

**La traducción automática:
Una comparación entre las traducciones de
Google Translate y DeepL del español al
holandés**

Trabajo Fin de Máster

Universidad de Utrecht

agosto del 2019

Nombre de estudiante: Li-Ann Croes

Número de estudiante: 5913128

Supervisora: Dr. D. Nieuwenhuijsen

Segunda lectora: S. Rosano Ochoa

Código de tesina: TLMV16205

Índice

Índice	2
Resumen	3
Introducción. Problemática y relevancia	4
Capítulo 1: La traducción automática	6
Capítulo 2: Problemas y limitaciones de la TA	14
Capítulo 3: Traducir mediante la TA y evaluaciones de la TA.	19
Capítulo 4: La metodología de la investigación	24
Capítulo 5: El análisis de la investigación	26
Capítulo 6: Conclusión	49
Capítulo 7: Discusión	51
Referencias bibliográficas	52
Anexo	54

Resumen

En el presente trabajo se investiga la calidad de las traducciones realizadas por Google Translate y DeepL. La traducción automática es algo que empezó a investigarse hace mucho tiempo, sin embargo, muchos no creían que podía ser algo exitoso. Hoy en día se conocen múltiples máquinas traductoras que han hecho y siguen haciendo posible que haya información disponible en diferentes idiomas.

Las traducciones realizadas mediante máquinas traductoras siguen siendo algo interesante y siguen conociendo avances y mejoras. Es por eso que en esta tesina tenemos como meta investigar a dos de ellos. Google Translate, que ha ganado mucha popularidad y que sigue ganando popularidad y DeepL, que traduce a una cantidad relativamente baja de idiomas pero que también se ha vuelto muy popular a lo largo del tiempo. Lo que muchos quieren lograr desde hace tiempo con una máquina de TA es que se realicen traducciones automáticas de alta calidad (FAHQT), pero como vamos a ver en esta investigación esto es casi imposible. Hay que tener en cuenta muchos factores por los cuales una máquina de TA no puede llegar a una traducción automática de alta calidad o mejor dicho de la calidad de un traductor humano. Es aquí en donde un traductor humano y un traductor automático llegan a trabajar conjuntamente para producir un texto de alta calidad.

Como veremos, dependiendo del tipo de texto logramos conseguir una traducción de buena o mala calidad. Esto es lo que investigamos en esta tesina y con la ayuda de expertos logramos analizar las traducciones propuestas por las dos máquinas anteriormente mencionadas.

Palabras claves: traducción automática, FAHQT, tipología textual.

Introducción. Problemática y relevancia

La traducción automática ha sido desde hace mucho tiempo un sueño de muchos que a partir del siglo XX se ha ido concretizando. (Hutchins, 2003; Cherágui, 2012). Fueron diferentes los motivos por los cuales hoy conocemos la traducción automática como la conocemos. Lo que empezó como un intento para crear un diccionario mecánico hoy nos lleva a un mundo en donde es posible conseguir traducciones de textos al instante.

Este fenómeno conoce una larga historia y muchos avances, los cuales hoy nos llevan a conocer diferentes máquinas traductoras, de las cuales algunas se pueden usar gratuitamente, a pesar de que no siempre se pueda disfrutar de la mejor calidad de output.

Durante el Máster de Traducción, se ha tratado este tema y personalmente a mi siempre me ha fascinado el hecho de que dispongamos de medios que nos permitan tener acceso a información en la lengua que queramos, por ejemplo, mediante Google Translate.

Google Translate es una herramienta conocida y usada por muchas personas. Es gratis y los textos, frases o palabras se traducen al instante. Aparte de Google Translate, que traduce a más o menos 130 lenguas, existen muchas otras máquinas traductoras como por ejemplo DeepL, que tan solo traduce a 9 lenguas.

DeepL es una herramienta de traducción que poco a poco ha ido ganando popularidad. Es utilizada por mucha gente y generalmente se dice que el output es sorprendente. Considerando que en el pasado ya se había investigado Google Translate, y que DeepL es relativamente nueva, quiero investigar cómo funcionan ambas herramientas y luego quiero ponerlas a prueba a través de diferentes géneros de textos.

Esta investigación resulta relevante ya que se está usando la traducción automática cada vez más. El internet ofrece muchas posibilidades para traducir textos en muy poco tiempo, y en innumerables lenguas. Se basan en un banco de datos muy grande para así poder realizar traducciones.

Por un lado, se puede considerar el hecho de que se pueda contar con una máquina traductora en el mundo de la traducción como un avance, pero por otro lado la traducción automática tiene sus problemas. Por ejemplo, se puede pensar en el hecho de que una máquina traductora no posee el mismo conocimiento que un ser humano. Es una máquina. Por este motivo el tema de mi tesina se centra en la traducción automática.

La pregunta principal de esta tesina es:

¿Cómo se diferencian las traducciones producidas por los traductores automáticos Google Translate y DeepL? Una comparación entre dos herramientas de traducciones automáticas que traducen del español al neerlandés.

Para contestar mi pregunta principal, primero nos enfocaremos en la historia de la traducción automática para así poder entender cómo funciona y cómo hemos llegado hasta donde la traducción automática hoy día nos permite traducir, con sus límites obviamente. En la segunda parte explicaremos la metodología empleada para investigar la calidad de las traducciones automáticas propuestas por Google Translate y DeepL. Aquí nos basamos en el conocimiento y criterio de expertos traductores, que se mueven este ámbito como profesionales. Por el hecho de ser profesionales podemos confiar en sus criterios de evaluar una traducción producida automáticamente como mala o buena. En la tercera parte analizaremos los resultados obtenidos a partir de la investigación, para poder llegar a nuestra respuesta a la pregunta principal de esta tesina.

Capítulo 1 La traducción automática

En el primer capítulo explicaremos brevemente qué es la traducción automática y se hará un resumen global de la historia de la traducción automática. Para entender cómo se llevan a cabo las traducciones automáticas, le prestaremos atención a los diferentes sistemas de traducciones en los siguientes apartados de este capítulo. Consideramos importante destacar estos datos, para poder formar una idea clara de lo que es la traducción automática y cómo funciona.

1.1 ¿Qué es la traducción automática?

En este apartado discutiremos “la traducción automática” como término. La traducción automática, conocida también como TA o MT (del inglés; machine translation), se puede describir como la traducción de un texto mediante un programa en un ordenador. Es decir, es una herramienta con el que se traduce un texto desde un lenguaje fuente a un lenguaje meta sin la ayuda de un traductor humano (Forcada, 2010).

En concreto se puede decir que en la traducción automática se puede distinguir entre dos enfoques principales: el enfoque basado en normas y el enfoque basado en un corpus.

El enfoque basado en normas

El enfoque basado en normas se puede dividir en tres enfoques tradicionales. Estos enfoques tradicionales se conocen como: el sistema de traducción directa, el sistema de interlingua y el sistema de transferencia.

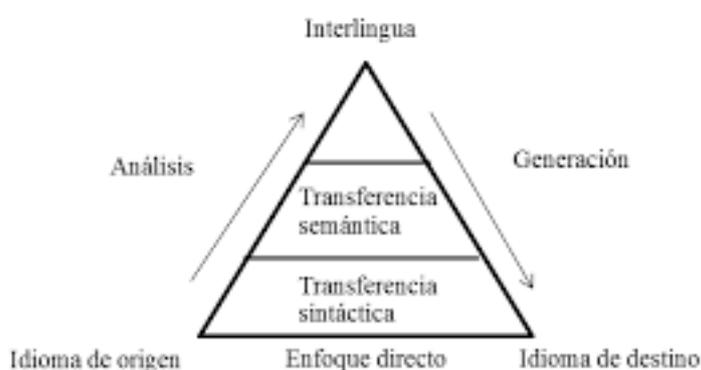


Figura 1: Sistemas de la traducción automática basado en normas (Hutchins, 2009; Cherágui, 2009)

El sistema de traducción directa consiste en la traducción inmediata del texto fuente al texto meta. En este sistema se analiza el vocabulario del texto fuente para saber que vocabulario emplear en el texto meta, y para poder resolver ambigüedades como también para identificar expresiones correctamente en el texto meta. En la

traducción directa se traduce mediante un sistema específico del orden de palabras en el texto meta. Normalmente este sistema funciona eliminando las variaciones morfológicas para obtener las raíces de las palabras y así buscarlas en diccionarios bilingües. En este sistema se emplean grandes diccionarios bilingües y un programa específico para analizar el léxico y la morfología para producir traducciones.

El sistema de interlingua consiste en dos partes: análisis y generación. Durante el análisis se transforma el texto fuente al sistema de interlingua y en el proceso de la generación se pasa del sistema de interlingua al texto meta. El sistema de interlingua se convierte en la primera fase del proceso, Se traduce el texto fuente a un sistema interlingua, es decir un 'lenguaje neutral' que es un lenguaje independiente de cualquier lenguaje fuente o lenguaje meta. El texto fuente se crea en la segunda fase en base a la interlingua.

El sistema de transferencia consiste en un proceso de tres etapas. En la primera etapa (análisis) se analiza el texto basándose en un conjunto de reglas lingüísticas. A partir de este análisis se definen las estructuras del lenguaje fuente. La segunda etapa es la transferencia, que consiste en convertir las estructuras del lenguaje fuente, como anteriormente mencionado, en estructuras del lenguaje meta. En la tercera y última etapa, llamada "generación", se genera una traducción a partir de las estructuras del lenguaje meta que fueron creadas en la segunda etapa.

La traducción automática basado en corpus

La traducción automática basada en corpus se puede dividir en dos enfoques: la traducción automática basada en ejemplos y la traducción automática basada en estadística. Es importante remarcar que estos dos enfoques se basan en la disponibilidad de textos paralelos (Carl & Way, 2003).

La traducción automática basada en ejemplos se conoce en inglés como Example-based machine translation (EBMT) y consiste en diferentes etapas. La idea detrás de este enfoque es básicamente reutilizar los resultados, es decir recuperar expresiones o frases que ya han sido traducidos anteriormente (Forcada, 2010).

Como dicho anteriormente este enfoque consiste en tres etapas: encontrar correspondencias, alinear, y recombinar. (Hutchins: 2006, Tertoolen: 2010, Sánchez-Martinez: 2012). A partir de textos paralelos, el sistema analiza la frase que se tiene que traducir, luego la divide en pequeños segmentos cuyas traducciones están presentes en textos paralelos. En la fase final del proceso estos textos encontrados son combinados para producir una nueva traducción.

La traducción automática basada en estadística se basa en la generación de traducciones usando métodos estadísticos. Las traducciones son generadas de la siguiente forma: se alinean las frases, grupos de palabras y palabras individuales de textos paralelos. De estos textos se calculan las probabilidades de que una frase del

lenguaje A corresponda a una palabra en una frase en el lenguaje B, en este caso el lenguaje A y B están alineadas. La tarea es en este caso encontrar una frase en el lenguaje meta (lenguaje B) que tenga la mayor probabilidad de ser la traducción correcta. Este enfoque requiere una gran base de datos (Koehn, 2009).

1.2 La historia de la traducción automática

La historia de la traducción automática según Hutchins (2009) data del siglo XVII. En esos tiempos se basaba en la idea de la creación de diccionarios mecánicos universales. No fue hasta el siglo XX cuando se estima que se hicieron las primeras sugerencias. En el 1933 se crearon dos patentes, una en Francia y una en Rusia por Georges Artsrouni y Trojanskij respectivamente. La patente de Artrouni era inicialmente una máquina para uso general de un diccionario mecánico que también podía funcionar como un diccionario mecánico multilingüe. Trojanskij presentó una patente, que fue considerada la más significativa. Básicamente podía ser un diccionario mecánico, solo que incluía propuestas para interpretar y codificar funciones gramaticales usando símbolos universales (partiendo del esperanto) en una máquina traductora multilingüe.

Andrew Booth y Warren Weaver (ingeniero eléctrico, físico y científico que construyó alguno de los primeros ordenadores del Reino Unido e investigador y director del Rockefeller Foundation respectivamente) desconocían estas ideas de diccionarios mecánicos. Fue en el año 1946 cuando se conocieron y siguieron desarrollando en el 1947 la idea de usar ordenadores para traducir lenguas naturales. Luego en el 1948 trabajaron tanto Booth como Richard Richens (director del "Commonwealth Bureau of Plant Breeding and Genetics" de la Cambridge University) en un análisis morfológico para un diccionario mecánico.

En este tiempo ya a muchos les había surgido la idea de la traducción mecánica y en el 1949 fue específicamente Warren Weaver (director del Rockefeller Foundation), quien propuso ideas para solucionar los problemas de ambigüedad, basándose en su conocimiento sobre la criptografía, estadística, teoría informática, lógica y lenguajes universales. Esto influyó en la iniciación de investigaciones sobre la traducción automática en los Estados Unidos. Después de algunos años el filósofo, matemático y lingüista Yehoshua Bar-Hillel fue asignado para hacer investigaciones en el Massachusetts Institute of Technology (MIT).

Se propusieron muchas ideas y sugerencias para el futuro y tuvo lugar la primera demostración de un sistema de TA en el año 1954. A pesar de que la demostración tenía poco valor científico, los resultados eran lo suficientemente sorprendentes para lograr que se invirtiera en una investigación de la TA en los Estados Unidos. Al mismo tiempo se iniciaron proyectos de TA en otras partes del mundo, sobre todo en la Unión Soviética. Viendo la calidad de los resultados, quedó desde ese tiempo claro que con la automatización de la traducción no se iba a lograr una traducción

de alta calidad y que era necesario la intervención humana ya sea antes o después de que el ordenador hiciera su parte, es decir en la pre-edición y/o post-edición. Muchas personas tenían la esperanza de que iba a ser una medida provisional mientras que la mayoría esperaba que la intervención humana siempre fuera a ser algo necesaria.

1956-1966: De optimismos y desilusiones

Cuando iniciaron las investigaciones de la TA, había poca ayuda de parte de lingüistas. Por consiguiente, las investigaciones entre los años 50 y 60 se polarizaron entre, por un lado por, investigaciones empíricas en donde se aplicaba métodos estadísticos para el descubrimiento de regularidades gramaticales y léxicas que se podían aplicar computacionalmente y, por otro lado, enfoques teóricos que constaban de investigaciones de lingüística, lo que hoy conocemos como lingüística computacional.

Estos métodos contrastivos, fueron descritos como 'la fuerza bruta' y 'el perfeccionismo' respectivamente. La primera tenía como meta desarrollar un sistema que produjera traducciones que se podían usar en el futuro, aunque eran 'crudas', es decir no en su formato final, y el segundo se enfocaba en desarrollar un sistema en donde se produjeran traducciones sin la necesidad de ayuda humana o con la más mínima ayuda humana posible.

Los empiristas escogieron el sistema de traducción directa, usando la mayoría de las veces análisis estadísticos de textos para así poder derivar reglas del diccionario, con poco o ninguna base teórica. Los perfeccionistas en cambio se guiaban por la teoría, y prestaban atención sobre todo a métodos de análisis sintácticos. Otros grupos apoyaban el sistema de traducción interlingual y creían que solamente mediante investigaciones en el proceso del pensamiento humano (que más tarde se conoció como inteligencia artificial) se solucionarían los problemas de la traducción automática. El grupo de los pragmáticos se concentraba en un modelo más simple basado en el método de transfer sintáctico, dejando los problemas pragmáticos para ser resueltos en otra etapa de la traducción. Pues en este tiempo se empezaron a usar los diferentes sistemas de traducción (véase §1.1, pagina 6).

En este tiempo la gran mayoría de investigaciones del EE. UU. se basaban en las traducciones del ruso al inglés, y la mayoría de las investigaciones de la Unión Soviética estaban enfocadas en sistemas inglés-ruso, por motivos políticos y militares.

El informe ALPAC y sus consecuencias

A partir de una encuesta muy influyente, Bar-Hillel expresó su crítica en el año 1960 en donde opinaba sobre las investigaciones hacia la traducción automática que se supone que deberían crear productos FAHQ (fully automatic high quality

translation), es decir traducciones automáticas de alta calidad. Según Bar-Hillel (Hutchins, 2009) no solo era algo idealista sino imposible de lograr, considerando el estatus del conocimiento lingüístico y los sistemas computarizados que existían en ese tiempo. A pesar de que su fuerte comentario hacia la traducción automática tenía mucho peso en el ámbito, Hutchins (2009, p. 5) comenta que después, con los desarrollos de la inteligencia artificial que se basaba en TA a partir de una base de conocimientos, ha sido demostrado que su comentario pesimista hacia las máquinas traductoras no era justificable del todo. La idea de obtener una MT (máquina traductora) de buena calidad estaba disminuyendo y en el 1964 acuden los que invirtieron dinero en la TA en los EE. UU. (mayormente instancias militares y de inteligencia) a la National Science Foundation, para instalar un comité ALPAC (Automatic Language Processing Advisory Committee) para que examinaran la situación.

Después de analizar la situación, salió según comenta (Hutchins, 2009), el famoso ALPAC-report, en donde se concluyó que la TA es lenta, menos preciso y cuesta dos veces más que una traductora humana. También llegaron a la conclusión de que no hay ningún indicio en el futuro cercano o no es predecible que exista una máquina traductora útil. De ahí que no siguieron invirtiendo en investigaciones de TA. Aunque muchos condenaron el reporte ALPAC de ser corto de vista y parcial, su influencia fue notable, tanto, como que ocasionó el fin de investigaciones en el campo de la TA en los EE. UU. por más de una década, lo que en su lugar ocasionó que ya no se investigara más en otras partes del mundo.

La década silenciosa (1967-1976)

Aunque denominada la década silenciosa, las investigaciones no pararon completamente. Por ejemplo, en los EE. UU. siguieron las investigaciones durante algunos años. Además, se produjo un cambio de dirección en el sistema que se empleaba para traducir. Mientras que entre los de la primera generación de investigadores, los de la época pre-ALPAC (1956-1966) dominaba mayoritariamente el sistema de traducción directa, entre los investigadores de la segunda generación, post-ALPAC, dominaba el método 'indirecto', que son tanto los métodos de traducción interlingual como métodos de transferencia.

En los EE. UU se concentraban en la traducción al inglés de textos científicos y técnicos escritos en ruso, mientras que en Canadá y Europa las necesidades eran un poco diferentes. Con la política bicultural del gobierno canadiense, hubo una demanda de información que fuera disponible tanto en inglés como en francés. Algo similar pasaba en ese tiempo en Europa, con la demanda creciente de traducciones de textos científicos, técnicos, administrativos y legales de y a las diferentes lenguas dentro de la comunidad europea. En el año 1970 se inició una investigación basada en un sistema de transferencia sintáctica para la traducción del par de lenguas inglés- francés. En el proyecto TAUM (Traduction Automatique de l'Université de Montréal) se han creado los sistemas Q-system y Météo, con la que se traducen

boletines meteorológicos. Ha sido diseñado con un vocabulario estricto y unas reglas sintácticas restringidas que sólo se utilizaba en boletines meteorológicos. A mediados de los años '70 el futuro del sistema interlingual era incierto. Los problemas principales se debían a la rigidez de los niveles de análisis, y particularmente la pérdida de información del texto fuente. Es por esto por lo que a muchos les parecía que en ese tiempo el sistema menos ambicioso, 'transferencia', ofrecía mejores soluciones.

Sistemas comerciales (1976-1989)

Durante la década después del informe ALPAC, llegaron a existir otros sistemas que funcionaban a partir de un sublenguaje: TITUS (1970) y CULT (1972). Titus fue introducido por el Institut Textile de France y era un sistema multilingüe para traducir resúmenes escritos en un lenguaje controlado. CULT, de la Universidad china de Hong Kong se diseñó específicamente para traducir textos de matemáticas del chino al inglés.

Lo más significativo de esta época fue el sistema SYSTRAN, que fue instalado en el 1970 y que traducía entre el ruso y el inglés. En el transcurso de los años SYSTRAN ha sido instalado en numerables institutos gubernamentales. También Xerox utilizó el SYSTRAN, eliminando la parte de post-edición y controlando el vocabulario y las estructuras para los manuales técnicos traducidos del inglés al francés, alemán, italiano, español, portugués etc.

Desde los principios del 1980 el rival de SYSTRAN era el sistema LOGOS Corporation, un sistema vietnamita-inglesa creado por Bernard Scott. Este sistema se siguió desarrollando con otros pares de lenguas como por ejemplo alemán-inglés.

A finales de 1980 apareció METAL, del par alemán-ingles, que en un principio utilizaba el sistema pero que después optó por cambiar al sistema de transferencia. Otros sistemas creados fueron los de el Pan American Health Organisation, un sistema del español al inglés (SPANAM) y otro del inglés al español (ENGSPAN). Durante el 1980 hubo mucha actividad comercial en Japón, en donde muchas empresas que se dedicaban a fabricar ordenadores crearon un sistema para traducciones automáticas primordialmente destinados al par de lenguas del mercado japonés-ingles y ingles-japonés, aunque no ignoraron la necesidad de traducciones de y al coreano, chino y otras lenguas. Lo más importante de esta época fue el proyecto Grenoble, cuyo objetivo era crear un sistema de transferencia basada en la lingüística. Según el artículo de Hutchins (2009), uno de los proyectos más conocido es el Eurotra, de las comunidades europeas. Su meta era crear un sistema de transferencia multilingüe para la traducción entre los países de la Comunidad Europea. Otros proyectos que han sido realizados partiendo del interés por un sistema interlingual, y motivado en parte por investigaciones de inteligencia artificial y lingüística cognitiva, fueron el sistema DLT (Distributed Language

Translation) en Utrecht y el proyecto Rosetta. Con el sistema DLT proponían crear un sistema multilingüe interactivo basado en redes informáticas, en donde cada red solo traduce de una sola lengua a otra lengua. El proyecto Rosetta, cuyo enfoque era la reversibilidad de las gramáticas, es decir, el conjunto de reglas gramaticales que funcionaba en una dirección para el análisis sintáctico y semántico funcionaba, funcionaba en otra dirección para la generación de frases correctas de una lengua. (Hutchins, 2005)

Desde el año 1989

Las investigaciones de la traducción automática hasta finales del 1980 básicamente se basaban en reglas lingüísticas de diferentes tipos, ya sea reglas de análisis sintácticas, reglas para la transferencia del léxico y reglas de morfología. Formaba la base para muchos sistemas interlinguales, sin embargo, desde el 1989 las investigaciones basadas en reglas pasaron a un segundo plano, mientras que los métodos basados en corpus cobraban popularidad y se volvían importantes.

Enfoques basados en corpus:

Este fue el desarrollo más dramático, la reactivación de los enfoques basados en estadísticas. Se consideraba como un retorno al 'empirismo' de la primera década y al mismo tiempo como un desafío a lo que predominaba en ese momento, el 'racionalismo' de los 1970 y 1980. Este método funcionaba de la siguiente forma: alinear frases, grupos de palabras y palabras de textos paralelos. Luego se calculaba el resultado y se alineaban los textos siguiendo una traducción de palabra por palabra. Lo más sorprendente para los investigadores fueron los resultados, casi la mitad de las frases traducidas o coincidían exactamente con la traducción en el corpus o expresaba lo mismo, pero, en otras palabras.

El segundo enfoque basado en corpus es la traducción automática basada en ejemplos. Para que se pueda traducir mediante este sistema es importante contar con una base de datos de corpus textuales. Este enfoque consiste en que se busquen ejemplos de traducciones de una expresión en particular. Se basa en un método de procesos de extraer y seleccionar frases equivalentes provenientes de un banco de datos de textos bilingües paralelos. (Hutchins, 2005; Forcada, 2005) Aunque lo más nuevo desde el 1990 haya sido el crecimiento del enfoque basado en corpus, el enfoque basado en reglas siguió estando presente en sistemas de traducciones de transferencia y de interlingua.

TA y el internet

Desde mediados de los años 1990, la evolución de la traducción automática se ha visto fuertemente influenciado por el internet. Al principio se pudo ver la aparición de sistemas de traducción automática para traducir páginas de internet y correos. Luego, ha habido mucha demanda para el servicio de traducciones inmediatas. El pionero en este tiempo en ofrecer este servicio era SYSTRAN. Después de un

tiempo, se lanzó Babelfish, que traducía francés, alemán, español al inglés y del inglés como también muchas otras lenguas.

La calidad de las máquinas traductoras de internet es muchas veces de mala calidad, dado a que en una lengua existe por ejemplo el lenguaje coloquial incorporado muchas veces en los textos. Sin embargo, responde a la demanda de tener un servicio de traducción inmediata en el lenguaje del usuario.

Según Hutchins (2009) la traducción automática sigue teniendo desafíos respecto al uso de sistemas para traducir a lenguas que no dominan los usuarios muy bien.

Pues en el internet el lenguaje usado es coloquial, contiene muchas frases agramaticales, con abreviaturas, chistes etc. Esto se ve sobre todo en correos plataformas digitales etc. Este empleo del lenguaje es totalmente diferente del lenguaje empleado en textos técnicos y científicos, lo que originalmente era la intención por el cual crearon sistemas de traducción automática al principio del siglo XX.

1.3 Conclusión

La traducción automática conoce una larga trayectoria de investigaciones y evoluciones. A pesar de que al principio la traducción automática no tenía mucha popularidad ni tampoco muchas personas que confiaban en el proyecto, supo sorprender a muchos gracias a los resultados que obtuvieron los investigadores. Esto conllevó a que se empezara a invertir dinero en este proyecto. Así pasó el tiempo, con altibajos, con muchos intentos de crear cada vez algo nuevo que mejorara la calidad de las traducciones. Sin embargo, se puede concluir que por sorprendentes que hayan sido los resultados, lograr una traducción automática perfecta ha sido hasta el momento imposible. Esto se ve en los resultados de por ejemplo las traducciones usando sistemas basadas en reglas, en ejemplos o en estadísticas no han llevado a una traducción perfecta. A partir de estos sistemas de traducción, se hizo paso para lo que hoy conocemos como “sistemas híbridos”, que muchas veces consiste en una combinación de mas de un sistema para traducir.

Capítulo 2: Problemas y limitaciones de la TA

Como se ha explicado anteriormente, los sistemas de TA se han creado basándose en una serie de enfoques o sistemas creado por lingüistas y otros expertos en el ámbito, donde se preparan una serie de reglas que el sistema de TA tiene que seguir y donde el sistema funciona basándose en una teoría. Es decir que, lo que no está descrito en las reglas, difícilmente podrá ser ejecutado por un sistema de TA.

Esto puede ocasionar que no todo sea traducido correctamente, algo que hoy día aún podemos ver a pesar de los avances tecnológicos y de conocimiento. Es por esto por lo que en este capítulo le dedicaremos atención a los problemas y las limitaciones que pueden haber dentro de la TA. Nos basaremos en Mogas Ribas (2016), Alarcón Navío (2003), Krauwer (2003) y Arnold (2003).

2.1 Problemas y limitaciones de la TA

En su artículo, Arnold (2003) explica que la traducción es algo que en si es difícil hasta para los traductores humanos. Según él, la traducción es una habilidad que va mucho más allá de solo ser competente en dos lenguas. Él comenta que es una tarea difícil, considerando que no todo puede traducirse exactamente igual. Hay lenguas en donde existen palabras que no tienen un equivalente en otro idioma. Otro punto que destaca Arnold (2003) es que traducir es un trabajo creativo. Se espera de un traductor que sea capaz de inventar nuevos términos para términos que aparecen en el texto fuente y que en realidad no existen en el lenguaje meta. Luego se espera de un traductor que sepa actuar como un intermediario de la cultura.

Los ordenadores son máquinas que solo siguen reglas, con cierta rapidez y precisión. Como comentamos anteriormente, a pesar de que la TA haya sido investigado durante muchos años y que se sigue investigado, aún existen problemas. Arnold (2003) menciona los siguientes problemas con relación a la incapacidad de los ordenadores:

- Realizar tareas no específicas

Para que un ordenador pueda traducir de un lenguaje a otro, es necesario que este ordenador tenga reglas específicas. El problema radica en que no siempre se puede encontrar con una regla específica precisa, siempre hay excepciones. Este es el caso no solamente para una máquina en TA, pero también en el lenguaje natural. Aquí lo problemático es que una regla dicte por ejemplo que: “*generalmente* el sujeto precede al verbo” (Arnold, 2003). Esto significa que esto no siempre es así, y este es justamente el problema. Un ordenador necesita que las reglas sean precisas.

- Aprender cosas nuevas:

aprender cosas nuevas es difícil para un ordenador y esto tiene muchos motivos. Uno de los motivos es que para poder aprender cosas nuevas uno a veces tiene que ser creativo. Ser creativo en situaciones que quizás requieran no seguir reglas. Esto es algo que para un ordenador es difícil ya que funciona usando reglas que le ha sido impuesto por humanos.

- Realizar razonamiento común:

Arnold (2003) menciona explícitamente que ordenadores no pueden realizar razonamiento común. Para poder realizar razonamiento común es necesario poseer información sobre hechos en el mundo. Por ejemplo: el agua es húmeda, los hombres no pueden embarazarse, la mayoría de las personas tienen dos pies etc. Todo lo que entendemos por razonamiento común es difícil de entender para un ordenador.

- Solucionar problemas cuando existe un gran número de soluciones posibles:

La cuarta problema que considera Arnold (2003) en cuanto a ordenadores es la incapacidad de enfrentarse a una gran cantidad de soluciones, a pesar de tener los problemas de traducción especificados. Arnold se refiere a este problema como "*the problem of aerial explosion*" (2003: 122). El TA debe escoger una solución entre esas infinitas posibilidades y allí es donde falla.

Arnold (2003) y Krauwer (2003) consideran tres dificultades principales cuando se trata de traducir mediante una máquina TA: la robustez, la ambigüedad y la diferencia entre lenguas.

Según Arnold (2003) cuando se trata de la robustez de un traductor automático, el traductor automático recibe un texto fuente que muchas veces contiene frases que no están correctamente escritas, según las reglas para analizarlas automáticamente. Con esto podemos pensar en por ejemplo la capacidad de reconocer palabras mal escritas, inflexiones etc.

Luego continúa comentando que en muchos casos un texto puede ser ambiguo. Esto puede ocasionar que para el traductor automático sea difícil determinar que es lo que el texto quiere decir realmente. Sobre la ambigüedad hablaremos detalladamente más adelante en este capítulo.

Krauwer (2003) destaca otra dificultad para un traductor automático, la diferencia entre lenguas. Con esto se refiere Krauwer (2003) que no todo se dice de igual forma en todas las lenguas. Hay veces en una lengua dices algo con una palabra, mientras que para decir lo mismo en otra lengua tienes que utilizar una frase. Lo que también se ve es que hay diferencias en el uso del tipo de palabra.

Un ejemplo claro es el que menciona Tertoolen (2010):

- ¿Qué hora es?
- a. Wat is het uur?
 - b. Wat is de tijd?
 - c. Hoe laat is het?

Aquí vemos como hay diferentes maneras de preguntar ¿Qué hora es?, que en realidad no contienen errores. Sin embargo, la opción a y b no son idiomáticos, no se usan. La opción idiomáticamente correcta es la: c. Hoe laat is het?'

2.2 Problemas lingüísticos

Partiendo de Mogas Ribas (2016), Krauwer (2003), Alarcón Navío (2003) y Arnold (2003), nos centraremos en problemas lingüísticos que se encuentran en la TA.

Mogas Ribas (2016) distingue entre tres diferentes problemas lingüísticos. A saber, problemas morfológicos, ambigüedades y anáforas.

Según Mogas Ribas (2016) la morfología no debería considerarse un problema en sí, ya que nos ayuda a identificar la palabra raíz por medio de las flexiones. Sin embargo, esto es difícil de analizar para un programa de TA, al igual que palabras compuestas, por ejemplo.

Otro problema lingüístico que destacan Mogas Ribas (2016) y Krauwer (2003) es la ambigüedad. Ellos distinguen entre la ambigüedad léxica y la ambigüedad estructural.

Se habla de ambigüedad léxica cuando las palabras pueden tener más de una interpretación (Mogas Ribas, 2006; Krauwer, 2003, Arnold, 2003). El problema de la ambigüedad es que siempre habrá posibilidad de que una palabra corresponda a una u otra significado, contrario del contexto en el que se encuentra (Arnold, 2003).

Generalmente los lectores no tienen problemas de este tipo, porque por lo general entienden lo que están leyendo. Para máquinas traductoras es esto un poco más difícil, porque ellos desconocen el contexto y justamente eso es lo que es importante para poder deducir el significado correcto de una palabra (Krauwer, 2003).

Por ejemplo:

un banco:

Un banco puede ser un asiento, con respaldo o sin él, en que pueden sentarse varias personas o se puede referir a un establecimiento público de crédito, constituido en sociedad por acciones. (fuente: RAE definiciones)

una carta:

Una carta puede ser un papel escrito, y ordinariamente cerrado, que una persona envía a otra para comunicarse con ella o por carta también se puede entender que son cartulinas rectangulares que, cubiertas con una cara uniforme y otra con dibujos determinados, se usan en distintos juegos de azar. (fuente: RAE definiciones)

Existen diferentes tipos de ambigüedades léxicas. El más común es la ambigüedad categorial, que es cuando existe la posibilidad de asignar más de una categoría gramatical o sintáctica, dependiendo del contexto, a una palabra (Somers: 2003). Por ejemplo: “Yo pido la cuenta (de rekening, sustantivo)”, y “Ella cuenta una historia (vertellen, verbo)”.

Homografía y polisemia

Se habla de otro tipo de ambigüedad léxica cuando una palabra tiene dos o más significados posibles. Pueden ser palabras homógrafas o polisémicas.

Palabras homógrafas son aquellas palabras con significados muy diferentes pero escritas de la misma forma. Palabras polisémicas (estas palabras tienen una variedad de significados, pero están relacionadas de alguna forma.)

Un ejemplo de una palabra homógrafa es la palabra vino. Vino puede categorizarse como un sustantivo, pero también como el pretérito indefinido del verbo venir en la 3ª persona singular.

Un ejemplo de frases polisémicas es el siguiente.

- a. La **blanca** flor se marchitó esta mañana
- b. **Blanca**, mi vecina, falleció el año pasado.

Se trata de la palabra “blanca”. En el caso de la frase a, se trata de un adjetivo y en el caso de b. se trata de un sustantivo. Aquí hay que preguntarse ¿Qué información me da el contexto?

La ambigüedad estructural se presenta cuando se puede analizar la estructura sintáctica de una frase de más de una manera. Para los sistemas de TA es difícil, ya que la mayoría de las veces traducen las frases de forma aislada, pues no recurren al contexto (Mogas Ribas, 2006; Somers, 2003).

Por ejemplo:

- a. “Eligió un coche rápido.”

Se puede entender esta frase de dos formas. Puede referirse a que eligió rápidamente (adverbio) un coche o que el coche que eligió es rápido (adjetivo)

- b. “Juan se encontró con Jorge para calmar su preocupación.”

Se puede entender de dos maneras, la preocupación puede ser de Juan o de Jorge.

Referencias anafóricas y catafóricas

Una anáfora es lo que los lingüistas definen como “una referencia indirecta a una entidad mencionada de forma explícita en otro lugar del texto.” (Mogas Ribas, 2006; Somers: 2003)

Por ejemplo:

- a. **María** no se rinde, **ella** es una luchadora.

Al contrario de una referencia anafórica, en una referencia catafórica va primero el pronombre después el nombre al que se refiere.

Por ejemplo:

- a. **Le dije a Juan** que estudiara.

Según Krauwer (2003), que también trata este fenómeno en su artículo, es difícil tener una solución para traducir las anáforas mediante un sistema de TA. Según él sería posible solamente si se pudiera incluir la información de como interpretamos nosotros los humanos esta información en un sistema automático. Según Somers (2003), para resolver los problemas anafóricos, se necesita el mismo tipo de conocimiento que se necesita para resolver un problema de ambigüedad.

2.3 Conclusión

Queda claro que la traducción es algo difícil. Se considera como una habilidad que va mucho más allá del dominio de los lenguajes. Hay muchas cosas que los ordenadores no son capaces de hacer. Por ejemplo, los ordenadores funcionan basándose en reglas. Si no figura una regla en el sistema de un ordenador, el ordenador no va a poder cumplir con la tarea. Un ordenador no puede ser creativo y pensar una solución creativa, lo que un traductor humano si logra hacer. Un traductor automático produce una traducción basándose en el texto fuente que recibe. Se enfoca en el texto y no toma en cuenta el contexto, posibles ambigüedades a nivel de léxico, sintaxis etc.

Capítulo 3: Traducir mediante la TA y evaluaciones de la TA

Después de habernos enfocado en los capítulos anteriores en la historia de la TA, problemas que pueden existir al traducir con una máquina traductora, vamos a enfocarnos desde este momento en adelante a comentar las características y los posibles problemas de los textos con los que vamos a experimentar para llegar a una respuesta a nuestra pregunta principal. Es importante saber qué características tienen estos textos, como se traducen normalmente y también saber que tipo de problemas se encuentra un traductor al momento de traducir estos textos. Para el experimento que vamos a llevar a cabo, se han seleccionado una serie de textos partiendo de la teoría de Reiss, citado en Munday (2006).

3.1 Los textos a traducir

Reiss distingue entre tres funciones de textos: textos con la función informativa, expresiva y operativa. Un texto informativo tiene como meta comunicar. Es por esto que el contenido debe ser transmitido totalmente. Lo que se transmite son hechos, con el fin de comunicar. Ejemplos son noticias, información, opiniones, etc. Textos expresivos son mas creativos y se le da contenido al texto de forma artística. El autor crea el contenido y busca la manera de comunicarse con su publico lector de una manera artística y creativa. Estos textos están estructurados de forma semántica y sintáctica como también artística. Por ultimo, el texto operativo tiene como función estimular al lector para que reaccione o haga algo. Es un texto que tiene una función lingüística y psicológica.

Tipo de texto	Informativo	Expresivo	Operativo
Función de lenguaje	Representar hechos	Expresas desde el punto de vista del emisor	Inducir al receptor del texto
Dimensión de lenguaje	Lógica	Estética	Dialógica
Foco del texto	Contenido	Forma	Apelativo
El texto meta debe:	Transmitir el contenido referencial	Transmitir la forma estética	Producir el efecto deseado
Estrategia de traducción	Sin redundancia y con explicación si es necesaria	Adoptar la perspectiva del autor del texto fuente	Crear efecto de equivalencia

Tabla 1: características tipología textual Reiss (Munday, 2008; Tertoolen, 2010)

Para poder elegir los textos específicos pertenecientes a cada función textual, nos hemos basado en la pirámide de Chesterman que presenta diferentes textos (géneros) categorizados por la función textual.

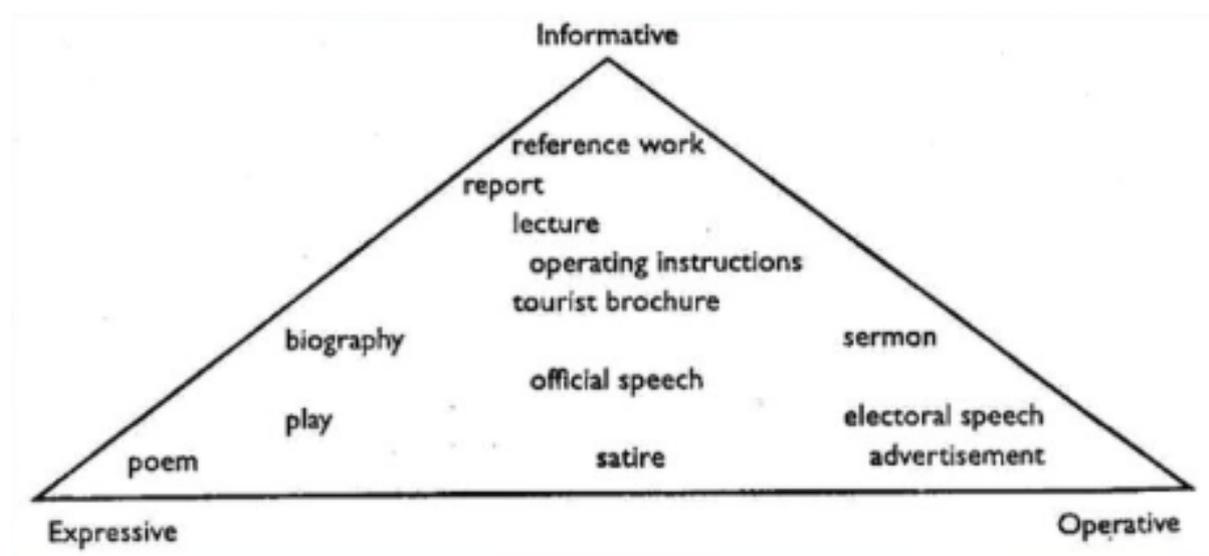


Figura 2: géneros textuales en la pirámide de Chesterman (Munday, 2008)

A continuación, hablaremos sobre los textos específicos que vamos a usar mas adelante para nuestro experimento. Comentaremos el texto en sí, como también posibles problemas que puedan presentar estos tipos de textos. En la siguiente figura podemos ver dónde figuran los textos escogidos en la pirámide de Chesterman.

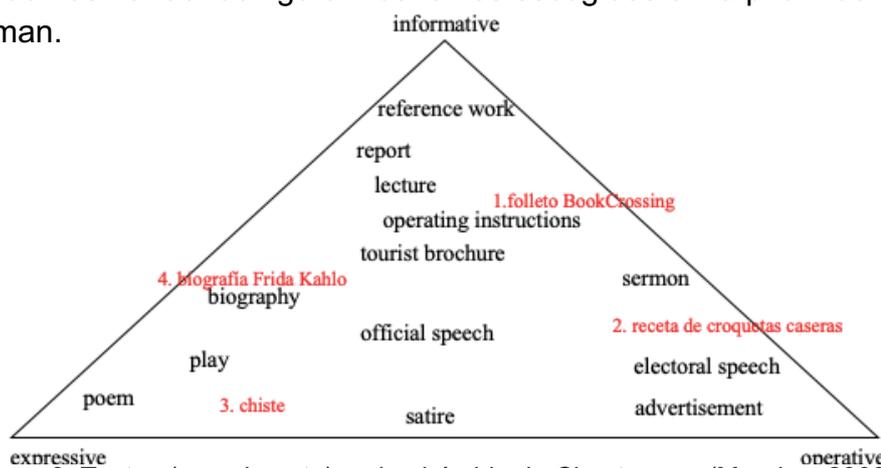


Figura 3: Textos (experimento) en la pirámide de Chesterman (Munday, 2008)

Texto 1: folleto BookCrossing (texto informativo)

El folleto de BookCrossing¹ fue creado para informar al público que estuvo presente durante el Día del Libro 2019 del nuevo proyecto que inició ese día en la biblioteca del Instituto Cervantes. Este texto está escrito de forma neutral y en el texto se explica que es el BookCrossing y como se puede participar en este Proyecto. Se

¹ El texto completo se encuentra en el anexo 1

puede concluir que este texto tiene una función informativa, que de igual forma se debería mantener en la traducción, sin redundancias.

Texto 2: Recetas de croquetas caseras (texto operativo)

Esta receta², sacada de una pagina de internet española, nos explica paso a paso como podemos hacer croquetas en casa. Es decir que este texto cumple con su meta de querer activar al lector para que haga algo, es un texto instructivo. En términos de Reiss, la traducción de este texto deberá cumplir con la misma función que cumplía en el texto fuente, el de inducir al receptor a que se anime a hacer algo.

Texto 3: chistes (texto expresivo)

Para analizar la traducción de textos expresivos he seleccionado un total de tres chistes cortos³. Este texto no figura en la pirámide de Reiss, aun así, es un proceso un poco difícil según describe Vandaele (2010). Vandaele describe el humor como algo que entretiene, que trae felicidad, una sonrisa espontanea o que ocasiona risas. Es considerado como un reto para traductores. Vandaele considera que el problema con la traducción del humor es por el hecho de que el humor tiene mucho que ver con el conocimiento implícito. Es más, puede haber grupos que piensen diferente de un tipo de humor u otro. Otro problema que menciona es que por ejemplo un problema cultural puede convertirse en un problema de ética y política. Un traductor puede verse confrontado con algo que a él le parezca culturalmente inapropiado. En el aspecto lingüístico, Vandaele destaca problemas relacionadas con la denotaciones y connotaciones lingüísticos. Se puede decir que, para traducir un chiste, se tiene que tener en cuenta muchas cosas. Sin embargo, como destaca Reiss, hay que mantener el aspecto artístico y creativo, dentro de lo posible.

Texto 4: biografía Frida Kahlo (texto expresivo)

Una biografía⁴ es considerada como un texto con función híbrida, como lo describe Reiss (citado en Munday, 2008), es considerada un texto informativo-expresivo. Según lo describe Reiss, el texto informa al lector, pero al mismo tiempo cumple la función artística de una obra de literatura. Reiss dice algo que es importante destacar y tener en cuenta al momento de traducir. Según Reiss, a pesar de la existencia de textos con funciones híbridas, la función del texto fuente es lo que determina que función va a tener el texto meta.

² El texto completo se encuentra en el anexo 1

³ El texto completo se encuentra en el anexo 1

⁴ El texto completo se encuentra en el anexo 1

3.2 Evaluaciones de la TA

En este apartado vamos a discutir los puntos de vista sobre las evaluaciones de las máquinas traductoras y vamos a ver más detalladamente las diferentes técnicas de evaluación, basándonos en White (2003) y Hansen (2010).

Hansen (2010) comenta en su artículo que se pueden clasificar errores de diferentes formas. En textos escritos podemos distinguir entre errores pragmáticas, semánticas, idiomáticas, ortográficas, lingüísticas o estilísticas. Sin embargo, él lanza la pregunta: ¿Qué es un error? Desde la perspectiva de la traducción podemos definirlo como lo que falta o difiere entre el texto fuente y el texto meta.

“Errores” de traducción pueden suceder porque no se ha entendido el contenido del texto fuente. Por no haber logrado conseguir el significado exacto del texto fuente, puede haber errores al momento de describir hechos, puede haber errores de terminología o de estilística o porque hay muchas interferencias entre el TF y el TM. Con una interferencia se refiere Hansen (2010) a que hay soluciones no deseadas en textos meta por la aparente simetría entre dos lenguajes o culturas que parecen ser muy aparentes pero que no lo son.

Diciendo que un error de traducción origina de la relación entre un TF y un TM, es un punto crucial. Determinar qué se considera como error y qué no, depende del enfoque teórico que se esté usando como base para evaluar una traducción.

Si hablamos de la evaluación en la TA, debemos recordar ese momento en la historia que hoy nos llevó a tener diferentes métodos para evaluar la TA. La TA es según White (2003) una de las primeras aplicaciones que se usó en un ordenador, y era algo que se contemplaba desde los 1930 y 1940. El primer gran experimento tuvo lugar en el 1954. En esos tiempos el hecho de que un ordenador fuera capaz de hacer eso, se percibía como algo muy novedoso y esto daba de entender que todo el proceso automático era muy poderoso.

Probablemente haya sido el reporte ALPAC (véase capítulo 1.2) el que ocasionó que el foco se colocara en la evaluación de la TA. Después de haber salido el reporte ALPAC, que sacó una serie de críticas hacia la TA, los financiamientos de los proyectos de la TA bajaron drásticamente y sobre todo en los Estados Unidos.

White (2003) comenta que la evaluación del output de una máquina traductora es difícil. Para poder lograr una evaluación se necesita señalar una traducción “correcta” o “ideal” para así poder hacer comparaciones. Los traductores más que nadie son conscientes de que dos personas jamás traducirían un texto igual. White (2003) distingue entre una serie de métodos de evaluación. Entre esos métodos distingue White (2003) entre el método de “la caja negra” vs. “la caja de vidrio”.

El método de “la caja negra” se refiere a que para evaluar no tomamos en cuenta los procesos del traductor automático. Solo analizamos el texto fuente y su traducción automática. En nuestro experimento es esto lo que queremos analizar, las traducciones realizadas. A partir del texto fuente y las dos traducciones vamos a ver como traducen Google Translate y DeepL.

Podemos distinguir el método de investigación con la visión de la “caja negra” y de la “caja de vidrio”. En el caso de la visión de la “caja negra” no tomamos en cuenta los mecanismos del traductor automático, sino que solo analizamos el output (la calidad de ello).

En la bibliografía existente se habla sobre diversos métodos para evaluar la TA. Uno de los sistemas mas recientes es el método BLEU (Papineni et al, 2001). El punto de partida de una evaluación automática es que los humanos no son capaces de analizar las cosas de una forma objetiva y estos análisis tardan mucho mas haciéndolas un humano que un ordenador. Este método se basa en la hipótesis de que, aunque haya diferentes formas de traducir un texto “correctamente”, la mayoría de las veces entre esas formas hay ciertas frases en común. Este método requiere múltiples traducciones humanas para formar parte del corpus.

3.3 Conclusión

En este capítulo hemos discutido los puntos importantes que han llevado a escoger el material para llevar a cabo nuestro experimento. Basándonos en Reiss, quien distingue entre tres funciones textuales, hemos elegido nuestros textos de los que hablaremos en el siguiente capítulo más detalladamente.

Como Reiss indica, es muy importante tener claro cuál es la función de un texto para así saber que tener en cuenta al momento de traducir. A parte de funciones específicas, existen funciones híbridas que son textos que cumplen con una doble función, como por ejemplo la biografía de Frida Kahlo que vamos a usar más adelante. La traducción de un chiste es considerada algo difícil y se ve como un reto, ya que los chistes pueden estar muy relacionados a una cultura. Esto puede traer problemas por ejemplo cuando el público del texto fuente y del texto meta no tienen similitudes culturales.

Existen diferentes métodos de evaluación de traducciones, tanto automáticas como no automáticas. Esto se considera como algo difícil de hacer, ya que se necesita una norma para poder determinar que es una mala o una buena traducción. Esto no siempre es objetivo, ya que de un texto fuente pueden surgir diferentes traducciones. Claro está que un texto traducido por dos personas nos lleva a conocer dos versiones diferentes de traducciones.

Capítulo 4: La metodología de la investigación

Para poder llegar a una conclusión respecto a mi pregunta de investigación, he llevado a cabo una encuesta. He optado por analizar los datos de forma cuantitativa. He diseñado una encuesta que se rellenó por medio de SurveyHero, una herramienta donde se registran las respuestas de forma anónima.

4.1 Encuesta y participantes

4.1.1 La encuesta

La encuesta consiste en cuatro preguntas. Cada pregunta consiste en una imagen con un texto en español. Luego, debajo de cada imagen hay una serie de frases traducidas por Google Translate y luego por DeepL. Cada frase se puede evaluar de 1 a 5. 1 corresponde a una traducción muy mala y 5 a una traducción muy buena.

Cada texto ha sido seleccionado siguiendo las pautas de Reiss (citado por Munday, 2016), quien distingue 3 funciones textuales. Basándome en estas funciones, a saber: texto informativo, texto operativo y texto expresivo (véase capítulo 3). He escogido como método evaluativo esta escala, ya que permite a los encuestados dar una valoración exacta, sobre todo por el hecho de que nos interesa analizar la calidad de las traducciones, lo que se conoce como el método de la “caja negra”. (véase capítulo 3)

Para poder analizar una variedad de traducciones automáticas, me he basado en la tipología de textos según Reiss y sus funciones. Considerando que existen diferentes tipos de texto, como mencionado anteriormente, he optado por escoger cuatro textos basándome en sus funciones. Lo que quería lograr con estos textos es que fueran textos diferentes y que no pertenecieran a la misma función. Cada texto es igual de extenso, de cada texto utilicé 180 palabras, es decir, una parte de los textos completos. El motivo por el cual opté por no usar los textos completos es porque quería analizar la misma cantidad de “líneas” para cada texto y de haber analizado los textos enteros, la encuesta iba a ser muchísimo mas largo de lo que ya es.

Para el primer texto he seleccionado el párrafo donde se explica que es el BookCrossing y como se puede participar en este proyecto. Considero que de aquí se puede sacar la mayoría parte de información puesto que la función del texto es informar. Con esta parte se puede decir que el texto cumplió con su función.

Para el segundo texto, la receta de las croquetas caseras he incluido la parte en donde se explica paso a paso como hacer las croquetas. Considerando que la función de este texto es operativa, explicando los pasos para hacer las croquetas, se cumple con la función del texto.

El tercer texto y el cuarto texto son ambos textos expresivos. Elegí chistes y una biografía, considerando que los chistes eran muy cortos para analizar por si solos. Opté por analizar una biografía ya que punto una biografía también tiene características expresivas a pesar de ser un texto informativo también. Por el hecho de ser un texto con una función híbrida, es interesante ver si eso influye en la calidad de la traducción. Los chistes han sido analizados completamente y de la biografía he escogido el núcleo del texto, ya que esto me proporcionaba soluciones interesantes de Google Translate y DeepL.

4.1.2 Los participantes

Para esta investigación, he elegido como participantes a expertos en el ámbito de la traducción. Considerando que se trata de una investigación que trata de contribuir con conocimiento existente sobre la traducción automática y considerando que lo interesante es ver qué opinan los expertos (y usuarios potenciales) sobre el output de las TA, le he pedido a 10 traductores que rellenaran el cuestionario. De estos 10 traductores conozco a tres personalmente. Los otros traductores también son profesionales y he llegado a ellos a través de la plataforma ProZ. Las edades de los participantes varían de más o menos 30 a 50 años. Según los datos de los perfiles hay algunos con mucha experiencia y otros con menos experiencia, algunos viven en el exterior y otros residen en los Países Bajos. No se puede decir con seguridad si todos son nativos holandeses, pero la mayoría de los participantes si eran nativos.

4.2 Conclusión

La encuesta se ha diseñado de tal forma que los encuestados pueden leer el texto fuente y luego comparar las traducciones y dar su valoración a las frases producidas por Google Translate y DeepL por medio de una cifra. Esta cifra puede ser un uno (muy mala) hasta un cinco (muy buena). Para este experimento he escogido a personas que son traductores ya sea en los Países Bajos o en el exterior. La encuesta es totalmente anónima, lo que significa que yo solo puedo ver resultados mas no puedo ver quien llenó cual encuesta.

Capítulo 5: El análisis de la investigación

En este capítulo presentaremos los resultados del experimento e intentaremos contestar la pregunta: ¿Cómo se diferencian las traducciones producidas por los traductores automáticos Google Translate y DeepL? Partiendo de los trozos de textos utilizados en la encuesta, vamos a evaluar los textos desde un punto de vista lingüístico. En el marco teórico hemos dedicado atención a problemas lingüísticos que se pueden presentar en la traducción de textos mediante una máquina TA, y vamos a intentar señalar los errores que conseguimos en los textos de Google Translate y DeepL. Por otro lado, vamos a incluir las evaluaciones de los expertos. Gracias a estas valoraciones y nuestro análisis lingüístico podemos relacionar lo que dicen los expertos con las muestras que nos salgan de nuestro análisis lingüístico. También incluiremos una reflexión de lo que un traductor optaría por cambiar a partir de las traducciones presentadas por Google Translate y DeepL

5.1.1 Texto 1 - folleto BookCrossing

El primer texto es un folleto del Instituto Cervantes sobre el proyecto BookCrossing. En el cuadro del Día del Libro 2019, se redactó un folleto. Este folleto fue creado para informar al público de qué se trataba y cómo podían participar en este proyecto. La función de este texto es informar. Conforme las características de este género, el foco de este texto es el contenido y la traducción debe transmitir el contenido sin redundancias y con explicación si se considera necesaria. En general se puede decir que los textos metas se pueden entender, se transmite el mensaje general pero los textos contienen errores que podríamos clasificar como malas escogencias del traductor automático, adaptaciones no necesarias, frases raras etc.

Google Translate

Het is een wereldbibliotheek. [1]

Het is het web van een intelligent sociaal netwerk. [2]

Het is de partij van literatuur en een plek waar boeken een nieuw leven krijgen. [3]

BookCrossing is de handeling van het geven van een boek een unieke identiteit, zodat, zoals het boek doorgegeven van de lezer aan de lezer kan volgen en met elkaar te verbinden zijn lezers. [4]

Op dit moment zijn er 1.940.532 BookCrossers en 12.844.916 boeken die door 132 landen reizen. [5]

Onze gemeenschap verandert de wereld en de levens van boeken in boek. [6]

Onze missie: mensen verbinden door boeken. [7]

Hoe een boek achterlaten in een BookCrossing-gebied? [8]

1. Registreer u op www.bookcrossing.com (dit kan slechts 2 of 3 minuten duren). [9]

2. Registreer uw boeken: <https://www.bookcrossing.com/register>. Ze geven je een BCID-code (identificatienummer van het boek) die je bij een label in het boek zult voegen. [10]

3. Maak je boek vrij "in de jungle". Daarna moet je het bevrijdingsgebied kiezen waar je je boek kunt vinden: bibliotheek Cervantes Utrecht. [11]

Als u geen zin heeft om u te registreren bij BookCrossing, geeft de bibliotheek uw boeken voor u vrij! [12]

DeepL

Het is een wereldwijde bibliotheek. [1]

Het is het web van een intelligent sociaal netwerk. [2]

Het is het festival van de literatuur en een plaats waar boeken een nieuw leven krijgen. [3]

BookCrossing is het geven van een unieke identiteit aan een boek, zodat het boek, wanneer het van lezer tot lezer wordt doorgegeven, gevolgd en verbonden kan worden met de lezers. [4]

Op dit moment zijn er 1.940.532 BookCrossers en 12.844.916 boeken die door 132 landen reizen. [5]

Onze gemeenschap verandert de wereld en leeft van boek tot boek. [6]

Onze missie: mensen met elkaar verbinden door middel van boeken. [7]

Hoe laat ik een boek achter in een BookCrossing gebied? [8]

1. Schrijf u in op www.bookcrossing.com (dit kan slechts 2 of 3 minuten duren). [9]

2. Registreer uw boeken: <https://www.bookcrossing.com/register>. Ze geven je een BCID-code (identificatienummer van het boek) die je bij een label in het boek zult voegen. [10]

3. Geef je boek "in de jungle" uit. Dan moet je de release area kiezen waar je boek te vinden is: Instituto Cervantes Utrecht bibliotheek. [11]

Als u zich niet wilt registreren in BookCrossing, zal de bibliotheek uw boeken voor u vrijgeven! [12]

Podemos decir que en general esta traducción contiene muchos errores de ambigüedad (diferentes tipos), de "palabras añadidas" y de frases incorrectas en cuanto a las reglas sintácticas o el empleo de funciones gramaticales o lexicales erróneas.

Por ejemplo, la primera frase es un ejemplo del empleo de una función léxico errónea:

Google Translate:

TF⁵: Es una Biblioteca mundial

TM⁶: Het is een wereldbibliotheek.

El traductor automático ha elegido *wereldbibliotheek*, que se puede considerar como una traducción literal de la palabra. Esta palabra tiene una noción diferente a lo que se refieren en el texto. Da a entender como si *wereld* (sustantivo, mundo) es un lugar específico, mientras que el texto fuente quiere decir que es una biblioteca que se encuentra en muchos lugares en el mundo, lo que en holandés se puede describir como *wereldwijd*. Aquí vemos un ejemplo de cómo la TA empleó una palabra con función de sustantivo, mientras que podía usar un adjetivo. Esto se conoce como un caso de ambigüedad categorial.

⁵ TF: Texto Fuente

⁶ TM: Texto Meta

DeepL

TF: Es una Biblioteca mundial

TM: Het is een wereldwijde bibliotheek.

En este caso el traductor DeepL ha traducido biblioteca mundial como *wereldwijde bibliotheek*, y como veremos más adelante, los expertos han evaluado esta traducción mejor que la traducción de Google (Una media de 4,31 → bien).

Google Translate / DeepL

En este caso tanto Google Translate como DeepL ha traducido el texto fuente de la misma forma.

TF: Es la web de una red social inteligente.

TM: Het is het web van een intelligent sociaal netwerk.

Ambas han sido evaluadas como malas traducciones. Personalmente pienso que esta frase es un poco confusa analizando el tema del texto entero. BookCrossing es una iniciativa en donde uno puede tomar libros prestados y dejar libros atrás en una zona de BookCrossing, en el Instituto Cervantes, por ejemplo. Para poder participar en esta iniciativa la persona tiene que crear una cuenta para registrar los libros y así conseguir códigos específicos para los libros. Existe un “online community” (así está descrito cuando se lee sobre BookCrossing) de BookCrossing, que consiste en personas que forman parte del BookCrossing y al que nuevo BookCrossers pueden unirse. Creo que esta es la intención de la frase: señalar que BookCrossing es una iniciativa, pero también puedes formar parte de la “red social” (por medio de la página web) de ellos, que lo describen como inteligente. Pienso que las máquinas de TA optaron por traducir palabra por palabra, y como hemos discutido más adelante no tomaron en cuenta cualquier información adicional a lo que tenían que traducir (no son capaces). Pienso que se podría traducir añadiendo algo de información o buscando un sinónimo, por ejemplo: *Het is het web van een slim en sociaal netwerk*, y quizás un sinónimo para *slim* puede ser *handig*. Otra opción podría ser, dependiendo de cuanta libertad la traductora tenga, unir la primera y la segunda frase de forma que podría existir la siguiente frase: *Het is een wereldwijde bibliotheek en het is een handig en sociaal online platform*.

El siguiente ejemplo que vamos a analizar contiene sobre todo errores léxicas y otras soluciones que podrían haberse traducido de otra forma, pero que no dificultan la comprensión de lo escrito. El mensaje llega y no es tan confuso, uno puede deducir que es lo que realmente el traductor quiere decir.

Google Translate

TF: Es la fiesta de la literatura y un lugar donde los libros cobran una nueva vida.

TM: Het is de partij van literatuur en een plek waar boeken een nieuw leven krijgen.

En este caso podemos decir que Google Translate tradujo fiesta como *partij*, quizás partiendo del inglés, pero que es la palabra errónea. *Partij* tiene otro significado y aquí no se puede decir que se trata de algún tipo de ambigüedad. *Partij van literatuur* suena muy raro decirlo, sin embargo, *van literatuur no está mal traducido, sino que no se usa*. Una mejor traducción sería buscar un equivalente de eventos con libros y usar ese término. En este caso podemos pensar en por ejemplo un *boekenfeest*. Como anteriormente hablamos, para una máquina TA es difícil pensar de forma creativa cuando se trata de traducir algo que en realidad se dice de otra forma que no tiene mucho que ver con lo que está escrito literalmente en el texto fuente.

DeepL

TF: Es la fiesta de la literatura y un lugar donde los libros cobran una nueva vida.

TM: Het is het festival van de literatuur in een plaats waar boeken een nieuw leven krijgen.

DeepL ha escogido traducir la fiesta como *festival* lo que en realidad se puede considerar como una mejor opción que *partij*, solo que en este caso no se trata de un festival. Pero como veníamos comentando anteriormente, un TA no es capaz de basarse en el contexto, sino que se enfoca en traducir lo que está escrito.

Una cosa que a veces puede ser difícil es la traducción de colocaciones, ya que esto puede ser diferente en los diferentes idiomas, sin embargo, tanto Google Translate han traducido cobrar una nueva vida como *een nieuw leven krijgen*. Esta es la colocación en holandés, pero fácilmente el traductor pudo analizar “cobrar” como un elemento aparte y dar otra traducción.

El siguiente ejemplo es una frase compleja. En vez de dos frases se hizo una larga que se une por medio de “de manera que”. Como esta frase se corta y después es retomada, es interesante ver que errores se pueden sacar.

Google Translate

TF: BookCrossing es el acto de dar a un libro una identidad única, de manera que, como el libro pasa de lector a lector, se pueda seguirlo e interconectar a sus lectores.

TM: BookCrossing is de handeling van het geven van een boek een unieke identiteit, zodat, zoals het boek doorgegeven van de lezer aan de lezer kan volgen en met elkaar te verbinden zijn lezers.

Aquí vemos que el TA quiso quedar lo más fiel posible al texto fuente, creando una frase que no es idiomático. El acto de dar es *het geven (van iets)*. Luego continuamos con *BookCrossing is het geven van een boek een unieke identiteit*. En realidad, esta frase se puede hacer activa, en vez de usar *is het geven*. Lo que queremos decir en este caso es que los libros que participan en el BookCrossing consiguen una identidad única. Como el libro va de lector a lector, se puede seguir el libro con este código vaya donde vaya y se puede conectar con los lectores. Esta

frase la podemos cortar y cambiar de manera que quede la información mucho más clara de lo que tradujo la TA, opino yo estando en los zapatos de una traductora.

Si la orden de la frase del texto fuente fuera diferente, quizás se pudiera haber evitado la mala traducción en este contexto de “de manera que, como...”. Aquí lo problemático es el cómo. Google Translate lo analizó como una unidad separada cuando en realidad tuvo que analizarlo como “como el libro pasa de lector a lector”. En este sentido “como” ha sido traducido erróneamente.

DeepL:

TF: BookCrossing es el acto de dar a un libro una identidad única, de manera que, como el libro pasa de lector a lector, se pueda seguirlo e interconectar a sus lectores.

TM: BookCrossing is het geven van een unieke identiteit aan een boek, zodat het boek, wanneer het van lezer tot lezer wordt doorgegeven, gevolgd en verbonden kan worden met de lezers.

Una opción podría ser: *Met BookCrossing geef je een boek een uniek nummer, waarmee lezers het boek kunnen volgen. Omdat het boek van de ene lezer naar de andere lezer gaat, kan het boek worden gevolgd en worden de lezers met elkaar verbonden.*

Google Translate / DeepL

Como hemos visto anteriormente, tanto Google Translate como DeepL tradujeron la frase igual.

TF: Ahora mismo hay **1.940.532 BookCrossers** y **12.844.916 libros** viajando por **132 países**.

TM: Op dit moment zijn er 1.940.532 BookCrossers en 12.844.916 boeken die door 132 landen reizen.

No hay mucho que comentar, ya que esta frase ha sido traducida de igual forma por Google Translate y DeepL y no contiene graves errores. Esto se refleja también en la evaluación de los expertos, que evaluaron esta frase con una media de 4,06 (buena calidad). A pesar de que la traducción es correcta, viajando se ha traducido como reizen, lo cual también es correcto, pero yo optaría por usar un otro verbo menos específico. *Reizen* significa en holandés het trekken van een plaats naar een andere, que significa ir (moverse) de un lugar a otro. En todo caso se trataría de la personificación de los libros y es por ese motivo que yo usaría algo un poco más neutral, como por ejemplo *rond gaan*.

Las traducciones de la siguiente frase han sido evaluadas con un 2,00 de media y un 2,50 de media para Google Translate y DeepL respectivamente. Esta frase contiene tres puntos débiles que queremos destacar en este análisis. Primero vamos a ver la traducción de Google Translate y luego la de DeepL

Google Translate

TF: Nuestra comunidad está cambiando el mundo y las vidas de libro en libro

TM: Onze gemeenschap verandert de wereld en de levens van boeken in boek

DeepL

TF: Nuestra comunidad está cambiando el mundo y las vidas de libro en libro.

TM: Onze gemeenschap verandert de wereld en leeft van boek tot boek.

Tanto Google Translate como DeepL tradujeron comunidad como *gemeenschap*. Según el diccionario español-holandes Van Dale Online esta traducción es correcta. Sin embargo, las connotaciones y denotaciones con las que nos encontramos al buscar en el diccionario holandés nos indica que *gemeenschap* se utiliza más en el sentido de tener algún tipo de relación o vínculo con otra persona. Esto no es realmente lo que estamos buscando para este contexto. Aquí podemos hablar de un problema de léxico (palabra errónea).

En el texto fuente está escrito: “está cambiando el mundo y las vidas de libro en libro”. El escritor se refiere al mundo como algo general y que la vida de la gente que participa en el BookCrossing (aunque no lo menciona explícitamente) también está siendo cambiadas. Google Translate tradujo lo último exactamente igual al holandés: *de levens*. Esto hace alusión a que BookCrossing está cambiando vidas, pero como si fuera de una sola persona. En holandés es más correcto explicitar lo dicho. En español se entiende perfectamente, pero una construcción así en holandés es inusual. La última parte “de libro en libro” quiere expresar que, con cada libro, ya sea prestado para leer o entregado para que otros lo lean, cambia la vida de la gente. Google Translate lo tradujo como *van boeken in boek* y DeepL lo tradujo como *van boek tot boek*.

Google Translate empleó un plural mientras que esto no era necesario, pues no está en plural en el texto fuente. DeepL optó por traducir “de libro en libro” como *van boek tot boek*. Esto no es incorrecto, sin embargo, aquí nuevamente vemos que no es lo más idiomático que existe. Esto es difícil de determinar para una máquina traductora.

A pesar de estos errores no se puede decir que no se entiende nada de las traducciones. Para solucionar el problema del uso de singular-plural para referirnos a conceptos generales, yo propondría traducir la frase de la siguiente forma: *Onze community verandert met elk boek de wereld en het leven van mensen. Het leven* puede ser un término bastante amplio y puede referir a diferentes tipos de vidas, es por eso que incluiría *van mensen*, con el fin de especificar.

La traducción de DeepL de la siguiente frase ha sido mejor evaluada que la traducción de Google Translate.

TF: Nuestra **misión**: conectar personas a través de los libros.

TMGoogle: Onze missie: mensen verbinden door boeken.

TMDDeepL: Onze missie: mensen met elkaar verbinden door middel van boeken.

Si tuviera que elegir una mejor y una peor traducción, elegiría a la traducción de DeepL como la de buena calidad y la de Google Translate como la de mala calidad. Este “error” o “diferencia” tiene que ver con la traducción de preposiciones. Se puede motivar la traducción de Google Translate argumentando que “a través” se puede ver como un sinónimo de mediante, sin embargo, la traducción exacta de a través es *door middel van*, que sí fue utilizado por DeepL. En segundo lugar, también vemos que una construcción corta como “conectar (a) personas” no existe en holandés. Se tiene que recurrir a una colocación gramaticalmente correcta, *mensen met elkaar verbinden*. Google Translate no incluyó esto en su traducción. Aunque en la traducción de Google Translate no este esa parte, no se puede decir que la frase no se entiende.

Google Translate

TF: ¿Cómo dejar un libro en una zona BookCrossing?

TM: Hoe een boek achterlaten in een BookCrossing-gebied?

En esta traducción propuesta por Google Translate vemos que se ha mantenido el verbo dejar en la frase holandés (*achterlaten*), y que también el orden de la frase es incorrecto: ¿*Hoe een boek achterlaten in een BookCrossing-gebied?* Normalmente sería “*hoe + verbo*”. En este caso el verbo no está conjugado y el traductor ha optado por poner el verbo detrás del sustantivo. Otro elemento que podemos remarcar y relatar a la teoría de Arnold (2003), es la escogencia de *BookCrossing-gebied*. Arnold dice que para la máquina es difícil aprender cosas nuevas. De ser un traductor humano que si puede acceder a información nueva de una forma u otra, el traductor podría ver qué zona BookCrossing tiene su traducción exacta que también se usa oficialmente, este es el *BookCrossing zone*.

TF: ¿Cómo dejar un libro en una zona BookCrossing?

TM: Hoe laat ik een boek achter in een BookCrossing gebied?

Esta traducción la produjo DeepL y cómo podemos ver el orden de la frase de DeepL sí es según las reglas sintácticas del holandés. Donde Google Translate no conjugó el verbo, DeepL sí lo hizo. Una duda que surge aquí es cómo DeepL si es capaz de por ejemplo notar que, aunque en español se use el verbo, en holandés no se dice así, sino que recurrimos al *ik-persoon*.

Google Translate

TF: Regístrate en www.bookcrossing.com (esto te puede llevar tan sólo unos 2 o 3 minutos).

TM: 1. Registreer u op www.bookcrossing.com (dit kan slechts 2 of 3 minuten duren).

DeepL

TF: Regístrate en www.bookcrossing.com (esto te puede llevar tan sólo unos 2 o 3 minutos).

TM: Schrijf u in op www.bookcrossing.com (dit kan slechts 2 of 3 minuten duren)

Esta frase contiene pocos errores. Lo que podemos ver es que por ejemplo el TA traduce la forma de usted, mientras que el texto fuente está escrito en la segunda persona singular (tú). Google Translate y DeepL han traducido “Regístrate” como *Registreer u op (registreer)* y *Schrijf u in op (inschrijven)*. Estas traducciones no son adecuadas, ya que lo que estamos buscando es que creen una cuenta, lo que se traduce como *een account aanmaken*. La traducción podría ser: *Maak een account aan op www.bookcrossing.com (dit kan slechts 2 of 3 minuten duren)*.

Google Translate /DeepL

TF: Registra tus libros: <https://www.bookcrossing.com/register>. Te darán un código BCID (nº de identificación del libro) que pegarás con una etiqueta en el libro.

TM: Registreer uw boeken: <https://www.bookcrossing.com/register>. Ze geven je een BCID-code (identificatienummer van het boek) die je bij een label in het boek zult voegen.

Las traducciones de Google Translate y de DeepL son iguales. Lo que podemos observar aquí es por ejemplo cómo *Te darán un código BCID* se traduce como “Ze geven je een BCID-code”, mientras que hay formas más idiomáticas. Por ejemplo, *je ontvangt/krijgt een BCID-code*. La traducción de “que pegarás con una etiqueta en el libro” fue traducido por Google Translate y DeepL como *die je bij een label in het boek zult toevoegen*. La idea es pegar el código en forma de una etiqueta en el libro. Para evitar confusiones se podría traducir esta frase como: *dat op het boek geplakt moet worden*. Es decir, omitir la parte de “que pegarás con una etiqueta”. Al leer que se tiene que pegar algo en un libro, automáticamente pensamos que puede ser una etiqueta o algún material que no se despegue (realizar razonamiento común).

Google Translate

TF: Libera tu libro “en la jungla”. Después tendrás que escoger la zona de liberación donde se podrá encontrar tu libro: biblioteca instituto Cervantes Utrecht.

TM: Maak je boek vrij "in de jungle". Daarna moet je het bevrijdingsgebied kiezen waar je je boek kunt vinden: bibliotheek Cervantes Utrecht.

DeepL

TF: Libera tu libro “en la jungla”. Después tendrás que escoger la zona de liberación donde se podrá encontrar tu libro: biblioteca instituto Cervantes Utrecht.

TM: Geef je boek "in de jungle" uit. Dan moet je de release area kiezen waar je boek te vinden is: Instituto Cervantes Utrecht bibliotheek.

Si analizamos estas traducciones, Google Translate propone “Maak je boek vrij “in de jungle” y DeepL propone “Geef je boek “in de jungle” uit. *Uitgeven* es un ejemplo

de ambigüedad, ya que *uitgeven* puede referirse a *een boek uitgeven* (publicar un libro), *geld uitgeven* (sacar dinero para pagar algo), sin embargo, lo utilizan aquí con el significado de liberar. La traducción de Google Translate tampoco es correcta. *Vrij maken* tiene otro significado que lo que representa en la traducción de DeepL.

Otro error que comete DeepL es la traducción de después como *dan*, mientras que tiene que ser *daarna*. *Dan* se usa para referirse a un momento concreto⁷, mientras que *daarna* significa después. Luego tenemos el problema de “zona de liberación”. Suena como un término muy específico que se tiene que traducir al pie de la letra, mientras que con esta zona se refieren a una Zona BookCrossing. Para evitar confusiones, pienso que sería mejor utilizar Zona BookCrossing. Una posible traducción podría ser: *Leg je boek “in het wild” neer. Na het registreren kun je de BookCrossing zone kiezen waar je je boek achter wilt laten / waar je boek te vinden is: Bibliotheek Instituto Cervantes.*

Google Translate

TF: ¡Si no te apetece registrarte en BookCrossing, la biblioteca liberará tus libros por ti!

TM: Als u geen zin heeft om u te registreren bij BookCrossing, geeft de bibliotheek uw boeken voor u vrij!

Lo primero que vemos aquí nuevamente es que en el texto fuente el texto se dirige al lector usando la forma de tú, pero al momento de traducir, el TA lo cambia a usted. Como comentado anteriormente, liberar en este ejemplo si es traducido como *vrijgeven*.

DeepL

TF: ¡Si no te apetece registrarte en BookCrossing, la biblioteca liberará tus libros por ti!

TM: Als u zich niet wilt registreren in BookCrossing, zal de bibliotheek uw boeken voor u vrijgeven!

DeepL optó por traducir “Si no te apetece registrarte” como *Als u zich niet wilt registreren*. Con DeepL pasa lo mismo que con Google Translate. El texto fuente está escrito en segunda persona singular, pero al momento de traducir se traduce en la forma de usted.

Google utiliza *registreren bij BookCrossing*, mientras que DeepL utiliza *registreren in BookCrossing*. Lo que en el texto no aparece es que se trata de una cuenta. Cuando un traductor humano traduce esto, puede deducirlo. Mediante búsquedas en internet, se dará cuenta que se trata de crear una cuenta en BookCrossing para hacer todo esto. Claramente, como comentamos en el marco teórico, esto es algo que un ordenador no puede hacer. El TA consigue por un lado un texto y por otro lado lo traduce usando las reglas que a él se le ha impuesto.

⁷ www.taaladvies.net (dan/daarna)

Ambas frases son correctas gramaticalmente, sin embargo, si se tratara de lo que es más idiomático o no, escogería otras palabras para la traducción de “Si no te apetece regístrate”. Yo cambiaría la frase y lo haría en forma de pregunta, por ejemplo: *Wil je je niet registreren voor een BookCrossing account? De bibliotheek kan je boeken voor je vrijgeven.*

Como sabemos que ambos sistemas funcionan en base a un gran corpus en donde mediante un análisis estadístico se escoge la traducción que parece ser la más adecuada, siempre basándose en ejemplos de su corpus, podemos pensar que esta una posible declaración de que se traduzca la segunda forma singular en la tercera forma.

5.1.2 Texto 2 - receta de croquetas caseras

El segundo texto es una receta de croquetas caseras. Es un texto operativo, porque en este texto se nos explica paso a paso como podemos hacer estas croquetas. Estos tipos de texto se escriben con el fin inducir al público lector a hacer algo (a tomar acción). Según los criterios de Reiss (citado en Munday, 2008), en la traducción de un texto operativo, el texto meta debe producir el efecto deseado. En este caso, esto se refiere a que el texto meta debe de “invitar” o “motivar” al lector a hacer croquetas, y el traductor debe tratar de crear un efecto de equivalencia en el texto meta. Antes de analizar las traducciones profundamente, vamos a comentar posibles problemas que se nos puede presentar al momento de traducir este texto. En este caso los ingredientes son simples y son conocidos en el lenguaje meta y en la cultura meta. En otros casos se debería aplicar estrategias para que se pueda traducir correctamente al lenguaje meta. Aquí podemos pensar en por ejemplo la teoría de Grit, que nos explica cómo podemos traducir aun cuando hay terminología que son muy de la cultura fuente y cómo podemos solucionar esto. Estas instrucciones son bastantes simples, es cuestión de que la TA traduzca correctamente tal y como se dirían las instrucciones en holandés.

En general podemos decir que este texto se puede seguir muy bien, a pesar de los errores cometidos por las máquinas de TA. Estos errores tienen que ver sobre todo con errores de ambigüedad. Esto conllevó a que la traducción del mensaje si se haya podido entender a pesar de que las traducciones contenían errores.

Google Translate

1. Snijd de ham en kip in heel kleine taco's. Book. [1]
2. Voeg de boter in een braadpan toe totdat deze een temperatuur bereikt, en dan worden de prei en uien geraspt tot ze goudbruin zijn. [2]
3. Voeg twee teentjes knoflook toe, heel klein gesneden. Laat een paar minuten. [3]
4. Voeg de bloem toe, met behulp van een aantal staafjes en roer zonder te stoppen. Voeg vervolgens de melk en room toe en blijf schuiven met de staven. [4]
5. Voeg een beetje geraspte nootmuskaat, zout en peper toe. [5]
6. Voeg de kip en de ham toe en blijf in vuur zetten zonder nog eens tien minuten te roeren. [6]
7. Leg opzij en laat 24 uur afkoelen in de koelkast. [7]
8. Leg het deeg op de werktafel, samen met een bord met bloem, een ander met losgeklopt ei en een laatste met broodkruim. [8]
9. Neem met behulp van een lepel (of met uw handen) porties van het deeg en meel ze in vorm. Doop ze in het ei en smeer ze in met broodkruimels. [9]
10. Bak met zeer hete zonnebloemolie om te voorkomen dat ze uit elkaar vallen. [10]

DeepL

1. Snijd de ham en de kip in zeer kleine blokjes. Zet hem opzij. [1]
2. Voeg de boter toe in een braadpan tot het warm wordt, en dan worden de prei en ui geraspt tot ze goudbruin zijn. [2]
3. Voeg twee zeer kleine gehakte teentjes knoflook toe. Vertrek voor een paar minuten. [3]
4. Voeg de bloem toe met behulp van een paar stokjes en roer non-stop. Voeg vervolgens de melk en de room toe, terwijl u verder mengt met de stokjes. [4]
5. Voeg een beetje geraspte nootmuskaat, zout en peper toe. [5]
6. Voeg de kip en de ham toe en blijf nog tien minuten roeren. [6]
7. Zet het geheel apart en laat het 24 uur afkoelen in de koelkast. [7]
8. Leg het deeg op de werktafel, samen met een bord met bloem, een ander met geklopt ei en een laatste met paneermeel. [8]
9. Neem met behulp van een lepel (of met je handen) porties van het deeg en bloem ze vorm te geven. Dompel ze in het ei en bestrijk ze vervolgens met paneermeel. [9]
10. Bak ze in zeer hete zonnebloemolie om te voorkomen dat ze uit elkaar vallen. [10]

Según los resultados generales, la traducción de Google Translate tiene una media de 2,23 mientras que la de DeepL tiene una 3,13.

Google Translate

1. Cortar el jamón y el pollo en tacos muy pequeños. Reservar.
1. Snijd de ham en kip in heel kleine taco's. Book.

La traducción de Google Translate presenta dos problemas en concreto. Por un lado, podemos argumentar que el uso de tacos para *stukjes*, puede ser un poco confuso, ya que no es una palabra común o mejor dicho muy usual en este contexto. Si buscamos en el diccionario en línea de RAE, la definición de tacos en este contexto se explica en el décimo segundo lugar. Con averiguar esto podemos concluir que muy difícilmente Google Translate lo hubiese podido traducir correctamente. En este caso, esta palabra no se tradujo, se mantuvo igual que en el texto fuente. Esto lo podemos clasificar como un problema de ambigüedad lexical. Google Translate ha traducido “reservar” erróneamente. La traducción que ofrece Google Translate es en inglés, reservar en inglés es *book*. Esto lo podemos clasificar como una palabra errónea que produjo Google Translate.

La traducción de DeepL ofrece como traducción de “tacos muy pequeños”, *zeer kleine blokjes*. Se puede decir que DeepL si lo tradujo correctamente, tomando el contexto en cuenta. En comparación con Google Translate, DeepL ofrece *zet hem opzij* como traducción de “reservar”. Una alternativa podría ser, para que la frase fuera más fluida, unirlos. Sería *snijd de ham en de kip in zeer kleine blokjes en zet/leg opzij*.

En la siguiente frase vemos que nuevamente que Google Translate ha traducido literalmente y también vemos que al final de la frase Google Translate ha cometido un error, con el resultado que el TF dice una cosa y el TM dice otra. En conclusión, aparte del error cometido por GT, opino que hay formas más idiomáticas para traducir esta frase en cuestión.

TF: Añadir la mantequilla en una cazuela hasta que coja temperatura, y después el puerro y la cebolla rallados hasta que doren.

TM Google Translate: Voeg de boter in een braadpan toe totdat deze een temperatuur bereikt, en dan worden de prei en uien geraspt tot ze goudbruin zijn.

TM DeepL: Voeg de boter toe in een braadpan tot het warm wordt, en dan worden de prei en ui geraspt tot ze goudbruin zijn.

Esta frase resulta un poco confusa para la máquina de TA de descifrar. La parte de “y después el puerro y la cebolla rallados hasta que doren” se refiere a que hay que añadirlos después en el proceso de dorar, pero que ya tenían que estar rallados. Las máquinas de TA (tanto Google Translate como DeepL) lo han traducido como *en dan worden de prei en uien geraspt tot ze goudbruin zijn* y *en dan worden de prei en ui geraspt tot ze goudbruin zijn*. Esto da a entender que rallar el puerro y la cebolla sucede después, mientras que lo que quiere expresar la frase es que se fríe el puerro y la cebolla después de que la mantequilla se caliente. Una mejor alternativa y más idiomática traducción podría ser: *Verhit de boter in een braadpan⁸ en voeg daarna geraspte prei en ui toe*. Yo diría que faltaría un verbo o añadir algo a la frase, por ejemplo: *Verhit de boter in een braadpan en voeg daarna geraspte prei en ui toe. Bak de geraspte prei en ui totdat het goudbruin wordt. Con verhitten* (calentar), ya el lector sabe más o menos que tiene que esperar a que la mantequilla coja temperatura para poder freír algo.

TF: Añadir dos dientes de ajo cortados muy pequeños. Dejar unos minutos.

TM Google Translate: Voeg twee teentjes knoflook toe, heel klein gesneden. Laat een paar minuten.

TM DeepL: Voeg twee zeer kleine gehakte teentjes knoflook toe. Vertrek voor een paar minuten.

Lo más problemático de esta frase es la parte final. En la primera parte se podría hablar de escoger una traducción más idiomática que lo que han ofrecido Google Translate y DeepL. Pero tanto *klein gesneden* como *klein gehakte* son correctas. “Dejar unos minutos” como frase puede ser muy ambiguo, no está claro qué hay que dejar unos minutos. Un traductor humano si puede deducir de la frase anterior que se trata de unos dientes de ajo. En este caso Google tradujo “dejar” con *laat* y DeepL lo tradujo con *vertrek*. Laat (Google Translate) no es incorrecto, solo que yo como traductora o adaptaría la frase o añadiría un sujeto a la frase, por ejemplo: *Voeg twee klein gesneden/gehakte knoflook teentjes toe en laat gedurende een paar minuten bakken*.

TF: Añadir la harina, con ayuda de unas varillas y remover sin parar. Seguidamente añadir la leche y la nata sin dejar de desliar con las varillas.

⁸ (verhitten wordt heel vaak gebruikt in recepten en dit geeft dan duidelijk het proces dat in het Spaans helemaal omschreven wordt).

TM Google Translate: Voeg de bloem toe, met behulp van een aantal staafjes en roer zonder te stoppen. Voeg vervolgens de melk en room toe en blijf schuiven met de staven.
TM DeepL: Voeg de bloem toe met behulp van een paar stokjes en roer non-stop. Voeg vervolgens de melk en de room toe, terwijl u verder mengt met de stokjes.

Problemas que podemos notar en esta frase es el uso de palabras erróneas y un orden erróneo de las frases. “Varillas” ha sido traducido por Google Translate y DeepL como *een aantal staafjes* y *een paar stokjes* respectivamente. Ambas traducciones son incorrectas, y esto quizás se debe a que en el texto fuente esta palabra este en plural y vaya acompañado del artículo “unas”. De ahí las traducciones *aantal* y *een paar*. Palabras correctas para varilla son: *klopper*, *garde*, *klutser*. Desliar⁹ significa en este contexto sin dejar de “menguar” la mezcla y la máquina TA que mejor lo tradujo fue DeepL. Una posible traducción podría ser: *Voeg de bloem toe en roer met behulp van klopper/garde door. Voeg vervolgens de melk en de room toe, terwijl u/je verder mengt met behulp van de klopper. O Voeg al roerende (sin dejar de desliar) de melk en de room toe.*

TF: Añadir un poco de nuez moscada rallada, sal y pimienta.

TM: Google Translate/DeepL: Voeg een beetje geraspte nootmuskaat, zout en peper toe.

Esta frase ha sido traducida igual por ambas máquinas de TA. Algo que podía ir mal al momento de traducción era si las máquinas no conocían la combinación “nuez moscada” + “rallada”. Se considera como una buena traducción, y esto se refleja en la evaluación de los expertos que han evaluado esta traducción con un 4,25 de media (entre bien y muy bien).

TF: Añadir el pollo y el jamón y mantener al fuego sin parar de remover durante diez minutos más.

TM Google Translate: Voeg de kip en de ham toe en blijf in vuur zetten zonder nog eens tien minuten te roeren.

TM DeepL: Voeg de kip en de ham toe en blijf nog tien minuten roeren.

Aquí vemos, por una parte, que la frase construida por Google Translate no es correcta y que a la frase de DeepL se puede considerar incompleta. Sin embargo, el mensaje si llega. Queda claro qué hay que hacer y esto lo vemos también en los resultados de la encuesta. Google Translate tradujo “mantener al fuego” como *in vuur zetten*, mientras que podía optar por también omitir esa parte, ya que diciendo *blijf nog eens tien minuten roeren en kook door* (por ejemplo), indica que se mantiene la comida al fuego.

⁹ Desliar: desatar, separar, *undo* en inglés

TF: Apartar y dejar enfriar en la nevera durante 24 horas.

TM Google Translate: Leg opzij en laat 24 uur afkoelen in de koelkast.

TM DeepL: Zet het geheel apart en laat het 24 uur afkoelen in de koelkast.

Aquí vemos nuevamente que el sujeto en este caso no se menciona, pues es difícil para una máquina TA saber qué es lo que hay que apartar y dejar enfriar, esto puede traer al momento de traducir problemas de ambigüedad. Ambas máquinas de TA omitieron la traducción de “durante”, *gedurende*. Las frases no están mal traducidas, pero si hay mejores opciones que las que ofrecen Google Translate y DeepL. Por ejemplo, considerando que se habla de una masa, yo emplearía ese término en holandés (e incluso en el texto fuente dado el caso que esta opinión se toma en cuenta. Por ejemplo: *Zet het deeg apart (in een apart bakje) en laat gedurende 24 uur afkoelen*.

TF: Colocar la masa sobre la mesa de trabajo, junto con un plato con harina, otro con huevo batido y un último con pan rallado.

TM: Leg het deeg op de werktafel, samen met een bord met bloem, een ander met losgeklopt ei en een laatste met broodkruim.

TM: Leg het deeg op de werktafel, samen met een bord met bloem, een ander met geklopt ei en een laatste met paneermeel.

Aquí vemos que se han traducido las frases casi iguales, a excepción de dos palabras. Google Translate tradujo “huevo batido” como *losgeklopt ei*, mientras que DeepL lo tradujo como *geklopt ei*. Luego “pan rallado” se ha traducido como *broodkruim* y *paneermeel* respectivamente. En esta frase, aunque se usa *andere* refiriéndose al plato, las máquinas de TA lo han captado y traducido correctamente. Sin embargo, podríamos hablar de algunas adaptaciones en cuanto a las traducciones ofrecidas. Por ejemplo: *Leg het deeg op de werktafel, samen met een bord met bloem, een bord met losgeklopte ei en een bord met paneermeel. Broodkruim no está mal, pero usar paneermeel es más común*.

TF: Con ayuda de una cuchara (o con las manos) coger porciones de la masa y enharinarlas dándoles forma. Mojarlas en el huevo y embadurnarlas después con pan rallado.

TM: Neem met behulp van een lepel (of met uw handen) porties van het deeg en meel ze in vorm. Doop ze in het ei en smeer ze in met broodkruimels.

TM: Neem met behulp van een lepel (of met je handen) porties van het deeg en bloem ze vorm te geven. Dompel ze in het ei en bestrijk ze vervolgens met paneermeel.

La primera parte de la frase no contiene errores o traducciones problemáticas, sin embargo, como las veces anteriores Google Translate traduce *uw handen* mientras que DeepL traduce *je handen*. Luego tenemos los siguientes pasos que son “enharinarlas dándoles forma”, “mojarlas en el huevo” y “embadurnarlas después con pan rallado”.

Google Translate traduce “enharinarlas dándoles forma” como *meel ze in vorm* y DeepL lo traduce como *bloem ze in vorm*. *Melen y bloemen son sinónimos y ambos significan según Van Dale: met bloem bestrooien*. Sin embargo, yo no categorizaría estas traducciones como buenas, hay mejores opciones que son más idiomáticas y fluyen mejor. Por ejemplo, se podría traducir como: *Neem met behulp van een lepel of maak gebruik van je handen om porties van het deeg door meel te halen. Vorm daar kroketten van. Haal de kroketten vervolgens door het geklopte ei en daarna door paneermeel*. Yo escogería verbos más neutrales, que suelen ser lo más usual, solamente que el orden de la frase cambiaría un poco. Esto resulta menos problemático que traducir cada verbo por lo que es y crear frases que no son correctas o idiomáticas.

TF: Freír con aceite de girasol muy caliente, para evitar que se deshagan.

TM Google Translate: Bak met zeer hete zonnebloemolie om te voorkomen dat ze uit elkaar vallen.

TM DeepL: Bak ze in zeer hete zonnebloemolie om te voorkomen dat ze uit elkaar vallen.

En esta última frase vemos que por ejemplo Google Translate ha traducido literalmente “freír con” al holandés como “*Bak met*”, mientras que esto en holandés es incorrecto, la preposición correcta es *in*, *bakken in*. DeepL si lo tradujo de tal forma. En este caso podemos hablar del uso de una preposición errónea de parte de Google Translate.

5.1.3 Texto 3: chistes

El tercer genero de texto que vamos consiste en tres chistes cortos, que consideramos principalmente expresivos. Como vimos en la teoría de la traducción del humor (Vandaele, 2010)¹⁰ y en los aspectos de textos expresivos (Reiss, citado en Munday, 2008), es importante que no solo se traduzca el contenido, sino que es importante la forma, la estética, como dice Reiss. Lo que hay que tener en cuenta es que hay normas que pueden ser diferentes en el LM y en el LF. Un texto expresivo La traducción de un texto expresivo debe adaptar la perspectiva del autor del texto fuente. En comparación con los otros dos textos, estos chistes vamos a analizarlos completos. Desde ahí vamos a ver qué tipo de errores conseguimos y que cosas podrían haber sido traducido diferentes según nuestro criterio.

Google Translate

Ze gaan twee blinds verderop de straat voorbij en zeggen:

- Ik wou dat het regende!
- Ik wou dat ik het ook deed!

Een moeder vertelt haar zoon:

- Een vogeltje zei dat ik je drugs moest geven!
- Degene die drugs gebruikt, is jij die spreekt met vogels!

Ze waren een kreupele man, een dove man, een blinde man en een kale man die om aalmoezen vroeg. In dat een politieagent zou aankomen en de dove man zegt:

- Ik hoor stappen.

De blinde ziet de politie en de lamme man wegrennen en de kale man vertelt de politieagent:

- De doven hoorden het, de blinde man zag het en de lamme liep weg, voor mij namen ze mijn haar.

DeepL

Er lopen twee kleine blinde mensen door de straat in het heetst van de strijd en ze zeggen:

- Ik wou dat het regende!
- Ik wou dat ik dat ook deed!

Een moeder zegt tegen haar zoon:

- Een klein vogeltje zegt me dat je aan de drugs bent!
- Jij bent degene die drugs gebruikt, jij bent degene die met kleine vogels praat!

Het waren een kreupele, dove, blinde en kale man die om een aalmoes smeekte.

Een politieagent zou aankomen en de dove man zou zeggen:

- Ik hoor voetstappen.

De blinde man ziet de politie en de kreupele man vlucht, en de kale man vertelt het aan de politieman:

- De doven hoorden het, de blinde man zag het en de kreupele man liep weg, voor mij werd mijn haar verwijderd.

TF:

Van dos cieguitos por la calle pasando calor y dicen:

- ¡Ojalá lloviera!
- ¡Ojalá yo también!

TM Google Translate:

Ze gaan twee blinds verderop de straat voorbij en zeggen:

- Ik wou dat het regende!
- Ik wou dat ik het ook deed!

TM DeepL:

Er lopen twee kleine blinde mensen door de straat in het heetst van de strijd en ze zeggen:

- Ik wou dat het regende!
- Ik wou dat ik dat ook deed!

Si analizamos este texto bien, nos damos cuenta de que hay más aspectos importantes que solamente la traducción. Este chiste tiene un componente ambiguo que no hemos abordado en el marco teórico, ya que considerábamos que esto no tenía mucha relevancia para esta investigación.

Esto es un ejemplo claro de un chiste en donde se usa un léxico ambiguo, cuyo término exacto es la homofonía. La homofonía se produce entre palabras que se pronuncian igual, e escriben de forma diferente y se refieren a dos conceptos diferentes. <<Ojalá lloviera (*regenen*)>> / <<Ojalá yo viera (*zien*)>>.

En este caso ya vemos que desde el principio Google Translate se fue por el sentido demasiado literal de la traducción, casi podemos decir que tradujo según el principio “*word-for-word*” (palabra por palabra).

“Van” lo tradujo GT como *ze gaan* (*ellos, ellas, ustedes* → *3a persona plural*), mientras que en este caso lo que se necesita es la voz pasiva en holandés y no una frase con la 3a persona singular, es decir una frase que empiece por *Er* + verbo.

Luego vemos que Google Translate incluyó en su traducción “*verderop de straat voorbij*”, mientras que en el texto fuente no hay nada que indique “*verderop*”. Esto es un error por parte de Google Translate porque esto cambia el significado.

DeepL si tradujo “van dos cieguitos...” como “Er lopen”, sin embargo, cometió un error morfológico. Interpretó el significado de cieguitos, -os, como personas pequeñas ciegas mientras que aquí se utiliza cieguitos como un diminutivo de afecto.

La otra parte, “pasando calor y dicen”, ha sido traducido por Google Translate como *verderop de straat voorbij en zeggen*. Esto se debe a que Google Translate cogió el significado de pasando solamente y no la combinación pasando + calor. DeepL tampoco lo tradujo correctamente. El output de DeepL fue: *in het heetst van de strijd*, mientras que en este texto no aparece nada de ninguna pelea (*strijd*).

Sobre el pequeño diálogo, hay que decir que gramaticalmente y referente al contenido traducido, fue traducido correctamente. Simplemente, el juego de palabras en holandés no tiene el mismo efecto que en español. Como explicado anteriormente aquí se trata de dos palabras totalmente distintos al momento de pronunciar. Podríamos considerar presentar una traducción correcta, la cual podría ser:

Er lopen twee blinde mensen door de straat, die het warm hebben, en zeggen ze:

-Ik wou dat het regende!

- Ik wou dat ik dat took deed!

Al traducir la última parte así, quedamos fieles al texto fuente, pero el chiste se pierde.

El segundo chiste, vemos a continuación el output de Google Translate y de DeepL:

TF: Le dice una madre a su hijo:

- ¡Me ha dicho un pajarito que te drogas!

- ¡La que se droga eres tú que hablas con pajaritos!

TM Google Translate: Een moeder vertelt haar zoon:

- Een vogeltje zei dat ik je drugs moest geven!

- Degene die drugs gebruikt, is jij die spreekt met vogels!

TM DeepL: Een moeder zegt tegen haar zoon:

- Een klein vogeltje zegt me dat je aan de drugs bent!

- Jij bent degene die drugs gebruikt, jij bent degene die met kleine vogels praat!

Cabe destacar que DeepL tuvo mucho mayor evaluación que Google Translate. La traducción de DeepL es muy buena a pesar de tener algunos errores, o traducciones que podríamos adaptar.

Google Translate empezó bien, *Een moeder vertelt haar zoon*, podría ser traducido como zegt tegen haar zoon, ya que vertellen es más parecido a contar algo (una historia). Luego vemos que la madre dice:

- ¡Me ha dicho un pajarito que te drogas!

Es decir: Un pajarito le dijo a la madre que el hijo de esa madre se droga. Esto ha sido mal interpretado y traducido por Google Translate. Google Translate cambió la perspectiva traduciendo esta frase con el significado de que el pajarito le dijo a una mama que ella debía darle drogas a su hijo. Esto es totalmente incorrecto y contrario a lo que el chiste dice en español.

Luego, la respuesta del hijo a la madre ha sido interpretada correctamente, solamente que contiene errores gramaticales y sintácticas:

Degene die drugs gebruikt, is jij die spreekt met vogels! Es incorrecto, ya que una frase en holandés debería empezar con el sujeto – predicado. Es decir, Jij **bent** (envez de is) degene die drugs gebruikt. Jij praat met vogels!

DeepL

La acusación de la madre ha sido interpretada correctamente, sin embargo, podríamos adaptar esta traducción.

- Een klein vogeltje zegt me dat je aan de drugs bent!

Podríamos cambiar zegt por *vertelt/heeft mij verteld*.

Luego la respuesta del hijo es correcta, hasta podemos notar que se ha ampliado la frase de forma que tenemos dos veces “*jij bent...jij bent...*”

- Jij bent degene die drugs gebruikt, jij bent degene die met kleine vogels praat!

TF:

Estaban un cojo, un sordo, un ciego y un calvo pidiendo limosna. En eso que iba a llegar un policía y el sordo dice:

- Oigo pasos.

El ciego ve al policía y el cojo sale huyendo y en eso que el calvo le dice al policía:

- El sordo lo oyó, el ciego lo vio y el cojo se fue corriendo, para mí que me tomaron el pelo.

Respecto a este chiste, podemos decir que en general las traducciones de tanto Google Translate como DeepL han sido evaluadas como malas traducciones. A

continuación, vamos a ver detalladamente que han producido ambas máquinas traductoras.

TM Google Translate:

Ze waren een kreupele man, een dove man, een blinde man en een kale man die om aalmoezen vroeg.

In dat een politieagent zou aankomen en de dove man zegt:

- Ik hoor stappen.

De blinde ziet de politie en de lamme man wegrennen en de kale man vertelt de politieagent:

- De doven hoorden het, de blinde man zag het en de lamme liep weg, voor mij namen ze mijn haar.

TM DeepL:

Het waren een kreupele, dove, blinde en kale man die om een aalmoes smeekte.

Een politieagent zou aankomen en de dove man zou zeggen:

- Ik hoor voetstappen.

De blinde man ziet de politie en de kreupele man vlucht, en de kale man vertelt het aan de politieman:

- De doven hoorden het, de blinde man zag het en de kreupele man liep weg, voor mij werd mijn haar verwijderd.

Tanto en la traducción de Google Translate y DeepL se puede decir que hay errores de léxico debido a la ambigüedad que los TA no pueden evitar, considerando que ellos no se basan en el contexto para resolver ambigüedades. Otro problema es que por el hecho de no poder traducir creativamente las máquinas tradujeron muy literalmente. Otro problema con la que nos encontramos esta vez es la traducción del refrán “tomar el pelo” que en holandés significa *voor de gek houden*.

5.1.4 Texto 4: biografía Frida Kahlo

El último texto que vamos a analizar es una biografía, y como dicho anteriormente en el marco teórico, este texto tiene una función híbrida, es un texto informativo y expresivo. La función de un texto informativo-expresivo es informar al lector mientras que se debe mantener la función artística de una obra de literatura. Normalmente estos textos quieren hacer llegar información a los lectores de una forma artística, y quizás esto sea lo difícil de estas traducciones. Aparte de que el contenido del texto tiene que ser correcto, el estilo tiene que ser equivalente al estilo del texto fuente.

Google Translate

Toen ze 18 was, leed ze een ernstig ongeluk waardoor ze 9 maanden in bed lag en toen begon ze te schilderen.

In 1927 ontmoette hij de muralist Diego Rivera en twee jaar later werd hij zijn derde vrouw.

Het huwelijk van Diego en Frida was een beetje conventioneel en problematisch.

Naast de frustratie om geen kinderen te kunnen krijgen, gebeurden er ontrouwtjes aan beide kanten.

Ze scheidde in 1939 en hertrouwde een jaar later.

DeepL

Op 18-jarige leeftijd kreeg ze een ernstig ongeluk dat haar 9 maanden in bed hield en dat was toen ze begon te schilderen.

In 1927 ontmoette hij de muralist Diego Rivera en twee jaar later werd hij zijn derde vrouw.

Diego en Frida's huwelijk was een beetje conventioneel en problematisch.

Naast de frustratie van het niet kunnen krijgen van kinderen, volgden ontrouw aan beide kanten.

Ze scheidden in 1939 en hertrouwden een jaar later.

TF: A los 18 años sufrió un grave accidente que la mantuvo en cama 9 meses y fue entonces cuando comenzó a pintar.

TM Google Translate: Toen ze 18 was, leed ze een ernstig ongeluk waardoor ze 9 maanden in bed lag en toen begon ze te schilderen.

TM DeepL: Op 18-jarige leeftijd kreeg ze een ernstig ongeluk dat haar 9 maanden in bed hield en dat was toen ze begon te schilderen.

De esta frase podemos decir en primer lugar que vemos errores en la traducción de Google Translate. Este error tiene que ver con el uso de un verbo fijo al lado de un sustantivo. En este caso en español se dice sufrir un accidente. Pero en holandés se sufre por una enfermedad, por ejemplo. Uno dice en holandés que consiguió un accidente enés de que sufrió (lijden) un accidente. Es por eso por lo que se puede considerar esta traducción de Google Translate como errónea. Luego tenemos otro problema que puede resultar conflictivo. En el texto fuente está escrito que después de su accidente se mantuvo 9 meses en la cama, y es entonces (cuando eso sucedió) que ella empezó a pintar. Con la traducción de Google Translate se puede entender que ella empezó a pintar una vez pasado esos nueve meses, esto es un caso de ambigüedad ya que no se dice explícitamente en el texto holandés que ella empezó a pintar durante esos nueve meses.

Al contrario que Google Translate, en la traducción de DeepL si el TA optó por especificar por medio de *en dat ... te schilderen*. Concluyendo, podemos decir que en este caso la traducción de DeepL es más completa que la de Google Translate.

TF: En 1927 conoció al muralista Diego Rivera y dos años tarde se convirtió en su tercera esposa.

TM Google Translate: in 1927 ontmoette hij de muralist Diego Rivera en twee jaar later werd hij zijn derde vrouw

TM DeepL: In 1927 ontmoette hij de muralist Diego Rivera en twee jaar later werd hij zijn derde vrouw.

Aquí vemos un caso de ambigüedad en donde el TA (Google Translate) no es capaz de basarse en el contexto para elegir la traducción correcta. La conjugación "conoció" pertenece a la tercera persona singular (él/ella/usted). En este caso del

contexto se puede entender que se trata de ella, sin embargo, el TA lo tradujo como él. En el caso de DeepL sucede lo mismo al principio <<ella por *hij*>>, como también traducir ella por él en el holandés. Ambas máquinas de TA cometieron los mismos errores.

TF: El matrimonio de Diego y Frida fue un poco convencional y problemático.

TM Google Translate: Het huwelijk van Diego en Frida was een beetje conventioneel en problematisch.

TM DeepL: Diego en Frida's huwelijk was een beetje conventioneel en problematisch.

De las traducciones de Google Translate y DeepL podemos ver que se han traducido casi iguales. Aquí podríamos hablar de escogencias menos idiomáticas por parte de los TTAA. Yo como traductora optaría por usar el inicio de la frase de Google Translate y cambiaría *conventioneel* por *ouderwets*, que en este contexto suena mejor y es más usual.

TF: Además de la frustración de no poder tener hijos, las infidelidades se sucedieron por ambas partes.

TM Google Translate: Naast de frustratie om geen kinderen te kunnen krijgen, gebeurden er ontrouwtjes aan beide kanten.

TM DeepL: Naast de frustratie van het niet kunnen krijgen van kinderen, volgden ontrouw aan beide kanten.

Otro error de la TA se puede ver por ejemplo en esta última frase, traducido por Google Translate erróneamente. Al principio de la frase se traduce “se divorciaron” como *ze scheidden* (plural), mientras que más adelante en la frase DeepL mantiene la forma plural, pero Google Translate no. Google Translate tradujo “Se volvieron a casar” como *hertrouwde*, la forma plural necesitaría un -n al final de *hertrouwde(n)*. Esto es traducido correctamente por DeepL

5.2 La valoración de los expertos

Como parte del experimento, se le ha pedido a diez expertos que evalúen la calidad de las traducciones automáticas de Google Translate y DeepL. Para cada frase se podía evaluar mediante una escala de 1 a 5, con el 1 siendo una traducción de muy mala calidad y el 5 de muy buena calidad. Con los resultados obtenidos podemos ver como opinan los expertos de los textos en general, y también podemos ver en cada genero que les pareció a los expertos una buena y una mala traducción. Así mismo podemos ver la calidad de tanto Google Translate como DeepL desde el punto de vista de los expertos de las máquinas traductoras.

Hemos hecho un análisis muy amplio de los errores que hemos encontrado a lo largo de las traducciones de los textos. Lo que queremos analizar ahora es en base a estas cifras es como se han evaluado estas traducciones teniendo en cuenta la función de cada texto. También queremos ver qué diferencias hay entre las

valoraciones de una traducción y la otra, como también cuales frases han sido mejor y peor evaluadas. Junto con la evaluación lingüística estos datos nos llevarán a una conclusión a nuestra pregunta de investigación.

A partir de la valoración de los expertos, podemos llegar a la siguiente comparación:

Texto	Tipología	Google Translate	DeepL
1. BookCrossing	Texto informativo	2,48	3,06
2. Receta	Texto operativo	2,23	3,13
3. Chistes	Texto expresivo (1)	1,54	2,50
4. Biografía Frida Kahlo	Texto expresivo (2)	2,84	3,79

Tabla 2: Valoraciones medias por tipología de texto

Como ya se venía comentando a lo largo de este trabajo, los textos evaluados pertenecen a una determinada tipología textual: informativo, operativo y expresivo. En la tabla 2 presentamos las valoraciones medias de cada texto. Estos resultados se obtuvieron a partir de las valoraciones por frase (pudiendo escoger de 1 a 5, de muy mala a muy buena). Lo que podemos ver a primera vista es que generalmente DeepL ha sido mejor evaluado que Google Translate. Podemos ver que excepto el texto informativo, hay una diferencia evidente entre la calidad de Google Translate y de DeepL. De esta tabla podemos deducir que por ejemplo para Google Translate es más difícil traducir un texto expresivo y le resulta mejor traducir un texto híbrido o informativo. Por el otro lado DeepL también ha sido mejor evaluada para texto informativo y expresivo (híbrido).

En la siguiente tabla veremos los valores de las oraciones mejores y peores evaluadas de cada máquina TA y la diferencia entre la oración mejor y peor evaluada.

Texto	Google Translate			DeepL		
	mejor oración	peor oración	diferencia	mejor oración	peor oración	diferencia
1. BookCrossing	4,06	1,63	2,43	4,31	2,12	2,19
2. Receta	4,25	1,25	3,00	4,25	1,63	2,62
3. Chistes	1,69	1,31	0,38	4,00	1,63	2,37
4. Biografía Frida Kahlo	2,88	1,75	1,13	4,00	1,50	2,50

Tabla 3: Oraciones mejores y peores evaluadas Google Translate y DeepL

En la tabla 3 vemos por ejemplo que la frase mejor evaluada de Google Translate del primer texto tiene una valoración de 4,06 mientras que la frase peor evaluada tiene un 1,63 de valoración. En comparación con DeepL, la frase peor evaluada de DeepL ha sido valorada con un 2,12 y la mejor con un 4,31. Si vemos las valoraciones del tercer texto, Google Translate ha sido evaluado con un 1,69 para la mejor frase y un 1,31 para la peor. Lo que podemos deducir de aquí es que la calidad de este texto se considera muy mal. Por el otro lado la frase mejor evaluada del texto 3 de DeepL, tiene una valoración de 4 y la peor tiene una valoración de

2.37. Esto nos muestra que en comparación con Google Translate, DeepL ha producido mejor calidad de traducción en cuanto a la traducción de un texto expresivo. Si vemos por ejemplo en cuanto al texto operativo, la mejor frase de Google Translate ha sido evaluada con un 4,25 y la peor con un 1,25. DeepL ha sido evaluado con un 4,25 y 1,63 respectivamente. Esto nos da una indicación que Google Translate nuevamente ha producido traducciones de menos calidad que DeepL.

5.3 Conclusión

Para poder conseguir una vista más completa posible sobre la calidad de las traducciones de los traductores automáticos, nos hemos basado primero en la teoría investigada para poder ver qué cosas reconocemos de la teoría en las traducciones. Hemos hecho un análisis detallado de las cosas que nos encontramos en las traducciones, ya sea ambigüedad en el léxico de los textos, errores por el hecho de no poder recurrir al contexto o traducciones erróneas porque las máquinas de TA no están acostumbradas a pensar por sí solos, sino que solo siguen las reglas que conocen. A parte de este extenso análisis, hemos presentado y resumido las valoraciones de los expertos quienes han evaluado las traducciones de los cuatro textos.

A pesar de todo lo discutido en la literatura vemos que cada texto es diferente y por tanto presenta diferentes tipos de problemas para la TA. Estos problemas pueden ser por culpa de un léxico ambiguo, una frase construida de forma que el TA no lo pueda reconocer ni conseguir una traducción exacta. Esto a veces puede ocasionar que el TA omita una parte o añada información al texto que antes no estaba.

Capítulo 6: Conclusión

La traducción automática es un fenómeno muy interesante, tanto por su historia como por las posibilidades que ofrece. Con un sistema de TA, se ha logrado hacer accesible mucha información en diferentes idiomas, y este es un ámbito de la traducción que se sigue investigando. En este estudio hemos intentado investigar mediante textos de diferentes géneros, la calidad de traducciones que producen Google Translate y DeepL. Concretamente, queremos saber, como lo dice la pregunta principal de esta investigación, lo siguiente:

¿Cómo se diferencian las traducciones producidas por los traductores automáticos Google Translate y DeepL? Una comparación entre dos herramientas de traducciones automáticas que traducen del español al neerlandés.

Basándonos en los resultados del análisis, se puede concluir en primer lugar que las traducciones propuestas por DeepL, generalmente fueron mejor evaluadas que Google Translate. Esto no significa que las traducciones de DeepL no contenían errores, ya que la mayoría de las veces sí había mejores soluciones en comparación con lo que produjo DeepL.

Para poder contestar esta pregunta primero nos hemos dedicado a revisar lo que dice la teoría sobre la traducción automática. Luego nos hemos basado en la teoría de tipologías textuales, para saber cómo poner a prueba las dos máquinas que queríamos investigar. Lo que no pretendemos hacer con este experimento es decir que nuestra solución es correcta y la de la TA no lo es, sino que al analizar las traducciones nos detenemos a comentar las cosas que vemos y que hemos estudiado en el marco teórico. Gracias al marco teórico tenemos pautas de cosas que nos podemos encontrar en traducciones realizadas de forma automática.

Considerando que no podremos dar una respuesta que valga para la TA en su totalidad, nos hemos centrado en cuatro textos partiendo de la teoría de Reiss. Opinamos que esta teoría nos explica cómo se debe de traducir los textos basándonos en sus funciones. Estas funciones son bastante diferentes, pues pensamos que de esta forma ofrecemos una vista amplia y variada de diferentes textos traducidos por Google Translate y DeepL.

Primero hemos intentado sacar las irregularidades de las traducciones para analizarlas lingüísticamente para luego pasar a analizar las encuestas rellenas por expertos, en donde se les pedía una evaluación de la calidad de traducción que estaban teniendo enfrente. A lo largo del análisis lingüístico hemos podido concluir que, en primer lugar, la ambigüedad en una máquina de TA difícilmente se puede evitar. Es difícil que una máquina traductora siempre acierte al momento de traducir, peor aún no pudiendo recurrir al resto del texto. A pesar de que esto es algo muy difícil para el TA, hubo casos en donde la traducción sobre todo de DeepL acertó y tradujo correctamente dentro del contexto del texto. Esto con Google Translate se veía casi nada. En segundo lugar, hemos concluido que mientras más implícito sea el texto o el conocimiento que se requiera para entender un texto, menos correcta iba a ser la traducción de una máquina traductora, tal y como es el caso de los chistes.

Otra conclusión a la que llegamos es que, a la hora de traducir, las máquinas traductoras no se paran a pensar que es lo que realmente nos quiere transmitir el texto fuente. Nos enfocamos puramente en el aspecto lingüístico de la traducción, mientras que un poco de cultura y conocimiento a veces también es necesario para traducir correctamente. También nos dimos cuenta de que depende del tipo de texto, te saldrá una traducción mejor o peor.

Lo que queda claro es que a pesar de que por ejemplo la TA haya obtenido buenos resultados con algunos textos y algunos no, aún no se ha logrado que una máquina traductora logre una traducción 100% correcta. El componente de “post-editar” siempre será algo necesario. Con la máquina traductora avanzamos muchísimo, pero como cualquier traductor, sabemos que más que la traducción “pura”, es necesario que el texto encaje en diferentes aspectos del texto y público meta.

El sueño de los precursores de crear una máquina FAHQT (fully automated high quality translation), puede cumplirse en parte siempre y cuando estemos conscientes de que este proceso no puede realizarse solo de forma automática. Habrá que revisar el texto y adaptarlo a medida que sea necesario. Esto lo tiene que hacer un traductor humano.

Capítulo 7: Discusión

Para esta investigación requeríamos de una cantidad de personas para poder dar su opinión como expertos sobre la calidad de los TTAA. Esto puede ser difícil de lograr ya que no sabes con seguridad quien va a colaborar y por más que intentes delimitar tu grupo de encuestados, siempre habrá encuestados que por un motivo u otro no pueda colaborar, entonces queda buscar en otros lugares para poder llegar a la cantidad necesaria para crear una cantidad representativa de opiniones.

Un componente importante de esta investigación era investigar la calidad, y lo que se tenía pensado desde un principio era dejar a expertos corregir las traducciones y a partir de ahí analizar los cambios (o los errores en la vista de ellos) en las traducciones. Esto acompañado de la encuesta que los encuestados han rellenado iba a llevar mucho tiempo e iba a ser difícil ejecutar en una sola encuesta. Es por este motivo que se descartó esta idea. No obstante, sería interesante aun investigar esto por separado ya que nos podrá traer más claridad sobre lo que traductores humanos consideran como error y a partir de ahí hacer una clasificación para saber con exactitud en donde fallan los traductores automáticos.

Como habíamos visto en la teoría hay diferentes maneras de evaluar las traducciones realizadas. En el cuadro de esta investigación sería interesante investigar en un futuro la calidad de las traducciones después de hacer una retraducción. Es decir, comparar el texto fuente original con la segunda versión del texto en su lenguaje fuente.

Referencias bibliográficas

Chéragai, M. (2012). Theoretical overview of machine translation. En *Proceedings ICWIT*. Recuperado de <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.416.1463&rep=rep1&type=pdf#page=176> (mayo del 2019)

Criado, E. (2005). Expresión escrita A2-B1- *Enclave ELE*. Libro

Forcada, M. (2011). Machine Translation Today. En K. Malmkjær, & K. Windle (Eds.), *The Oxford Handbook of Translation Studies* (pp. 215–223). Recuperado de https://www.researchgate.net/publication/326381184_Handbook_of_Translation_Studies_-_Vol_1 (abril del 2019)

Forcada, M. (2017). Making sense of neural machine translation. En J. Benjamins, *International Journal of Translation* (pp. 5-20). Recuperado de <https://benjamins.com/catalog/ts.6.2.06for/fulltext/ts.6.2.06for.pdf> (abril del 2019)

Hutchins, J. (2006). Machine Translation: A Brief History. Recuperado de <https://www.sciencedirect.com/science/article/pii/B9780080425801500660> (abril del 2019)

Hutchins, J. (2001). Machine translation and human translation: in competition or in complementation?. Recuperado de <http://www.hutchinsweb.me.uk/IJT-2001.pdf> (abril del 2019)

Hutchins, J. (2003). The history of machine translation in a nutshell. Recuperado de <http://www.hutchinsweb.me.uk/Nutshell-2005.pdf> (abril del 2019)

Hutchins, J. (2007). Machine translation: a concise history. Recuperado de <http://www.hutchinsweb.me.uk/CUHK-2006.pdf> (abril del 2019)

Mogas Ribas, M. (2016). La traducción automática: ¿una amenaza para el traductor del siglo XXI? Recuperado de <https://ddd.uab.cat/record/160589> (junio del 2019)

Munday, J. (2008). Functional theories of translation. En *Introducing Translation Studies: Theories and Applications* (pp. 113-140). Recuperado de

<https://www.routledge.com/Introducing-Translation-Studies-Theories-and-Applications-4th-Edition/Munday/p/book/9781138912557> (abril del 2019)

Tertoolen, R. (2010). El alcance de la traducción automática. Recuperado de <https://dspace.library.uu.nl/handle/1874/191661> (abril del 2019)

Vandaele, J. (2010). Humor in Translation. En Gambier, A. & Van Doorslaer, L. (Eds.), *Handbook of Translation Studies* (pp. 147-152). Recuperado de <https://www.researchgate.net/publication/326381184> Handbook of Translation Studies - Vol 1 (junio del 2019)

Anexo I

Texto informativo (folleto BookCrossing de Instituto Cervantes)

Instituto Cervantes de Utrecht, ¡zona BookCrossing!

¿Qué es BookCrossing?

Es una Biblioteca mundial. Es la web de una red social inteligente. Es la fiesta de la literatura y un lugar donde los libros cobran una nueva vida. BookCrossing es el acto de dar a un libro una identidad única, de manera que, como el libro pasa de lector a lector, se pueda seguirlo e interconectar a sus lectores. Ahora mismo hay **1.940.532 BookCrossers** y **12.844.916 libros** viajando por **132 países**. Nuestra comunidad está cambiando el mundo y las vidas de libro en libro.

Nuestra **misión**: conectar personas a través de los libros

¿Cómo dejar un libro en una zona BookCrossing?

1. **Regístrate** en www.bookcrossing.com (esto te puede llevar tan sólo unos 2 o 3 minutos).
2. **Registra tus libros**: <https://www.bookcrossing.com/register>. Te darán un código BCID (nº de identificación del libro) que pegarás con una etiqueta en el libro.
3. **Libera tu libro “en la jungla”**. Después tendrás que escoger la zona de liberación donde se podrá encontrar tu libro: biblioteca instituto Cervantes Utrecht.

¡Si no te apetece registrarte en BookCrossing, la biblioteca liberará tus libros por ti!

¿Cómo sabrás si alguien ha encontrado un libro que tú has liberado?

Al estar registrado en www.bookcrossing.com, el sistema de esta página te enviará un mensaje cada vez que alguien escriba una entrada de diario en el libro que hayas liberado. También puedes mirar un apartado llamado “bookshelf” (estantería) dentro de tu perfil en BookCrossing.com para ver todas las entradas de diario de tus libros.

¿Cómo llevarse un libro de una zona BookCrossing?

Dirígete a www.bookcrossing.com (no necesitas estar registrado) e introduce el código BCID que encontrarás en el libro que hayas encontrado. Deja una breve

nota en la entrada de diario para que el antiguo dueño del libro sepa qué ha pasado con éste. Cuando te canses de él, ¡lo podrás liberar!

Fuente: Instituto Cervantes Utrecht, Biblioteca

Texto operativo (receta de croquetas caseras)

Croquetas de jamón y pollo

Ingredientes para elaborar 40 unidades

- > Leche, 1 L
- > Nata, 1 L
- > Dos cebollas
- > Un diente de ajo
- > Un muslo de pollo
- > Jamón ibérico, 200g
- > Dos huevos
- > Harina, 200 g
- > Harina, huevo y pan rallado para rebozar, c/s
- > Mantequilla, 100 g
- > Nuez moscada, c/s
- > Sal y pimienta, c/s



Paso a paso

1. Cortar el jamón y el pollo en tacos muy pequeños. Reservar.
2. Añadir la mantequilla en una cazuela hasta que coja temperatura, y después el puerro y la cebolla rallados hasta que doren.
3. Añadir dos dientes de ajo cortados muy pequeños. Dejar unos minutos.
4. Añadir la harina, con ayuda de unas varillas y remover sin parar. Seguidamente añadir la leche y la nata sin dejar de desliar con las varillas.
5. Añadir un poco de nuez moscada rallada, sal y pimienta.
6. Añadir el pollo y el jamón y mantener al fuego sin parar de remover durante diez minutos más.
7. Apartar y dejar enfriar en la nevera durante 24 horas.
8. Colocar la masa sobre la mesa de trabajo, junto con un plato con harina, otro con huevo batido y un último con pan rallado.
9. Con ayuda de una cuchara (o con las manos) coger porciones de la masa y enharinarlas dándoles forma. Mojarlas en el huevo y embadurnarlas después con pan rallado.
10. Freír con aceite de girasol muy caliente, para evitar que se deshagan.

Fuente: <https://www.lavanguardia.com/comer/20160617/402570331477/receta-croquetas-jamon-pollo.html>

Texto expresivo (chistes)

Le dice una madre a su hijo:

- ¡Me ha dicho un pajarito que te drogas!
- ¡La que se droga eres tú que hablas con pajaritos!

Van dos ciegos por la calle pasando calor y dicen:

- ¡Ojalá lloviera!
- ¡Ojalá yo también!

Estaban un cojo, un sordo, un ciego y un calvo pidiendo limosna. En eso que iba a llegar un policía y el sordo dice:

- Oigo pasos.

El ciego ve al policía y el cojo sale huyendo y en eso que el calvo le dice al policía:

- El sordo lo oyó, el ciego lo vio y el cojo se fue corriendo, para mí que me tomaron el pelo.

Fuente: (<https://chistescortos.yavendras.com/>)

Frida Kahlo

Frida Kahlo nació en Coyoacán en 1907. Curiosamente cambió su fecha de nacimiento a 1910 porque ese año comenzó la Revolución Mexicana. A los 18 años sufrió un grave accidente que la mantuvo en cama 9 meses y fue entonces cuando comenzó a pintar.

En 1927 conoció al muralista Diego Rivera y dos años tarde se convirtió en su tercera esposa.

El matrimonio de Diego y Frida fue un poco convencional y problemático. Además de la frustración de no poder tener hijos, las infidelidades se sucedieron por ambas partes. Se divorciaron en 1939 y se volvieron a casar un año después.

Diego pintaba gigantescos murales con temas políticos y Frida dedicó Buena parte de sus pinturas a autorretratos de pequeño tamaño.

Fuente: Eugenia Criado. (2005). Expresión escrita. *Enclave-ELE* (p. 96)

