, por ello, conocer su comportamiento puede ayudarnos a valorar la conveniencia, o no, de hacer uso de este tipo de herramientas.

3.1. Principales sistemas de traducción automática

Para entender la complejidad de la TA, es importante conocer los diferentes sistemas de traducción automática, ya que a lo largo de los años se han desarrollado diferentes enfoques con también diferentes resultados.

En este documento vamos a centrarnos únicamente en aquellos métodos que conviven principalmente en el mercado en la actualidad: la TA que utiliza información lingüística, comúnmente denominada como TA basada en reglas, la estadística y la basada en redes neuronales.

3.1.1. Traducción automática basada en reglas

El método basado en reglas (RBMT, por sus siglas en inglés, *Rule Based Machine Translation*) realiza «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» (Córdoba, Rico et al, 2015: 14).

Este sistema utiliza las reglas gramaticales y los diccionarios de cada lengua para realizar la traducción automática. El proceso constaría de las siguientes fases (figura 1):

- **Análisis:** el texto original se analiza morfológica y sintácticamente.
- **Transferencia:** se transfieren las estructuras gramaticales del texto origen a lengua meta (interlingua).
- **Generación:** en base a las estructuras gramaticales obtenidas, se genera el texto en lengua meta.

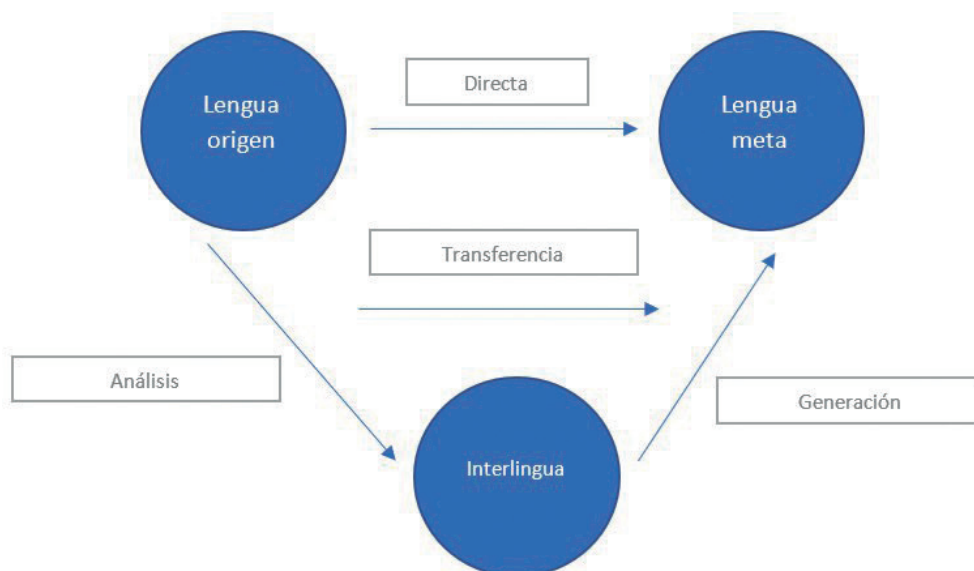


Fig. 1. Descripción gráfica del funcionamiento de un motor de TA basado en reglas (adaptación del gráfico consultado en Useless Paperclip, 2017).

Este método resulta muy útil para pares de lenguas parecidas (por ejemplo, español \leftrightarrow portugués) y, desde un punto de vista sintáctico, suele ofrecer traducciones correctas. Sin embargo, el tono general del texto meta puede resultar demasiado literal. Además, requiere de una gran inversión de tiempo y recursos: para desarrollar un único par de lenguas es necesario contar con gramáticas de la lengua origen y de la lengua meta, así como con diccionarios bilingües y reglas de transferencia. Además, el sistema no puede traducir estructuras lingüísticas que no estén incluidas en sus gramáticas o reglas de transferencia, ni palabras o expresiones que no se encuentren en sus diccionarios. Esto hace que el mantenimiento de los motores basados en reglas deba ser prácticamente continuo para garantizar que son capaces de traducir textos de un nuevo dominio, por ejemplo, o estructuras gramaticales que en un primer momento no fueron previstas.

Veamos un ejemplo de las limitaciones de este sistema de TA para la combinación lingüística inglés \rightarrow español (figura 2):

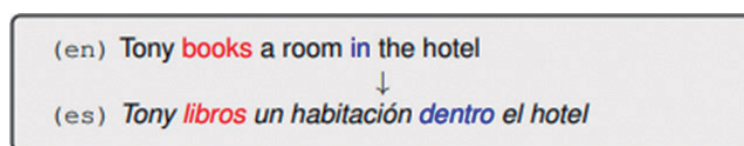


Fig. 2. Ejemplo de un problema de una traducción automática basada en reglas (Sánchez, F; Forcada, M.L. y Esplà M., 2020).

Book en inglés puede traducirse al español por un sustantivo (*libro*) o por un verbo (*reservar*). La preposición *in* puede traducirse en español por *en* o por *dentro*. Si el motor de TA no contempla esta posibilidad, no será capaz de resolver la ambigüedad y el texto resultante será incorrecto.

3.1.2. Traducción automática basada en estadística

La traducción automática estadística (SMT, por sus siglas en inglés, *Statistical Machine Translation*) no usa datos lingüísticos tradicionales. La esencia de este método radica en alinear frases, grupos de palabras y palabras individuales de textos paralelos y calcular las probabilidades de que una palabra, en una frase de una lengua, se corresponda con una palabra en una frase de una traducción con la que está alineada.

Es decir, utiliza un corpus basado en millones de documentos traducidos de internet en el idioma solicitado para realizar una traducción automática. Entre todos esos millones de documentos busca traducciones que ya se hayan hecho previamente sobre ese grupo de palabras y escoge la opción más utilizada.

Estos sistemas de traducción automática constan de tres componentes principales (figura 3):

- El modelo de lenguaje: se encarga de calcular la probabilidad de que una frase en la lengua de destino sea correcta. Es el encargado de la fluidez de la traducción y para entrenarlo se utiliza un corpus monolingüe de la lengua de destino lo más grande posible.
- El modelo de traducción: se encarga de establecer la correspondencia entre el idioma de origen y el idioma de destino y se entrena utilizando un corpus alineado a nivel oracional. Durante esa fase de entrenamiento, el sistema estima la probabilidad de una traducción a partir de las traducciones que aparecen en el corpus de entrenamiento.
- El decodificador: es el responsable de buscar dentro de todas las traducciones posibles la más probable en cada caso. Así, dado un modelo de lenguaje y un modelo de traducción, crea todas las traducciones posibles y propone la más probable.

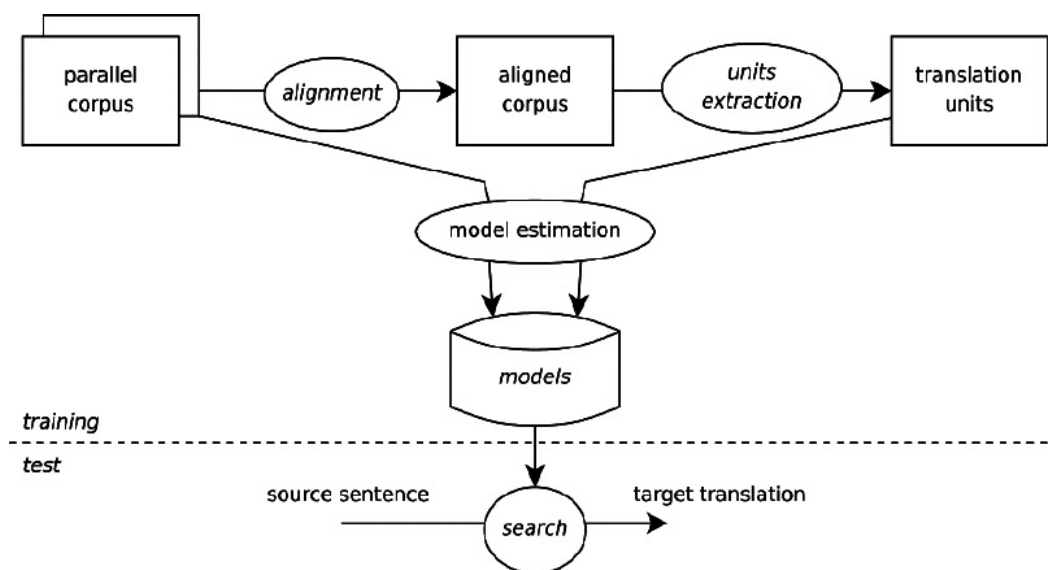


Fig. 3. Descripción gráfica de un sistema de traducción automática estadística (Singla, 2015:4).

Dado que la TA estadística genera sus traducciones a partir de métodos estadísticos basados en corpus de textos bilingües, la disponibilidad de un corpus grande de traducciones fiables es una característica esencial de este sistema.

A continuación, vamos a ilustrar, de forma muy simplificada, cómo funcionaría este sistema de TA para la combinación inglés > español (figuras 4 y 5).

Tomemos como ejemplo la oración *I don't know the answer*. En primer lugar, el motor segmenta la frase en grupo de palabras y busca su correspondencia en el corpus de textos paralelos del que está alimentado.

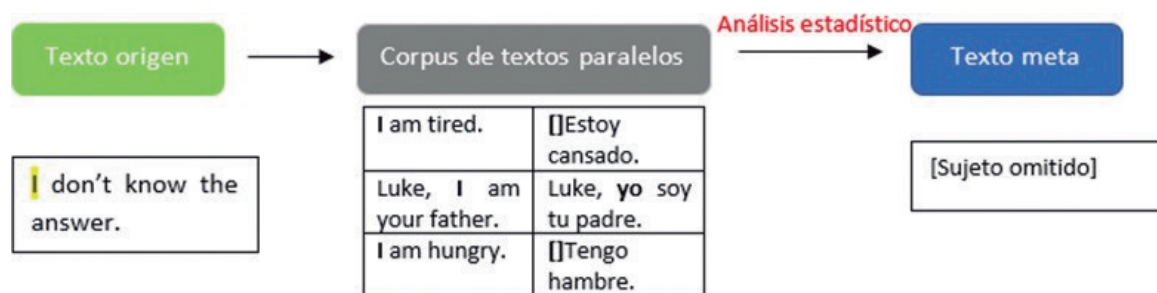


Fig. 4. Ejemplo de cómo se comporta un motor de traducción automática basada en estadística (adaptación del gráfico consultado en Leddy, 2019).

El motor realiza un análisis estadístico de sus textos paralelos y concluye que el pronombre *I* en inglés, en un porcentaje *N* de su corpus, se omite en español y considera esa respuesta la más adecuada. El mismo proceso se repetiría con el resto de grupos de palabras.

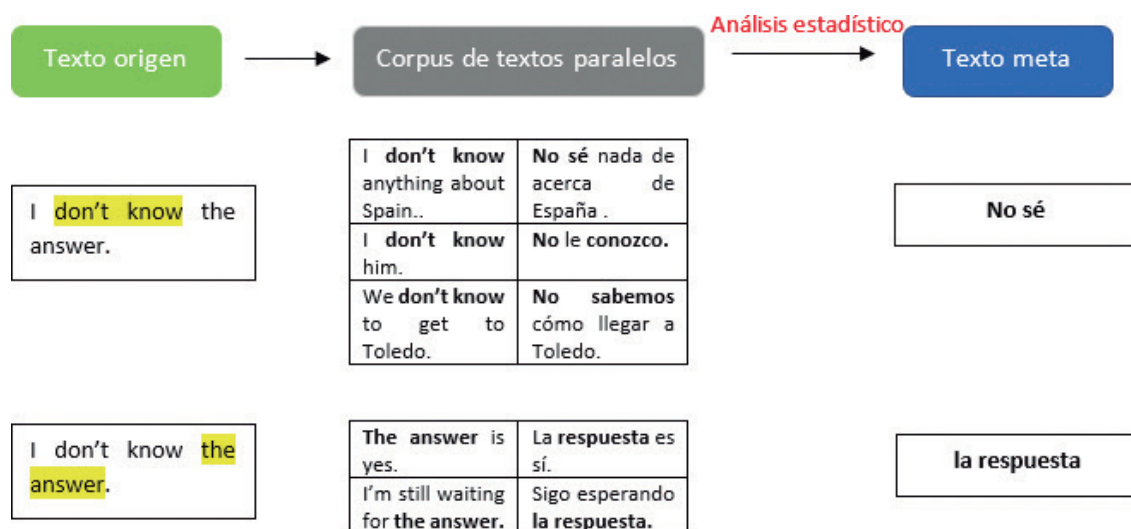


Fig. 5. Ejemplo de cómo se comporta un motor de traducción automática basada en estadística (adaptación del gráfico consultado en Leddy, 2019).

Tras analizar todos los grupos de palabras, el motor de TA determinaría que la traducción correcta para *I don't know the answer* sería *No sé la respuesta*.

A pesar de la rapidez y eficacia de la SMT, el resultado de las traducciones no es óptimo y se observan errores tanto léxicos como gramaticales, ya que no siempre el corpus de textos paralelos está bien alimentado y, por tanto, su análisis estadístico puede devolvernos la respuesta más probable, pero no por ello la más correcta.

Hasta hace poco, era el sistema de traducción automática por excelencia, y empresas como Microsoft o Google optaban por este método para sus programas.

3.1.3. Traducción automática basada en redes neuronales

Se trata de la última aparición en la familia de los paradigmas de traducción automática. Al igual que sucede con la TA estadística, los sistemas basados en redes neuronales (NMT, por sus siglas en inglés, *Neuronal Machine Translation*) requieren de grandes corpus de textos paralelos para su correcto funcionamiento. Puesto que una explicación detallada de su funcionamiento requiere adentrarnos en materias muy ajenas a la que nos ocupa, vamos a tratar de explicar, a grandes rasgos, en qué consiste este sistema: las redes neuronales artificiales pretenden emular cómo funcionan las neuronas de nuestro cerebro. Del mismo modo en que nuestras neuronas reciben información y realizan conexiones entre sí, los componentes del lenguaje se asocian con otra información subyacente para formar asociaciones y generar traducciones. Así, utilizando técnicas de aprendizaje automático, el ordenador *aprende* a traducir a partir de grandes cantidades de textos paralelos que, además, incluyen todo tipo de información lingüística y no lingüística (figura 6).

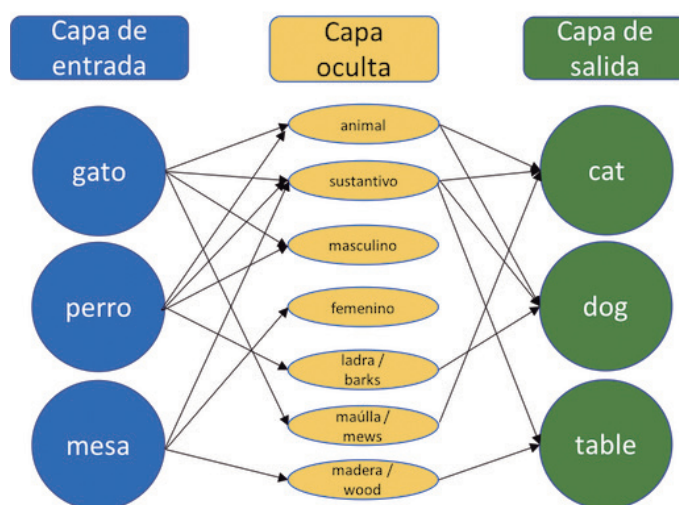


Fig. 6. Funcionamiento red neuronal artificial (Parra, 2018).

Google Traductor, Bing Microsoft Translator y el traductor de DeepL ya utilizan este tipo de software.

3.2. Problemas más comunes de la traducción automática

Los beneficios que aporta la TA son innegables, tanto para traductores como para usuarios. Sin embargo, como ya hemos adelantado, no debemos perder de vista que los programas de traducción automática no son infalibles y los diferentes motores de TA siguen resultando imprecisos a la hora de reconocer sinónimos, colocaciones o frases hechas, entre otros.

Los motores de traducción automática se han convertido en una solución rápida para quienes desean trasladar un texto de una lengua a otra, puesto que agilizan y economizan el proceso de traducción. Los traductores profesionales no son ajenos a estas herramientas ya que, dependiendo del tipo de texto, se pueden conseguir unos resultados aceptables. Sin embargo, un uso indiscriminado de la TA puede poner en tela de juicio el valor de la traducción resultante.

A continuación, detallamos, a título ilustrativo pero no exhaustivo, alguno de los problemas más comunes que nos encontramos al hacer uso de motores de traducción automática. Los hemos agrupado en cuatro categorías.

3.2.1. Errores léxicos

- a) Ambigüedades léxicas: en los casos en los que existe una pluralidad de acepciones para una palabra, bien se deba a un caso de polisemia, bien a un caso de homonimia, el contexto se torna clave. Los motores de TA son, de momento, incapaces de entender y diferenciar los contextos, lo que puede desembocar en una traducción errónea de un determinado término. Veamos algunos ejemplos:
 - Polisemia: imaginemos que, en el marco de la asignatura de Ciencias Naturales, que ahora en muchos centros educativos se imparte en inglés, un alumno se encuentra con el término *fish school*, que desconoce. Decide resolver su duda buscando el equivalente en español en Google Traductor y se encuentra con lo siguiente (figura 7):

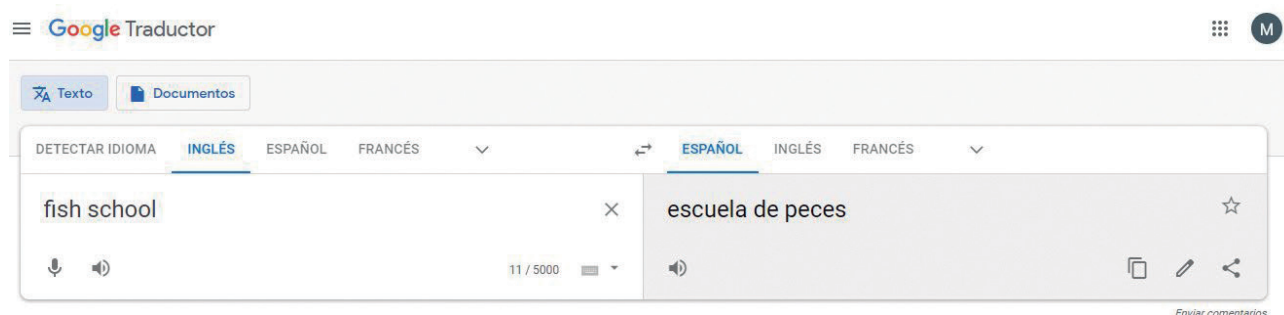


Fig. 7. Ejemplo de error léxico de Google Traductor por polisemia.

En inglés, la palabra *school* hace referencia no solo a *escuela* sino también a *cardumen* (*banco de peces*). Sin embargo, Google Traductor no capta el contexto.

- Homonimia: nos hemos topado con el siguiente titular en una entrada de una página web: *How to hold a charity ball*. Copiamos y pegamos en Bing Microsoft Translator (figura 8):

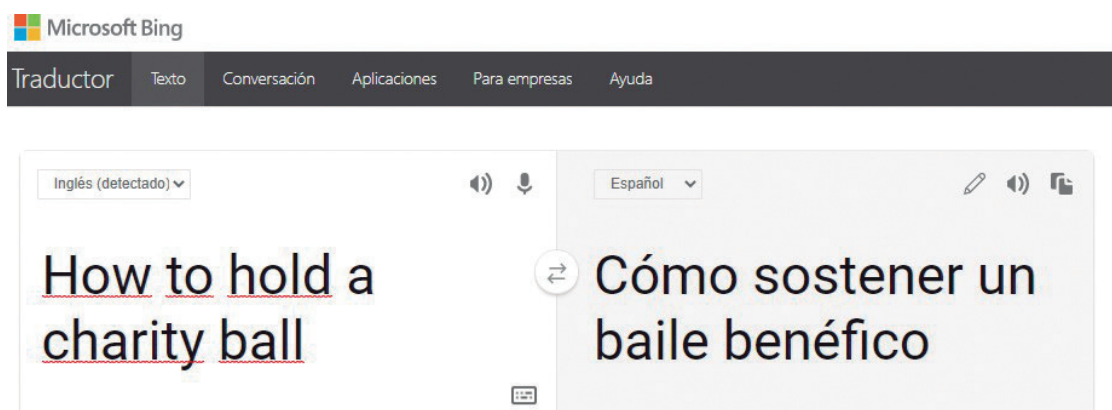


Fig. 8. Ejemplo de error léxico de Bing Microsoft Translator por homonimia.

Aunque se puede intuir el sentido de la frase, el motor de TA no ha generado una solución precisa. En inglés, *ball* es una manifestación de homonimia, que puede hacer referencia tanto a un objeto (*pelota*) como a un festejo (*baile*). Ambas se pueden combinar con el verbo *hold*. Así, *hold a ball* puede significar *sujetar* (*sostener*) una *pelota* u *organizar un baile*.

b) Fraseología: las frases hechas, locuciones figuradas, metáforas y comparaciones fijadas, modismos y refranes existentes en una lengua contienen muchos matices culturales. Ya de por sí suponen una dificultad para la traducción humana puesto que, aunque en algunos casos sí existe una traducción exacta, en otros quizás haya que recurrir a una expresión con un sentido equivalente o, directamente, no existe equivalente y hay que optar por otra solución. La traducción automática, sin embargo, no siempre percibe estos matices. Veamos algunos casos:

- *Knowledge is no burden* es un refrán inglés. El traductor de DeepL (figura 9) lo traduce al español de forma literal: *el conocimiento no es una carga** (y en la parte inferior de la pantalla, incluso, ofrece una alternativa, también desacertada: *El saber no es una carga**). Sin embargo, sí contamos en español con un refrán equivalente: *El saber no ocupa lugar*.

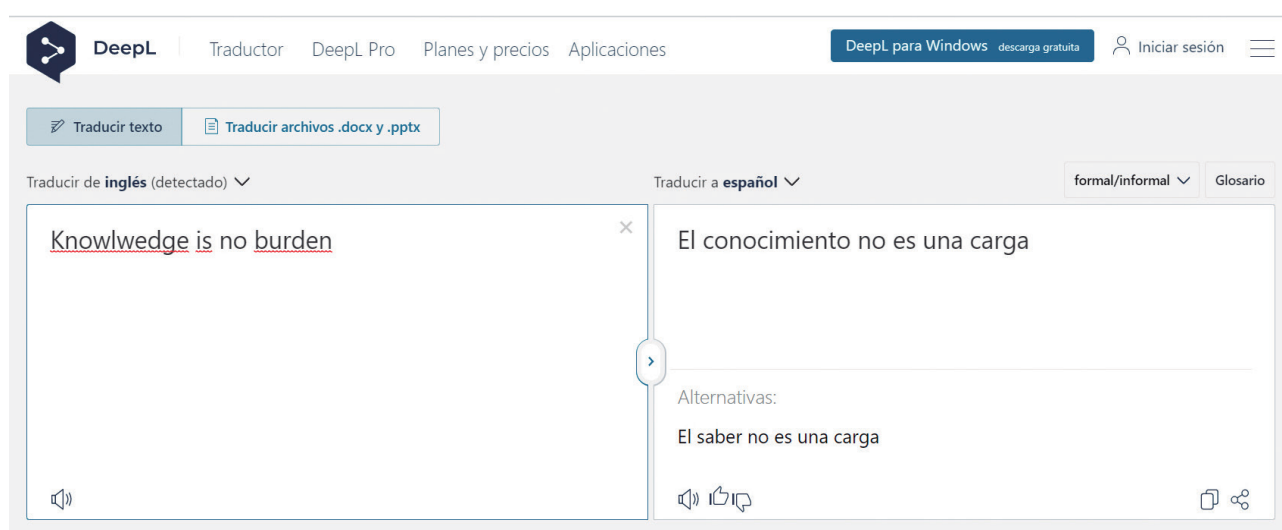


Fig. 9. Ejemplo de error del traductor de DeepL al traducir un refrán.

- *Irse de la lengua* es una locución verbal coloquial que se utiliza para referir que alguien ha revelado algo que no debería. Sin embargo, Google Traductor devuelve una traducción literal (figura 10). En inglés, existe también una expresión similar: *let the cat out the bag*.

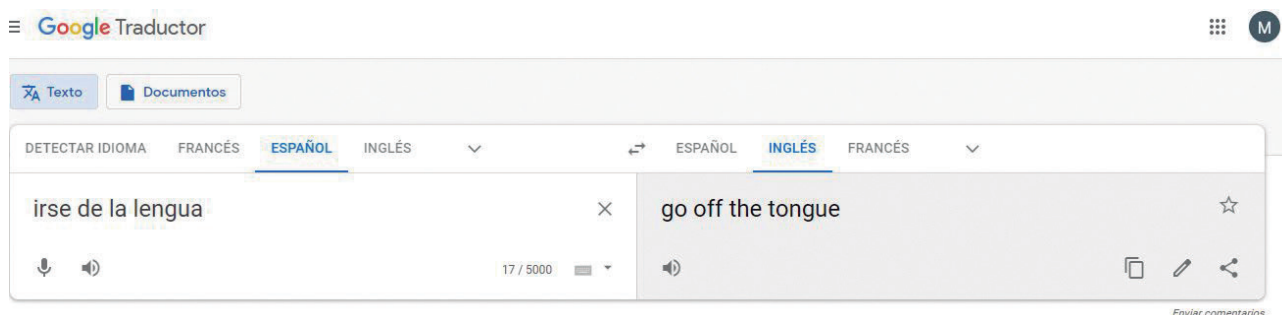


Fig. 10. Ejemplo de error de Google Traductor al traducir una locución verbal coloquial.

- En el caso del modismo *a troche y moche*, la herramienta Bing Microsoft Translator no reconoce la expresión y devuelve una frase que no significa nada en inglés (figura 11):

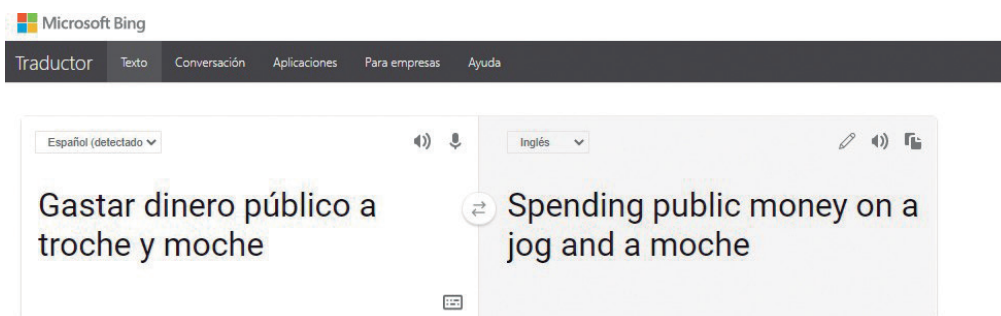


Fig. 11. Ejemplo de error de Bing Microsoft Translator al traducir un modismo.

- c) Extranjerismos innecesarios: los extranjerismos forman parte del caudal léxico de un idioma y, en ocasiones, son necesarios. En otros casos existe un equivalente en la lengua meta para referirnos a ellos y, por lo tanto, el uso del extranjerismo está desaconsejado. Sin embargo, no son pocas las veces que los motores de traducción automática dejan en el texto meta el extranjerismo (figura 12):

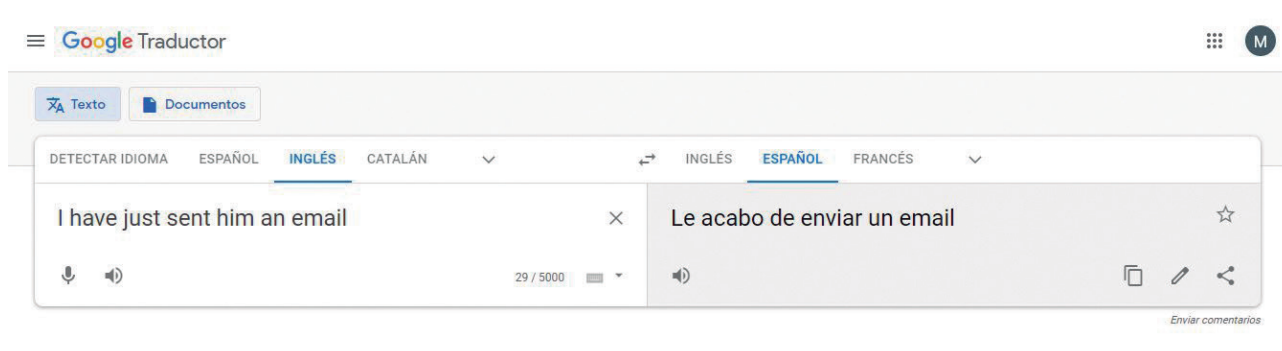


Fig. 12. Ejemplo de cómo Google Traductor utiliza extranjerismos innecesarios.

- d) Calcos léxicos: cabría incluir en esta categoría todos aquellos errores que derivan de la adopción de un significado extranjero para una palabra ya existente en una lengua. En la figura 13 vemos cómo en español ya se ha extendido el uso del adjetivo *doméstico* con el sentido de *nacional*, por influencia de la palabra análoga en inglés *domestic* (de hecho, en el *Diccionario de americanismos* ya aparece así recogido). Si bien no es un uso censurable, en el caso de la variante de español de España, no resultaría natural y quizás sería preferible utilizar el adjetivo *interna* o *nacional*.

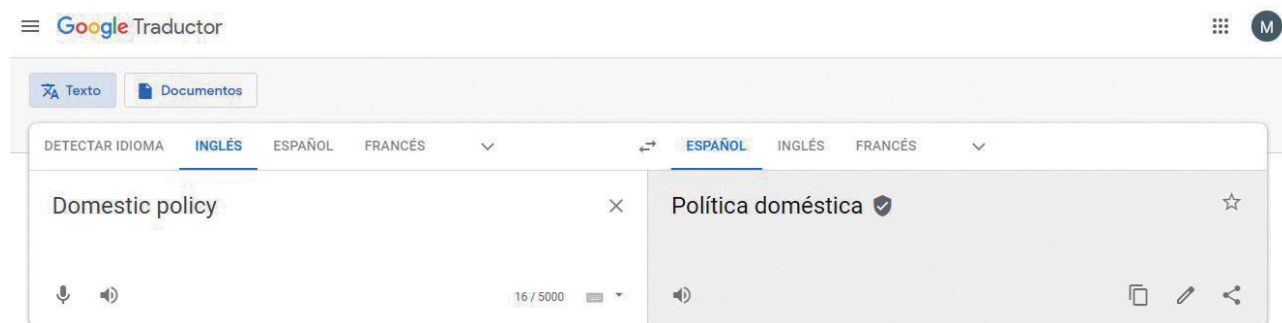


Fig. 13. Ejemplo de calco léxico en el uso de Google Traductor.

3.2.2. Errores sintácticos

- a) Uso incorrecto de preposiciones: el uso de preposiciones suele ser uno de los mayores desafíos para quienes estudian un segundo idioma. Por ejemplo, la preposición *with* del inglés puede traducirse en español por *con*, *de*, *en*, o *según*, lo que puede dificultar el proceso de traducción automática y devolvernos un texto que no resulte natural o fluido. En la figura 14, podemos apreciar cómo Bing Microsoft Translator genera una tra-

ducción en la que el uso de la preposición no es adecuado. Traduce *work with an advertising company* por *trabajar con una empresa de publicidad** y en español sería más adecuado hablar de *trabajar en una empresa de publicidad*.

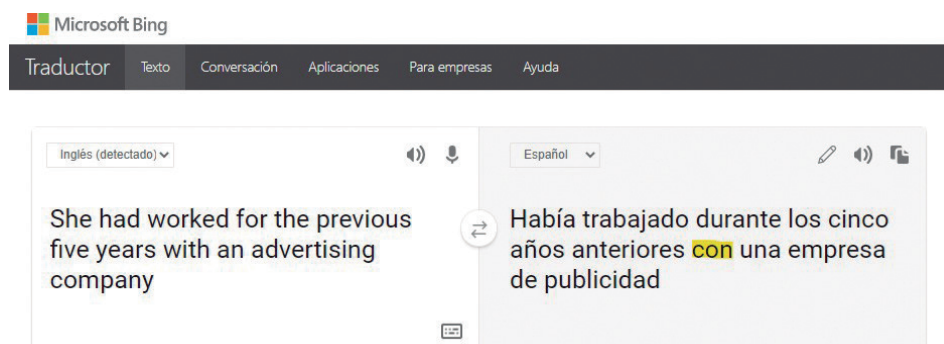


Fig. 14. Ejemplo de uso incorrecto de preposiciones en Bing Microsoft Translator.

- b) Errores de concordancia: observamos que, en ocasiones, la TA tiene dificultades para determinar la concordancia entre sujeto y verbo, una expresión anafórica y su antecedente, etc. El traductor de DeepL (figura 15) no ha sabido identificar que los referentes de *este* y *aquel* son Pepe y Juan respectivamente y, en vez de utilizar una estructura similar en inglés (*the former* y *the latter*), repite en dos ocasiones uno de los nombres propios.

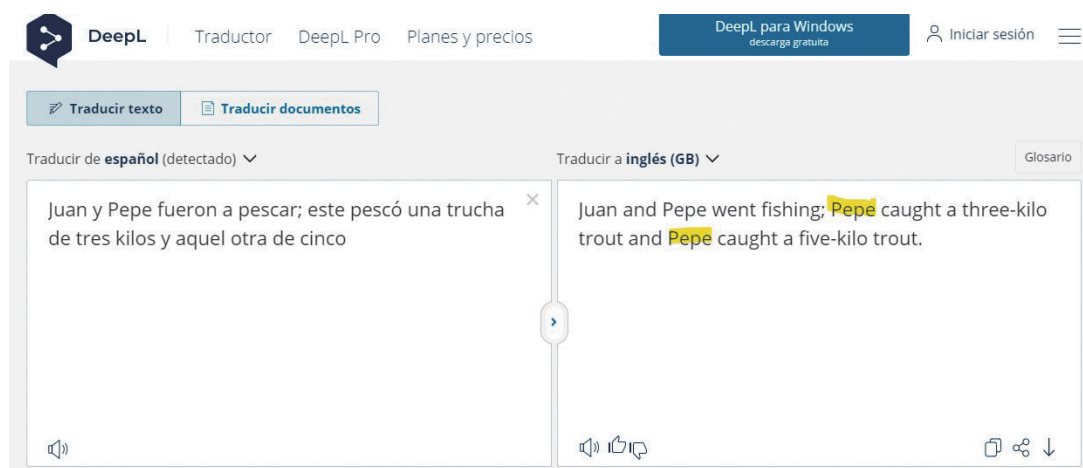


Fig. 15. Ejemplo de error de concordancia en el traductor de DeepL.

- c) Uso incorrecto de tiempos verbales: el uso de los tiempos verbales difiere entre idiomas y el contexto, una vez más, resulta determinante. Como hemos mencionado, las herramientas de TA no siempre están capacitadas para interpretar dicho contexto. En este caso (figura 16), una traducción más natural en el español de España para la expresión *I am not feeling well* sería *No me encuentro bien*. El motor de TA ha traducido el tiempo verbal de manera literal, sin tener en consideración que lo que en inglés se conoce como *present continuous* equivale en español a un presente de indicativo.

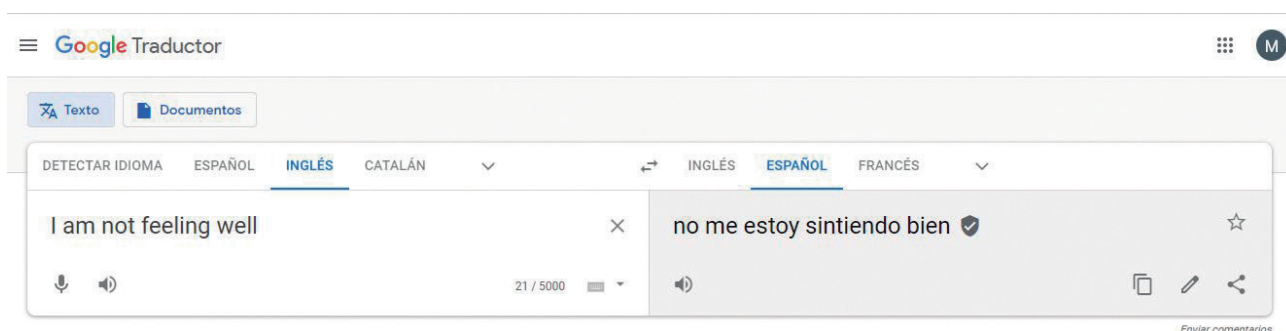


Fig. 16. Ejemplo de error sintáctico al usar Google Traductor.

3.2.3. Errores semánticos

Son aquellos errores que impiden la correcta comprensión del texto, los denominados *falsos sentidos*. Este error es común en textos de temas específicos, ya que el motor de TA no puede entender la naturaleza del texto. Por ejemplo, en terminología médica, *compromise* significa *insuficiencia*. Como se puede observar en la figura 17, si buscamos en el traductor de DeepL el término *respiratory compromise*, obtenemos una traducción errónea: *compromiso respiratorio**.



Fig. 17. Ejemplo de error semántico al usar el traductor de DeepL.

3.2.4. Errores ortotipográficos

- a) Alteración de la puntuación: las reglas de puntuación son diferentes en cada idioma y muchas veces los motores de TA no adaptan las reglas de puntuación de la lengua origen a la lengua meta.

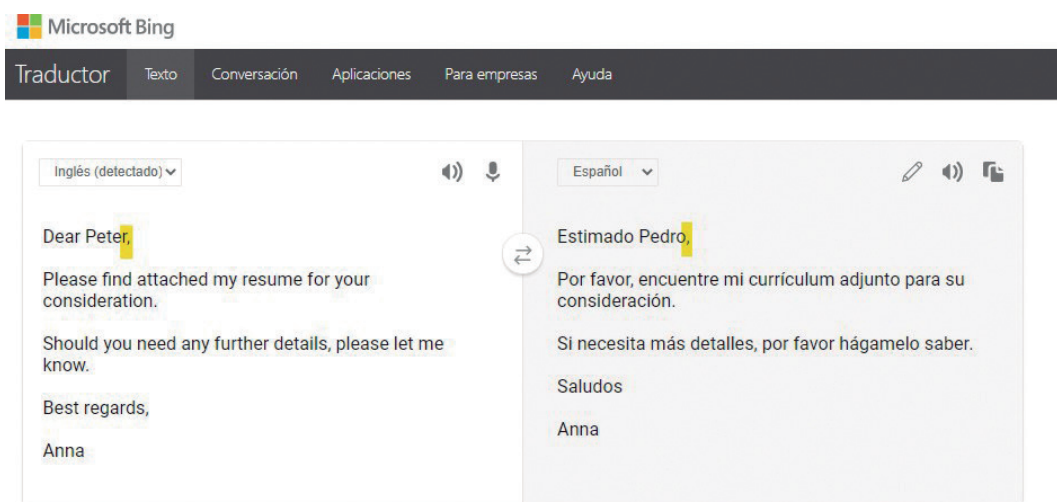


Fig. 18. Ejemplo de alteración de la puntuación al usar Bing Microsoft Translator.

Como se ve en la figura 18, el motor de TA no ha adaptado las normas ortográficas para el encabezamiento de una carta, que en español se cierra tradicionalmente con dos puntos y no con coma, como ocurre en inglés.

- b) Uso incorrecto de mayúsculas y minúsculas: al igual que en el punto anterior, este error sucede cuando los motores de TA no adecúan el uso de mayúsculas de la lengua origen a la lengua meta. Suele suceder, por ejemplo, en el caso de los títulos en inglés, donde todas las palabras (excepto preposiciones), se escriben en mayúscula. En la siguiente imagen (figura 19), sin entrar a valorar la corrección de la traducción, se observa que el uso de mayúsculas en el texto meta es incorrecto ya que, en los títulos de textos de presentaciones, reuniones, conferencias, etc., solo se escriben con mayúscula la primera palabra y aquellas otras cuya naturaleza lo exija.

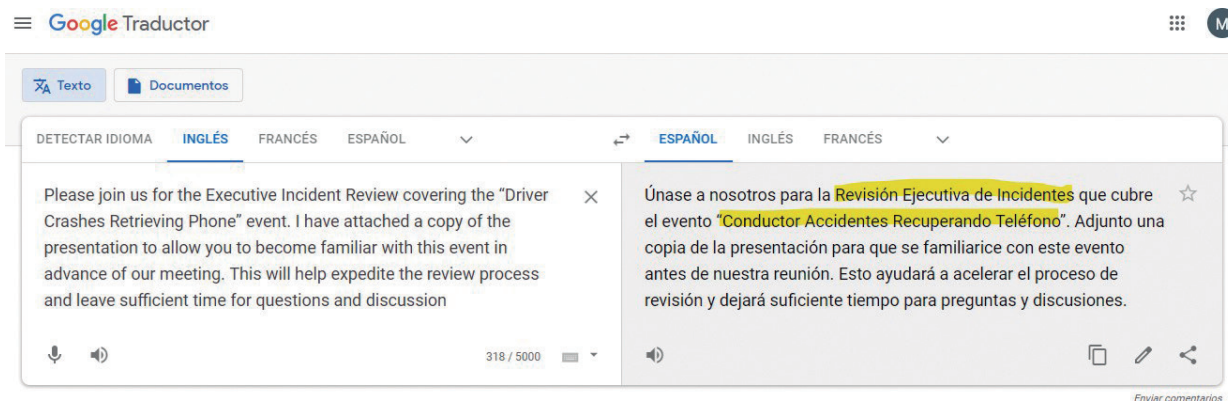


Fig. 19. Ejemplo de uso incorrecto de mayúsculas en Google Traductor.

Estos errores pueden resultar anecdóticos y no ir más allá de una captura de pantalla que corre como la pólvora por internet durante unos días y con la misma rapidez cae en el olvido. Probablemente ya nadie recuerde a estas alturas el caso del Ayuntamiento de Santander (España), que recurrió a la traducción automática para su página web de turismo y tradujo el Centro Botín, centro de arte ubicado en dicha ciudad, por *Loot Center*, entendiendo *botín* como nombre común en vez del apellido que da nombre al centro (Blanco, 2018). Sin embargo, en otros contextos, una incongruencia en un texto puede invalidar un documento o causarnos un malentendido profesional con consecuencias serias.

No obstante lo anterior, la calidad de la traducción generada mediante motores de TA dependerá mucho del tipo de textos y de la combinación lingüística con la que estemos trabajando, como explicaremos a continuación.

3.3. Funcionamiento de Google Traductor, Bing Microsoft Translator y el traductor de DeepL

En los últimos años, debido, entre otros factores, a la irrupción de las TIC, el número de traductores automáticos en línea gratuitos ha aumentado considerablemente. Se podría elaborar una lista con todos ellos pero, para el trabajo que nos ocupa, vamos a centrarnos en aquellos cuyo uso está más extendido: Google Traductor, Bing Microsoft Translator y el traductor de DeepL.

- Google Traductor es un sistema multilingüe de traducción automática, desarrollado y proporcionado por Google, para traducir texto, voz, imágenes o vídeo en tiempo real de un idioma a otro. Google posee la capacidad de traducir 109 idiomas, a través de un sistema de TA basado en la estadística. En 2016 incorporó el método de traducción automática neuronal. Por ahora, ofrece este sistema solo en algunos idiomas: inglés, francés, alemán, portugués, español, chino, japonés y turco.
- Por su parte, Bing Microsoft Translator es un servicio proporcionado por Microsoft que permite a los usuarios traducir textos o páginas web completas en más de 70 idiomas. Al igual que sucede con Google Traductor, ya ha comenzado su transición de traducción automática estadística a neuronal.
- El traductor de DeepL es un servicio de traducción automática en línea de DeepL GmbH, empresa que desarrolló *Linguee*. Permite la traducción de 11 idiomas (alemán, chino, español, francés, inglés, italiano, japonés, neerlandés, polaco, portugués y ruso). Utiliza traducción automática neuronal.

Podemos diferenciar dos niveles de uso de los motores de TA: como herramienta de traducción automática en textos (un aspecto que no abordaremos en este trabajo, y que depende, entre otros, de factores tan diversos como la combinación lingüística y cuál sea la lengua elegida como lengua pivote, el tipo de texto o su longitud) y como herramienta lexicográfica bilingüe para unidades léxicas. En ambos casos, desde la experiencia del usuario, el funcionamiento de estas herramientas de TA es bastante sencillo e intuitivo. Como ya hemos visto al describir los problemas más comunes en la TA (apartado 3.2.), escribimos en un recuadro el texto o la palabra que queremos traducir, elegimos el par de idiomas con el que estamos trabajando y el sistema nos devuelve el resultado generado en la lengua meta.

4. Análisis de uso de TA en unidades léxicas

4.1. Una muestra de análisis

Aunque en sus inicios la TA no se planteó con este fin, la realidad es que cada vez se recurre más a los motores de TA como herramienta lexicográfica para consultas bilingües de unidades léxicas, haciendo las veces de diccionarios. Ante este uso, los motores de TA no solo nos devuelven la unidad léxica en la lengua meta, como ocurre con la TA de textos, sino que aportan información adicional.

Para ilustrar de un modo más exhaustivo el comportamiento de los motores de TA en unidades léxicas, vamos a tomar como ejemplo la unidad léxica *juego*, con la que podemos estar refiriéndonos a:

- Un ejercicio recreativo (*juego del parchís*)
- El resultado de un partido de tenis (*Nadal va tres juegos a dos*)
- Un conjunto formado por un determinado número de ciertas cosas similares y para un mismo fin (*juego de café*)
- La primera persona de singular del presente de indicativo del verbo *jugar*

Como podemos observar en las figuras 20, 21 y 22, los tres motores de TA analizados nos devuelven lo que consideran que es la opción más correcta, en este caso, *game*, en función del análisis de sus corpus. Y, además, ofrecen también una serie de datos adicionales (diferentes traducciones a la lengua meta, ejemplos de uso, definiciones en la lengua origen, etc.).

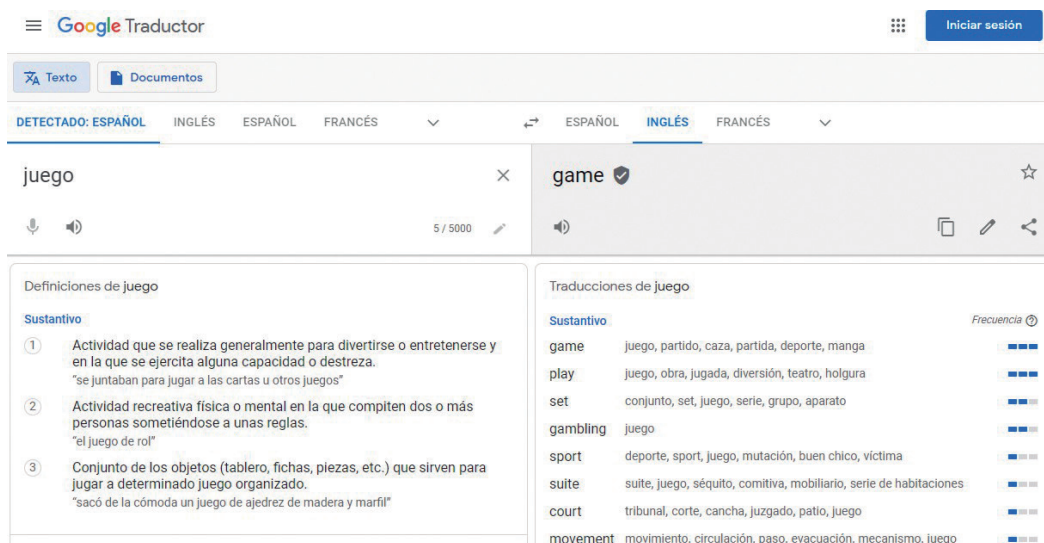


Fig. 20. Resultado obtenido al buscar *juego* en Google Traductor, donde figura información monolingüe del término en lengua origen y posibles traducciones en lengua meta.

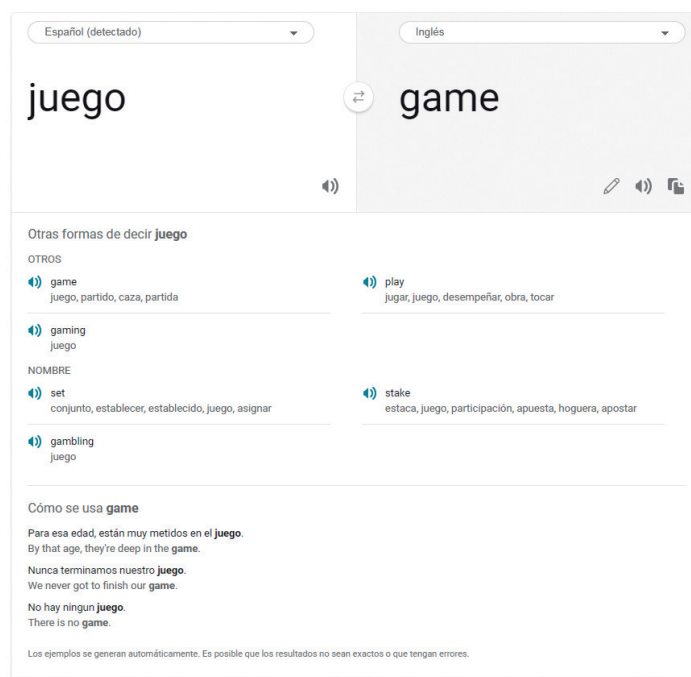


Fig. 21. Captura de pantalla de una búsqueda en Bing Microsoft Translator, donde figura un catálogo de sinónimos del término en lengua meta, con las traducciones en la lengua origen.

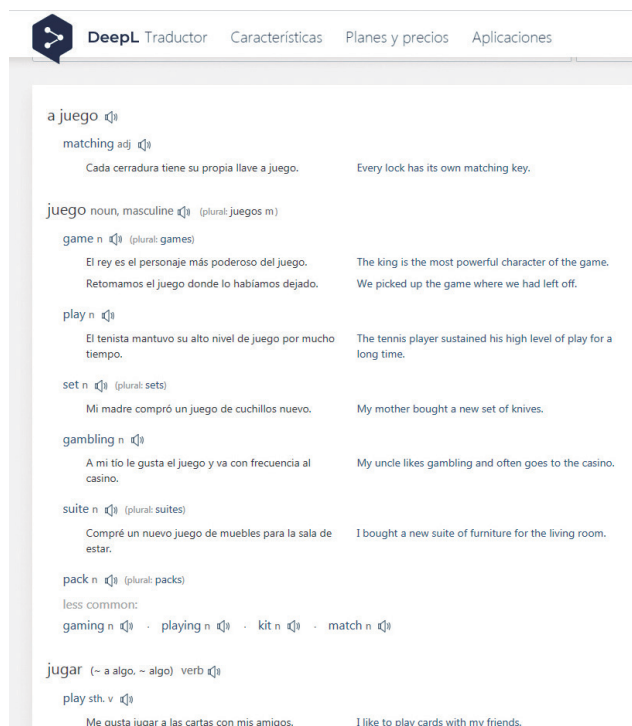


Fig. 22. Captura de pantalla de una búsqueda en el traductor de DeepL, donde se ofrecen diferentes traducciones en la lengua meta con ejemplos.

Sin embargo, esta información, lejos de ser esclarecedora, puede tener el efecto contrario. Por ejemplo, Google Traductor nos proporciona posibles traducciones de *juego* en la lengua meta (figura 23). Cada traducción, a su vez, ofrece un catálogo de términos de nuevo en la lengua origen. De esta forma, ante la consulta de *juego*, el motor de TA nos dice que *sport* es una posible traducción de *juego* para, a su vez, ofrecernos traducciones al español de *sport* (*buen chico* o *víctima*), que resultan irrelevantes para nuestra consulta.

Traducciones de juego		
Sustantivo		Frecuencia
game	juego, partido, caza, partida, deporte, manga	100%
play	juego, obra, jugada, diversión, teatro, holgura	100%
set	conjunto, set, juego, serie, grupo, aparato	100%
gambling	juego	100%
sport	deporte, sport, juego, mutación, buen chico , víctima	100%
suite	suite, juego, séquito, comitiva, mobiliario, serie de habitaciones	100%
court	tribunal, corte, cancha, juzgado, patio, juego	100%
movement	movimiento, circulación, paso, evacuación, mecanismo, juego	100%
service	servicio, oficio, entrega, uso, utilidad, juego	100%

Fig. 23. Catálogo de traducciones propuestas por Google Traductor para la voz «juego».

En cuanto a las indicaciones de uso, la TA genera los ejemplos de manera automática, de manera que, como los propios motores de TA advierten (figura 24), pueden contener errores o no ser correctos.

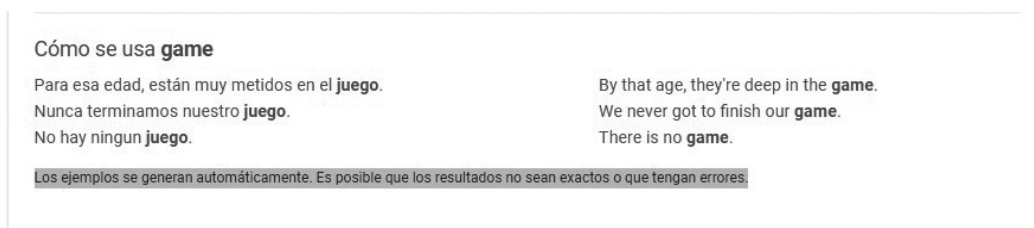


Fig. 24. Advertencia de Bing Microsoft Translator, ante los posibles errores generados por su sistema de TA.

Nos preguntamos si toda esta información supone una ayuda real al usuario ya que, por muy cómodo que resulte consultar las diferentes acepciones de un término concreto en la lengua origen sin tener que abrir una nueva ventana en el explorador, no podemos perder de vista que los usuarios han acudido al traductor automático para solventar una duda bilingüe. ¿Sirve de algo ofrecer, por ejemplo, una definición en la lengua origen si no está vinculada con la traducción en la lengua meta? De hecho, con la información proporcionada por los motores de TA, no hemos podido determinar el equivalente en inglés para las cuatro acepciones mencionadas de *juego*.

Por otra parte, al contrario de lo que sucedía con los diccionarios en papel una vez impresos, una de las grandes ventajas que ofrecen este tipo de herramientas en línea es que se puede actualizar y subsanar errores prácticamente al instante. Hemos observado que los motores de TA analizados se nutren de la comunidad de usuarios para identificar estos errores, verificar la idoneidad de las traducciones propuestas, etc. Si tomamos como muestra la página web para ayudar a mejorar Google Traductor, basta con tener una cuenta de correo electrónico para convertirse en colaborador y, por un lado, aportar nuevas traducciones de frases o palabras y, por el otro, verificar las aportaciones de otros usuarios (figuras 25, 26 y 27), sin que estos datos pasen ningún tipo de filtro lexicográfico.

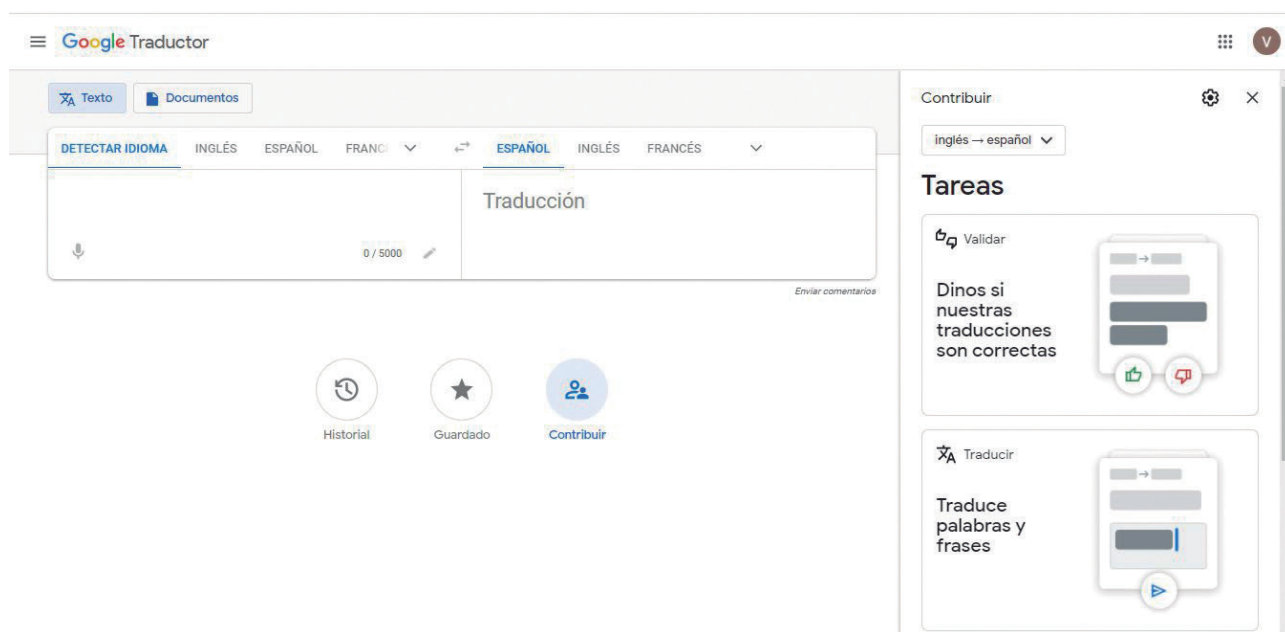


Fig. 25. Tareas que ofrece Google Traductor a los usuarios para ayudar a mejorar la herramienta.

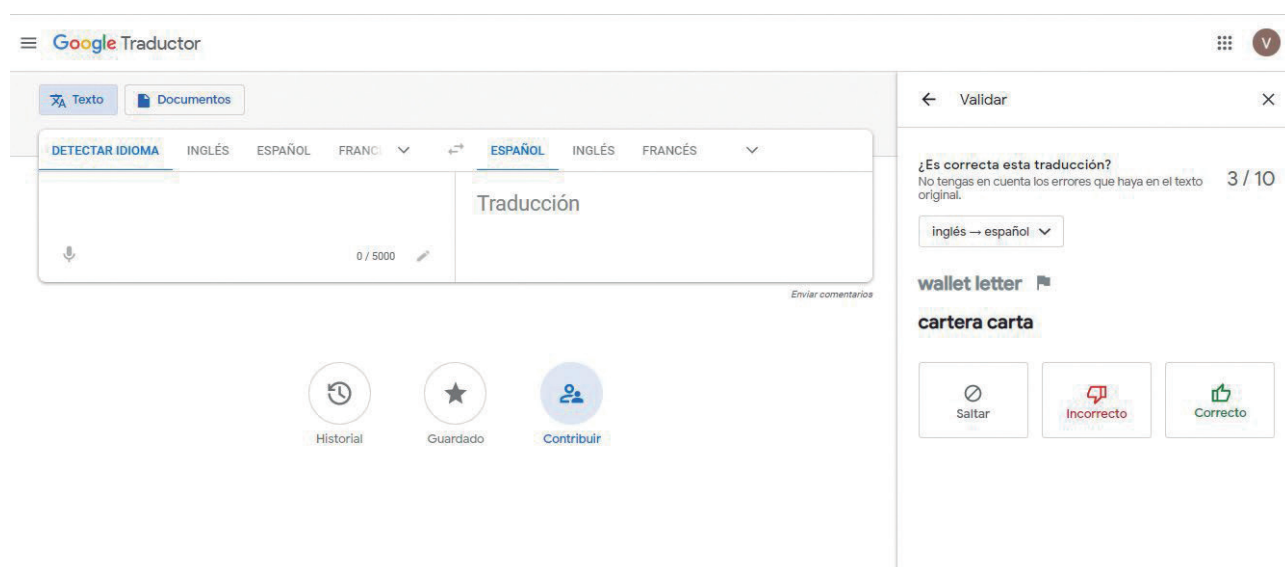


Fig. 26. Tarea de validación traducciones en Google Traductor con el fin de ayudar a mejorar la herramienta.

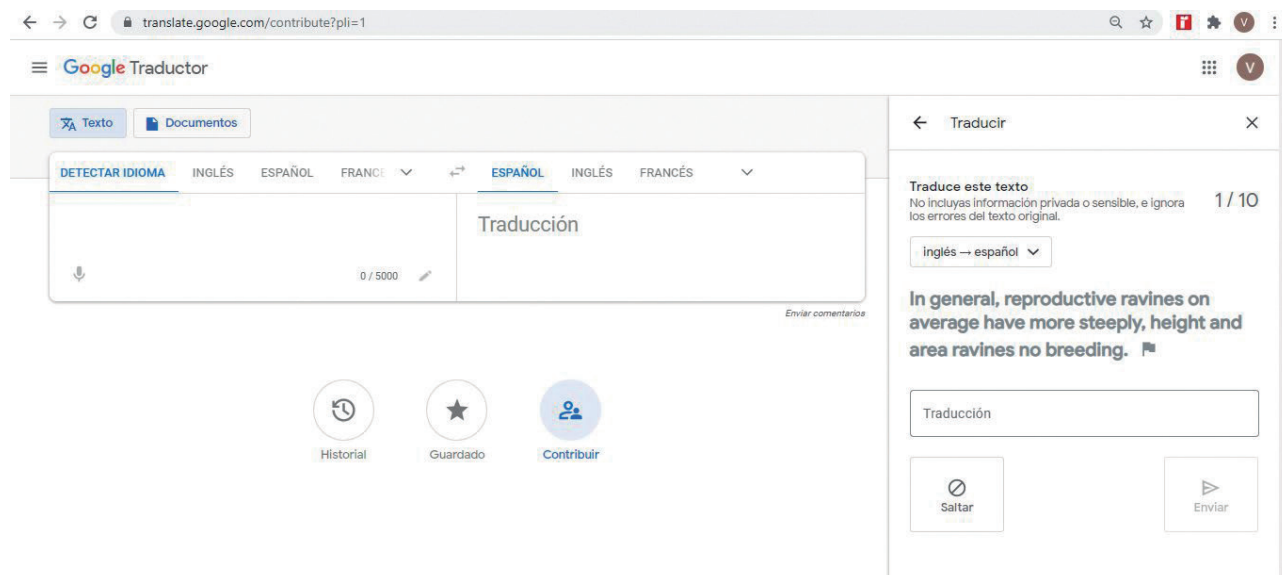


Fig. 27. Tarea de traducción de frases o palabras en Google Traductor con el fin de ayudar a mejorar la herramienta.

Con la información que, de primeras, nos proporcionan los motores de TA no podremos determinar el equivalente en inglés para las cuatro acepciones mencionadas.

En ocasiones, este problema se puede solventar si *ayudamos* al motor de TA a centrar su búsqueda. Y, para ello, podemos servirnos de la sintaxis. Las palabras se desgranán en rasgos de significado que determinan su comportamiento sintáctico. Y del mismo modo, la función de una determinada palabra determinará su definición (de Miguel, 2016). Igualmente, las palabras que la rodean nos ayudarán a proporcionar contexto.

Pongamos que queremos consultar en Google Traductor cómo se dice [Yo] *juego* en inglés. Para indicar a la herramienta que me estoy refiriendo a la forma verbal, bastaría con añadir el sujeto a la búsqueda (figura 28):

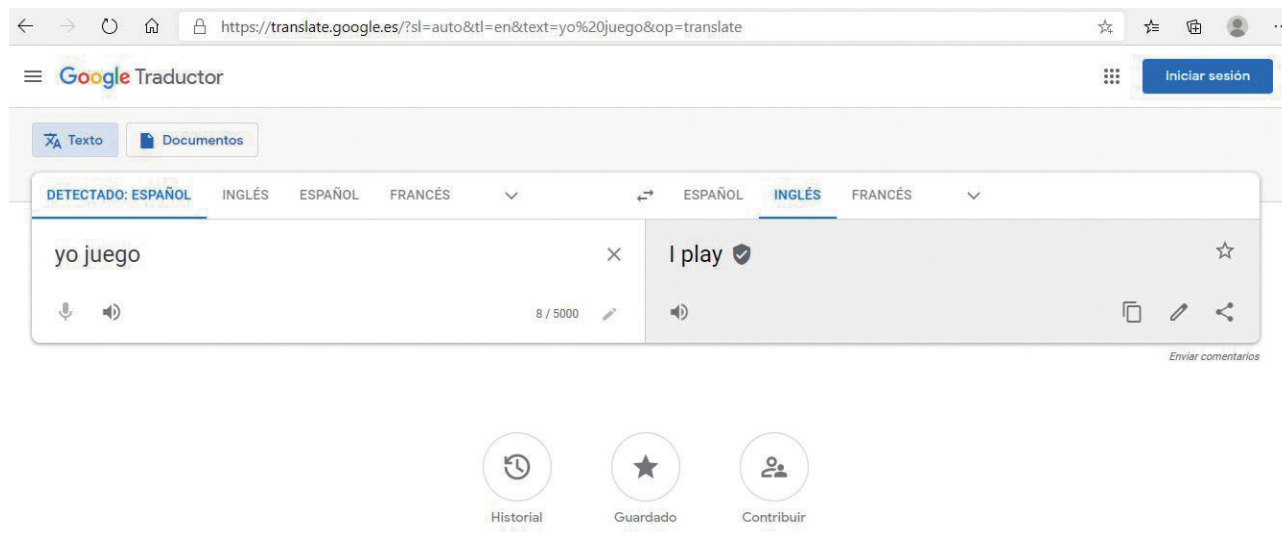


Fig. 28. Ejemplo de uso de Google Traductor para evitar ambigüedades.

Lo mismo sucede con la ambigüedad de la palabra *juego* como sustantivo. Si lo que queremos es saber qué termino se usa en inglés para designar un conjunto formado por un determinado número de ciertas cosas similares, resultará de gran ayuda aportar dicha información al motor (figura 29):

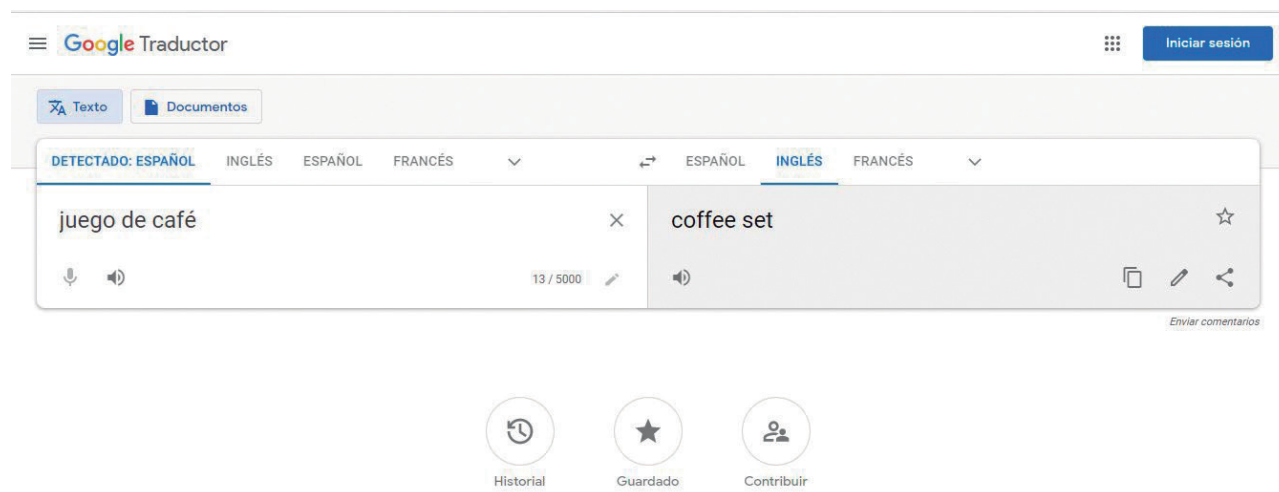


Fig. 29. Ejemplo de uso de Google Traductor para facilitar contexto y acotar los resultados.

Sin embargo, ¿qué ocurre cuando no es tan sencillo *ayudar* al motor de TA? Tomemos ahora como ejemplo de polisemia el verbo *dar*, un verbo de apoyo para el que se recogen 53 acepciones en el diccionario académico (ASALE, 2014).

Queremos comprobar cómo se comporta el motor de TA ante una consulta ES > EN acerca del verbo *dar* en su acepción de «rendir fruto». La estructura en inglés se construye con verbos como *to bear/to yield* (según la edición digital del diccionario Merriam Webster, 2021).

Introducimos directamente el verbo *dar*. La respuesta que nos devuelve el motor de TA no resulta muy esclarecedora (figura 30).

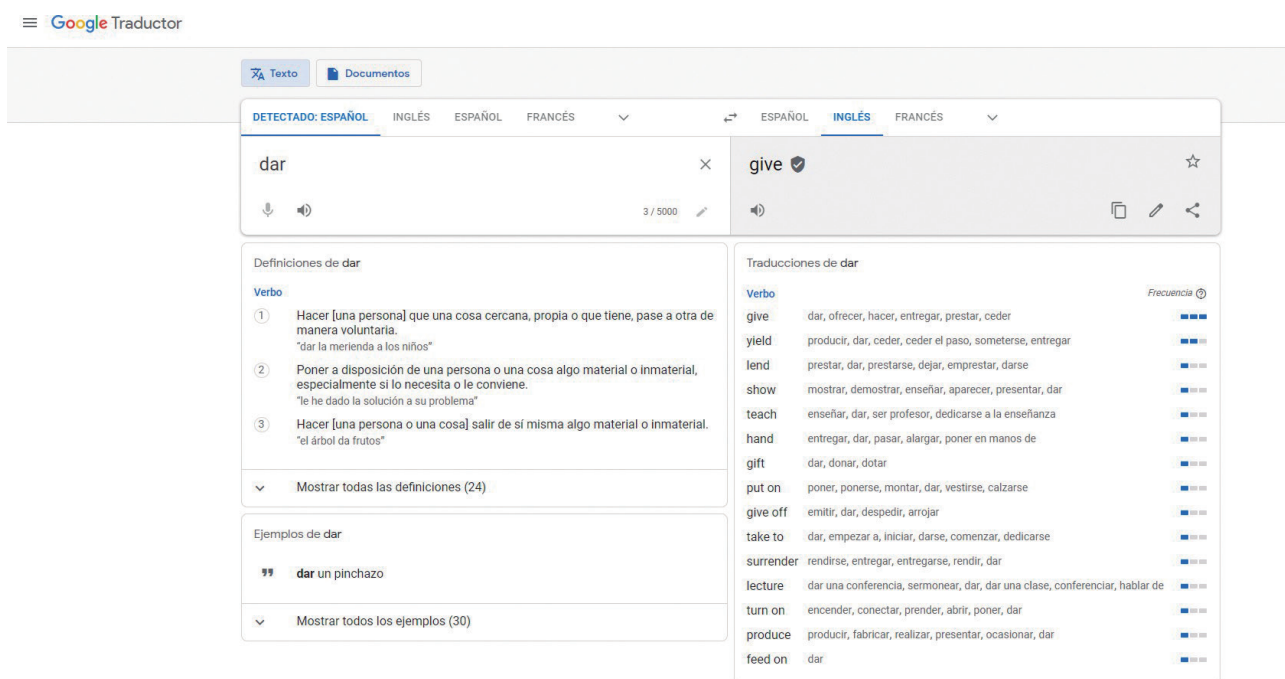


Fig. 30. Ejemplo de uso de Google Traductor ante la búsqueda de un verbo con numerosas acepciones.

Probamos, entonces, a *ayudar* a la TA, a ver si con ello mejora su rendimiento (figuras 31, 32 y 33):

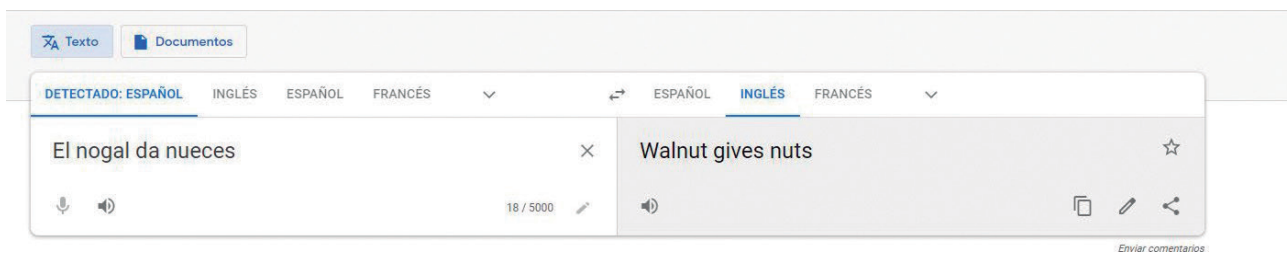


Fig. 31. Ejemplo de uso de Google Traductor para ayudar al motor de TA a contextualizar la búsqueda.

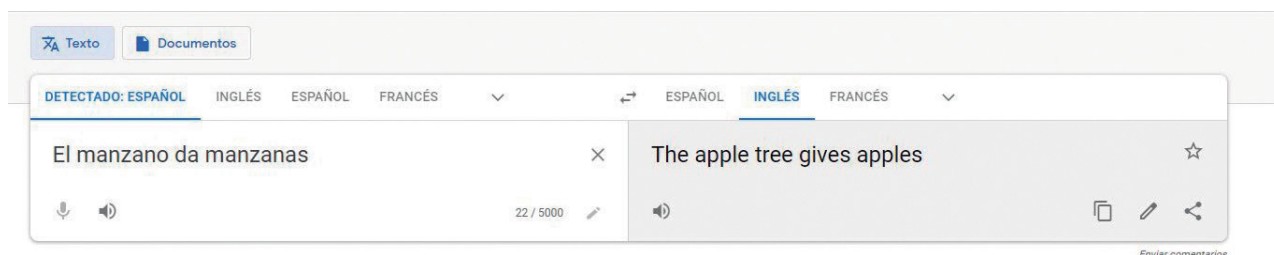


Fig. 32. Ejemplo de uso de Google Traductor para ayudar al motor de TA a contextualizar la búsqueda.

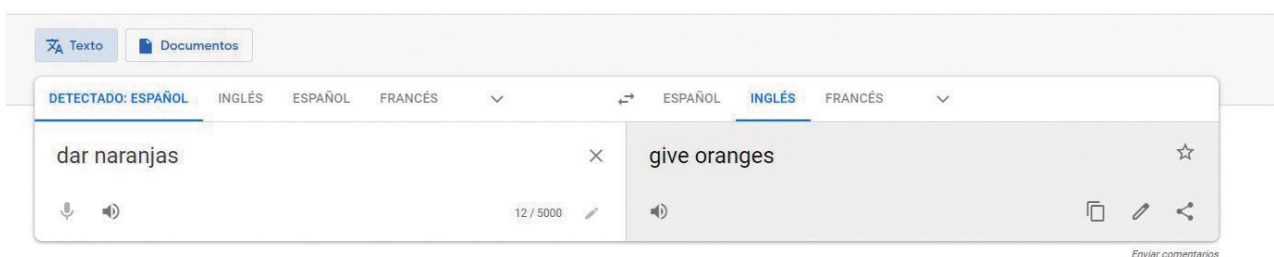


Fig. 33. Ejemplo de uso de Google Traductor para ayudar al motor de TA a contextualizar la búsqueda.

Los tres primeros intentos han sido en vano. Al cuarto intento (figura 34), en cambio, encontramos la respuesta (*to bear*):

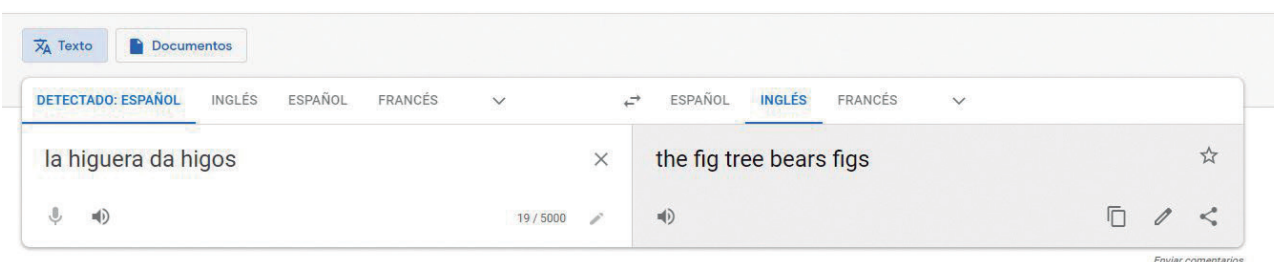
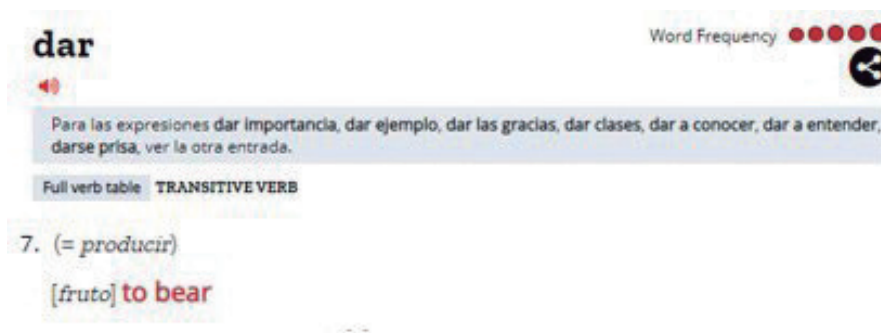


Fig. 34. Ejemplo de uso de Google Traductor para ayudar al motor de TA a contextualizar la búsqueda.

¿Qué sucedería si el usuario, que a priori asumimos desconoce la respuesta en inglés y por eso realiza la consulta, se fía del primer resultado que le ofrece la traducción automática? Que, sin saberlo, su búsqueda habría sido infructuosa.

En el caso que acabamos de citar, acudir a una herramienta de TA no ha resultado ni más rápido ni más efectivo que consultar un diccionario bilingüe, como veremos a continuación. Por ejemplo, el *Collins Spanish online dictionary* (figura 35) recoge 18 acepciones para el verbo *dar*, entre las que se encuentra la acepción «rendir fruto» que buscábamos:

Fig. 35. Consulta del verbo *dar* en el *Collins Spanish online dictionary*.

Si el usuario decidiese optar por un diccionario colaborativo, como el *WordReference English-Spanish Dictionary* (figura 36) su búsqueda también habría resultado satisfactoria.

Fig. 36. Consulta del verbo *dar* en *WordReference English-Spanish Dictionary*.

4.2. Valoración de resultados

En estos dos breves ejercicios prácticos que acabamos de realizar, hemos podido comprobar que el rendimiento de los motores de TA presenta carencias ante la consulta de unidades léxicas, aunque si *ayudamos* al motor de TA, el funcionamiento de este puede mejorar.

Hemos visto que los motores de TA usados en unidades léxicas ofrecen la traducción del término que consideran más adecuada en función del análisis de su corpus de textos paralelos, sin aportar mayor contexto, y la acompañan de una serie de datos adicionales. Estos datos no guardan relación entre sí sino que son informaciones independientes que figuran en la misma interfaz y que aportan muy poco a la consulta: definiciones en la lengua origen que no remiten a las propuestas de traducción, ejemplos fuera de contexto, catálogo de sinónimos donde figuran términos sin ningún tipo de lógica que resultan irrelevantes para nuestra búsqueda, etc.

En los diccionarios, en cambio, no solo encontramos el término equivalente en la lengua meta, sino que obtenemos un amplio abanico de informaciones, con un orden y una estructura, acerca de otros aspectos de la lengua (categoría gramatical, información geográfica complementaria, indicaciones de uso...). Esta combinación de datos que conforman los artículos de los diccionarios es la que proporciona esos pequeños matices que son esenciales para transmitir el significado deseado y no uno parecido.

Creemos, pues, que queda demostrado que, para consultas de unidades léxicas, el diccionario resulta más eficiente y eficaz que los motores de TA y que, si estos últimos quieren asumir el papel de los diccionarios bilingües, deberían, cuanto menos, nutrirse de datos lexicográficos que mejorarían su rendimiento.

5. Conclusiones

Los avances en el campo de las Tecnologías de la Información y la Comunicación han propiciado que los actuales hábitos de consulta lingüística de los usuarios se caractericen por la necesidad de encontrar respuestas rápidas a las búsquedas realizadas. Como consecuencia, se observa una creciente demanda de diferentes recursos lexicográficos en línea (diccionarios, portales y plataforma multilingües, etc.). En este trabajo, hemos querido centrarnos en el funcionamiento de los traductores automáticos, que se han posicionado como uno de los recursos en línea más demandados por usuarios no profesionales, no solo para traducir textos sino también para realizar consultas bilingües de unidades léxicas.

En el éxito de los motores de traducción automática han confluído diversos factores. Por un lado, la globalización ha impuesto el inglés como lengua franca de los negocios, lo que ha propiciado que muchos usuarios tengan que desarrollar labores de traducción o redacción en otros idiomas. Por otro, los hábitos han cambiado con la irrupción de internet: los internautas se han acostumbrado a la inmediatez, a obtener la información a golpe de un clic. Cuanto más rápido, mejor. Esta tendencia hacia el consumo rápido de contenidos hace que, en ocasiones, los usuarios no se detengan a analizar o verificar la calidad de la información obtenida.

Con nuestro análisis de uso de motores de TA en unidades léxicas, hemos podido comprobar que el rendimiento de los motores de TA aún presenta numerosas carencias. Ofrecen la traducción que consideran más adecuada en función del análisis de su corpus de textos paralelos, sin aportar mayor contexto, y la acompañan de una serie de datos adicionales, sin interconexión entre sí, que figuran en la misma interfaz y que aportan muy poco a la consulta: definiciones en la lengua origen que no guardan relación con las propuestas de traducción, ejemplos fuera de contexto, catálogos de sinónimos donde figuran términos sin ningún tipo de lógica y que resultan irrelevantes para nuestra consulta, etc.

En contraste con lo anterior, en los diccionarios bilingües consultados no solo encontramos el término equivalente en la lengua meta, sino que obtenemos una serie de informaciones de otros aspectos de la lengua (categoría gramatical, información geográfica complementaria, indicaciones de uso...), que proporciona esos pequeños matices que son esenciales para transmitir el significado deseado.

No pretendemos con nuestro análisis demonizar los avances en el campo de la traducción automática, ya que eso implicaría ir en contra del progreso. Pero sí creemos que la inteligencia artificial en el ámbito lingüístico, de momento, no debería pretender sustituir a la inteligencia humana, sino nutrirse de ella para que los datos y el contenido que ofrece a sus usuarios puedan considerarse una fuente de consulta primaria.

El uso de los traductores automáticos como herramientas lexicográficas, lejos de ser una moda pasajera, es una apuesta en firme de los gigantes de las TIC, que además de los esfuerzos por perfeccionar los motores de traducción en línea, están ampliando horizontes con el desarrollo de aplicaciones para los dispositivos móviles.

Es decir, tenemos al alcance de la mano, en sentido literal y figurado, herramientas que, sin estar concebidas como obras lexicográficas, están ocupando su lugar. Y esto, no solo tiene consecuencias para la lexicografía profesional tal y como se conocía hasta ahora, sino también para los usuarios no profesionales, aunque, en ciertos casos, su percepción acerca de estas herramientas tienda a ser más benévola y menos crítica.

Estamos asistiendo además a un fenómeno sin precedentes. La aparición de las redes sociales y la proliferación de los dispositivos móviles han propiciado un inmenso número de posibilidades comunicativas. Las distancias se han acortado, podemos comunicarnos de forma rápida, tenemos acceso a la información a golpe de clic y con la misma inmediatez podemos verter nuestras propias opiniones.

Si nos detenemos a observar el uso del lenguaje que se está haciendo en dichos medios sociales, parece evidente que todas las facilidades comunicativas que proporcionan las TIC han derivado en un uso más descuidado del lenguaje: faltas de ortografía, abuso de anglicismos, sintaxis descuidada... Según un estudio realizado en 2017 por la Universidad de Alcalá (Gumiel, 2019), el 90% de los jóvenes encuestados, con edades comprendidas entre los 14 y 30 años, admitía cometer faltas de ortografía cuando escribían en las redes sociales y otro 88,5% asumía que no cuidaban la elaboración de sus textos al escribir en dispositivos móviles como lo harían en otro formato.

Parece claro que, en estos tiempos, la rapidez prima y un texto bien escrito requiere dedicarle tiempo, ya sea para evitar faltas de ortografía, ya sea para encontrar el vocablo adecuado.

Esto nos lleva a la siguiente reflexión: si un usuario común, especialmente el nativo digital, ha cambiado su forma de comunicarse por escrito, ¿qué nos hace pensar que no ocurre lo mismo con sus hábitos de consulta lexicográfica? Si

a golpe de aplicación móvil dan por resuelta su duda, ¿se va a invertir tiempo en contrastar la respuesta, en cuestionar si la fuente es fiable? Nuestra experiencia personal es que no siempre. En ocasiones, se confía ciegamente en que una empresa X, líder en su sector, va a proporcionar información de primera. Prima el envoltorio sobre el contenido.

Nuestro trabajo ha demostrado que los diccionarios (o los datos en ellos contenidos) siguen siendo necesarios, pero la realidad es que los diccionarios tradicionales, como se entendían hasta ahora, han dejado de percibirse como una fuente de autoridad (Maldonado, 2019). Los usuarios quieren respuestas inmediatas y las quieren ya. Entonces, si Mahoma (usuarios) no va a la montaña (diccionarios), quizá la montaña (diccionario y, por extensión datos lexicográficos) tenga que ir a Mahoma.

De la misma manera que en los últimos tiempos se ha abierto el debate acerca de cómo se puede controlar la desinformación en la red (las denominadas *noticias falsas*), nos preguntamos si no es posible *regular* qué requisitos han de cumplir aquellas herramientas de consulta que se suben a internet para uso abierto. Es decir, pongamos que un motor de traducción ofrece, entre sus funcionalidades, la consulta de términos aislados, como si de un diccionario bilingüe se tratara. Además, le añade una columna con información del término en la lengua de partida, haciendo las veces de diccionario monolingüe.

Para garantizar que esa herramienta *sirve* realmente, ¿por qué no requerir que los datos lexicográficos que se muestran vengan de la mano de profesionales? Ya hemos visto un ejemplo en este documento: al introducir en el campo de búsqueda el verbo *juego*, el traductor automático nos devuelve el término *sport* en inglés y, otros sinónimos para *sport* que no guardan relación con nuestra consulta inicial. La información se ofrece de manera inconexa y desordenada, de forma que el usuario se encuentra con un batiburrillo de datos, como en la sección de oportunidades de unos grandes almacenes: si rebuscamos mucho, encontraremos, entre el montón de ropa, la prenda que nos queda como un guante. Si no dedicamos tiempo, es probable que nos llevemos una prenda con tara. Nos podrá valer, sí, pero no nos sentará igual.

Llegados a este punto, creemos que se hace imperante el compromiso de las grandes empresas que desarrollan estos motores de TA. Si se ofrecen determinadas funcionalidades, el contenido de estas debería ser fiable pero este índice de fiabilidad difícilmente se alcanzará si cualquier usuario puede crear, actualizar o corregir palabras o frases y publicarlas sin que estos datos pasen ningún tipo de filtro lexicográfico que garantice, por ejemplo, que se adaptan a las normas gramaticales y ortográficas de la lengua meta o que el contenido es adecuado.

No estamos con esto censurando la participación de los usuarios en la mejora de las herramientas gratuitas en línea. Sin embargo, sí creemos necesario que esta *lexicografía colaborativa* cuente con unos procedimientos que garanticen que el contenido generado por los usuarios pase ciertos filtros de calidad antes de ser accesible al público general. Existe un precedente en la editorial PONS (Cop, 2019) que, en sus diccionarios en línea, se permite a los usuarios crear sus propias entradas. Después de haber creado la entrada, la envían a la redacción pulsando un botón. Seguidamente, la redacción de PONS lleva a cabo un control de calidad antes de publicar la entrada.

Creemos firmemente que un proceso similar podría resultar muy beneficioso para todas las partes. Por un lado, los gigantes de las TIC continuarían desarrollando sus aplicaciones, con el valor añadido de ampliar sus miras y abrirse paso como herramienta didáctica válida. Por otro, a la lexicografía comercial se le abriría un nuevo nicho de mercado. Desde hace años no se abordan en las editoriales nuevos proyectos lexicográficos. Los hábitos de consulta han cambiado y el diccionario ya no se percibe como una fuente de autoridad. Con este nuevo paradigma, cambiarían los medios, pero el fin sigue siendo el mismo: se necesita de la lexicografía más que nunca para preservar la calidad de nuestro análisis léxico. Por último, los usuarios no verían sus hábitos de consulta modificados, podrían seguir formando parte de una plataforma colaborativa, tan en auge últimamente y, he aquí la gran diferencia, accederían a contenido fiable, que redundaría en su beneficio.

Somos conscientes de que pasar de la teoría a la práctica no es tan sencillo y que emprender una tarea de este tipo requiere muchos recursos humanos y económicos. Quizás, incluso sea una cuestión de la que haya que hacer partícipes a las autoridades competentes porque velar por la calidad del léxico no deja de ser una labor social y educativa: los diccionarios, y por extensión los datos lexicográficos, transmiten una visión de la sociedad que habla la lengua en ellos descritas. El lenguaje, y por ende las palabras, refleja lo que existe, crea la realidad y configura nuestra visión del mundo. Como decía Alex Grijelmo (2000), «pensamos con palabras». Es decir, según qué palabras utilicemos así formaremos nuestro pensamiento. Podría decirse que es una suerte de *patrimonio inmaterial*. Si a todos nos parece una atrocidad romper en añicos las vidrieras de una catedral, ¿por qué no poner todos los medios necesarios para proteger las lenguas?

Bibliografía

- Álvarez, M. (2018). «Los desafíos digitales en el mercado de la traducción». *Caracteres. Estudios culturales y críticos de la esfera digital*, 7(2): 193-215 [Disponible en <http://revistacaracteres.net/wp-content/uploads/2018/11/Caracteresvol7n2noviembre2018-traduccion.pdf> (Última consulta: 12 de junio de 2021)].
- ASALE (2010). *Diccionario de americanismos*. Madrid: Santillana.
- ASALE (2010). *Ortografía de la lengua española*. Madrid: Espasa.
- ASALE (2011). *Nueva gramática básica de la lengua española*. Madrid: Espasa.

- ASALE (2014): *Diccionario de la lengua española*, 23.^a ed., [versión 23.4 en línea]. [Disponible en: <https://dle.rae.es> (Última consulta: 11 de junio de 2021)].
- Blanco, P. R. (2018). «La bochornosa traducción con Google Translate de la web oficial de turismo de Santander». *El País* [Disponible en https://elpais.com/politica/2018/01/17/diario_de_espana/1516183473_130112.html (Última consulta: de 2021)].
- Cambridge University Press (n.d.). «Cambridge English-Spanish Dictionary» [Disponible en <https://dictionary.cambridge.org/dictionary/english-spanish/> (Última consulta: 11 de junio de 2021)].
- Centro de Traducción de los Órganos de la Unión Europea (2021). «IATE. European Union Terminology» [Disponible en <https://iate.europa.eu/home> (Última consulta: 11 de junio de 2021)].
- Chen, Z. (2020). «Un estudio empírico del alcance de la traducción automática del español al chino. Caso de estudio de GNMT sobre las expresiones metafóricas y metonímicas». *Círculo De Lingüística Aplicada a La Comunicación*, 83, 1-24. [Disponible en <https://doi.org/10.5209/clac.70560> (Última consulta: 11 de junio de 2021)].
- Cop, M. (2019). «Los nuevos diccionarios bilingües en línea: *castells* colaborativos. El ejemplo de www.pons.com». En: *JEFE-VI I: Contribuciones a las Primeras Jornadas de Español para Fines Específicos de Viena*. Consejería de Educación en Suiza y Austria, abril 2019: 2-17 [Disponible en https://sede.educacion.gob.es/publiventa/descarga.action?f_codigo_agc=20079 (Última consulta: 11 de junio de 2021)].
- Córdoba, I.; Rico C. *et al* (2015). «Estudio de viabilidad para la implantación de la traducción automática en la empresa VITAE». [Disponible en https://www.academia.edu/13019044/Estudio_de_viabilidad_para_la_implantaci%C3%B3n_de_la_traduccion%C3%B3n_autom%C3%A1tica (Última consulta: 11 de junio de 2021)].
- De Miguel, E. (2016). «Lexicología». En Gutiérrez-Rexach (dir.) *Enciclopedia de la lingüística Hispánica*. Londres/Nueva York: Routledge, 173-185.
- DeepL (2021). «El Traductor de DeepL» [Disponible en <https://www.deepl.com/translator.html> (Última consulta: 11 de junio de 2021)].
- DeepL (2021). «Linguee» [Disponible en <https://www.linguee.es/> (Última consulta: 11 de junio de 2021)].
- Diéguez, M.I. y Riedemann, K. (2001). «Aciertos y errores en la traducción automática: metodología de la enseñanza-aprendizaje de la traducción humana». *Onomázein*, 6: 203-221 [Disponible en <https://www.redalyc.org/pdf/1345/134518177011.pdf> (Última consulta: 11 de junio de 2021)].
- Domínguez, M.J. (2019). «Recursos lexicográficos electrónicos multilingües y plurilingües: definición y clasificación tipológico-descriptiva». *Revista Internacional de Lenguas Extranjeras*, 10: 49-74 [Disponible en <http://dx.doi.org/10.17345/rile10.49-74> (Última consulta: 11 de junio de 2021)].
- Domínguez, M.J. y C. Valcárcel (2015). «Hábitos de uso de los diccionarios entre los estudiantes universitarios europeos: ¿nuevas tendencias?». En Domínguez, M.J., X. Gómez y C. Valcárcel (eds.) *Lexicografía de las lenguas románicas II. Aproximaciones a la lexicografía contemporánea contrastiva*. Berlín: de Gruyter, 165-189.
- Espasa-Calpe (2001). «Gran diccionario español-portugués portugués-español» [Disponible en <https://www.wordreference.com/espt/> (Última consulta: 11 de junio de 2021)].
- Farlex (2021). *The Free Dictionary* [Disponible en <https://es.thefreedictionary.com/> (Última consulta: 11 de junio de 2021)].
- Google (2021). «Google Traductor». [Disponible en <https://translate.google.com/intl/es/about/> (Última consulta: 11 de junio de 2021)].
- Grijelmo, A. (2000). *La seducción de las palabras*. Madrid: Taurus.
- Gumiel, S. (2019). «Objetivo del año: escribir y hablar mejor». *Diario digital*, Universidad de Alcalá [Disponible en <http://portalcomunicacion.uah.es/diario-digital/entrevista/escribir-correctamente-algo-mas-que-una-cuestion-de-ortografia-fotos?n=11> (Última consulta: 11 de junio de 2021)].
- Hernández, H.; Maldonado, C.; Tarp, S. y Nomdedeu-Rull, A. (2019). «El estatus de la lexicografía. Nuevas y variadas respuestas a una antigua cuestión». *La investigación en lexicografía hoy (Volumen II): Diccionarios bilingües, lingüística y uso del diccionario*: 699-733.
- Larousse (n.d.). «*Dictionnaire Larousse français-espagnol*» [Disponible en <https://www.larousse.fr/dictionnaires/francais-espagnol> (Última consulta: 11 de junio de 2021)].
- Leddy, M. (2019). «Your Questions About Machine Translation, Answered (and no, the robots haven't won...yet)». [Disponible en <https://blog.weglot.com/your-questions-about-machine-translation-answered-and-no-the-robots-havent-won-yet/> (Última consulta: 11 de junio de 2021)].
- Maldonado, M.C. (2013): «La entrada léxica en el discurso lexicográfico digital». *Círculo de Lingüística aplicada a la comunicación* 56:26-52, [Disponible en <http://revistas.ucm.es/index.php/CLAC/article/view/43866/41467> (Última consulta: 11 de junio de 2021)].
- Maldonado, M.C. (2019): «Las investigaciones de mercado en lexicografía comercial: un aprendizaje para el mundo académico e investigador». *Revista Internacional de Lenguas Extranjeras*, 10: 101- 118 [Disponible en <https://doi.org/10.17345/rile10.2557> (Última consulta: 11 de junio de 2021)].
- Merriam-Webster (2021). «Dictionary by Merriam-Webster». [Disponible en <https://www.merriam-webster.com/> (Última consulta: 11 de junio de 2021)].
- Microsoft (2021). «Microsoft Bing» [Disponible en <https://www.bing.com/translator/> (Última consulta: 11 de junio de 2021)].
- Parra, C. (2018). «Evolución de la traducción automática». *La linterna del Traductor. La revista multilingüe de Asetrad*, 16: 20-27 [Disponible en http://www.lalinternadeltraductor.org/pdf/lalinterna_n16.pdf (Última consulta: 11 de junio de 2021)].
- PONS (2021). «Diccionario PONS español-francés» [Disponible en <https://es.pons.com/traduccion/francés-español> (Última consulta: 11 de junio de 2021)].
- PONS (2021). «Diccionario PONS inglés-español» [Disponible en <https://es.pons.com/traduccion/español-inglés> (Última consulta: 11 de junio de 2021)].

- Sánchez, F; Forcada, M.L. y Esplà M. (2020). «Tecnologías de la Traducción: técnicas de traducción automática». [Disponible en: <https://rua.ua.es/dspace/bitstream/10045/106980/1/slides-ta-es.pdf> (Última consulta: 11 de junio de 2021)].
- Singla, K. (2015). «Methods for Leveraging Lexical Information in SMT». Trabajo final de Máster en Lingüística Computacional [Disponible en https://www.researchgate.net/publication/279181014_Methods_for_Leveraging_Lexical_Information_in_SMT (Última consulta: 11 de junio de 2021)].
- Tarp, S. (2019). «La ventana al futuro: despidiéndose de los diccionarios para abrazar la lexicografía». *RILEX. Revista Sobre Investigaciones léxicas*, 2(2), 5-36. [Disponible en <https://doi.org/10.17561/rilex.v2.n2.1> (Última consulta: 11 de junio de 2021)].
- Useless paperclip: Blog de informática y traducción (2017). «Traducción automática basada en reglas». [Disponible en: <https://uselesspaperclip2.wordpress.com/2017/09/20/traduccion-automatica-basada-en-reglas/> (Última consulta: 11 de junio de 2021)].
- WordReference (2021). «Diccionario WordReference Francés-Español» [Disponible en <https://www.wordreference.com/esfr/> (Última consulta: 11 de junio de 2021)].
- WordReference (2021). «WordReference English-Spanish Dictionary» [Disponible en <https://www.wordreference.com/es/en/translation.asp?spen=> (Última consulta: 11 de junio de 2021)].

Copyright of CIRCULO de Linguistica Aplicada a la Comunicacion is the property of Universidad Complutense de Madrid and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.