

LA TRADUCCIÓN AUTOMÁTICA DE LOCUCIONES NOMINALES DEL ESPAÑOL AL INGLÉS: A PAIN IN THE NECK?

Noun Idioms and Machine Translation from Spanish into English: A Pain in the Neck

María Teresa ORTEGO-ANTÓN

Universidad de Valladolid

mariateresa.ortego@uva.es

Miriam SEGHIRI

Universidad de Málaga

seghiri@uma.es

RESUMEN: Las unidades fraseológicas y, más específicamente, la identificación y traducción de las locuciones nominales, constituyen uno de los obstáculos a la hora de desarrollar herramientas basadas en el procesamiento del lenguaje natural (PLN) y en TAO. De hecho, muchas de las dificultades están relacionadas directamente con el trasvase de las locuciones. Por tanto, en este trabajo pretendemos abordar los patrones de comportamiento que se desprenden del trasvase interlingüístico del español al inglés que una selección de motores de TA (Zaretskaya *et al.*, 2015, 2016 y 2018) ofrece de una muestra de locuciones nominales. Los resultados obtenidos nos permitirán, en fases posteriores, implementar las mencionadas herramientas, así como mejorar la formación de traductores y estudiantes de Traducción para conseguir una mayor productividad durante la posesición.

Palabras clave: locución nominal; traducción automática; inglés; español; equivalentes, fraseología.

1. LAS LOCUCIONES NOMINALES Y SU TRADUCCIÓN

Las unidades fraseológicas o unidades multiverbales pueden definirse como «meaningful lexical units made of two or more words in which at least one of them is restricted by linguistic conventions in the sense that it is not freely chosen» (Monti *et al.* 2018, 1), es decir, el significado de una unidad multiverbal no siempre se deriva del significado de cada uno de sus componentes ni puede determinarse por las reglas utilizadas para combinarlas. De hecho, Baldwin y Kim (2010, 268) señalan que estas unidades se caracterizan por poderse descomponer en múltiples lexemas y mostrar una idiomatidad léxica, sintáctica, semántica, pragmática o estadística.

Conviene apuntar que dependiendo del enfoque desde el que se aborden estas unidades, coexisten dos términos para denominarlas; a saber, unidades fraseológicas, que se emplea en fraseología (Corpas Pastor 2013, 9), y unidades multiverbales, que es la denominación utilizada en el campo del procesamiento del lenguaje natural (PLN).

Aunque son muchos los estudios que se centran en describir las propiedades de las unidades multiverbales y su identificación monolingüe (Baldwin y Villavicendio 2002), el análisis interlingüístico de estas unidades y la extracción automática de su traducción es un campo que adolece de estudios (Mirkov

2016, 280), probablemente debido a las diferencias culturales y a los anisomorfismos lingüísticos, unido a los diversos grados de equivalencia en los diferentes niveles del sistema y del texto.

Asimismo, ante la existencia de múltiples categorizaciones de las unidades multiverbales, en este trabajo vamos a basarnos en la categorización propuesta por Corpas Pastor (2000, 484) que, en función del grado de fijación y de la independencia a nivel textual y oracional, divide las unidades multiverbales en tres categorías: colocaciones, locuciones y expresiones o fórmulas rutinarias. Nos centraremos en el estudio de las locuciones, definidas por Corpas Pastor (1996, 88) como unidades multiverbales con una serie de características distintivas, tales como la fijación interna, la unidad de significado y la fijación externa pasemática. Estas unidades no constituyen enunciados completos y, generalmente, funcionan como elementos oracionales. Además, pueden clasificarse atendiendo a la función oracional y al núcleo de su sintagma, así que acotaremos el análisis a las locuciones nominales formadas en español por «sustantivo + preposición + sustantivo».

Por lo que respecta a los grados de equivalencia de las locuciones, Corpas Pastor (2003, 206-209) distingue cuatro tipos:

- Equivalencia total: dos locuciones presentan relación de equivalencia total cuando ambas forman un translema en todo momento, es decir, cuando ambas se traducen mutuamente mediante la técnica de transferencia denominada *sustitución*. En este caso, ambas unidades presentan el mismo significado denotativo y connotativo (expresivo), dándose en los mismos niveles diastráticos, diafásicos y diatópicos.
- Equivalencia parcial: una locución de la lengua origen (LO) se traduce por una locución en la lengua meta (LM) que presenta diferencias de significado denominativo o connotativo, o que pertenece a una variedad (diastrática diafásica y diatópica) distinta, o es traducida por una unidad léxica simple carente de los valores expresivos de la locución en cuestión.
- Equivalencia nula: no existe un equivalente de traducción en la LM. El traductor debe optar por la paráfrasis o por otra técnica de transferencia para expresar el significado de la unidad, perdiéndose, generalmente, las connotaciones correspondientes.
- Equivalencia aparente: se dan en aquellas unidades de la LO y de la LM que presentan similitud formal con respecto a sus elementos constitutivos, pero diferencia semántica; por ejemplo, los falsos amigos.

En este sentido, un buen tratamiento de las unidades multiverbales es esencial para el PLN y, en consecuencia, para la TA y las tecnologías de la traducción, dado que, si no se detectan automáticamente las unidades multiverbales, el producto resultante comprometerá la actuación de las aplicaciones (Monti *et al.* 2018, 3). En este sentido, en estudios previos (Monti 2013, Ramisch *et al.* 2013, Barreiro *et al.* 2014, entre otros) se pone de manifiesto que los motores de TA fallan a la hora de detectar y traducir las locuciones, por lo que se han convertido en la china en el zapato para los expertos en PLN.

Por tanto, en este trabajo pretendemos evaluar hasta qué punto el producto resultante de la traducción del español al inglés de una selección de locuciones proporcionado por los motores de TA más utilizados por los traductores profesionales (Zaretskaya *et al.* 2015, 2016, 2018) se corresponde con los equivalentes establecidos o con traducciones literales que no son adecuadas en la LM.

2. LOS MOTORES DE TRADUCCIÓN AUTOMÁTICA

La TA puede definirse, según la Norma ISO 17100/2015 de Servicios de traducción como la «traducción automatizada del texto o discurso de un lenguaje natural a otro mediante el uso de un sistema informático». En función del sistema empleado, los motores de TA pueden clasificarse en la actualidad en cuatro tipos:

- Sistemas de TA basados en reglas (TAR), donde el proceso de traducción consiste en aplicar las reglas necesarias sobre una frase dada para obtener la traducción (Babylon o SDL Free Translation).
- Sistemas de TA basados en estadística (TAE); a partir del uso de los datos del corpus de entrenamiento se desarrollan modelos probabilísticos que asignan una probabilidad alta a aquellas frases destino que son traducciones reales de la frase dada y una probabilidad baja o nula a aquellas frases que no lo son (Bing o MyMemory).
- Sistemas de TA basados en motores híbridos (TAH), que son aquellos que combinan los sistemas de TA basados en reglas y en estadística (SYSTRANet).
- Sistemas de TA neuronal (TAN), que es una nueva tecnología de aprendizaje de máquinas llamada *Deep Learning* o *Deep Neural Networks*, que trata de imitar el funcionamiento del cerebro humano. Las redes neuronales para la TA se hicieron posibles recientemente y, aunque aún se encuentran en sus primeras etapas de desarrollo, ya ofrecen mejores traducciones que los motores de TAE en muchos idiomas. Uno de los principales ejemplos de TAN lo encontramos en Google Translate o en el novedoso DeepL, creado en 2017, que se perfila como el mayor competidor del anterior.

Actualmente, el aumento de la demanda de traducción automática y de posesición por parte de los proveedores de servicios lingüísticos es evidente, como constatan Torres Hostench *et al.* (2016, 16). Estas autoras reconocen que el 47,3 % de las empresas proveedoras de servicios lingüísticos participantes utilizan la TA y la posesición durante el proceso de traducción. En esta línea, el estudio conducido por Zaretskaya *et al.* (2015, 2016, 2018) reveló (Gráfico 1) que el programa de TA gratuito más utilizado por los traductores fue Google Translate (73,3 %), basado en TAN. Le siguen en uso, aunque de lejos, My Memory (TAE), SDL Free Translation, (TAR), Babylon (TAR), Bing (TAE) y SYSTRANet (TAH). En este sentido, cabe apuntar que el estudio supraindicado se llevó a cabo durante los años 2015 y 2016, motivo por el que no se recogió DeepL.

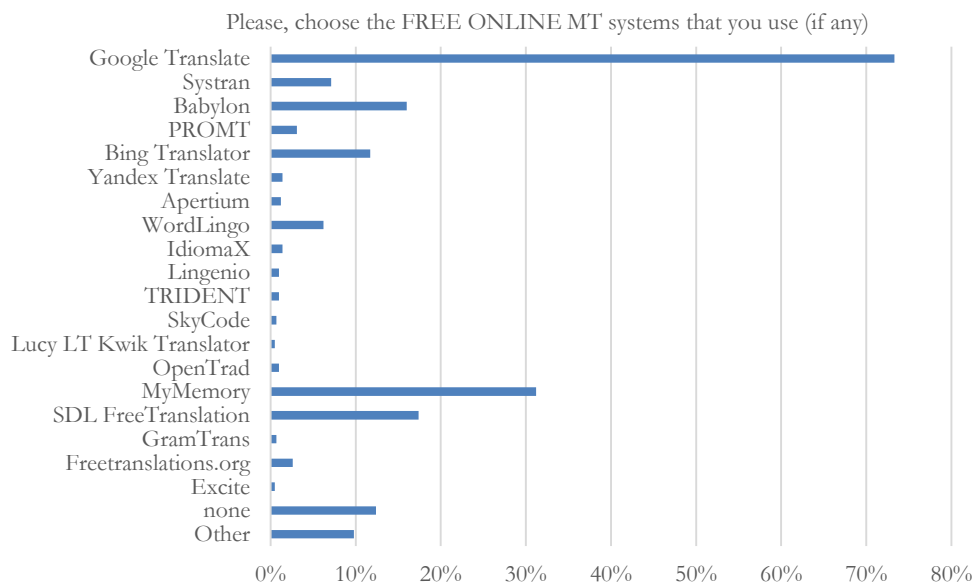


Gráfico 1: Programas de TA más utilizados por los usuarios.

Por tanto, basándonos en los resultados del estudio de Zaretskaya *et al.* (2015, 2016, 2018), analizaremos los siguientes motores de TA:

- Google Translate¹ (GT) es un motor de TAN disponible en más de 100 lenguas que no solo traduce textos, sino también otros formatos, como páginas web.
- MyMemory² (MM) es un motor de TAE que contiene segmentos traducidos de forma profesional. Estos segmentos se han creado recopilando memorias de traducción de la Unión Europea, la ONU y páginas web multilingües.
- SDL FreeTranslation³ (SDL) es un motor TAE que permite a los usuarios obtener traducciones gratuitas de textos, documentos y páginas web.
- Babylon⁴ (BB) es un motor basado en ejemplos que incluye 1700 diccionarios, glosarios, tesaurus, enciclopedias y léxicos en más de 77 lenguas.
- Bing⁵ (BI) es el motor de TA desarrollado por Microsoft que combina la TA neuronal y la TAE. Está disponible en más de 60 lenguas.
- SYSTRANet⁶ (SY) es un motor híbrido que combina las reglas con la estadística. Proporciona traducciones en más de 40 pares de lenguas.
- DeepL⁷ (DL) es un motor de TA neuronal gratuito desarrollado en Alemania lanzado en agosto de 2017.

Una vez acotada la selección de motores de TA, procedemos a detallar la metodología de análisis que vamos a emplear en nuestro estudio.

3. METODOLOGÍA DE ANÁLISIS

En primer lugar, en cada motor de TA introducimos la locución nominal aislada de contexto. La muestra de locuciones nominales del tipo «sustantivo + preposición + sustantivo» la hemos obtenido de Corpas Pastor (1996, 95). A continuación, hemos extraído un ejemplo de la locución empleada en contexto de CORPES XXI⁸, un corpus monolingüe de referencia, como se detalla en la Tabla 1.

LOCUCIÓN NOMINAL	LOCUCIÓN NOMINAL EN CONTEXTO
cero a la izquierda	Es un cero a la izquierda.
alma de cántaro	¿Dónde estás, alma de cántaro?
cuesta de enero	El dinero ayudará a superar la difícil cuesta de enero.
cortina de humo	El remedio es, en realidad, una cortina de humo.
patas de gallo	Hidratar, limpiar o cuidarse las patas de gallo ya no es exclusivo de las mujeres.
huevo de Colón	La marca personal es el huevo de Colón.
lágrimas de cocodrilo	Llora con lágrimas de cocodrilo.
pañó de lágrimas	Unos usan las banderas como pañó de lágrimas.
conejiño de indias	Allí fue el conejiño de indias del experimento.
tabla de salvación	Estados Unidos y México se han convertido en la tabla de salvación de los dos gigantes españoles.

- 1 <https://translate.google.com/?hl=es>.
- 2 <https://mymemory.translated.net/espa%C3%B1ol/>.
- 3 <http://www.freetranslation.com/es/traducir-ingles-espanol>.
- 4 <http://traductor.babylon-software.com/>.
- 5 <https://www.bing.com/translator>.
- 6 <http://www.SYSTRANet.com/translate>.
- 7 <https://www.deepl.com/translator>.
- 8 <http://www.rae.es/recursos/banco-de-datos/corpes-xxi>.

Tabla 1: Muestra de análisis.

El siguiente paso es extraer el producto derivado de la traducción de la locución nominal con y sin contexto en cada motor de TA y analizar la técnica de traducción empleada siguiendo la clasificación propuesta por Molina y Hurtado Albir (2002, 509-511). Además, comprobamos cómo se ha traducido la locución en contexto y, por último, verificamos si el equivalente propuesto sin contexto es idéntico al ofrecido en contexto y contrastamos los resultados.

A continuación, procedemos a estudiar las diez locuciones nominales presentadas en la Tabla 1.

4. ANÁLISIS Y RESULTADOS

4.1 *Cero a la izquierda*

No hemos encontrado una locución equivalente en lengua inglesa que signifique «persona inútil, o que no vale para nada» (RAE, 2018). De hecho, la mayoría de los motores trasvasan esta locución nominal por una traducción literal, como se muestra en el Gráfico 2.

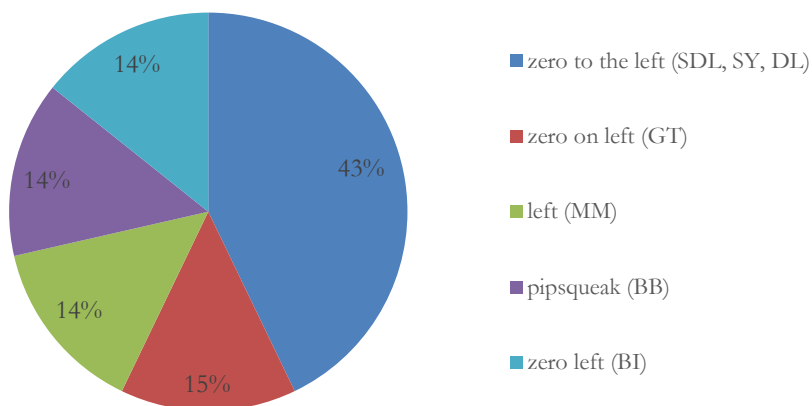


Gráfico 2: Equivalentes de *cero a la izquierda*.

Babylon ofrece un equivalente con un significado similar (*pipsqueak*), pero este motor de TA proporciona una opción diferente cuando trasvasa la locución nominal en contexto: «It is a zero on the left». Otros sistemas de TA también cambian la preposición que se emplea con «zero» cuando la locución está en contexto. Por ejemplo, el producto de DeepL se corresponde con «zero to the left» sin contexto y por «It's a zero on the left» cuando la locución está en contexto.

4.2 *Alma de cántaro*

El significado de esta locución se corresponde con «persona sumamente ingenua o pasmada» (RAE, 2018). Como se muestra en el Gráfico 3, los motores de TA ofrecen una traducción literal, así que el significado idiomático se pierde durante el trasvase interlingüístico del español al inglés.

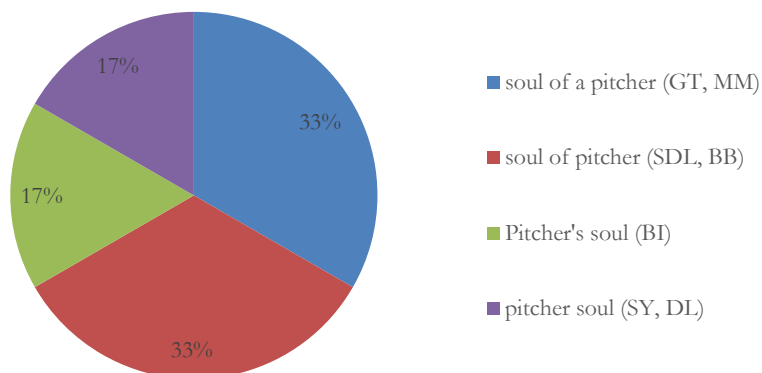


Gráfico 3: Equivalentes de *alma de cántaro*.

Asimismo, se utilizan diferentes equivalentes cuando la locución se emplea en contexto. Por ejemplo, Google Translate y My Memory no traducen la locución («And where are you, alma de cántaro?»), Bing ofrece un equivalente diferente «And where are you, soul of crock?» en vez de «Pitcher's soul» y DeepL propone «And where are you, sweetheart?», que tiene un significado completamente diferente al expresado en la LO.

4.3 *Cuesta de enero*

Esta locución puede transferirse al inglés por «post-holiday budget crunch» o «January's belt-tightening», pero los motores analizados producen una traducción literal, como se puede observar en el Gráfico 4. Por tanto, el significado y las connotaciones de esta locución se pierden durante la transferencia interlingüística.

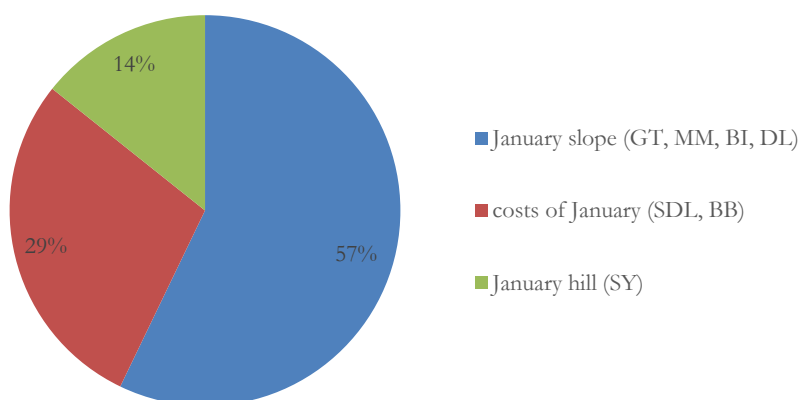


Gráfico 4: Equivalentes de *cuesta de enero*.

Por lo que respecta a los equivalentes en contexto, cuatro motores (Google Translate, My Memory, Bing y DeepL) ofrecen la siguiente opción: «[...] the difficult cost of January», mientras que SDL Free Translation y Babylon optan por la traducción literal: «the difficult slope of January». La traducción literal es también la técnica empleada por SYSTRANet: «the difficult hill of January». En consecuencia, los motores de TA no están lo suficientemente desarrollados para detectar y trasvasar adecuadamente esta locución nominal.

4.4 *Cortina de humo*

El empleo de esta locución nominal está ampliamente extendido porque puede clasificarse como un *européanismo* (Corpas Pastor 2000, 491), así que el producto ofrecido por la mayoría de los sistemas de traducción (86 %) se corresponde con el equivalente establecido, salvo en el caso de SDL Free Translation, motor que brinda una traducción literal palabra por palabra (*curtain of smoke*), como se recoge en el Gráfico 5.

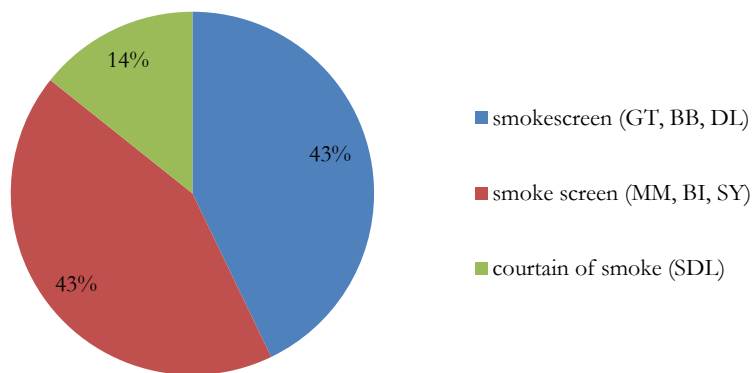


Gráfico 5: Equivalentes de *cortina de humo*.

Sin embargo, cuando la locución nominal se emplea en contexto, tanto SDL Free Translation como Babylon utilizan «curtain of smoke», como se aprecia en el siguiente ejemplo: «the remedy is in reality a curtain of smoke». Aunque en la descripción que SDL Free Translation reza sobre su motor afirma que se sustenta en un sistema TAN, tenemos dudas sobre esta afirmación puesto que la traducción que presenta se caracteriza por ser palabra por palabra.

4.5 *Patas de gallo*

El equivalente establecido de esta locución nominal en lengua inglesa se corresponde con «crow's feet», que coincide con el producto generado por Google Translate, My Memory, SDL Free Translator, Bing y DeepL. Por otro lado, Babylon ofrece la locución en singular, aunque hayamos introducido el texto en plural en la LO. Además, SYSTRANet proporciona una traducción literal, «houndstooth checks», que hace referencia a la parte inferior del animal, como puede observarse en el Gráfico 6.

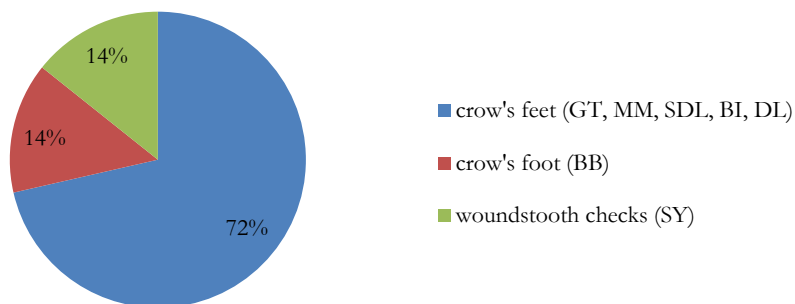


Gráfico 6: Equivalentes de *patas de gallo*.

En lo que respecta al trasvase de las locuciones en contexto, solamente Google Translate y DeepL coinciden en emplear el mismo equivalente que el propuesto sin contexto. El resto de los sistemas de TA ofrecen diversas posibilidades que se caracterizan por no traducir alguno de los constituyentes de la locución, como se puede comprobar en los siguientes ejemplos: «take care of feet of gallo» (SDL Free Translation y Babylon), «taking care of las patas de Gallo» (Bing) o «take care of themselves las patas of rooster» (SYSTRANet). Por tanto, el producto resultante es incomprensible para los usuarios de la LM.

4.6 *Huevo de Colón*

Esta locución nominal tiene un equivalente establecido en lengua inglesa, aunque la ortografía varía según los resultados proporcionados por cada sistema de TA.

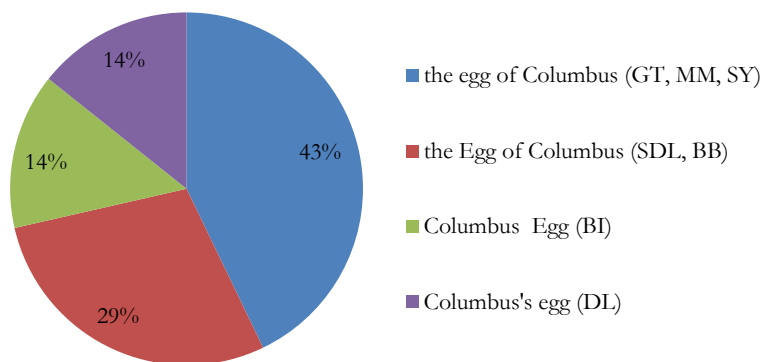


Gráfico 7: Equivalentes de *huevo de Colón*.

El equivalente propuesto coincide con el empleado cuando la locución está en contexto, salvo en los casos de Bing y de DeepL, puesto que Bing transfiere por «The personal mark is the egg of Columbus» en vez de emplear «Columbus egg». En el caso de DeepL, el equivalente propuesto tiene un error, puesto que duplica la “s” en el uso del genitivo sajón (*Columbus’s egg), si bien este error se subsana cuando se presenta la locución en contexto: «The personal mark is Columbus’ egg».

4.7 *Lágrimas de cocodrilo*

En el caso de esta locución, todos los motores de TA analizados proporcionan el mismo equivalente, que se corresponde con el equivalente establecido (*crocodile tears*). Además, cuando se trasvasa la locución en contexto, no se producen cambios.

4.8 *Paño de lágrimas*

Esta locución significa «persona en quien se encuentra frecuentemente atención, consuelo o ayuda» (RAE, 2018). Sin embargo, los motores de TA analizados ofrecen una traducción literal, como se puede observar en el Gráfico 8.

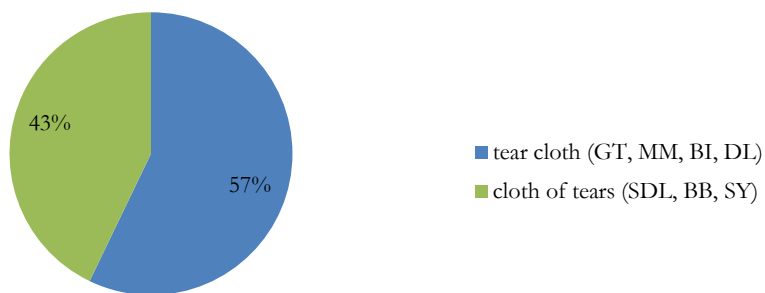


Gráfico 8: Equivalentes de *paño de lágrimas*.

Por lo que respecta al uso en contexto, solo tres de los motores analizados, a saber, SDL Free Translation, Babylon y SYSTRANet utilizan el mismo equivalente de forma aislada. Google Translate, My Memory y Bing optan por utilizar la técnica de postmodificación («Some use flags as a cloth of tears») en tanto que DeepL ofrece otra locución diferente («Some use the flags as a teardrop cloth»), cuyo significado no se adecua a la locución original en español.

4.9 *Conejillo de indias*

Tal y como sucede con «pata de gallo», las traducciones propuestas para «conejillo de indias» corresponden con equivalentes establecidos salvo en el caso de MyMemory, motor que no genera resultados. Asimismo, observamos diferencias gráficas a la hora de utilizar mayúsculas en «Guinea», como queda registrado en el Gráfico 9.

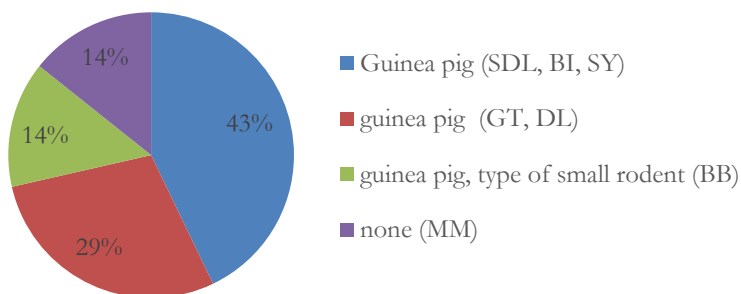


Gráfico 9: Equivalentes de *conejillo de indias*.

Además, los motores seleccionados presentan el mismo equivalente en contexto empleando minúsculas, salvo SYSTRANet, que propone «Guinea pig».

4.10 *Tabla de salvación*

Solo uno de los motores de TA, que se corresponde con DeepL, brinda un equivalente que muestra el significado de locución en lengua inglesa: «lifeline». El resto de los motores analizados proporciona una traducción literal, así que el significado de la locución no se transfiere de la LO a la LM, como se puede observar en el Gráfico 10.



Gráfico 10: Equivalentes de *tabla de salvación*.

Sin embargo, cuando la locución se trasvasa en contexto, seis de los motores (Google Translate, My Memory, SDL Free Translation, Babylon, Bing y DeepL) traducen empleando el término «lifeline», como por ejemplo en la oración «the U.S. and Mexico have become the lifeline [...]». Además, SYSTRANet usa un equivalente diferente al propuesto cuando trasvasa la locución en contexto, como se puede apreciar en el siguiente ejemplo: «the U.S. and Mexico have become the table of salvation».

5. DISCUSIÓN DE RESULTADOS

Una vez analizadas las locuciones nominales seleccionadas, los datos obtenidos nos permiten afirmar que los motores de TA que producen traducciones adecuadas se corresponden con DeepL y Google Translate, como se observa en los datos del Gráfico 11.

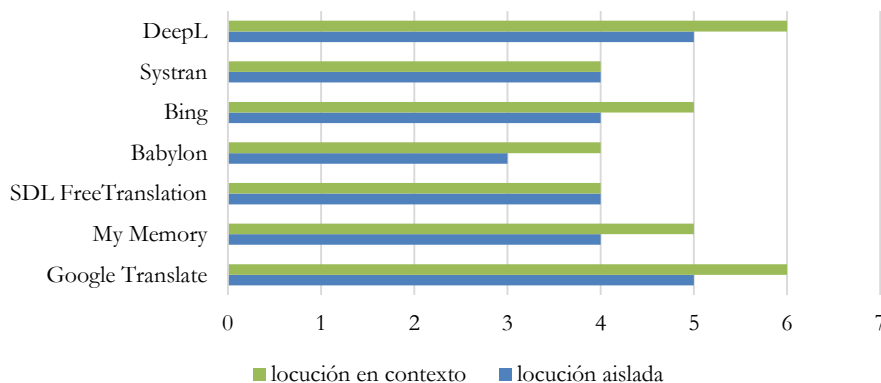


Gráfico 11: Resultados por motores de TA.

De hecho, estos dos motores de TA están basados en redes neuronales y ofrecen mejores resultados cuando la locución aparece en contexto en seis de los diez casos analizados. Además, también proporcionan equivalentes aceptables en el 50 % de los casos analizados.

Por el contrario, el motor de TA que presenta los peores resultados es Babylon, basado en estadística, puesto que la gran base de datos no ha sido alimentada con locuciones y, en consecuencia, no es capaz de detectarlas.

6. CONCLUSIONES

El producto que los motores de TA proporcionan durante el trasvase interlingüístico del español al inglés de las locuciones nominales depende de diversas variables. Parece obvio que el contexto en el que se utiliza una locución determina el equivalente propuesto. Sin embargo, si buscamos una locución sin contexto, los motores de TA basados en reglas y en estadística producen traducciones literales que no tienen significado en la lengua meta. Sin embargo, las locuciones relacionadas con animales sí que se trasvasan y utilizan el equivalente establecido, por ejemplo «crow's feet», «crocodile tears», «guinea pig» o «egg of Columbus». Este modelo también se reproduce en el caso de los europeísmos como «telón de acero». Las locuciones que denotan referencias culturales como «cuesta de enero», «alma de cántaro», «cero a la izquierda» o «pañó de lágrimas» se trasvasan utilizando traducciones literales o, incluso, una mezcla de ambas lenguas, que no es comprensible en la LM. A ello se suma que los motores de TA basados en redes neuronales proporcionan resultados más precisos que aquellos basados en otras tecnologías.

Puesto que todavía queda mucho por hacer antes de que el PLN y los sistemas de TA procesen con éxito las unidades multiverbales (Monti *et al.* 2018, 1.), consideramos que es necesario mejorar los motores de TA, por una parte, incluyendo los equivalentes de las locuciones nominales y, por otro lado, incorporando datos para que dichas unidades sean reconocidas como locuciones y puedan transferirse adecuadamente.

7. AGRADECIMIENTOS

El presente trabajo ha sido realizado en el marco de los proyectos “POSTrad II: La virtualización de la formación de traductores en traducción automática y posesición” de la Universidad de Valladolid (ref. PID nº 121, Convocatoria UVa 2018-2019), VIP (ref. FFI2016-75831) e INTERPRETA 2.0 (ref. PIE17-015 UMA, 2017-2019), además de en el seno de la red temática TRAJUTEC y la red docente de excelencia TACTRAD (ref. 719/2018), coordinadas por la Universidad de Málaga (España).

BIBLIOGRAFÍA

- Baldwin, Timothy y Aline Villavicencio. 2002. «Extracting the Unextractable: A Case Study on Verb-Particles». En *Proceedings of the 6th Conference on Natural Language Learning (CONLL'2002)*, 98–104.
- Baldwin, Timothy y Su Nam Kim. 2010. «Multiword Expressions». En *Handbook of Natural Language Processing*, ed. por N. Indurkha y F. J. Damerau. Boca Raton, FL: Taylor & Francis, 267–292.
- Barreiro, Anabela, Johanna Monti, Brigitte Orliac, Susanne Preuß, Kutz Arrieta, Wang Ling, Fernando Batista e Isabel Trancoso. 2014. «Linguistic Evaluation of Support Verb Constructions by OpenLogos and Google Translate». En *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC2014)*. Islandia, Reykjavik, 35–40.
- Corpas Pastor, Gloria. 1996. *Manual de fraseología española*. Madrid: Gredos.

- Corpas Pastor, Gloria. 2000. «Acerca de la (in)traducibilidad de la fraseología». En *Las lenguas de Europa: Estudios de Fraseología, Fraseografía y Traducción*, ed. por Gloria Corpas Pastor. Granada: Comares, 483–522.
- Corpas Pastor, Gloria. 2003. *Diez años de investigación en fraseología: análisis sintáctico-semánticos, contrastivos y traductológicos*. Madrid: Iberoamericana.
- Corpas Pastor, Gloria. 2013. «All that Glitters is not Gold when Translating phraseological Units». En *Workshop Proceedings for: Multi-word Units in Machine Translation and Translation Technologies*, ed. por Johanna Monti *et al.* Allschwil: The European Association for Machine Translation, 9–10.
- Mitkov, Ruslan. 2016. «Computational Phraseology Light: Automatic Translation of Multiword Expressions without Translation Resources». *Yearbook of Phraseology* 7(1): 149–166.
- Molina, Lucía y Amparo Hurtado Albir. 2002. «Translation Techniques Revisited: A Dynamic and Functionalist Approach». *Métra* 47(4): 498–512.
- Monti, Johanna. 2013. *Multi-Word Unit Processing in Machine Translation*. Tesis doctoral. Universidad de Salerno.
- Monti, Johanna, Violeta Seretan, Gloria Corpas Pastor y Ruslan Mitkov. 2018. *Multitword Units in Machine Translation and Translation Technology*. Amsterdam/Filadélfia: John Benjamins.
- RAE. 2018. *Diccionario de la Lengua Española*. Fecha de acceso 15 de octubre de 2018. <http://dle.rae.es/>.
- Ramisch, Carlos, Aline Villavicencio y Valia Kordoni. 2013. «Introduction to the Special Issue on Multiword Expressions: From Theory to Practice and Use». *ACM Transactions on Speech and Language Processing* 10 (2): 1–10.
- Torres-Hostench, Olga, Marisa Presas y Pilar Cid-Leal (coords.) 2016. *El uso de traducción automática y posesión en las empresas de servicios lingüísticos españolas: Informe de investigación ProjeTA 2015*. Bellaterra: UAB.
- Zaretskaya, Anna, Gloria Corpas Pastor y Míriam Seghiri. 2015. «Translators' Requirements for Translation Technologies: a User Survey». En *Nuevos horizontes en los Estudios de Traducción e Interpretación (Trabajos completos)*, ed. por Gloria Corpas Pastor *et al.* Ginebra: Tradulex, 247–254.
- Zaretskaya, Anna, Gloria Corpas Pastor y Míriam Seghiri. 2016. «Corpora in Computer Assisted Translation: a Users' View». En *Corpus-based Approaches to Translation and Interpreting: from Theory to Applications*, ed. por Gloria Corpas Pastor y Míriam Seghiri. Frankfurt: Peter Lang, 253–276.
- Zaretskaya, Anna, Gloria Corpas Pastor y Míriam Seghiri. 2018. «User Perspective on Translation Tools: Findings of a User Survey». En *Trends in e-Tools and Resources for Translators and Interpreters*, ed. por Gloria Corpas Pastor e Isabel Durán Muñoz. Leiden/Boston: Brill, 37–56.