

Google Translate: Análisis de errores en traducciones español-inglés en textos sobre nutrición

Google Translate: Error Analysis in Spanish-to-English Translation in Texts about Nutrition

**Carolina Clerici, Alcides Juan Diego Caballero, Ma. Florencia Becerra, Carolina
Chirino, Danisa Bastida, Valérie Gänswain, Francisco Migueles, Elisa Naef**

Universidad Nacional de Entre Ríos, Argentina

E-mail: carolina.clerici@uner.edu.ar; caballeroalcidesjuandiego@gmail.com; florencia.
becerra@uner.edu.ar; carolina.chirino@uner.edu.ar; danisa.bastida@uner.edu.ar; valerie.
ganswein@uner.edu.ar; francisco.migueles@uner.edu.ar; elisa.naef@uner.edu.ar

Fecha de recepción: 1 de Diciembre 2022 • Aceptado: 11 de Diciembre 2023

CLERICI, C.; CABALLERO, A.J. D.; BECERRA, M. F.; CHIRINO, C.; BASTIDA, D.; GÄNSWEIN, V.; MIGUELES, F. P. Á.;
NAEF, E.. (2023). Google Translate: Análisis de errores en traducciones español-inglés en textos sobre nutrición *Virtualidad,
Educación y Ciencia*, 27 (14), pp. 24-34.

Resumen

La traducción automática en línea (TAL) surge como una herramienta para solucionar la problemática del científico no angloparlante a la hora de publicar sus investigaciones. Una de estas herramientas es Google Translate (GT), por ello es necesario conocer su funcionamiento en textos específicos de nutrición para explotar su potencial y conocer sus limitaciones. Se realizó un análisis exploratorio-descriptivo de errores de traducción automática con GT, utilizando artículos de una revista argentina especializada en nutrición. La investigación fue llevada a cabo en tres etapas: registro de errores, cuantificación y clasificación. Se observó que muchos ocurrieron debido a una redacción poco clara del artículo original. Los errores más frecuentes fueron la traducción de siglas y los tecnicismos. Se registró una media de 54 errores por artículo. A partir de ello, se construyó una matriz de análisis de tipos de errores más comunes y se registró evidencia empírica de errores de TAL en el área específica de la nutrición.

Palabras clave: traducción automática en línea; errores; Google Translate; nutrición; artículos científicos

Abstract

Online Machine Translation (MT) emerges as a tool to solve the problem that non-English-speaking scientists face when publishing their research. One of these tools is Google Translate (GT), but it is necessary to know how it works in specific nutrition texts to exploit its potential and recognize its limitations. An exploratory-descriptive analysis of automatic translation errors with GT was performed, using articles from an Argentinian nutrition journal. It was carried out in three stages: error recording, quantification and classification. Many errors were noted to have occurred due to unclear wording of the original article. The most frequent errors were the translation of acronyms and technical terms. An average of 54 errors were recorded per article. The study led to the construction of a matrix of analysis of the most common types of errors and it was possible to record empirical evidence of MT errors in the specific area of nutrition.

Keywords: Online machine translation; errors; Google Translate; nutrition; scientific articles.

Introducción

En la actualidad, el inglés se ha convertido en la lengua de la ciencia, no sólo por la influencia de los países angloparlantes en el campo de la investigación científica sino también por razones de carácter lingüístico, como su relativa facilidad de uso, la flexibilidad, el amplio vocabulario y su brevedad (Barraclough, 2009).

Publicar trabajos en inglés posibilita llegar a la comunidad investigadora global mediante publicaciones en revistas de alto impacto. Por este motivo, a pesar de que sea relevante el descubrimiento, si el investigador no puede expresarse de manera clara, su investigación carecerá de impacto. A esto se suma la presión de publicar trabajos científicos de forma rápida y continua para sostener la carrera investigadora. Esta circunstancia nos conduce a pensar que, para obtener reconocimiento en el campo disciplinar de pertenencia, resulta imprescindible publicar en inglés; hecho que se convierte en un obstáculo para gran parte de la comunidad científica que no domina ese idioma. Por otro lado, la traducción humana manual, además de ser costosa, demanda mayor cantidad de tiempo. Según Craciunescu et al. (2007), para producir una traducción de calidad de un texto complejo, un traductor no puede traducir más de 4 a 6 páginas o 2000 palabras por día. En ese contexto, la traducción automática en línea (TAL) resulta una herramienta más atractiva, ya que permite solucionar la problemática que se le presenta al científico no angloparlante a la hora de publicar y dinamizar los procesos tradicionales de traducción.

Google Translate (GT) surgió en octubre de 2007 como un servicio de TAL gratuito que no precisa descarga ni instalación previa. Además, está disponible para computadoras y celulares. Para traducir un texto, GT analiza documentos en busca de la traducción más apropiada entre patrones de textos traducidos por usuarios (Ghasemi et al., 2016), lo que se conoce como traducción automática estadística (conocida por su sigla en inglés SMT, Statistical Machine Translation). La calidad de la traducción depende del número de textos disponibles (Karami, 2014) y se puede observar que desde sus inicios la calidad de las traducciones de español a inglés ha mejorado considerablemente.

Diversos estudios se han enfocado en los servicios de TAL con el objetivo de conocer su funcionamiento, evaluar y comparar la calidad de los resultados que proporcionan, así como también los errores que éstos puedan arrojar (Ghasemi et al. 2016). Ablanedo et al. (2007) realizaron un estudio sobre problemas en la traducción automática español-inglés. Estos autores destacaron que las principales ventajas de la TAL en comparación con la traducción humana son la velocidad, el costo, la disponibilidad y la consistencia; y probaron que la TAL puede ser hasta 195 veces más rápida que la traducción humana. Por su parte, Aiken et al. (2009) realizaron un estudio comparativo de traducciones inglés-alemán y alemán-inglés entre cuatro servicios de TAL (Google Translate, Yahoo Systran, Applied Language y X10) y observaron que GT es el que menos errores registró. Concluyeron que, aunque las traducciones no fueron perfectas, el nivel de comprensión fue elevado.

En su investigación, González Boluda (2010) realizó un estudio comparativo entre tres programas de TAL (Google, Systran y Reverso) que fueron analizados con base en los siguientes tres parámetros de evaluación: número de lenguas, enfoque de traducción, y análisis de traducción de frases y oraciones breves. Domínguez et al. (2013) condujeron un estudio de carácter experimental basados en un único servicio de TAL (Google), en el cual describen su experiencia para la traducción y análisis de textos. Entre las opciones disponibles en línea, se destaca GT (Aiken et al., 2009), inclusive en el ámbito

académico universitario (Clerici, 2016, Groves y Mundtl., 2015).

Por lo antes expuesto, se llevó a cabo una investigación en el marco de la convocatoria de Proyectos de Investigación desde las cátedras de la Facultad de Bromatología, UNER. El equipo de investigación pertenece a la cátedra Inglés Técnico de la Licenciatura en Nutrición. Se propuso, entonces, conocer cómo funciona GT en textos específicos de nutrición, analizar sus errores y evaluar sus resultados con el propósito de explotar su potencial y conocer sus limitaciones con el objetivo de diseñar un manual de apoyo a la traducción para estudiantes e investigadores en el área de nutrición.

Materiales y métodos

Materiales

Se realizó un análisis exploratorio-descriptivo de errores de traducción automática utilizando sólo GT por ser el más usado a nivel mundial (Ghasemi et al., 2016).

Se trabajó con una revista argentina especializada en nutrición que permite el acceso gratuito a sus artículos y que es un referente en el área y en la región. Se analizaron los artículos de 2018.

Procedimiento

Se descargaron los artículos de los volúmenes seleccionados en formato PDF y se convirtieron a documentos de texto sin formato. Se realizaron los cambios necesarios en el nuevo documento para evitar posibles errores de traducción producidos por la conversión de archivos, particularmente por el espaciado. Se descargaron y analizaron seis artículos en total.

Se creó una matriz de cuatro columnas para cada artículo. En la primera se copió el texto original (omitiendo los datos de los autores, tablas, figuras, referencias y agradecimientos). En las tres siguientes, se importaron tres traducciones del texto hechas en GT utilizando tres computadoras distintas, ya que, en experiencias preliminares, se observó que GT se adapta al perfil de usuario, generando versiones diferentes. La traducción se realizó por oraciones debido a la practicidad y por ser la unidad mínima de sentido a los fines de su cuantificación posterior.

Análisis de datos

El análisis de los datos se realizó en dos etapas. En un primer momento, se analizó la precisión de las traducciones automáticas a través de un panel de expertos formado por profesores de inglés, traductores y especialistas en nutrición, todos ellos bilingües. Se contrastaron las tres versiones de GT y se registraron errores presentes en una o más de las tres versiones. Se consideraron como errores aquellos originados por la traducción automática y no los que surgieron de una redacción deficiente en el texto original. En esta etapa se trabajó con categorías tomadas de estudios previos que sirvieron como marco teórico para la presente investigación, tales como tecnicismos, construcciones en voz pasiva, siglas y nombres propios (Ablanedo et al., 2007, González Boluda, 2010, Groves y Mundt, 2015). Además, se agregaron otras categorías que se construyeron durante el análisis de los datos. Finalmente, las categorías de análisis fueron: errores totales, errores en texto original y errores de GT. Esta última se subdividió en sigla, tecnicismo, voz pasiva refleja, expresión de edad, sujeto tácito, pronombre, nombre propio, orden de palabras, preposición, uso de artículo, elipsis y significado implícito, y alteración sujeto-verbo.

En una segunda etapa de análisis de datos se cuantificaron los errores totales y su frecuencia.

Resultados

Durante la etapa de identificación de errores, se contrastaron las tres columnas de la matriz de traducciones. En casos de versiones diferentes, el panel de expertos se reunió para determinar si se trataba de maneras diferentes de decir lo mismo o de errores de traducción. La tabla 1 muestra algunos ejemplos de las diferencias entre las tres traducciones obtenidas. Se destacan en negritas las diferencias en cada caso.

Tabla 1: Diferencias entre traducciones obtenidas para la misma frase en diferentes dispositivos

Texto original	Traducción 1	Traducción 2	Traducción 3
posterior a la intervención se observó descenso en el consumo de sodio	after the intervention , there was a decrease in the sodium consumption	post-intervention was reduced in sodium consumption	after the intervention , there was a decrease in the sodium consumption
De este modo , las cantidades máximas de endulzantes no nutritivos que pueden consumir niños varones de 6 años de edad según su respectiva IDA es la siguiente:	In this way , the maximum amounts of non-nutritive sweeteners that can be consumed by 6-year-old boys according to their respective ADI is the following:	In this way , the maximum amounts of non-nutritive sweeteners that can be consumed by 6-year-old boys according to their respective ADI is the following:	Thus , the maximum amounts of non-nutritive sweeteners that can be consumed by male children 6 years of age according to their respective ADI is the following:
según las pautas de la OMS que considera a una población como de " riesgo leve " si su prevalencia está entre 5,0-19,9%.	according to WHO guidelines that consider a population as " low risk " if its prevalence is between 5.0-19.9 %	according to WHO guidelines that consider a population as " Slight risk " if its prevalence is between 5.0-19.9%	according to WHO guidelines that consider a population as " low risk " if its prevalence is between 5.0-19.9 %
mujeres y niños de tres países en vías de desarrollo .	women and children from three developing countries .	women and children from three countries in process of development .	women and children from three developing countries .
Los resultados no mostraron diferencias significativas	The results did not show significant differences	The results showed no significant differences	The results showed no significant differences

Fuente: elaboración propia

Una vez establecida la corrección o el error de traducción en cada oración, se procedió a su análisis por tipo.

Análisis cualitativo de errores y construcción de categorías

Para la identificación y clasificación de los tipos de error, se consideró por separado el error ocasionado por la traducción con GT y aquellos ocasionados por errores en la versión original dado que desde el panel de expertos se determinó que un error en el original posiblemente conduciría a un error de traducción. En estos casos se realizó una prueba de edición del texto original y se obtuvo una traducción correcta. Como muestra el siguiente ejemplo:

Versión original

La existencia de un exceso de aporte de sodio en la dieta en pacientes con enfermedad renal crónica (ERC) estadio 5D impacta en forma negativa sobre la presión arterial (PA), el volumen extracelular (VEC) y la ganancia de peso interdiálisis el cual se debería a limitaciones en el aprendizaje y el mantenimiento de un plan alimentario acorde a la situación clínica de cada paciente.

Traducción con GT

The existence of an excess of sodium intake in the diet in patients with chronic kidney disease (CKD) stage 5D negatively impacts on blood pressure (BP), extracellular volume (ECV) and interdialysis weight gain which is should to limitations in learning and maintenance of a food plan according to the clinical situation of each patient.

Versión sugerida (se agregó una coma)

La existencia de un exceso de aporte de sodio en la dieta en pacientes con enfermedad renal crónica (ERC) estadio 5D impacta en forma negativa sobre la presión arterial (PA), el volumen extracelular (VEC) y la ganancia de peso interdiálisis, el cual se debería a limitaciones en el aprendizaje y el mantenimiento de un plan alimentario acorde a la situación clínica de cada paciente.

Traducción con GT

The existence of an excess of sodium intake in the diet in patients with chronic kidney disease (CKD) stage 5D negatively impacts on blood pressure (BP), extracellular volume (ECV) and interdialysis weight gain, which it would be due to limitations in the learning and maintenance of a food plan according to the clinical situation of each patient.

Una vez identificados los errores que el panel de expertos atribuyó a la TAL, se clasificaron por tipo. En la tabla 2 (en la página siguiente), se presentan ejemplos de los errores analizados, teniendo en cuenta el tipo y la traducción ofrecida por GT.

Como se observa en la tabla 2, las siglas no fueron traducidas al inglés. Sin embargo, en algunos casos la TAL se realizó en forma correcta. Esto se observó en siglas de público conocimiento (por ejemplo, OMS, es registrada como WHO) o en aquellos casos en los que estas estaban acompañadas de la expresión completa.

Versión original

Los resultados del análisis de la composición química de las harinas se expresaron como promedios y desvío estándar (DE).

Traducción con GT

The results of the analysis of the chemical composition of the flours were expressed as averages and standard deviation (SD).

Versión original

La HTA es uno de los factores más importantes pronósticos de progresión de la ERC y conocido factor de riesgo cardiovascular.

Traducción con GT

Hypertension is one of the most important prognostic factors for the progression of CKD and a known cardiovascular risk factor.

Tabla 2: Ejemplos de errores más frecuentes

Tipo de error	Versión original	Traducción GT
Siglas	El control del VEC con restricción de la ingesta de sal, junto con diálisis adecuada, colabora con el control de la TA .	The control of the VEC with the restriction of the salt intake, together with adequate dialysis, the collaboration with the control of the TA .
Tecnicismo y regionalismo	Los ensayos realizados permitieron establecer las proporciones de los ingredientes necesarios para la elaboración de las prepizzas .	The tests carried out allowed to establish the proportions of the ingredients necessary for the preparation of the prepices .
Voz pasiva refleja	Se midió sin calzado ni objetos en la cabeza.	It was measured without shoes or objects on the head.
Expresión de edad	Resultados: se evaluaron 147 niños de 6-59 meses, 170 de 5-11 años , 70 de 12-14 años y 533 ≥ 15 años de edad, de ambos sexos.	Results: 147 children of 6-59 months, 170 of 5-11 years , 70 of 12-14 years and 533 ≥ 15 years of age , of both sexes, were evaluated.
Sujeto tácito	Correlaciona de forma negativa.	Correlates in a negative way.

Fuente: elaboración propia

Los tecnicismos y regionalismos presentan dificultades para la traducción, ya que se trata de vocablos que son de uso específico de ciertas áreas disciplinares o regiones, o bien términos de uso corriente que tienen un significado diferente dentro de un campo o ámbito profesional en particular. En la tabla 2 se incluye el ejemplo del término pre-pizza que, en la variedad de español utilizada en

Argentina, Chile y Uruguay, refiere a una masa plana ligeramente cocida, cubierta generalmente de tomate, que se usa como base para preparar una pizza (RAE, 2014). Otros ejemplos de este tipo fueron alfajor, evaluación sensorial, formulación, kiosco saludable, lactancia y obra social.

En cuanto a la voz pasiva refleja, el problema radica en que se trata de una estructura propia del español que no existe en inglés. En la tabla 2, el ejemplo se midió resulta en una traducción incorrecta, además, por la omisión de la información referida a quienes fueron medidos. Estas omisiones o elipsis, propias de la cohesión de los textos, resultan en errores de la TAL. Algo similar ocurre con el sujeto tácito, siendo que en español el sujeto puede ser omitido en varias estructuras, mientras que en inglés no es tan usual.

La referencia a la edad es también un error frecuente. Se observó inconsistencia en el uso de traducciones como *years*, *years of age*, entre otras. Sumado a esto, la palabra años puede referir tanto a años de vida, como a un período de tiempo. Este tipo de traducción no resultó correcta en todos los casos.

Por último, en ambigüedad léxica se incluyen aquellos verbos que en español se dicen de igual manera en tiempo pasado y presente, como por ejemplo utilizamos.

Análisis cuantitativo y frecuencia

El segundo nivel de análisis consistió en cuantificar errores por texto, oraciones y palabras, de lo cual se obtuvo una media de 54 errores por artículo. Respecto a la determinación de errores por oraciones y palabras, se registró un error cada 2,5 oraciones o cada 64,8 palabras. Cabe señalar que estos errores fueron cuantificados cada vez que aparecieron, sin discriminar aquellos que se repitieron a lo largo del texto. Esta decisión se sustentó en que la repetición es uno de los rasgos de la cohesión de los textos y, por lo tanto, en una traducción automática, inevitablemente se repetirá cada error tantas veces como aparezca en el texto original.

La tabla 3 (en la página siguiente) muestra la cantidad de oraciones y palabras analizadas en cada artículo y los resultados clasificados por tipo de error. Como puede observarse, los errores más frecuentes fueron la traducción de siglas (106) y los tecnicismos (81), en comparación con los demás tipos de errores (14 o menos). Esa cantidad puede deberse a que es habitual que tanto las siglas como los términos propios de cada disciplina se repitan en el texto.

Tabla 3: Cantidad y tipo de error por texto

	Texto 1	Texto 2	Texto 3	Texto 4	Texto 5	Texto 6	Tota 1
Cantidad de oraciones	108	209	86	174	95	141	813
Cantidad de palabras	2686	4282	2170	4806	2589	4469	2100 2
Errores totales	49	83	36	46	41	69	324
- Errores en texto original	4	10	4	12	9	12	51
- Errores de GT	45	73	33	34	32	57	274
- Sigla	17	35	3	3	12	36	106
- Tecnicismo	4	21	19	14	10	13	81
- Voz pasiva refleja	3	3	2	5	1	0	14
- Expresión de edad	11	1	0	0	0	0	12
- Sujeto tácito	0	7	1	2	0	3	13
- Pronombre	0	0	2	4	1	2	9
- Nombre propio	5	1	0	0	5	0	11
- Orden de palabras	2	1	2	0	0	1	6
- Preposición	0	2	2	1	0	1	6
- Uso de artículo	2	0	0	3	1	0	6
- Elipsis y significados implícitos	1	0	1	2	0	1	5
- Alteración sujeto-verbo	0	2	0	0	2	0	4

Fuente: elaboración propia

Discusión y conclusión

La lengua inglesa es, desde hace mucho tiempo, el idioma que se utiliza universalmente para comunicar contenido académico, por lo que resulta imprescindible trabajar con este si se desea tener relevancia en distintos campos, particularmente en el científico.

Gracias a los motores de TAL, como lo es GT, la ardua tarea de comprender y redactar en otro idioma se ve facilitada. A pesar de tratarse de un estudio exploratorio-descriptivo y con un muestra reducida, fue posible registrar los errores más comunes que pueden ser encontrados en GT cuando es utilizado para la traducción de textos en español de un área o disciplina específica, como en este caso la nutrición, entre los que se destacan principalmente el uso de siglas y tecnicismos, seguidos luego y en menor medida por errores en voz pasiva refleja, sujeto tácito y expresión de edad.

Es importante señalar que en distintas ocasiones GT pudo resolver correctamente errores de redacción, siglas y otras faltas comunes mencionadas en el presente trabajo. Dado que el propósito del estudio fue identificar errores, no se centró en generar traducciones correctas aún en caso de

versiones originales erróneas.

A juzgar por experiencias preliminares a este estudio y en línea con Groves y Mundt (2015), es de esperar que la calidad de GT mejore con el tiempo, a medida que aumenta la base de datos.

Se puede concluir entonces que, para lograr una traducción adecuada, no es suficiente copiar y pegar un texto a un TAL, sino que es necesario un trabajo previo: es de suma importancia que el escrito sea redactado con lenguaje apropiado en el idioma original, considerando todas las reglas gramaticales, incluyendo coherencia y cohesión; a su vez, se deben observar los términos específicos y siglas que puedan ser traducidos incorrectamente, o con un equivalente poco adecuado al contexto. Se considera que una mejora en la redacción no ocasionaría el mismo error de traducción.

Entre las limitaciones del estudio se destaca el número reducido de artículos de una sola revista y, por ende, la poca variación de temáticas abordadas aún dentro del campo específico de la nutrición. Además, como GT se actualiza con el uso, es posible que los resultados de esta investigación cambien en el futuro. No obstante, el estudio permitió la construcción de una matriz de análisis de tipos de errores más comunes para su tratamiento en futuros estudios, a la vez que registró evidencia empírica de errores de TAL en el área de la nutrición que puede utilizarse para la creación de un manual de guía en este ámbito.

Referencias bibliográficas

- ABLANEDO, J., AIKEN, M., y VANJANI, M. (2007). Efficacy of English to Spanish automatic translation. *International Journal Information and Operations Management Education*, 2(2), 194-210. doi: <https://doi.org/10.1504/IJIOME.2007.015284>
- AIKEN, M., GHOSH, K., WEE, J., y VANJANI, M. (2009). An Evaluation of the Accuracy of Online Translation Systems. *Communications of the IIMA*, 9(4), 66-84. <https://scholarworks.lib.csusb.edu/ciima/vol9/iss4/6>
- BARRACLOUGH, F. (2009). La lingua franca de la ciencia. *Metode Science Studies Journal*. <https://metode.es/revistas-metode/article-revistas/la-lingua-franca-de-la-ciencia-2.html>
- CLERICI, C. (2016). Traducción y pronunciación con TIC: Una experiencia de uso de traductores en línea y simuladores de voz en Inglés. *Virtualidad, Educación y Ciencia*, 7(13), 102-109. <https://revistas.unc.edu.ar/index.php/vesc/article/view/16209>
- CRACIUNESCU, O., GERDING-SALAS, C., y STRINGER-O'KEEFFE, S. (2007). Machine translation and computer-assisted translation: a new way of translating? *Translation Journal*, 8, <https://api.semanticscholar.org/CorpusID:2481868>
- DOMÍNGUEZ, M. B., LAURENTI, L., y AGUIRRE CÉLIZ, C. (2013). Google Translate: una experiencia con alumnos de inglés técnico en el nivel superior. *Virtualidad, Educación y Ciencia*, 6(4), 44-53. <https://revistas.unc.edu.ar/index.php/vesc/article/view/4646>
- GHASEMI, H., y HASHEMIAN, M. (2016). A Comparative Study of "Google Translate" Translations: An Error Analysis of English-to-Persian and Persian-to-English Translations. *English Language Teaching*, 9(3), 13-17. <http://dx.doi.org/10.5539/elt.v9n3p13>
- GONZÁLEZ BOLUDA, M. (2010). Estudio Comparativo de Traductores Automáticos en Línea: Systran, Reverso y Google. *Núcleo*, 27, 187-216. http://saber.ucv.ve/ojs/index.php/rev_n/article/view/2565

GROVES, M. y MUNDT, K. (2015). Friend or foe? Google Translate in language for academic purposes. *English for Specific Purposes*, 37, 112–121. <https://doi.org/10.1016/j.esp.2014.09.001>

KARAMI, O. (2014). The brief view on Google Translate machine. Artículo presentado en el Seminario de Inteligencia Artificial sobre Lengua Natural, Alemania. <https://pdfs.semanticscholar.org/c6f9/5d543c0b34c4026b9e6cf64decd94b793823.pdf>

REAL ACADEMIA ESPAÑOLA [RAE] (2014). *Diccionario de la lengua española* (23ª edición). Espasa.